THE STANDARDIZATION OF AN APTITUDE TEST BATTERY FOR INDIAN PUPILS IN STANDARDS 6 TO 8 FOR INTEGRATION WITH A VOCATIONAL GUIDANCE PROGRAM

by

STANLEY OOSTHUIZEN, M.A.

Thesis presented in fulfilment of the requirements for the degree of

DOCTOR PHILOSOPHIAE

at the

POTCHEFSTROOM UNIVERSITY FOR CHRISTIAN HIGHER EDUCATION

POTCHEFSTROOM

JANUARY 1972

ACKNOWLEDGEMENTS

At the completion of this thesis I feel the need to express my thanks and appreciation to all who have assisted me so willingly, namely:

* My promoter, Prof. L.A. Gouws, for his expert guidance, his interest in my progress with the project and his friendliness at all times;

* The President of the HSRC for permitting me to use research data of the HSRC in Chapters 3 and 4 of this thesis, as well as Dr. A.J. van Rooy and Dr. J.H. Robbertse for assisting me in this respect;

* Dr. F.W.O. Heinichen and Mr. A.D. van Staden for their extremely valuable hints on technical matters;

* Mr. C.M. Stimie for editing the manuscript;

* Mrs. E. Swart for typing the manuscript under pressure at such short notice;

* All my colleagues and relatives who encouraged me;

* My parents and parents-in-law who assisted me financially;

and, last but not least,

* My wife and three children who have patiently allowed me many hours of study.

PRETORIA,
JANUARY 1972

S. OOSTHUIZEN

REMARKS

According to instructions, no HSRC test, manual, scoring key or related instruments may form part of any thesis since these are confidential and may only be made available to authorized persons. For this reason it is not possible to include copies of any such instruments in this thesis. The relevant tests, manual, computer results, etc. are, however, available for perusal at the offices of the Human Sciences Research Council, subject to the approval of the President.

Further, all opinions expressed in this thesis are those of the author and will not necessarily coincide with those of the Human Sciences Research Council.

I hereby declare that this is my own original work.

S. Oosthuizen

CONTENTS

Page

		Page

FIGURES

TABLES

TABLE		Page

TABLE		Page

CHAPTER 1

INTRODUCTION, AIMS AND PRESENT POSITION OF APTITUDE TESTING WITH REGARD TO THE SOUTH AFRICAN INDIAN POPULATION

1.1 INTRODUCTION

1.1.1 General

It would probably surprise many a South African if the statement were made that the Indians in South Africa are racially and culturally a very heterogeneous group. Naidoo (1947, p.582) indicates the language differences, amongst other things, when he remarks: "These five languages; Tamil, Telegu, Hindustani and Urdu together with Gujarati were the tongues used by the Indian immigrants to Natal". Logue (1954, p.11) also indicates the position by saying: "The Indian community with its Tamil, Telegu, Hindi and Moslem-gujarati sections presents an array of people who differ not only in language and social custom but also perhaps even in racial origin". It must be mentioned, however, that this situation did not remain unchanged with the passing of time.

During the past decade or so the Indian community of South Africa has undergone a gradual change with regard to language, religion and so forth. As far as language is concerned, the position already exists that English is the medium of instruction at all Indian schools. In addition several Indian families have changed their home language to English because of the training received by their children at school and owing to environmental conditions. According to a survey made by the "Media and Communications Research, Pty., Ltd., Johannesburg" (Editorial article, Fiat Lux, May 1970, p.9): "English is the home language in nearly half of the households, with Tamil and then Hindi spoken in the bulk of the other half". Although many of the older members of Indian families to-day still speak the different dialects, they are rarely used in the written language.

Also in other domains the Indian people have gradually been westernized. For instance, whereas, the females of the eastern cultures play a minor role, the position is not exactly the same in South Africa, although this was the position initially. So, for instance, the number of Indian boys that attended school was initially much larger than the number of girls, since the view was taken that a girl's place is at home. This situation gradually changed and nowadays it is not only found that almost equal numbers of the two sexes attend school, but also that there is not much of a difference in the achievement of the two sexes. In the modern community it is not strange to find a large number of Indian girls at colleges or universities who are studying with a view to qualifying for highlevel vocations. From an editorial article in Fiat Lux (June 1968, p.151) it becomes evident that several Indian ladies took part in a 1968 graduation ceremony. In another editorial article of Fiat Lux (August 1970, p.33) the following statement is made: "More and more Indian women in Natal are taking up law, according to a "Natal Mercury" survey. Mrs. Naidoo's admission will bring the number of Indian women lawyers to six this year". The conclusion can thus be drawn that the position of the Indian woman in South Africa with regard to social and educational aspects is gradually changing.

In the domain of religion the rate of change was perhaps less impressive and the different religions are still faithfully adhered to. In the May 1970 edition of Fiat Lux (Editorial article, p.9) for example, the following statement is made: "On the religious side by far most are Hindu (nearly 80 per cent), 12 per cent are Muslim and eight per cent Christian".

In spite of the fact that the Indian in South Africa has gradually been westernized with the passing of time, it can nevertheless be stated that his culture still differs sufficiently from that of the European to justify separate investigation. To support the last argument reference can be made to the typical interests and abilities of the Indian people.

1.1.2 Development in the educational field

The first Indian immigrants arrived in Natal in 1860. Only eight years after the first Indians had arrived in *Natal*, attention was given to the education of children by awarding an allowance to Reverend Ralph Stott with a view to beginning an evening school for older children. In 1869 this allowance was extended so that the younger children could receive training during the day.

In 1878 the Indian Immigrant School Board was instituted with a view to providing assistance at Indian schools and to control the erection of Indian schools. In 1894 the educational affairs were delegated to a Department of Indian Education and in 1910 the administration of education in the provinces was delegated to the Provincial Councils. Until about 1940 the policy was that the Indian community would be responsible for the provision of buildings, whilst the Education Department would be responsible for the salaries of the teachers. At a later stage, however, the Provincial Administration contributed on a £ for £ basis to the provision of buildings (Ramphal, 1960, p.10).

In order to allow a maximum number of pupils to attend the schools the so-called "platoon"- or afternoon schools were later established. This implied that one group of pupils and teachers would use the buildings in the morning and another group would use them in the afternoon.

In the field of secondary education the Indian community also took the lead in Natal. During 1907 a school was begun in Pietermaritzburg and during 1911 one was started in Durban. Sastri College was opened in 1930 (Naude, 1960, p.215) and provided for secondary education as well as for the training of teachers. As a result of these developments there were already 554 Indian pupils at secondary schools during 1940

(Naude, 1960, p.216). Since school attendance was not compulsory for Indian pupils, however, the attendance was in general rather poor.

In *Transvaal* the education of Indian pupils was catered for by the Transvaal Education Department. The number of pupils attending schools also rose rapidly in this province. In 1931 there already existed 15 schools which accommodated 1637 pupils (Naude, 1960, p.245). Here again school attendance was not compulsory.

During 1965 a law was brought into force which provided for the control of education by the Department of Indian Education and on 1st April 1965, the first director of Indian Education was appointed. On 1st April 1966, the Department took over control of Indian Education in Natal and on 1st April 1967, of Indian Education in Transvaal.

The Department of Indian Education promptly gave attention to the need which arose for a Section for Psychological Services. The necessity for these services mainly arose as a result of the rapid increase in the number of Indian pupils. In 1927, for example, there were 9766 Indian pupils at school in Natal and this figure had increased by 81,4 per cent in 1933. In 1943 the number of pupils had increased by a further 74 per cent and in 1967 the number of pupils had risen to 133,817 according to figures of Logue (1954, p.6) and information from the Department of Indian Education.

During 1965 a full-time psychologist was appointed in the Section for Psychological Services and during 1966 a psychometrician was also appointed. The aims of the Section for Psychological Services can briefly be summarized as follows (Joubert, 1966, p.90):

(i) to determine each pupil's intelligence at least once during his school career;
(ii) to apply diagnostic tests;
(iii) to trace and assist mentally retarded pupils;
(iv) to treat problem cases;
(v) *to provide for a system of vocational guidance.*

It is especially the last aspect which is related to this project.

1.2 AIMS

From the foregoing discussion it can be gathered that the need for vocational guidance was associated with the development of the Indian in the educational field. After the Section for Psychological Services had been established it did not take long before the necessity for a vocational guidance program was realised. For the institution of vocational guidance, however, several psychological instruments are required. Joubert (1966, p.91) outlines the position as follows: "Standardised tests will be used to determine the ability and aptitudes of the individual pupils and an interest survey will help to determine the preference of the pupils. On the basis of this knowledge the pupils will be guided to choose a vocation in accordance with their interests and aptitudes". It is thus clear that one of the instruments required in a vocational guidance situation is an aptitude test battery. The production of an objective, reliable and valid test battery, which would be used mainly for vocational guidance purposes, was in fact the main function of this project.

The battery would be classified as a general aptitude test battery since it would serve more than one purpose. Although it would mainly be used for vocational guidance it could also be of use for guidance with regard to the choice of school subjects. The battery would be standardized for Standards 6 to 8 since the greatest need for such a battery existed in this range.

Since an aptitude test battery is used in conjunction with other instruments in a vocational guidance situation, special attention was also given to the sinchronization of the different instruments as far as possible. In practice for instance the counsellor often finds himself in a situation where he must make predictions on the basis of the results obtained on one instrument only, for example an interest questionnaire. Typical examples are the predictions which are made with regard to artistic- and mechanical vocations. In addition an attempt was also made to solve several other problems which confront the test constructor and counsellor.

The aims of this project may briefly be summarized as follows:

(i) to provide an objective, reliable and valid aptitude test battery which would mainly be used for vocational guidance in Indian schools;

(ii) to indicate the problems which are associated with the standardization process;

(iii) to indicate the problems associated with the guidance process, with special emphasis on the integration of the different instruments used in the guidance program and on the interpretation of test results.

In standardizing a test battery there are, of course, not only one or two hypotheses which are tested as the case usually is when conducting an experiment. The standardization of a test battery includes the investigation of a whole series of hypotheses and it would be impracticable to specify all these hypotheses here. As an illustration a few of the null hypotheses are supplied, however, namely: the achievement of Indian boys and Indian girls is the same in the different fields of aptitude; the factor structures for different standards are the same; the relation between achievement in the test battery and in school subjects is insignificant. However, the main hypotheses can be related to the aims which have already been referred to.

1.3 PRESENT POSITION OF APTITUDE TESTING WITH REGARD TO THE SOUTH AFRICAN INDIAN POPULATION

Although the Indians in South Africa have been living here for more than a century, relatively few psychological tests have been standardized for them until now. One of the first persons who made a contribution in this field was presumably Logue (M.Ed.

dissertation, 1954) who standardized an intelligence test for Indian pupils in Durban. Most of the other work in this field has been done by the National Bureau for Educational and Social Research (whose functions are now incorporated with the Human Sciences Research Council), and the Human Sciences Research Council. At the request of the Department of Indian Education several tests and questionnaires have been standardized by these organizations, amongst which are intelligence tests, proficiency tests and an interest questionnaire.

The intelligence tests are being used to indicate the general intellectual ability of the pupils, and to distinguish between normal and retarded pupils. They will in future probably also be used for streaming purposes. The proficiency tests will be used in conjunction with the intelligence tests to give an indication of scholastic ability. The interest questionnaire referred to is mainly being used for vocational guidance but can also be used for scholastic guidance and for purposes of adjustment. At present attention is being given to aptitude testing (standardization), personality testing (item analysis stage) and maturity testing (item analysis stage) of Indian pupils by the relevant section of the Human Sciences Research Council.

1.4 SUMMARY

Certain background information with regard to the nature of the Indian population and its development, especially in the educational field, is first presented in this chapter. Thereafter it is indicated how the need arose for the establishment of a Department of Indian Education and subsequently the institution of a Section for Psychological Services. In the light of the functions of the Section for Psychological Services and in the light of the rapid increase in the Indian population the need for psychological testing of Indian pupils is indicated.

One of the functions of the Section for Psychological Services is the provision of vocational guidance and one of the main instruments needed in the guidance process is an aptitude test battery. The provision of the latter instrument is also the main function of this investigation. It can be said that an instrument of this nature is required for the adequate use of available manpower and to direct the developing manpower into the right channels at an early stage. It is thus clear that a definite need exists for the use of an aptitude test battery both in a vocational guidance setting and in an educational guidance setting. Finally the present status of aptitude testing with regard to the Indian population is depicted. In this last discussion it is clearly indicated that no aptitude tests have yet been standardized for Indian South Africans.

CHAPTER 2

THE PSYCHOLOGY OF APTITUDE TESTING

2.1 THE NATURE OF APTITUDE

2.1.1 General

In the study of literature concerning aptitude it becomes clear that there is no uniformity with regard to the circumscription or definition of the concept "aptitude". Each person's circumscription corresponds to his view, which again corresponds to different periods in history. It can be asserted, however, that in general, aptitude is regarded as an aspect of personality which indicates potential ability in different intellectual fields.

The measurement of aptitude is closely related to the measurement of intelligence - to such an extent that the two concepts are regarded as synonymous by some writers. The assertion can be made, however, that the measurement of intelligence received attention long before the measurement of aptitude and that even to-day there is a fairly clear-cut line of division although intellectual factors are common to both aspects.

2.1.2 Different definitions or circumscriptions of aptitude

The following are some of the definitions or circumscriptions of aptitude which are found in literature:

2.1.2.1 Bingham (1937)

Just as Strong and Kuder are regarded as pioneers in the field of interest measurement, Bingham is regarded as one of the pioneers in the field of aptitude measurement. Bingham (1937, p.16) defines aptitude as "..... a condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with training some (usually specified) knowledge, skill or set of responses such as the ability to speak a language, to produce music, etc".

Gekoski (1964, p.42) criticizes Bingham's definition because it does not circumscribe aptitude scientifically enough and other people criticize his definition because they say it is not clear from the definition what the contribution of heredity and environment is towards aptitude.

2.1.2.2 <u>Remmers and Gage (1955)</u>

These investigators see aptitudes as: "... present traits considered as predictors of future achievements" (1955, p.218).

2.1.2.3 <u>English and English (1958)</u>

These writers (1958, p.39) define aptitude as "the capacity to acquire proficiency with a given amount of training, formal or informal". According to this definition aptitude is thus the ability to reach a certain level of proficiency when the necessary training is given.

2.1.2.4 <u>Guilford (1959)</u>

Guilford regards aptitude as composed of certain dimensions. According to him aptitudes are primary abilities appearing in the areas perceptual, psychomotor and intellectual. In 1956 he made a first attempt at systematizing the different intellectual traits. He depicts the organization of mental traits in three dimensions. Along one axis of his model (Figure 2.1.2.4) the basic processes performed by the mind are presented - the operations; along the second axis are the different contents which are employed for the operations and along the third axis are found the different products, which can result when the operations are executed with one or more of the contents. Altogether his model then consists of 120 factors. A graphic representation of his model is given below (Smith and Adams, 1966, p.249).

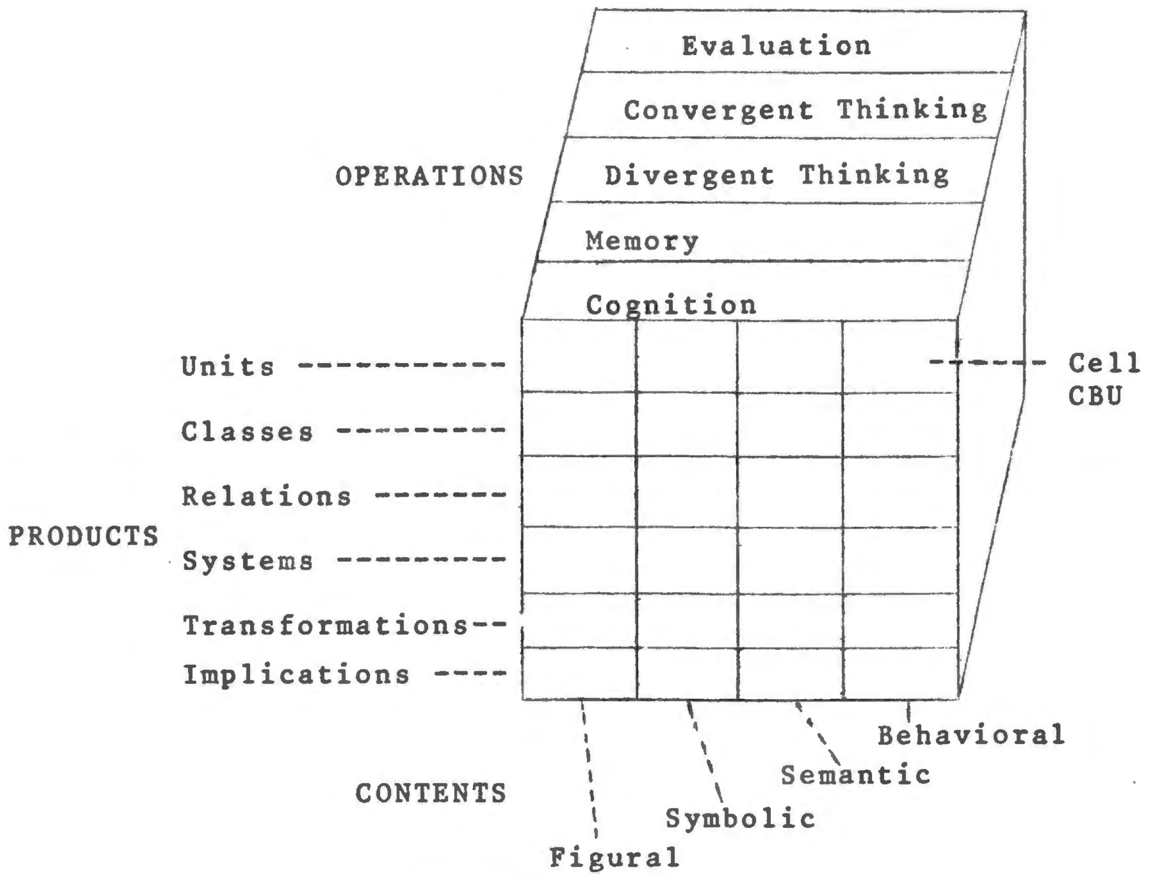


FIGURE 2.1.2.4 GUILFORD'S MODEL FOR REPRESENTING THE STRUCTURE OF INTELLECT

We get a very exact presentation of aptitude here, which reminds the author of the periodic table which is encountered in chemistry. By chance this analogy was also made by Helmstadter. He remarked (1964, p.111): "While admittedly a rough, first attempt, this classification scheme, somewhat analogous to the periodic table in chemistry, has provided for meaningful classification of the forty odd factors already developed". Whether the social sciences can be moulded in the same shape as the natural sciences is of course a debatable point.

2.1.2.5 Lyman (1963)

Lyman (1963, p.194) regards aptitude as ".... that combination of characteristics, both native and acquired, which indicate the capacity of a person to develop proficiency in some skill or subject matter after relevant training".

2.1.2.6 Gekoski (1964)

Gekoski (1964, p.41) regards aptitude as a central concept and sees all other aspects of personality as subordinate to aptitude. He remarks: "Whatever it is, be it a condition, a state of affairs - it is something which is associated with, gives rise to behavior, a composite of which may be described as successful performance in a job or pursuit". With his approach he overcomes the problem of giving class names to concepts like introversion-extroversion, sociability, etc., but he brings about a radical change with regard to the modern view of personality. He provides the following graphic presentation as an illustration of his view (1964, p.41).

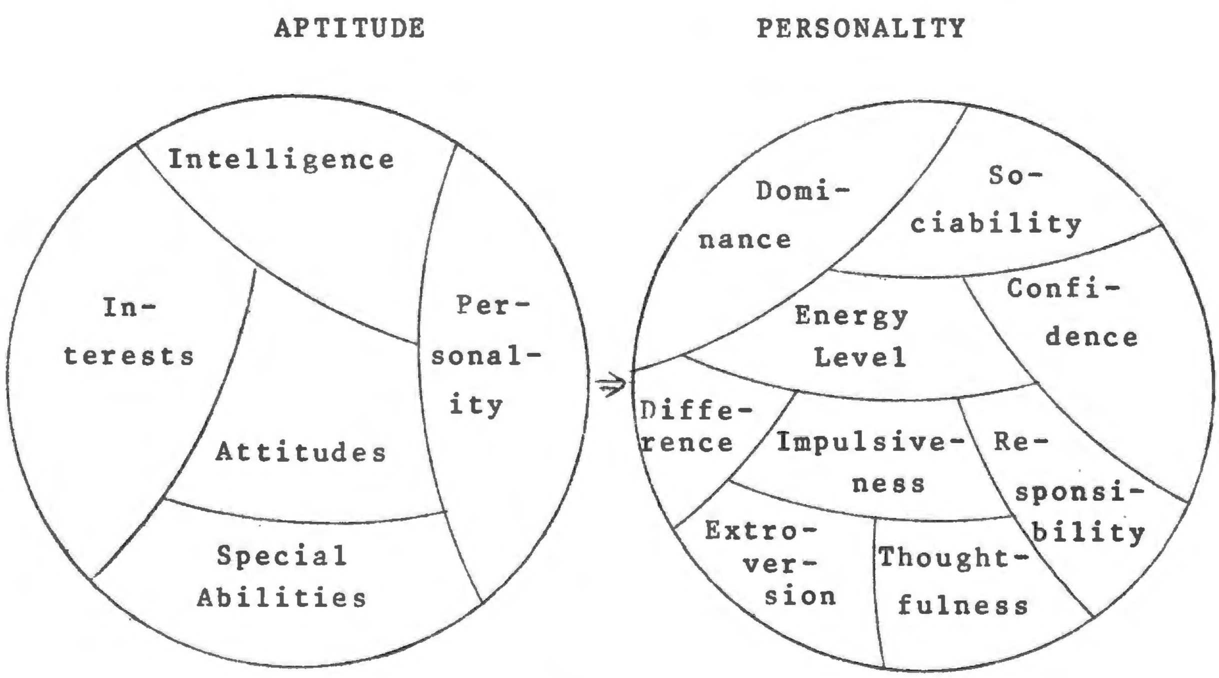


FIGURE 2.1.2.6 THE SUBDIVISION OF APTITUDE AND PERSONALITY ACCORDING TO GEKOSKI

2.1.2.7 Super and Crites (1965)

These investigators criticize the definitions of Bingham and English. According to them (1965, p.71) a requirement for the scientific study of aptitude or any other entity is "... that one be able to name it, describe it and locate it in a variety of individuals and situations. This means that it must be relatively constant in its nature and composition". According to Super and Crites a scientific definition of aptitude should include:

(i) specificness;

(ii) unitary composition;

(iii) potentiality for the learning of an activity or type of activity and

(iv) relative constancy.

The latter aspect will be of special importance for the prediction of success. It is clear that the approach of these writers largely corresponds with that of Guilford, and we also clearly observe the stamp of the factor analyst here.

From the preceding discussions it appears that there are actually two approaches to the concept of aptitude. One approach is presented by Bingham and Lyman who regard it as a system of characteristics, and the other approach by Super and Crites, and Guilford, who regard aptitudes as primary unitary dimensions. Super and Crites (1965, p.71) admit, however, that the approach of Bingham is not necessarily wrong: "In practice the requirement of unitary nature is frequently disregarded without ill effects".

In summary it can be stated that investigators like Super and Crites strive at factorial pureness, while this is not the case with the other investigators. Although a presentation like that of Guilford theoretically appears acceptable, it must be borne in mind that measurement in the social sciences is not of such an exact nature as measurement in the natural sciences. We are, for instance, aware of the influence of factors like age, standard, sex, motivation, socio-economic background, etc., in psychological measurement. We thus cannot always rely on linear relations and uniqueness in the social sciences. On the other hand, we can in the natural sciences, always predict the result of the reaction of one chemical compound on another, regardless of the place and conditions under which the experiment is conducted.

For the purpose of this investigation aptitude will be regarded as: (1) a personality aspect; (2) potential intellectual ability determined both by hereditary and environmental conditions and which enables a person to obtain a certain level of proficiency when the necessary training is received. The hereditary aspect is stressed, but the existence of environmental influences is not denied. A pure factor structure was not assumed in the theory of aptitude either.

2.2 THE MEASUREMENT OF APTITUDE

2.2.1 The meaning of the concept "aptitude test"

According to Freeman (1955, p.306) an aptitude test is ".... a device designed to indicate a person's potential ability for performance of a specialized kind and within a restricted range".

English and English (1958, p.40) describe an aptitude test as a set of tasks so chosen and standardized that they yield an estimate of a person's future performance on other tasks not necessarily having evident similarity to the test tasks". This description is closely related to the typical use of aptitude tests in test construction.

Finally Lindquist (1963, p.456) describes an aptitude test as ".... a device for measuring the capacity or potentiality of an individual for a particular kind of behaviour".

From the above descriptions it is clear that "predictive value" is the most distinguishing feature of an aptitude test. With an aptitude test we thus want to determine whether a person possesses the ability to complete a certain task successfully if he receives the necessary training.

2.2.2 Different kinds of aptitude tests

If aptitude tests are classified according to structure they can be subdivided into specific and multiple aptitude tests. Specific aptitude tests (also called special aptitude tests) predict for one sphere only, whereas multiple aptitude tests predict for different spheres. There are, of course, also other means of subdividing aptitude tests.

2.2.2.1 Specific aptitude tests

Before the second World War specificity of measurement was less known than after the war. So, for instance, it is known that intelligence was originally regarded as consisting of a g-factor, but was later divided into sub-factors (e.g. a verbal and non-verbal section). This type of extension gradually also came into being with aptitude tests. Originally specific aptitude tests like clerical aptitude tests and mechanical aptitude tests were separately constructed, but later a whole group of such tests formed part of a multiple aptitude test battery. Until to-day, however, both specific and multiple aptitude tests are found since they serve different functions.

When tests are presented in the specific form they are usually more comprehensive for the specific spheres than are tests of a multiple aptitude test battery covering the same spheres. Specific aptitude tests are also usually standardized for different groups. Some of the well-known specific aptitude tests which are related to the fields Clerical, Mechanical and Artistic are the following:

Clerical

(i) Benge's Clerical Aptitude Test (Benge, 1950, p.109)
(ii) Minnesota Clerical Test (Anastasi, 1969, p.363)
(iii) Short Tests of Clerical Ability (Anastasi, 1969, p.643)

Mechanical

(iv) Bennett Mechanical Comprehension Test (Anastasi, 1969, p.362)
(v) MacQuarrie Test for Mechanical Ability (Anastasi, 1969, p.643)
(vi) S R A Mechanical Aptitude Test (Anastasi, 1969, p.643)

Artistic

(vii) McAdory Art Test (Hahn and McLean, 1955, p.194)
(viii) Meier Art Test (Mortensen and Schmuller, 1960, p.196)
(ix) Lewerenz's Tests in Fundamental Abilities in Visual Arts (Anastasi, 1969, p.643)

2.2.2.2 Multiple Aptitude Tests

Multiple aptitude tests in contrast with intelligence tests for instance, follow a differential approach to the measurement of ability. Anastasi (1955, p.351) remarks in this connection: "Such instruments yield, not a single overall measure such as an I.Q. but a set of scores in different aptitudes. They thus provide a "psychograph" or intellectual profile showing the individual's characteristic strengths and weaknesses".

In contrast with specific aptitude tests the sphere of prediction is naturally much wider here and a multiple aptitude test battery can be used to give a picture of ability with regard to vocational- or scholastic fields. There is, of course, no reason why specific aptitude tests cannot be used as supplement to multiple aptitude test batteries especially if the latter cover a limited domain. The advantage of multiple aptitude tests is that they have been standardized for one group, whereas specific aptitude tests have usually been standardized for specific groups.

Some of the best known multiple aptitude tests, which have been used successfully in the past for the measurement of differential aptitudes are the following:

(i) Differential Aptitude Tests (DAT): (Shertzer and Peters, 1965, p.148)
(ii) Flanagan Aptitude Classification Tests (FACT): (Shertzer and Peters, 1965, p.146)
(iii) General Aptitude Test Battery (GATB) : (Shertzer and Peters, 1965, p.147)
(iv) Multiple Aptitude Tests (MAT): (Shertzer and Peters, 1965, p.147)

2.2.3 Objections to and criticism of specific and multiple aptitude tests

2.2.3.1 Objections to and criticism of specific aptitude tests

When too much attention is given to theoretical aspects and too little to practical aspects, the situation often arises that tests are criticised on theoretical grounds. This is exactly what happened with regard to the use and construction of specific aptitude tests. The measurement of artistic and musical aptitudes, for instance, is often criticised since these spheres are still relatively underdeveloped. According to Cronbach (1949, p.228) the reliability and validity of these types of tests is " ... far too low to permit *final* judgement of talent from test scores". Surely the existing tests must have a certain degree of practical value, however, and can always be improved.

Another objection to many of the specific aptitude tests is that they do not produce pure factors. Especially investigators like Guilford (as in Jackson and Messick, 1967, p.837) and Cattell (as in Jackson and Messick, 1967, p.837) are very particular about this point. Their view also becomes clear when their models and psychological tests are studied. Because of their factorial inclination, however, many investigators are forced to exclude a mechanical aptitude test from their test battery, since mechanical aptitude is usually decomposed into a reasoning factor and a spatial perception factor. The latter is, however, no valid reason for the rejection of this type of test, especially if it is borne in mind that the number of vocations bearing a relation to mechanics is very large. To illustrate this we need only think of all the different trades.

2.2.3.2 <u>Objections to and criticism of multiple aptitude tests</u>

In all the fields of psychological measurement one finds master critics. Bauernfeind (1963, p.192) is one of these critics and he mentions the following objections to multiple aptitude tests:

(i) Each time a new factor is identified there's a tendency to give it a certain "label", e.g. "verbal ability", "numerical ability". The problem is that different authors attach different *labels* to the same types of tests and this causes chaos, according to Bauernfeind.

The abovementioned objection can only be partially accepted. Although different writers attach different labels to the same type of test this is not necessarily wrong, nor does it necessarily cause chaos, since the attachment of a name to a factor is directly dependent on the contents of the test and the purpose it will serve. Thus one test with a numerical format may yield a numerical factor only and have preditive value for vocations like teller, clerk, etc. Another test with a numerical format, however, may yield a numerical factor and a reasoning factor, so that it can also be a predictor of general school achievement. To distinguish the two factors as well as the two tests from each other, they will have to be "labelled" differently. In other words, the main aspect here is that the test administrator must know his tests thoroughly, irrespective of the "labels" which are attached to them.

(ii) Although we are intent on dividing the human intellect into sections, our criteria do not always divide accordingly. After Thurstone's early contribution to factor analysis it was hoped that the division of the human intellect into independent components would be of value for the prediction of different types of criteria, so that verbal ability should predict achievement in reading and English; numerical ability should predict achievement in mathematics; reasoning ability should predict achievement in science, and spatial ability should predict achievement in geometry and trigonometry. Unfortunately these relations were not always found in practice. For example, Bauernfeind (1963, p.192) remarks: "In many studies, number ability scores provide the best predictions of grades in foreign languages (though in other studies they do not); verbal ability scores best predict grades in geometry (though not in other studies); and language-arts scores yield the most accurate predictions of success in mathematics (though in other studies they do not)". His conclusion is then: "It is clear that our criteria are often highly complex and that single measures of a unitary talent often do not predict the criteria well, despite the logical assumption that they might".

The second objection mentioned by Bauernfeind is valid to a large extent, but Bauernfeind himself admitted that the findings were not consistent. Several investigators also supplied logical explanations for apparent contradictions of this nature. Goldman (1962, p.328), for instance, tried to indicate why the correlations between certain variables were not higher. It must also be remembered that aptitudes and other constructs are usually more differentiated (i.e. less complex) with the lower standards than with the higher standards.

Apart from the abovementioned points of criticism raised by Bauernfeind (but actually adopted from Lamke), he also asserted that achievement tests are the best predictors of school success, and that it is doubtful whether multiple aptitude tests have much value for predicting vocational success. In contrast with this sharp criticism, however, we have enough proof of the success that has been obtained with well-known aptitude tests like the DAT, the Multi Aptitude Tests and the GATB. With regard to the GATB, for instance, we know that 22 vocational profiles representative of 500 vocations have been identified (Tyler, 1963, p.63). It is obvious that an achievement test cannot supply this type of information. It must also be borne in mind that an achievement test cannot determine *potential* ability like an aptitude test. In other words, before a certain amount of learning material has not been treated, evaluation or prediction cannot be made.

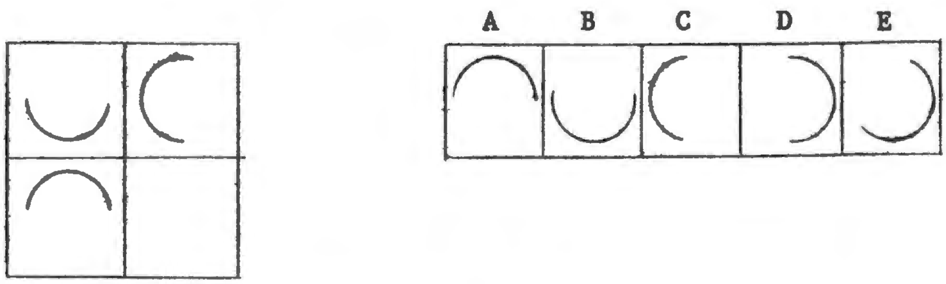
2.2.4 The construction of aptitude tests in general

The *first problem* encountered by the constructor of aptitude tests (and for that matter of any tests) is the choice of the basis of the tests. A group of tests can have a factorial basis, an empirical basis or a logical basis, to mention the most important bases. Whereas the last two bases are more related to practical value, the first is more related to theoretical value, and should be approached with due consideration of certain requirements.

Regardless of the basis of the tests the *next problem* encountered by the test constructor is usually the selection of the most suitable areas that have to be covered or represented by the tests. Here again different methods can be employed. Previous statistical results like those obtained from factor analyses can be used, or job classifications or job analyses can be studied, or the areas of existing tests having practical value can be compared.

Although there remain many other problems which are encountered in constructing aptitude tests, as e.g. differences in the results of the two sexes, and choice of appropriate statistical techniques, the next vital problem according to the writer is the avoidance of ambiguous items. This naturally also applies to other types of psychological tests. The ambiguity of certain items was pointed out by Gronlund (1965, p.152), amongst others. That considerable care should be taken to avoid ambiguity when constructing aptitude tests is illustrated by the following two items:

<u>Item 1</u>:



This is a pattern completion item and the item clearly has two answers, namely C or D, depending on whether the principle of mirror image or rotation is employed.

<u>Item 2</u>:

Which of the following pairs of words does not go with the others?

A	B	C	D	E
fear	blow	hard	light	sit
smear	slow	fast	night	pet

It is clear that either A, C or E can be taken as an answer depending on whether the number of letters in each member of a pair is stressed, or whether the last letters of members in a pair are considered or whether the rhyming aspect of a pair is considered.

Instead of constructing a test from scratch, an existing instrument can also be adapted, but this does not rule out all the above problems.

2.3. THE RELATION OF APTITUDE TESTS TO OTHER PERSONALITY TESTS AND QUESTIONNAIRES

It is very important to distinguish or to try to distinguish between tests like aptitude tests, achievement tests, interest questionnaires, etc. Aptitude tests attempt to determine the potentialities of a person, and are used to make predictions about prospective learning activities and behavioural patterns. Stress is placed on prediction and the future in this case. Aptitude is usually divided into two categories, namely scholastic aptitude and vocational aptitude. The former is actually very difficult to distinguish from the so-called intelligence tests, whereas the latter is clearly outlined. Aptitude tests must not be confused with other types of tests however. Subsequently a few other types of psychological tests will be discussed, and afterwards a number of comparisons will be drawn.

2.3.1 Achievement tests

According to Shertzer and Peters (1965, p.127) these tests measure "... the extent to which individuals have acquired certain knowledge, skills, concepts and data as a result of instruction. They measure the outcomes of instruction in selected areas, for example reading, spelling, language and computing". Training and acquired knowledge up to a certain level thus play the most important role here, and also the fact that an evaluation is made with regard to the amount of knowledge acquired in the *past*.

2.3.2 Interest Questionnaires

These questionnaires measure the interest of the individuals in a number of the most familiar interest fields. This is usually done by determining the responses on a number of activities which can be regarded as representative of these fields. Strong (1965, p.1) indicates the relative position of interest in psychological measurement very clearly when he makes the following comparison: "The relationship between interests and other types of mental measurement may very roughly be expressed in terms of a motor-boat. A motor-boat needs an engine, fuel, a rudder and a seaworthy hull. Without all four one will not get the boat across a lake. The engine is analogous to ability, the fuel to motivation, the rudder to interests, the good or leaky hull to character". From this description it is clear that interest is fairly clearly outlined and forms an entity.

2.3.3 Intelligence tests

These tests are used for predicting general school achievement and for diagnostic purposes. The future is thus also stressed in this case.

With these tests the environmental influences are usually accentuated, but the influence of heredity is not denied. These tests are also often related to learning curves.

2.3.4 Difference between aptitude tests and achievement tests

The most well-known difference between these two types of tests has been indicated already, namely that aptitude tests are designed for making predictions, while achievement tests are designed for measuring what the individual has learned in the past. Achievement tests measure actual learning, whereas aptitude tests measure potentiality of learning. In brief it can be said that, with achievement tests it is assumed that the examinees have had uniform learning experiences, but this is not a supposition with aptitude tests.

Tyler depicts the position more or less in the same manner as above by remarking (1963, p.56): "More useful at present than the distinction based on research history is the distinction based on purpose. If we plan to use a test chiefly as a predictor of how well individuals will perform in some area, we can consider it an aptitude test. If we plan to use it mainly to evaluate an individual's accomplishments or the adequacy of his education and experience it is for our purposes an achievement test".

2.3.5 Difference between aptitude tests and intelligence tests

The distinction between these two types of tests has become dimmer and dimmer owing to developments in the sphere of factor analysis. To illustrate this statement one need only think of Guilford's SI model, consisting of 120 factors, which is used to represent the different mental factors.

In a certain sense intelligence tests are less extensive or comprehensive than aptitude tests and more stress is placed on environmental conditions. Shertzer and Peters (1965, p.126) remark in this connection: "Most of the mental ability tests used in the schools emphasize verbal comprehension, numerical comprehension and reasoning factors. Some tests yield a single composite score, but the trend is toward tests that provide part scores - verbal and nonverbal or verbal and quantitative - which can be combined to make up a *total score*. Intelligence tests do not measure innate ability, but rather the extent to which human potential has been modified by environment".

Because of the structure of most intelligence tests the predictions made by means of intelligence tests are usually less specific than is the case with aptitude tests, where predictions are made for specific vocations or subjects. As mentioned previously, one of the well-known functions of intelligence tests is to predict general success in school work.

2.4 APTITUDE TESTS AND THE INDIAN POPULATION OF SOUTH AFRICA

In spite of the fact that the Indians in South Africa are a minority group it cannot be said that they make no attempts at progress. In the light of developments up to now it can also be claimed that the available manpower should benefit by the administration and use of aptitude tests and other psychological tests.

Owing to certain inherent characteristics it is reasonable to assume that the Indians will perform better in certain fields than in others. For example, it is well-known that the Indians are keen businessmen and excellent technicians. The objection may be raised that there is a lack of sufficient vocations and facilities in certain spheres. This could then imply that the particular psychological tests would have limited value. There seems to be enough proof, however, that the Indians are busy with selfdevelopment in most of the general vocational fields. So, for instance, there is proof that the Indians are performing well in the clerical and scientific (especially chemical) fields (Editorial article, Fiat Lux, March 1970, pp. 10-11). Their participation in social activities, commercial activities and mechanical activities can also be mentioned (Van der Spuy, 1968, p.26; Editorial article, Fiat Lux, May 1970).

2.5 MOTIVATION FOR THE CONSTRUCTION OF AN ORIGINAL APTITUDE TEST BATTERY FOR INDIAN PUPILS

The easiest way to provide a new test battery is to adapt an existing test battery. Owing to different reasons this is not always the best method, and can cause several problems when the two cultural groups which are involved, differ radically from each other. One of the reasons referred to is that when a battery is adapted one is dependent on the existing tests or fields of the particular battery. Another reason is that limited scope is usually allowed for alterations by the author of the original battery. Thus, there is a big difference between changing the norms of a battery only, and drastically changing certain items of the battery.

Since none of the existing aptitude test batteries satisfied certain requirements set forth by the author, it was decided to start a new battery from scratch and not to adapt an existing battery in any case. The requirements referred to are the following:

(i) An aptitude test battery which is used mainly for vocational guidance, should contain tests which can be related to other spheres of psychological measurement. This is regarded as a vital defect of most existing aptitude test batteries. Fields which cause problems in this relation are the artistic and mechanical fields. The mechanical field is found in certain test batteries, but then the items are usually of a two-choice or three-choice type as is the case in the DAT and the Multiple Aptitude Tests. In addition these tests are usually limited to items on leverage and forces. As far as the artistic field is concerned, tests relating to this field are only found as single tests (special tests) and not as part of a test battery.

(ii) The general difficulty level of the battery must be suitable for the cultural group concerned. However, it is known that especially certain fields cause problems in this connection. Two of these fields are the Mechanical Field and the 3-D Field (see Manual of the Multi-Aptitude Test, 1955, p.29; Technical Report of the HSRC, Owen, 1971, p.39).

(iii) The internal structure of the battery should be balanced so that it has predictive value for a wide spectrum. Most test batteries, however, attach more weight to certain fields than to others. So for instance the DAT (as in Cronbach, 1970, p.355) attaches considerable value to the measurement of language ability, whereas the GATB (as in Cronbach, 1970, p.358) attaches considerable value to the measurement of co-ordination and dexterity.

(iv) It is desirable that all the tests of a test battery consist of the same number of items, otherwise it will be difficult to determine whether the variation in test scores is due to differences in test lengths, or to actual differences in ability.

Two test batteries which satisfy the above requirements to a certain extent are the DAT and the Multiple Aptitude

Tests. It is doubtful, however, whether any of these batteries would be suitable for Indian pupils at the Standard 6 or 7 level, especially with regard to item difficulty level and to language difficulty level. It can be foreseen, for instance, that Indian South Africans for whom English is a second or even a third language, will find American tests rather difficult, because of the standard of English employed in these tests.

2.6 VIEWS HELD CONCERNING THE CONSTRUCTION OF THE JUNIOR APTITUDE TESTS FOR INDIAN SOUTH AFRICANS

In constructing the aptitude test battery for Indian pupils, hereafter called the JATISA *(Junior Aptitude Tests for Indian South Africans)*, the view was taken *first* of all, that a general factor does not form the basis of all intellectual abilities. This is in agreement with the later developments in the domain of factor analysis, and justifies the composition of a multiple aptitude test battery. For the determination of the abilities which are important for most of the general vocations, use was made of vocational classifications and descriptions, as well as factor analysis results. In addition direct comparisons were drawn between the tests of a number of well-known multiple aptitude test batteries.

Secondly, the view was taken that aptitude and other personality attributes should lend themselves to integration. So, for instance, aptitude test results and interest questionnaire results should be readily relatable. This shortcoming is regarded as common to most of the well-known test batteries. In Chapter I (p.6) reference was made to this problem by referring to the fields Mechanical and Artistic. It was also kept in mind that pairing of similar fields is not always possible. For instance, an interest field which is difficult to relate to a similar aptitude field is the scientific field. Some investigators combine a scientific and a mechanical test in their test batteries as occurs in the Multiple Aptitude Tests for instance (Manual, 1955, p.3), but an aptitude test solely for the measurement of scientific ability is seldom, if ever, found in practice. The possibility exists that scientific ability may consist of several components, in the same way as mechanical ability consists of more than one component, so that several tests may be necessary for predictive purposes. A recent development in this field is the finding that creative ability is closely related to scientific ability. Anastasi (as quoted by Ohlsen, 1964, p.300) remarks in this connection: "Thus creativity long regarded as the prime quality in artistic production is coming more and more to be recognized as a basis for scientific achievement as well". One possible explanation for this finding is that we may be dealing with a factor common to arts and science, namely "creating" or "experimenting".

Thirdly the view was taken that an aptitude test battery must not rely on factorial pureness for its existence and composition. The reason for this view is that a purer factor structure will not necessarily improve the predictive validity of a test, and a purer factor structure may cause a decrease in the predictive range of the test battery. With regard to the first aspect a test like the Mechanical Insight Test which is composed of more than one factor, may be excluded from a battery based on factorial pureness (Anastasi, 1969, p.336). It is then also clear why tests of this nature are not included with the Holzinger Crowder Unifactor Tests or the Guilford Aptitude Survey (as in Bauernfeind, 1963, p.202). It must also be remembered that when a test is dependent on factor analysis results for its composition, there are certain requirements which must be fulfilled. Guilford (as quoted by Jackson and Messick, 1967, pp.309-317) mentions a number of conditions under which factor analysis should not be conducted, or the results regarded with considerable caution in any case. The author is of the opinion that very few aptitude tests comply with the requirements referred to here, especially owing to the length of these test batteries. Reference will be made to these requirements again in Chapter 3 under the section dealing with factor analysis.

Finally the view was held that the interpretability of test results should be in a form which is acceptable to the counsellor.

2.7 SUMMARY

In this chapter the nature of aptitude is first described. In discussing the concept "aptitude" different views are presented, ranging from the oldest (those of Bingham and others) to the latest (those of Guilford and others). Several definitions are provided and the viewpoint of the author, with regard to the concept "aptitude" is also stated.

The measurement of aptitude is described next under a number of different headings. The meaning of the concept "aptitude test" is supplied and a distinction is made between specific (or special) aptitude tests and multiple aptitude tests. Several objections against the two categories of tests are also presented. Furthermore a brief discussion is presented of the construction of aptitude tests in general.

The difference between aptitude tests and other types of psychological tests is indicated next and also the need for aptitude testing with regard to the Indian population. Reasons are given why the JATISA was constructed from scratch instead of existing tests being used or modified. There seemed to be sufficient reason for the construction of a completely new test battery. Finally certain presuppositions underlying the construction of the JATISA are discussed.

CHAPTER 3

CONSTRUCTION OF THE ORIGINAL BATTERY AND APPLICATION FOR ITEM ANALYSIS

3.1 CONSTRUCTION OF THE ORIGINAL BATTERY

3.1.1 Selection of the tests in the test battery

There are several methods of determining the contents of a new test battery, largely depending upon a person's viewpoints and aims. In the case of the JATISA a number of the most well-known aptitude test batteries were compared in order to determine which tests should form "core"-tests of an aptitude test battery, and an extensive study was made of literature on aptitude and aptitude testing. Research results proved valuable in this connection.

In Table 3.1.1 the contents of the following test batteries are presented in comparative form:

(i) Differential Aptitude Test Battery (DATB)
(ii) Guilford-Zimmermann Aptitude Survey (GZAS)
(iii) General Aptitude Test Battery (GATB)
(iv) The Multi-Aptitude Test (MAT)
(v) Multiple Aptitude Tests
(vi) Holzinger Crowder Uni-Factor Tests.

On the basis of this table, on the basis of literature study (Anastasi, 1969, pp.360-380; Cronbach, 1965, pp.353-364, etc.), and on account of practical experience gained through vocational guidance and selection, it was decided to construct a preliminary test battery consisting of two forms with twelve tests in each form. The contents of these forms are given on page 32.

TABLE 3.1.1 CONTENTS OF A NUMBER OF WELL-KNOWN TEST BATTERIES

D.A.T.	G.Z.A.S.	G.A.T.B.	The Multi-Apt. Test	M.A.T.	Holzinger Crowder Uni-Factor Tests
1. Verbal Reasoning	General Reasoning	Verbal Aptitude	General Information	-	Reasoning (mixed series, figural changes, etc.)
2. Abstract Reasoning	General Reasoning	Verbal Reasoning	Figure Classification	-	-
3. Numerical Ability	Numerical Operations	Numerical Aptitude	Arithmetic, Figure Series	Arithmetical Reasoning and Computations	Numerical Aptitude
4. Spatial Relations	Spatial Orientation	Form Perception, Spatial Aptitude	Paper Formboard	Spatial Relations (2-D and 3-D)	Spatial Aptitude
5. Mechanical Reasoning	Mechanical Knowledge	-	Mechanical Comprehension	Applied Science and Mechanics	-
6. Clerical Speed and Accuracy	Perceptual Speed	Clerical Perception	Comparison	Routine Clerical Ability	-
7. Language Usage	Verbal Comprehension	Verbal Aptitude	Mixed Letters, Word Recognition, Vocabulary	Word Meaning, Paragraph Meaning, Language Usage	Verbal Aptitude
8. -	-	Motor Co-ordination, Finger Dexterity, Manual Dexterity	-	-	-

	FORM A	FORM B
1.	Verbal Reasoning	Verbal Reasoning
2.	Series Completion-1	Classification-1
3.	Series Completion-2	Classification-2
4.	Computations	Computations
5.	Spatial Perception (2-D)	Spatial Perception (2-D)
6.	Spatial Perception ($3\text{-}D_1$)	Spatial Perception ($3\text{-}D_1$)
7.	Spatial Perception ($3\text{-}D_2$)	Spatial Perception ($3\text{-}D_2$)
8.	Memory	Memory
9.	Clerical Speed and Accuracy	Clerical Speed and Accuracy
10.	Visual Arts	Visual Arts
11.	Word Building	Word Building
12.	Mechanical Insight	Mechanical Insight

Each test contained 25 items so that each form consisted of 25 x 12 = 300 items. In addition each test also contained five practice examples which were supposed to acquaint the pupils with the tests. The answers were of a five-choice type and for the answering of the tests an IBM-answer sheet was used.

3.1.2 Description of the tests in the battery and their function

The following is a brief description of the nine types of tests included in the preliminary test battery and the aptitudes which they were supposed to measure.

(i) Verbal Reasoning

The Verbal Reasoning Test is a reasoning test which consists of verbal material. The measurement of the ability to form verbal concepts and to manipulate them in a logical manner is involved here. Since this test measures the ability to reason inductively and deductively, the higher thought processes are involved. The test is tentatively subdivided into the following categories: analogies, classifications, series, letter codes and logical deductions.

The Verbal Reasoning Test ought to be a reasonable predictor of success in fields where complex verbal concepts play a role as is the case in most academical courses. In conjunction with other tests this test should give an indication of the suitability of a pupil for academic training.

The following is an example of an item analogous to the items appearing in the actual test:

Item:

SNOW is to WHITE as SOOT is to ....

A. PINK.
B. WHITE.
C. BLACK.
D. ORANGE.
E. PURPLE.

Answer: C

(ii) Classification and Series Completion Tests

These tests can also be regarded as non-verbal reasoning tests (Gronlund, 1968, p.234). The supposition is then, that it is not the reasoning process which is non-verbal (although more use may be made of spatial reasoning than verbal reasoning in these tests), but only the contents of the tests. These tests do not contain verbal material but consist of sets of figures. The measurement of the ability to identify relations on the basis of changes which a number of figures undergo, is involved here. The pupil's task is to determine the principle(s) according to which the figures undergo these changes. Only gross changes are involved so that fine perceptual ability is not required.

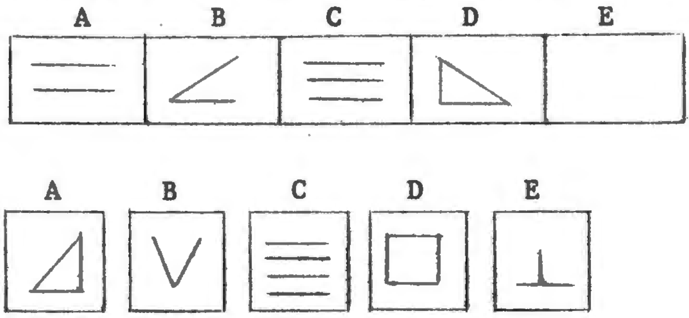
The non-verbal reasoning tests are presented in the form of Classification Tests and in the form of Series Completion Tests. In the Classification Tests the pupils have to classify a set of figures on the basis of a common feature like shape or size and then find a figure with a similar characteristic as the set of given figures amongst the distractors. In the Series Completion Tests a set of figures are presented, which form a series as a result of some determining factor like direction, size, number or position. The missing element of the series must then be indicated by the pupil.

The Non-Verbal Reasoning Test ought to be a reasonable predictor of success in fields where complex non-verbal concepts play a role. It should have predictive value for the academical as well as the technical field.

The following item is analogous to the items appearing in the actual test:

Item:

1. Complete the following series:



Answer: C

(iii) Computations

The Computations Test measures only *basic numerical ability*, since it only measures skill in the use of the four basic arithmetical operations - addition, subtraction, multiplication and division. In conjunction with the other reasoning tests it should give an indication of the general learning ability of a pupil. The computations are in general quite simple and should not be influenced much by school training. The author had certain doubts about the contents of this test and it was decided to change this test after item analysis had been done, if necessary.

Numerical ability should be of importance for the prediction of success in natural science domains like Physics, Chemistry, Mathematics and Astronomy. In addition this ability should be of importance for vocations like bookkeeper, accountant, clerk, statistician and engineer.

An example of an item analogous to the items appearing in the actual test is the following:

<u>Item:</u>

	A	B	C	D	E
1. (24-12) ÷ 0,3 =	3,6	40	4	9	21

Answer: B

(iv) <u>Spatial Perception (Two-dimensional and three-dimensional)</u>

The measurement of the ability to depict lines, figures and objects in the imagination, and to manipulate them is involved here.

(a) <u>Spatial Perception (2-D)</u>

This test measures the ability to make a two-dimensional representation of a figure (that means on a flat surface), and to rotate it in the imagination. All the items in this test refer to two-dimensional figures which may only be turned *around* in the imagination and not *over*. The pupil's task is to determine which one of five figures is an exact representation of a given figure, after rotation of these figures.

This test ought to be important for the prediction of success in fields where shaping and construction play a role, e.g. for professions like those of tailor, sheet metalworker and artist.

(b) <u>Spatial Perception (3-D)</u>

Here two types of tests are encountered. The first type measures the ability to represent a three-dimensional figure from a two-dimensional plan in the imagination and to manipulate it. The plan and figure which can be formed from the plan appear opposite each other, and the pupil must indicate which lines of the figure correspond with which lines of the plan.

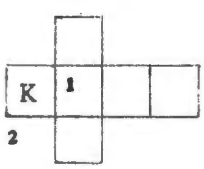
The second type of 3-D test measures the ability to make a three dimensional representation of a figure, that means in space, and to rotate it in the imagination. The pupil's task is to determine which one of five figures is an exact replica of a given figure, after rotation. It was decided that if enough justification could be found, only one of these tests would be retained in the final test battery.

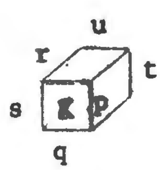
This test ought to be important for the prediction of success in many of the constructive fields like those of engineer, plumber and architect.

An example of two items analogous to those appearing in the actual test is given below:

<u>Items</u>:

Which borderlines of the figure on the right-hand side correspond with the numbered lines of the plan on the left-hand side?





	A	B	C	D	E
1.	t	p	q	r	s
	F	G	H	I	J
2.	q	u	t	r	s

Answers: 1 = B, 2 = F

(v) <u>Memory (for symbols and letters)</u>

This test measures the ability to recall accurately, verbal and non-verbal material which has been placed in the field of perception for a short period. One part of the test relates to the reproduction of certain *code words* which are linked to certain meaningful words. The other part of the test consists of the reproduction of certain *symbols* which are linked to certain verbal concepts.

The memory factor is a category which is also included by Guilford in his Structure-of-Intellect model. The item analysis application would shed further light on the importance of this factor. It was borne in mind that the emphasis is placed on vocational groups, and not on specific vocations when testing aptitude. The memory factor can differ for different vocational groups, however, so that the value of including such a test in an aptitude test battery may be restricted. It can be conceived that the material which is placed in the memory of artists, differs from that which is placed in the memory of linguists, and will thus not bear the same relation to the memory of symbols for instance.

This ability is a general ability and is thus not linked to specific vocational groups.

An example of a *few items* analogous to those appearing in the actual test are the following:

Items:

Study the following combinations since you will have to recall them later ....

1. fire = ssieb   2. black = riod

(vi) Clerical Speed and Accuracy

This test is also known by other names such as Checking Test or Comparison Test (as in Manual for the Multi-Aptitude Test, 1955, p.18). The relevant test measures the ability to perceive and reproduce fast and accurately. In contrast with the previous test, use is made of meaningful material in this case, and not of code words or symbols.

The Clerical Speed and Accuracy Test consists of the perception, comparison and sometimes classification of names and symbols. The test contains three sections. In the first section the classification of a number of concepts must be examined. In the second section of the test the arrangement of a number of names in alphabetical order must be perused. In the third section a group of letters, numbers or symbols on the left-hand side of the page have to be compared with groups of letters, numbers or symbols on the right-hand side of the page. In all three sections of the test, perception and comparison are therefore involved.

In contrast with the other tests of the battery, response speed plays an important part in this test.

Although this test will not have much value in the school situation, it should have considerable value in the vocational guidance program. It can serve as a valuable instrument for selecting people who possess the ability to perform routine tasks successfully. As can be concluded, it should serve as a predictor for clerical and purely mechanical (repetitive) vocations like those of clerk, sorter, factory worker.

The *following items* are examples of items analogous to those appearing in one section of the actual test:

Items:

Which of the following words are filed under the wrong headings?

Food			Furniture		Fluids		Vehicles		Sports	
1.	A.	Butter	B.	Desk	C.	Ice	D.	Motor-car	E.	Rugby
2.	F.	Shoes	G.	Chair	H.	Milk	I.	Bicy-cle	J.	Soccer

Answers: 1 = C, 2 = F

(vii) Visual Arts

This test measures the ability to perceive and interpret relations between lines, shades and dimensions. The ability to reproduce is unfortunately not measured in this test, but information in this respect can possibly be gained from the Clerical Test and the Memory Test. The test consists of four sections, namely:

(a) recognition of shape;
(b) observation of light and shade;
(c) problems concerning perspective of curves;
(d) problems concerning perspective of angles and parallel lines.

The first section is purely a comparative section, whereas in the other sections the pupil must indicate the mistakes which have been made in the drawings with regard to lines, shades, etc.

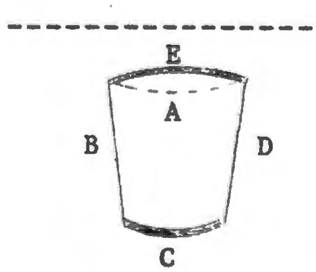
Although this ability has received attention for many years, it is a sphere which is still relatively undeveloped as is indicated by the following remark of Nunnally (1970, p.328): "The nature of art and artistic ability has been a matter of interest to psychologists for well over a hundred years, since before the time of Fechner. In spite of this prolonged interest, the measurement of artistic ability lags behind the testing of other ability functions". He continues by giving a number of explanations for this state of affairs. Cronbach (1960, p.318) reaches a more or less similar conclusion when he says: "Research on artistic abilities is still in a most primitive stage. No systematic research has been done using modern tests and adequate criteria. The nature of artistic aptitude remains an unsolved - and neglected - problem".

This ability ought to play an important role in vocations like those of artist, architect and sculptor.

The following is an example of an item analogous to the items appearing in one of the sections of the test:

<u>Item</u>:

1. Which line or curve in this sketch of a glass is incorrectly drawn?



Answer: A

(viii) <u>Word Building</u>

This test ought to give an indication of general language ability. It was surmised, however, that several other abilities like memory and creativity could also play a role in this test. A deviation was made from the traditional manner of measurement of this ability where vocabulary and comprehension sections are included (e.g. the DAT, p.25 and Multi Aptitude Tests, p. 2), to try to diminish the influence of knowledge. As will appear later, however, the proven ways were resorted to again with the second application.

This ability should be of importance in vocations like those of journalist, translator and writer.

An example of an item analogous to those appearing in the actual test is given below:

Item:

1. Which one of the five words on the right-hand side has nearly the same meaning as the word that can be formed by adding a certain letter to the three letters on the left-hand side?

clm

A. welcome
B. here
C. turn
D. cloud
E. serene

Answer: E

(ix) Mechanical Insight

Mechanical ability can be decomposed into two components, namely reasoning (with regard to mechanical material) and spatial perception. These two components are not only related to each other but are also related to intelligence as is evident from the correlations supplied in Table 3.2.4.3 (pp.58-60).

In factor analysis of the results of aptitude tests and interest questionnaires the mechanical field is usually one of the fields which causes many problems. Since the mechanical tests usually do not yield pure factors this type of test is not included in test batteries having factorial validity as a basis.

Owing to the important role of mechanics in the vocational domain, this field can in general not be excluded from an aptitude test battery. Further it must be noted that this type of test may still be replaced by other tests like verbal reasoning tests in the case of professional groups but definitely not in the case of technical groups. The following scores quoted from the DAT Manual (p. 60) serve to prove this statement:

<u>Degree-seeking students</u>

	N	VR 1)	NA	AR	SR	MR 2)	CSA	Sp
Engineer	70	<u>80</u>	86	80	81	<u>82</u>	67	68
Liberal Arts	68	<u>79</u>	75	78	61	<u>64</u>	75	78

<u>Non degree-seeking students</u>

	N	VR 1)	NA	AR	SR	MR 2)	CSA	Sp
Mechanical, Electrical and Building Trades	66	<u>34</u>	41	46	49	<u>56</u>	44	28

1) VR = Verbal Reasoning

2) MR = Mechanical Reasoning

Each item in this test is composed of a drawing of mechanical objects and a question based on the drawing. The items in the test cover the following aspects, amongst others: levers, forces, pressure and projectiles. Although previous experience can influence the score obtained in this test, this influence should not have a serious effect on the interpretation of the test scores.

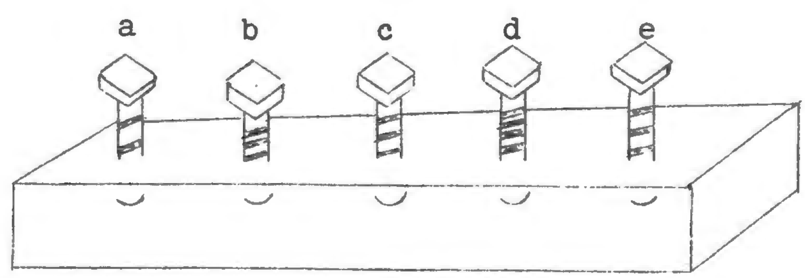
The Mechanical Insight Test in conjunction with the Spatial Perception Tests should be good predictors of success in trades on the one hand, and of success in professional vocations, like engineer and architect, on the other hand.

This test actually pertains more to boys than to girls, with the result that the performance of the girls on this test can influence the average of the whole group substantially. This can necessitate separate norms for boys and girls. However, the test will also have value for the girls since there are mechanical vocations which are mainly followed by girls, as for instance, in general factory work, in the clothing industry and in the jewellery industry. If, however, a girl wants to become an engineer, her test score will have to be converted to a standard score for boys, and if a boy wants to practise a vocation which is mainly practised by girls, his test score will have to be converted to a standard score for girls.

This ability ought to play an important role in vocations like those of engineer, pilot, radiotrician, electrician, fitter and turner, and operator.

An example of an item analogous to the items appearing in the actual test is given below:

Item:



Which bolt will require the least number of turns in an anti-clockwise direction before it is completely free from the metal-block?

A. bolt e
B. bolt d
C. bolt c
D. bolt b
E. bolt a

Answer: E

3.2 APPLICATION FOR ITEM ANALYSIS

3.2.1 Sample

The first application of the aptitude test battery took place during March and April 1969. The application in Natal was undertaken by the Department of Indian Education while the application in Transvaal was undertaken by officials of the HSRC.

Since the largest part of the Indian population of the Republic is concentrated in Natal and Transvaal, only schools from these two provinces were included in the sample. The samples were drawn separately for Standards 6, 7 and 8 with an estimated number of ± 600 pupils per standard.

The samples were stratified and drawn at random. The strata that were taken into consideration were: province, area and sex. The ratios for the provinces Natal and Transvaal for the separate standards were first determined. These were fixed at 5,4:1; 5,4:1 and 3,7:1 for Standards 6, 7 and 8 respectively. Thereafter the ratios for the areas, urban and rural, were determined. These were fixed at 1,8:1; 2,3:1 and 2,3:1 for the three standards respectively. Finally the ratios for boys and girls were determined. These were fixed at 1,2:1; 1,4:1 and 1,6:1 for the three standards respectively. On the basis of these ratios the exact number of pupils per province, area and sex for the sample could be determined. It was further decided to test approximately thirty four pupils, per standard, per school, so that the number of schools required to fill the quota could be determined and chosen at random. Owing to practical reasons small differences were found between the practical and theoretical distributions.

The table below gives an indication of the distribution according to standard, sex, province and area as actually tested, and Appendix A supplies a list of the schools included in the sample and the number of pupils tested in the different categories.

TABLE 3.2.1 NUMBER OF PUPILS IN THE SAMPLE ACCORDING TO STANDARD, SEX, PROVINCE AND AREA

Province and Area	Std. 6		Std. 7		Std. 8	
	Boys	Girls	Boys	Girls	Boys	Girls
Natal-Urban	172	144	200	140	198	126
Natal-Rural	112	92	100	70	100	60
Transvaal-Urban	44	36	48	36	66	42
Transvaal-Rural	6	6	8	6	10	6
Sub-total	334	278	356	252	374	234
Total	612		608		608	

With regard to the size of a sample, Helmstadter (1964, p. 162) remarks: "To achieve a satisfactory degree of stability in the several item indices described below, it is generally recommended that around 400 subjects be used".

The size of the schools (the enrollments) were not taken into consideration, when drawing the sample owing to practical reasons. A fixed number of pupils per school was decided upon, because it would not be practical, for instance, to send an examiner from Pretoria to Pietersburg to test only six pupils.

3.2.2 Time limits for the tests

Since the first application was made with a view to item analysis specifically, sufficient time had to be allowed for the completion of all the items in a test. Unrestricted times could not be allowed, however, especially with tests like the Clerical Speed Test so that 90% of the pupils in a group were allowed to complete a test. The testing times were kept the same for the three standards, otherwise it would not be possible to determine which percentage of the variance in test scores was due to differences in testing times, and which to actual differences in ability.

3.2.3 Scoring of tests

Each pupil used two IBM-answer sheets for the answering of the twelve tests. Initially the answer sheets for the Group Intelligence Tests are being used, but later an answer sheet specifically for the aptitude test battery will probably be constructed. For the scoring of the tests an optical scoring machine (IBM 1230) was used. The item responses and the totals of each test were punched directly on punch cards and prepared for calculations. The calculations were executed by an IBM 360 computer.

3.2.4 Discussion of results

3.2.4.1 Item analysis

(i) A problem of primary importance in connection with item analysis is that a generally accepted technique does not exist. Helmstadter (1964, p. 163) remarks for instance: "A recent review of the literature has indicated that there are at least some twenty-three different methods of item analysis". Gulliksen (1962, p.363) also remarks: "Twenty-three methods are described by Long and Sandiford (1935). Nineteen methods are summarized by Guilford (1936b) in Psychometric Methods, pages 426-456. With one or two exceptions these lists are essentially the same".

Although there are certain statistics which are common to most item analysis techniques, there are nevertheless large differences between the different techniques. Helmstadter even mentions a graphic method of item analysis. In this method use is made of the so-called Item Characteristic Curves. An item with perfect discrimination ability, for example, is represented by a vertical line which indicates that nobody who scored beneath the indicated score answered the item correctly while everyone with a higher score gave the correct answer. A horizontal line on the other hand would mean that the item is completely invalid, or, stated otherwise, that the probability of answering the item correctly is completely independent of the trait being measured (be it test or criterion). The figure given on the following page serves as an illustration of the graphic method.

According to the figure, item C will possess the best discriminative value.

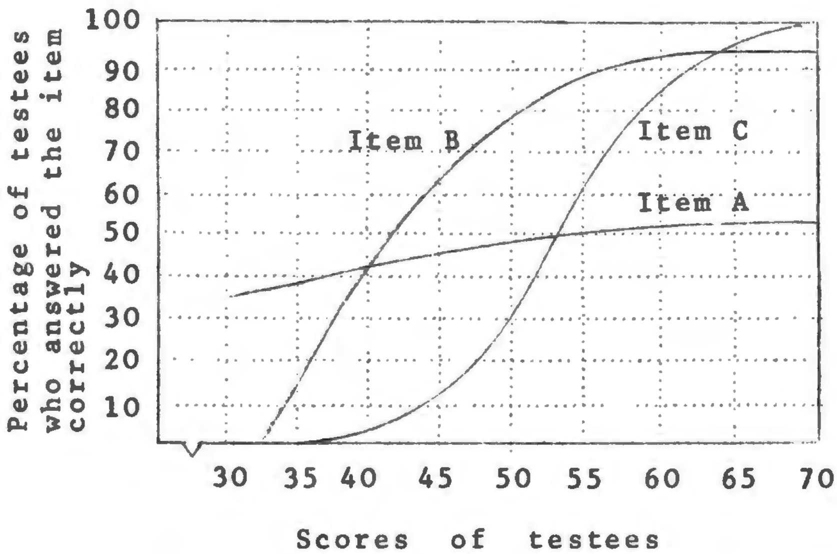


FIGURE 3.2.4.1 A GRAPHIC METHOD OF ITEM ANALYSIS

Another problem is that with many of the item analysis techniques there is no *definite* relation between the item selection procedure and some or other important parameter of the test. According to Gulliksen (1962, p. 364) the exceptions are the following: "........ the method of successive residuals (Horst, 1934); the use of a maximizing function (Horst, 1936); the L-method and its various modifications (Stoops, 1941)".

For this application item analysis was done according to a program which consists of a combination of Anstey's (1966, pp. 167-177) and Gulliksen's (1962, p. 365) item analysis methods. The information supplied by this program is the following:

(a) The percentage of testees that have answered an item;
(b) the proportion(p) of testees that have answered an item correctly;
(c) the percentage of responses drawn by each distractor;
(d) the average of the group that has chosen a specific distractor;
(e) the correlation between the item and the total score of a test (r_{it}-value or discrimination value);

(f) the product value which is composed of the product of the discrimination index and the standard deviation of the item (reliability index), and

(g) the reliability (r_{tt}), mean ($\bar{x}$), standard deviation (s), skewness and kurtosis of the test as a whole.

Although all the above-mentioned information was available, only the aspects to which most value was attached will be described in the paragraphs below.

(ii) Factors of primary importance in item analysis

(a) Correlation of items with external fields

To ensure the pureness (homogeneity, unitary nature) of the test, the condition was set that no item must correlate higher with two rationally chosen external fields than with their own field. For the determination of this statistic use was made of the following formula (Gekoski, 1964, p.168):

$$r_{xy} = \frac{\frac{\Sigma XY}{N} - M_x M_y}{(SD_x)\ (SD_y)}$$ where,

r_{xy} = the correlation coefficient,

Σ = sum of

X and Y = scores on item and external field respectively

M_x and M_y = the means, and

SD_x and SD_y = the standard deviations.

The external fields were determined with the aid of the intercorrelation matrix of the different tests. The two fields which correlated highest with a specific field were chosen as the external fields. No items were found that correlated higher with external fields than with their own field, not even in the fields that are closely related like the 2-D and 3-D fields.

The determination of the significance of differences between internal correlations and external correlations was also considered. However, since this would bring about 2 x 25 x 12 = 600 additional indices which would have to be evaluated, it was decided that the previous consideration would suffice.

(b) <u>Discrimination value</u> (r_{it}-value)

This statistic is used to determine if an item discriminates adequately between pupils with high scores on a test and those with low scores. For the determination of this value use was made of the formula for point biserial correlation (Helmstadter, 1964, p.164; Ghiselli, 1964, p. 162), namely:

$$r_{pbis} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}} \text{ where,}$$

r_{pbis} = point biserial correlation coefficient

M_p = mean score of all testees who have answered the item correctly

M_t = mean score of all testees who have answered the test

S_t = standard deviation of all total scores

p = proportion of testees who have answered the item correctly

q = 1 - p.

With regard to this value there is once again no uniformity, but the lower limit can apparently be fixed at 0,20. Garrett (1964, p. 368) remarks in this connection: "As a general rule items with validity indices of 0,20 or more are regarded as satisfactory". The assumption is, of course, that this cut-off point bears relation to a reasonably large sample of testees. Other investigators again supply more definite limits for this value. Furst (1958, p. 314) for example, supplies the following basis of interpretation:

0,50 and higher	- high positive
0,30 to 0,49	- medium positive
0,20 to 0,29	- borderline positive
0,00 to 0,19	- none to low positive
-0,01 and lower	- none to negative.

For this battery the minimum value was set at 0,20 ,since this value is significant at the 5% level for a group of 100 pupils according to tables of Peatman (1963, p. 406).

Another index which can be used for the same purpose as the point biserial coefficient is the biserial coefficient. The disadvantage of the biserial coefficient (as in Mehrens & Ebel, 1967, p. 390), however, is that the trait or ability underlying the dichotomous variable is assumed to be normally distributed.

(c) <u>Proportion of pupils who answered an item correctly</u>

There appears to be no uniformity with regard to the limits of proportion of correct answers for the retention of an item in a test. Garrett (1964, p. 363) remarks in this relation: "Other things being equal, items of moderate difficulty are to be preferred to those which are much easier or much harder". Hardaway (1964, p. 176) again remarks: "Some test designers eliminate or revise those items that were solved by less than 10% of the students and by more than 90% of the students". Finally Helmstadter (1964, p. 169) remarks: "There is some evidence (c.f. Lord 1953) that for maximum efficiency in selection, a test should be peaked; that is the average item difficulty should match the selection ratio".

A good viewpoint seems to be the relation of the limits or the optimum difficulty value to the nature of the group being tested and the nature of the instrument being used. For a highly selected group a range of 10% - 90% will not be appropriate while this range may be ideal for a very heterogeneous group. For this test battery the limits were set at 10% and 90% respectively owing to the range of the group being tested. In a few exceptional cases, however, items were also retained if they did not comply with this requirement for *only one of the three standards*. A summary of the items which did not satisfy the above requirements are supplied in Appendices B-1 and B-2.

Related to this criterion is the distribution of the choices of the distractors separate from the correct distractor. Items of which any of the distractors drew less than 5% of the responses were altered as far as possible. Items were not altered radically, however, since this might bring about a totally different response pattern which could again influence the expected average of the test.

3.2.4.2 Reliability

In simple non-technical language the reliability of an instrument can be regarded as the extent to which it accurately and constantly measures what has to be measured.

In a statistical sense reliability can be regarded as the proportion of true variance in the total test variance. Guilford (1954, p. 350) supplies the following formula in this connection:

$$r_{tt} = \frac{\sigma_{\infty}^{2}}{\sigma_{t}^{2}}$$

where, r_{tt} = reliability

σ_{t}^{2} = variance of test

σ_{∞}^{2} = true variance.

Whereas this definition is still very theoretical, Jackson and Messick (1967, p. 182) place the practical aspects much stronger in the foreground. They present the following definition of reliability: "Reliability refers to the accuracy (consistency and stability) of measurement by a test ...... Whereas "accuracy" is a general expression, the terms "consistency" and "stability" are needed to describe, respectively form-associated and time-associated reliability".

There are four basic techniques for studying the reliability of test scores. They are the following (Bauernfeind, 1963, p. 99; Gronlund 1968, p. 182):

(i) the test-retest technique
(ii) the alternative forms technique
(iii) the split-half technique
(iv) the Kuder-Richardson technique.

Owing to the extensiveness of the aptitude tests, and since the pupils who are tested the first time are not always available for the second application, only the internal consistencies of the tests of the JATISA were determined. The formula which was employed for the determination of the internal consistency (homogeneity) was the Kuder-Richardson formula 20. This formula was developed by Kuder and Richardson as a measure of test homogeneity. For the use of this formula the test is applied to the group once only and the distribution of scores obtained relative to the total length of the test is then analyzed statistically. This technique should be used for tests which are not speed tests and in cases where all items have been attempted. These requirements were satisfied except for the Clerical Speed Test. The intention was to make a test-retest application for this test at a later stage. Reliabilities were also determined by means of the split-half method but the results were very similar to those obtained by using the K-R formula since the two methods are only variants of each other.

The K-R formula 20 yields a coefficient which is an average and which is based on all the possible divisions of a test. The formula is as follows (Cronbach, 1970, p.161):

$$r_{tt} = \frac{N}{N-1} \left(1 - \frac{\Sigma pq}{s_t^2}\right)$$

where, N = the number of items

Σ = sum of

p = proportion of individuals who have answered an item correctly

q = 1-p and

s_t^2 = variance of total test scores.

When interpreting the K-R 20 reliability index it must be borne in mind that this index is a direct function of the variance of the test scores (the other factors in the equation having less influence). Bauernfeind (1963, p. 115) makes the following remark in this connection: "Referring again to the Kuder-Richardson techniques, we see that score reliability for a given group is a direct function of the spread of scores over the range of possible scores". This situation can cause problems since the nature of the *group* may determine the reliability instead of the nature of the *test*. The writer is of the opinion that for specific groups, the reliability indices of the fields which are relevant to them will be smaller than for general groups, owing to the differences in variance. Further the writer is also of the opinion that the smaller the variance of a test or field, the larger the difference will be between a homogeneity index and a test-retest index. The writer hopes to test both these hypotheses shortly.

Although the K-R index is a homogeneity index as stated above, Loevinger (as in Mehrens and Ebel, 1967, p.186) maintains that this index gives an estimate of item homogeneity as well as of test-retest reliability. This does not imply that large differences between indices actually determined according to the different methods cannot occur, however.

The K-R indices obtained for the two forms of the battery employed in the first application are given in Tables 3.2.4.2(a) and 3.2.4.2(b). The coefficients for Standards 6, 7 and 8 are supplied. In general the reliabilities of the tests are reasonably high except in the case of the Art Test and the Word Building Test. In the case of these tests the low reliability indices can be ascribed to the small variances of the tests. As mentioned previously, the smaller the variance of a test the smaller the corresponding K-R index (if the other variables in the formula do not differ largely).

Further reasons why homogeneity indices should be interpreted with caution are because an increase in homogeneity does not necessarily bring about an increase in predictive value (in other words validity), and because homogeneity indices as well as other reliability indices cannot be related to confidence intervals. With regard to the first aspect, Bauernfeind (1963, p.103) remarks: "...... a test with high reliability may have low validity in which case the high reliability is of no value. The value of an aptitude test should be assessed by its validity coefficients, not by its reliability". In addition Horst (as in Mehrens, 1967, p.25) proclaims: "...such procedures will tend to make the test more pure or homogeneous in the sense of conserving those items which have the largest intercorrelations. This is the only sense in which it may be said that the conserved items are more valid than the rejected ones". With regard to the second aspect it can be said that more and more attention is

TABLE 3.2.4.2(a) RELIABILITY COEFFICIENTS FOR FORM A (K-R 20) - FIRST APPLICATION

Std.	Test number											
	1[1)]	2	3	4	5	6	7	8	9	10	11	12
6	0,753	0,814	0,798	0,759	0,901	0,797	0,720	0,878	0,831	0,693	0,649	0,652
7	0,765	0,846	0,831	0,747	0,919	0,903	0,705	0,850	0,803	0,679	0,466	0,749
8	0,751	0,836	0,801	0,796	0,921	0,896	0,770	0,889	0,806	0,500	0,666	0,799

TABLE 3.2.4.2(b) RELIABILITY COEFFICIENTS FOR FORM B (K-R 20) - FIRST APPLICATION

Std.	Test number											
	1	2	3	4	5	6	7	8	9	10	11	12
6	0,765	0,699	0,662	0,762	0,877	0,807	0,759	0,901	0,848	0,667	0,658	0,647
7	0,746	0,791	0,623	0,715	0,885	0,885	0,743	0,846	0,757	0,641	0,567	0,647
8	0,715	0,819	0,615	0,743	0,903	0,893	0,780	0,886	0,747	0,468	0,546	0,724

Test 1 - Verbal Reasoning
Test 2 - } Non-Verbal Reasoning
Test 3 -
Test 4 - Computations
Test 5 - Spatial Perception 2-D
Test 6 - Spatial Perception $3\text{-}D_1$
Test 7 - Spatial Perception $3\text{-}D_2$
Test 8 - Memory
Test 9 - Clerical Speed
Test 10 - Visual Arts
Test 11 - Word Building
Test 12 - Mechanical Insight

being given to reliability in terms of errors of measurement and so-called 'bands'. The bands referred to here pertain to the interval which includes a distance of one or two errors of measurement on both sides of the obtained score. Cronbach and Gleser (1965, p.292) remark in this connection: "It is recommended for such batteries as SCAT and STEP that each score be plotted on the profile sheet not as a point but as a band extending 1 S.E. above and below the observed score". Goldman (1961, p.285) further remarks: "A promising sign, however, is the use of the band approach to score description in some recently published tests". And Whitla (1968, p.274) finally remarks: "The outstanding advantage of using the standard error is that it permits us to make certain probability statements about an individual's "true score", given the score he actually received".

3.2.4.3 Intercorrelations[1)]

(i) General

Correlations were computed by the product-moment formula of Galton and Pearson which is presented in any of the following forms (Gulliksen, 1962, p.424, Gekoski, 1964, p.168, Bruning & Kintz,1968, p.153):

$$r_{xy} = \frac{\frac{\Sigma XY}{N} - M_x M_y}{(SD_x)\ (SD_y)} = \frac{\Sigma XY - NM_x M_y}{\sqrt{\Sigma X^2 - NM_x{}^2}\sqrt{\Sigma Y^2 - NM_y{}^2}}$$

$$= \frac{N(\Sigma XY) - (\Sigma X)\ (\Sigma Y)}{\sqrt{(N\Sigma X^2 - (\Sigma X)^2)(N\Sigma Y^2 - (\Sigma Y)^2)}}$$

where, r_{xy} = the correlation coefficient

Σ = sum of

X and Y = raw scores with regard to the X-variable and the Y-variable.

---

1) The intercorrelations in the following paragraphs and tables refer to correlations between test scores.

N = number of examinees

M_x and M_y = means with regard to the X-variable and the Y-variable.

The above formula can also be briefly presented as follows (Hogg & Craig, 1963, p.62; Stilson, 1966, p.257):

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

where, r_{xy} = the correlation coefficient

σ_{xy} = the covariance of variables x and y

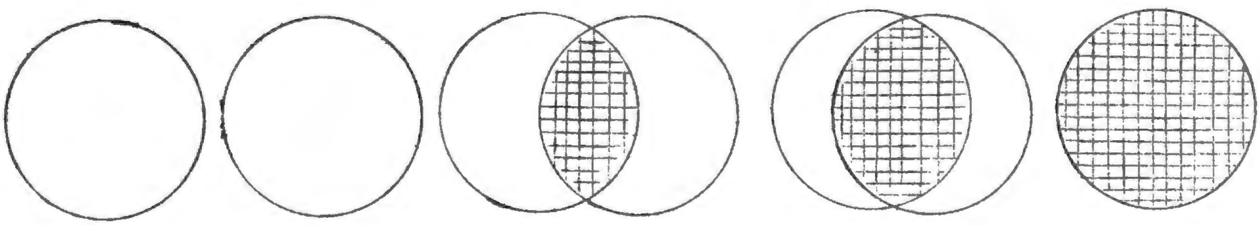
$\sigma_x \sigma_y$ = the product of the standard deviations of the variables x and y.

On first inspection of this formula it appears as if smaller variances of the separate variables (denominator) yield a larger correlation coefficient. This is not the case, however, since the term in the numerator also includes variance, namely co-variance which actually influences the formula as a whole *more* than the denominator (product of separate variances). Stilson (1966, p.257) outlines the position clearly when he remarks: "While variance is reduced by restriction of the range of either variable, the effect on the covariance is greater and the correlation coefficient is decreased in absolute magnitude". It is thus apparent that the same relation exists here as exists with a Kuder-Richardson formula, for instance. The importance of the latter aspect becomes clear when a Kuder-Richardson technique for determining reliability is compared with the test-retest technique.

(ii) Interpretation of correlation coefficients

Correlation coefficients are usually interpreted in one of two ways, namely, either by means of significance tables or by making use of the principle of common variance. In the first case the number of individuals in the sample plays an important role in determining the significance level of the correlation coefficient. In the second case no tables are required and the percentage of common variance is determined by simply squaring the correlation coefficient. Thus an r-value of 0,50 implies 25% common variance (Clarke, 1965, p.127; Bauernfeind, 1963, p.67; Helmstadter, 1964, p.119; Peatman, 1964, p.95) The percentage of common variance can also be depicted graphically. Bauernfeind (1963, pp.64-65) presented the extent of common variance graphically by making use of overlapping circles (or spheres). The figures below illustrate his method of interpretation of correlation coefficients:

r = 0,00	r = ± 0,40	r = ± 0,70	r = 1,00



Common variance = 0%	Common variance = ± 16%	Common variance = ±50%	Common variance = 100%

FIGURE 3.2.4.3 RELATION BETWEEN CORRELATION COEFFICIENT AND PERCENTAGE OF COMMON VARIANCE

In the case of the JATISA correlation coefficients were also interpreted in terms of percentage of common variance and 0,5 (approximately) was taken as lower cut-off point for the correlations which were regarded as significant. Clarke (1965, p.127) remarks in this connection: "In these terms correlations of less than 0,50 would appear to be quite low because they would indicate that less than 25 per cent of the Y variance is associated with X." If the correlation coefficients were to be interpreted according to significance (say on the 5% level) then all values larger than 0,20 (for the relevant N's) would have to be considered. This would unfortunately have the effect that most of the coefficients of the matrix would have to be regarded as significant.

(iii) Discussion of results

The intercorrelations for the 12 tests of Form A are presented in Tables 3.2.4.3(a), (b) and (c). Only the intercorrelations of Form A are reported since most of the tests of Form A were retained in the final test battery. The reason for the latter decision is given at a later stage.

From the correlation matrices for the different standards it is evident that the correlations are in general reasonably low, although there are of course several coefficients which are significant. The cut-off level for the correlations to which significance was attached, was $\pm$ 0,5 (lower limit) and the coefficients falling in this category are all underlined in the matrices.

The first correlations to which attention can be given are those between tests 1,2 and 3. The correlations between tests 2 and 3 are very high for all three standards, namely in the region of 0,7. This corresponds with the hypothesis that the tests are of a similar nature, namely, series completion tests. If test 1 is further regarded as a measure of reasoning, then tests 2 and 3 are apparently also reasoning tests on the basis of their correlations with test 1. Owing to the non-verbal nature of the two tests, the fact that both are related to a reasoning factor and the extensiveness of the battery it was decided to retain only one of these two tests in the final test battery.

In the second place the correlations between tests 5 and 6, and 5 and 7 can also be regarded as significant. Tests 5, 6 and 7 are all three spatial perception tests and a high correlation between tests 6 and 7 could also be expected. From a study of test 6, however, it becomes clear that it also contains a creative or structural aspect which is not present in test 7, and also that a three-dimensional object has to be formed from a two-dimensional plan in contrast with test 7.

TABLE 3.2.4.3(a) INTERCORRELATIONS OF TESTS IN FORM A (STD. 6)

(N = 612)

Test No.	1	2	3	4	5	6	7	8	9	10	11	12
1	1,000											
2	0,549	1,000										
3	0,574	0,719	1,000									
4	0,229	0,160	0,222	1,000								
5	0,312	0,418	0,407	0,161	1,000							
6	0,281	0,265	0,284	0,094	0,182	1,000						
7	0,318	0,389	0,327	0,180	0,470	0,241	1,000					
8	0,325	0,196	0,285	0,377	0,226	0,227	0,138	1,000				
9	0,455	0,470	0,512	0,291	0,293	0,346	0,269	0,411	1,000			
10	0,218	0,263	0,266	0,007	0,153	0,161	0,236	0,193	0,237	1,000		
11	0,441	0,307	0,275	0,209	0,235	0,201	0,208	0,236	0,345	0,187	1,000	
12	0,409	0,288	0,289	0,150	0,307	0,265	0,334	0,192	0,295	0,238	0,581	1,000

TABLE 3.2.4.3(b) INTERCORRELATIONS OF TESTS IN FORM A (STD. 7)

Test No.	1	2	3	4	5	6	7	8	9	10	11	12
1	1,000									(N = 608)		
2	0,525	1,000										
3	0,580	0,716	1,000									
4	0,311	0,294	0,273	1,000								
5	0,382	0,360	0,350	0,219	1,000							
6	0,278	0,429	0,457	0,346	0,275	1,000						
7	0,355	0,371	0,374	0,170	0,398	0,325	1,000					
8	0,304	0,389	0,368	0,340	0,200	0,361	0,149	1,000				
9	0,367	0,366	0,415	0,261	0,198	0,244	0,413	0,313	1,000			
10	0,298	0,383	0,389	0,135	0,280	0,299	0,270	0,207	0,190	1,000		
11	0,382	0,338	0,286	0,247	0,214	0,250	0,338	0,256	0,498	0,238	1,000	
12	0,357	0,461	0,416	0,184	0,289	0,255	0,294	0,160	0,298	0,388	0,233	1,000

TABLE 3.2.4.3(c) INTERCORRELATIONS OF TESTS IN FORM A (STD. 8)

(N = 608)

Test No.	1	2	3	4	5	6	7	8	9	10	11	12
1	1,000											
2	0,610	1,000										
3	0,544	0,741	1,000									
4	0,446	0,425	0,432	1,000								
5	0,336	0,372	0,312	0,193	1,000							
6	0,365	0,497	0,498	0,253	0,477	1,000						
7	0,258	0,353	0,303	0,132	0,398	0,340	1,000					
8	0,404	0,411	0,388	0,572	0,290	0,291	0,204	1,000				
9	0,325	0,313	0,364	0,272	0,174	0,238	0,588	0,311	1,000			
10	0,217	0,301	0,300	0,026	0,261	0,353	0,254	0,175	0,115	1,000		
11	0,332	0,341	0,319	0,238	0,225	0,252	0,507	0,259	0,659	0,119	1,000	
12	0,391	0,440	0,414	0,273	0,340	0,415	0,373	0,322	0,307	0,342	0,353	1,000

Test 1 - Verbal Reasoning
Test 2 - Series Completion (1)
Test 3 - Series Completion (2)
Test 4 - Computations
Test 5 - Spatial Perception 2-D
Test 6 - Spatial Perception 3-D_1
Test 7 - Spatial Perception 3-D_2
Test 8 - Memory
Test 9 - Clerical Speed and Accuracy
Test 10 - Visual Arts
Test 11 - Word Building
Test 12 - Mechanical Insight

In the third place a high correlation is found between tests 9 and 11 for two of the three standards. This can be ascribed to a language factor which is common to both tests and also to a classification or comparison process which is common to both tests.

Next the correlations between tests 2 and 6, and 3 and 6 for Standard 8 can probably be ascribed to a common perceptual or comparative aspect.

Further correlations of note are those between tests 2 and 9, 3 and 9, 4 and 8, and 7 and 9. These correlations can probably be ascribed to a memory aspect which is common to all these tests.

The only correlations which are difficult to explain are those between tests 11 and 12 for Standard 6, and 11 and 7 for Standard 8. Perhaps these correlations should be regarded with caution, since the difficulty level of these tests also caused problems.

Summarizing it can be said that most of the correlations in the tables are low. Although it is not a requirement that correlations must be low, such a situation is often advantageous when multiple correlations have to be executed with regard to an external criterion.

3.2.4.4 Factor Analysis

(i) Introduction

There was a period in history when the use of factor analysis in psychology was regarded as more or less equivalent to the use of nylon in chemistry. It is doubtful, however, whether factor analysis will be able to pass the test of the times as nylon has done it. By this is not meant that factor analysis serves no purpose. On the contrary, the author is of the opinion that factor analysis is extremely valuable if correctly applied, and for certain situations almost indispensable.

Factor analysis is described by Du Bois (1965, p.452) as follows: "Factor analysis is a branch of statistics concerned with the isolation and identification of a limited number of hypothetical variables underlying a group of observed variables". Viewed from another angle, factor analysis can be regarded as a technique for analyzing tables of correlations between test scores with a view to indicating the common factors among the tests.

Modern factor analysis developed from the work of Spearman, who sought motivation for a general factor plus specific factors underlying the intelligence tests. When it was found that his two-factor theory was insufficient to explain all correlation matrices, Thurstone et al. developed multiple factor analysis methods. Thereafter further development took place with regard to orthogonal rotation and oblique rotation, and computers were used to perform calculations in a few hours which normally took days or even months.

(ii) Description of concepts

Some of the most important concepts relevant to factor analysis are the following:

(a) Communalities

The communalities of a matrix bear relation to the percentage of common variance of the variables which is explained by the factors. In a factor matrix any specific communality is obtained by summing the squares of the factor loadings in a specific row.

These values are usually estimated, and are related to multiple correlation coefficients and reliability coefficients. Du Bois (1965, p.460) says in this connection: "In factor analysis, however, what is analyzed for each variable is the common variance or communality, generally placed in the diagonal of the correlation matrix. A lower limit is known to be the square of multiple R............. The upper limit of the communality is the reliability coefficient".

From the last statement it is also clear that the factor analyst will attach great value to high reliabilities.

(b) Characteristic roots, latent roots or eigenvalues

The characteristic roots of a matrix are usually obtained by solving the following: $|A-\lambda I| = 0$, for λ, where I is the identity matrix and A the given matrix. The value of characteristic roots in factor analysis is usually found in their relation to the number of factors that have to be extracted. So, for instance, there is Kaiser's (as in Cattell, 1966, p.207) technique where the number of roots with values larger than one are taken as determinants for the number of factors, and (ii) the "scree-test" of Cattell (1966, p.206) which is based on the plotting of factors (X-axis) against size of characteristic roots (Y-axis), so that it can be determined where the plotted points spread out to form a more or less horizontal line.

The number of factors that have to be extracted are also related to the rank of the matrix so that there must be a relation between characteristic roots and rank. Horn (Psychometrika, 1965, p.181) indicates this relation as follows: "Guttman has shown that the weakest of three lower bound estimates for the minimum rank of a Gramian R is given by the number of roots of R (with ones in the diagonal) which are greater than or equal to unity".

(iii) Errors which are made in factor analysis

It often happens in practice, that a psychological instrument is constructed, factor analysis is executed and considerable value attached to the results, without consideration of any of the conditions which are associated with the correct application of factor analysis. Guilford (as in Jackson and Messick, 1967, p.309), amongst others, regarded this matter in a serious light, and mentioned the following errors which are often made in factor analyses:

(a) Too many factors are extracted for the number of variables

He regards it as a good rule that three variables be available for each factor which has to be extracted. In other words, if there are nine tests, three factors have to be extracted.

(b) Too many experimental variables with a complex factorial composition are included

Such variables are often related to two or more factors.

(c) An insufficient number of "marker" tests are employed

A result of this is that some common factors which would otherwise have been extracted are now neglected.

(d) Not enough factors are extracted

According to Guilford one must rather be liberal in this respect since even the last factors include common variance as well as error variance. According to him the best criterion for ceasing the extraction of factors is the size of the factor-loadings.

(e) Correlation coefficients which are spurious are used

Guilford points out, for instance, that Thurstone has often stressed that correlations which are used in a factor analysis should be based on variables which are linearly independent.

(f) Correlations based on ipsative scores are often included in factor analyses

Whereas the R-technique can be used for normative data it is not applicable to ipsative data. This is actually a matter of prime importance which is disregarded by certain reputable psychologists.

(g) Insufficient attention is given to the requirements for correlation coefficients

For the correlations of a sample to be representative of the correlations of the population certain requirements are set namely, the distributions should not be significantly skewed, truncated or multimodal. The latter aspects are again linked to the difficulty values of the tests.

(h) <u>The population on which the analysis is based should not be too heterogeneous</u>

As an example, different populations are often united without determining the influence thereof on intercorrelations.

(iv) <u>Requirements for an efficient factor analysis</u>

Just before these requirements are mentioned the relevant notation will be introduced. R_o is an observed correlation matrix in which communalities determined by some or other method are placed in the diagonal. R is an approximation of R_o which is obtained by multiplying the factor matrix F with its transposed ($FF^T=R$).

There are at least three requirements for an efficient factor analysis, namely, (Du Bois 1965, p.465):

"(a) The residual coefficients (obtained from R_o-R) must be as small as possible.

(b) The solution must have a low rank; in other words the number of factors as represented by the number of columns in F must be small.

(c) The F-matrix must possess simple structure. This means that the F-matrix must have high loadings on the 'marker' variables and low positive or nil loadings on the other variables. If all elements in F are nought or positive it is said that the factor solution has 'positive manifold'".

Harman (as in Whitla, 1968, p.148) also cites a number of additional conditions, amongst which are:

"(a) Each row of the factor matrix should have at least one zero.

(b) If there are *m* common factors, each column of the factor matrix should have at least *m* zeros".

(v) Discussion of results

Factor analysis was performed on the twelve variables of each form. The method employed was Principal Factor Analysis and the axes were rotated according to the Varimax criterion. Since the data of Form A were used almost exclusively for the construction of the final test battery, only the results of Form A will be given in the tables below.

TABLE 3.2.4.4(a) FACTOR ANALYSIS OF TESTS IN FORM A (STD. 6)

Test no.	Test 1) name	Communality	Factor loading					
			1	2	3	4	5	6
1	Verb. Reas.	0,550	0,506	0,375	0,218	0,108	-0,260	-0,161
2	Series Compl.(1)	0,689	0,726	0,166	0,059	0,277	-0,221	-0,073
3	Series Compl.(2)	0,707	0,741	0,130	0,094	0,211	-0,203	-0,215
4	Comput.	0,348	0,090	0,099	0,516	0,071	-0,051	-0,236
5	2-D	0,445	0,253	0,133	0,272	0,490	-0,223	0,009
6	$3\text{-}D_1$	0,232	0,158	0,141	0,065	0,111	-0,375	-0,172
7	$3\text{-}D_2$	0,462	0,198	0,175	0,003	0,590	-0,131	-0,163
8	Memory	0,432	0,131	0,123	0,224	0,099	-0,094	-0,575
9	Cler. Speed	0,508	0,390	0,180	0,184	0,094	-0,327	-0,418
10	Visual Arts	0,213	0,148	0,125	0,024	0,113	-0,402	-0,025
11	Word Building	0,538	0,183	0,670	0,128	0,071	-0,123	-0,138
12	Mech. Insight	0,555	0,111	0,651	0,040	0,245	-0,226	-0,078
	Latent root :		3,825	0,605	0,523	0,355	0,204	0,165

1) See page 32 for full names

TABLE 3.2.4.4(b) FACTOR ANALYSIS OF TESTS IN FORM A (STD. 7)

Test no.	Communality	Factor loading				
		1	2	3	4	5
1	0,478	0,422	0,317	-0,286	0,282	0,194
2	0,679	0,651	0,211	-0,354	0,242	0,165
3	0,730	0,708	0,221	-0,356	0,218	-0,069
4	0,306	0,078	0,161	-0,504	0,142	0,012
5	0,356	0,203	0,104	-0,219	0,505	0,038
6	0,393	0,247	0,084	-0,509	0,189	0,174
7	0,408	0,188	0,373	-0,100	0,473	-0,006
8	0,361	0,215	0,187	-0,523	0,044	-0,064
9	0,539	0,202	0,638	-0,205	0,119	0,188
10	0,312	0,359	0,110	-0,098	0,347	0,201
11	0,441	0,143	0,561	-0,226	0,194	-0,127
12	0,372	0,438	0,178	-0,098	0,363	-0,088
Latent root :		4,052	0,475	0,442	0,237	0,170

TABLE 3.2.4.4(c) FACTOR ANALYSIS OF TESTS IN FORM A (STD.8)

Test no.	Communality	Factor loading				
		1	2	3	4	5
1	0,517	0,388	0,203	0,223	-0,515	-0,107
2	0,730	0,281	0,193	0,385	-0,680	0,065
3	0,685	0,272	0,203	0,302	-0,691	-0,037
4	0,569	0,684	0,129	0,087	-0,244	0,131
5	0,433	0,199	0,146	0,602	-0,091	0,036
6	0,508	0,172	0,128	0,552	-0,322	-0,231
7	0,597	0,023	0,619	0,370	-0,095	-0,260
8	0,529	0,641	0,159	0,169	-0,195	-0,160
9	0,693	0,190	0,778	0,099	-0,150	0,135
10	0,277	-0,013	0,068	0,489	-0,184	0,013
11	0,566	0,155	0,705	0,112	-0,174	-0,046
12	0,390	0,209	0,267	0,450	-0,256	-0,082
Latent root :		4,379	0,936	0,691	0,330	0,158

On studying the results there are a few points which must be borne in mind. In the *first place* it must be mentioned that 'marker' variables were not used when constructing the test battery. In the *second place* it will be noted that the communalities of some of the tests are exceptionally low, with the result that the loadings of such tests on the relevant factors must be regarded with great caution. *Finally* it must be pointed out that, on inspection, the factor structures for all three standards appear to be different. This aspect was not statistically investigated however (for example by making use of target rotations) since the battery was not based on factor analysis for its construction.

Although the factor structures for the three groups had to be regarded with caution, the following factors were nevertheless identified tentatively:

(a) For Standard 6, six factors were identified. Factor one was identified as a general *reasoning factor* owing to the loadings of tests one, two and three on this factor. Factor two was vaguely identified as a *construction* factor since tests eleven and twelve had distinct loadings on this factor. Factor three was identified as a *numerical* factor owing to the loading of test four on this factor, but it will be observed that the communality of this test is extremely low. Next factor four was identified as a *spatial perception* factor owing to the loadings of tests five and seven on this factor. Penultimately factor five was identified as an *art* factor owing to the loading of test ten on this factor. The test with the second highest loading on this factor is a spatial perception test, which is explicable since spatial perception certainly plays an important role in the area of art. Once again, however, it must be pointed out that the communalities of the tests referred to are extremely low. Ultimately factor six was identified as a *memory* factor owing to the loadings of tests eight and nine on this factor.

(b) For Standard 7 four factors were tentatively identified, namely a *general reasoning* factor, a *verbal* *comprehension* or *classification* factor, a *memory* factor and a *spatial perception* factor.

(c) For Standard 8 the same four factors were identified, namely a *general reasoning* factor, a *classification* factor, a *memory* factor and a *spatial perception* factor.

(vi) Summary

For all three standards a reasoning factor, a classification or construction factor, a spatial perception factor and a memory factor were identified. In addition a numerical factor and an art factor were also identified for Standard 6. The latter finding seems to indicate that abilities are finer differentiated in the lower standards.

3.2.4.5 Validity

The original and most well-known definition of validity is: Validity is the extent to which a test measures what it is supposed to measure. With the passing of time this definition fell into disuse because of its generality. In this connection Loevinger (as in Jackson and Messick, 1967, p.79) remarks: "This definition is too vague, too remote from actual measuring operations, to be useful; it is consistent with all the current meanings of validity".

The definition of validity which has been in general use in psychometrics until recently, and which is described by Loevinger (as in Jackson and Messick, 1967, p.79) as the classical definition of validity is the following: "The validity of a test is the correlation of a test with some criterion. In this sense a test has a great many different "validities" .....". On the basis of this definition, validity cannot be regarded as a fixed or unitary feature of a test. However, this definition has also been criticised by investigators like Anastasi (1969, pp.111-114), Cattell (as quoted by Jackson and Messick, 1967, p.82), and Loevinger (as quoted by Jackson and Messick, (1967, p.82). As an illustration, Cattell remarked: "Particularized validation is not only devoid of proper scientific interest but deceptive in its promise of practical economy ...... Its absurdity is most cogently argued by the demands of practical economy and efficiency alone ......". Loevinger (as in Jackson and Messick, 1967, p.82) expressed herself even more sharply when she maintained: "The argument against classical criterion-oriented psychometrics is thus twofold: it contributes no more to the science of psychology than rules for boiling an egg contribute to the science of chemistry, and the number of genuine eggboiling decisions is small. A direct result of the previous objections to predictive validity and so forth was naturally that construct validity and factorial validity became more prominent.

In contrast with the above view there are investigators like Du Bois (1965, p.403), Gulliksen (1965, p.403) and Lord (as in Jackson and Messick, 1967, p.82), who are still adherents of the specific-criterion view. As an illustration Loevinger (as in Jackson and Messick, 1967, p.82) says of Lord: "Lord strongly defends the importance of specific validity, stating that the concept of "overall validity" is not basic to psychometrics. The discriminating power of the test for a specified examinee is the truly basic concept". The latter is also the view of the author in this investigation, owing to the importance which is attached to practical value, predictive value, etc. However, the author does not deny the value of trait-oriented validity for certain purposes, especially of a theoretical nature.

Apart from the two poles mentioned above, it also happens that validity is approached in other ways. One of the latest presentations in this connection is that of Stilson (1966, p.126). He discusses validity under two headings: (i) external validity, and (ii) internal validity. External validity is further subdivided into (a) concurrent validity, (b) predictive validity, and (c) construct validity, whereas internal validity pertains to the discussion of nominal scales, ordinal scales, interval scales and ratio scales. Since scaling plays an important role in the sphere of personality measurement, interest measurement and attitude measurement, amongst other things, it is perhaps of value that validity is also approached in this manner.

The types of validity which received attention in this investigation are the following:

(i) <u>Content validity</u>

As is known, content validity is not supplied in the form of a correlation coefficient but as a judgement by experts of the relation between the contents of a test and its aims.

Content validity received attention in this investigation in the form of a detailed analysis of the contents of the tests. This was done by presenting the items to a panel of judges.

(ii) <u>Predictive validity</u>

This type of validity refers to the prediction of future behaviour with the aid of test scores. Bauernfeind (1963, p.84) puts it as follows: "Predictive validity refers to the power of specific test scores to predict important future behaviours".

It is especially this type of validity which plays an important part in aptitude testing. The reason for this is that aptitude tests are constructed specifically with a view to prediction. With regard to the test battery for Indian pupils predictive validity can be investigated by relating test results to vocational choices and success, and also by ascertaining the relation between test scores and school examination results. Since the former aspect, however, is a longterm project with certain economic implications, it could not be investigated immediately. The latter aspect received due consideration.

(iii) <u>Construct validity</u>

Construct validity can be regarded as the extent to which a test measures a theoretical "construct" or trait (Anastasi, 1969, p.114). The concept "construct validity" first appeared in reports of the APA-committee and was later extended by Cronbach and Meehl (Helmstadter, 1964, p.134).

With regard to the JATISA it can be said that, if the construct "aptitude" is measured with regard to certain fields, then the pupils ought to receive high scores in the fields which are of importance for their proposed vocations. In addition, since the construct aptitude involves the higher mental processes, there should be a reasonable correlation between scores obtained in the JATISA and scores obtained in an intelligence test battery or examination marks. Proof of the latter aspect has been obtained already. There should also be a relation between aptitude and interest but this aspect will be investigated at a later stage.

(iv) Factorial validity

Certain investigators like Kleinmuntz (1967, p.45) and Anastasi (1969, p.116) regard factorial validity as an example of construct validity, while others like Garrett (1964, p.356) and Guilford (1965, p.471) discuss it separately. As a result of the fact that factor analysis is a reasonably extensive technique, the author personally prefers to discuss factorial validity separately although it is not denied that factor analysis can be used to prove the construct validity of a test.

In factor analysis the intercorrelations of a large number of tests are examined and if possible the intercorrelations are explicated in terms of a much smaller number of more general factors or trait categories. These factors are interwoven with the complex abilities which are measured by the different tests. Thus it may be found that four or five factors supply an explanation for the intercorrelations of twelve or more tests. The tests must be chosen selectively, however, otherwise the intercorrelations for twelve tests can be explained in terms of one or two factors.

The factorial validity of a specific test is usually determined by its factor-loadings as given by the correlation of the test with each factor. If an arithmetic test correlates 0,71 with the numerical factor which has been identified in the battery, then this coefficient indicates the factorial validity of the test.

In general it can be said that the larger the loading of a test on a certain factor, the purer criterion the test is of that factor. Theoretically the ideal situation would be that each test or pair of tests has a high loading on one factor only. Unfortunately this situation is seldom found in practice and even if it is fcund, it does not necessarily have the effect of most practical value (Jackson and Messick, 1967, p.317). It also happens that tests have relatively high loadings on the different factors, but then the communalities of the tests may be very low. It would have little value, for example, if the communality for a certain test is 0,32 and the test has a loading of 0,51 on a certain factor, since only about 30% of the variance of the test is explained by the factors.

The factor analyses of the tests of Form A have been discussed in paragraph 3.2.4.4. In general it may be said that the tests possess a fair amount of factorial validity, but owing to the fact that factorial validity was not used as a basis for the construction of the battery, pureness of factors cannot be claimed or expected. The main problem with factorial validity is that a test battery must be constructed in a specific way to possess high factorial validity.

In conclusion it can be mentioned that concepts which could also have received attention in the area of validity are "moderator" variables and "suppressor" variables. Attention will be given to these aspects in follow-up studies.

3.3 SUMMARY

In summary it can be said that this chapter pertains to the first application of the JATISA. For the first application a test battery consisting of two forms with twelve tests in each form was constructed. A brief description of the nature and purpose of the tests included in these forms is supplied. The composition of the sample employed for the first application is set forth, and a list of the names of the schools included in the sample are given in Appendix A.

The results of the first application are discussed, but before this is done a few methods of item analysis are presented. It is indicated that numerous methods of item analysis exist, ranging from the method where most value is attached to difficulty value and discrimination value, to the one which makes use of graphs. Apart from item analysis results, reference is also made to aspects like reliability, validity and factor analysis. Both the theoretical and practical sides of these aspects are presented and developments regarding these concepts are also indicated.

CHAPTER 4

CONSTRUCTION OF THE FINAL TEST BATTERY AND APPLICATION FOR NORM CALCULATION

4.1 CONSTRUCTION OF THE FINAL TEST BATTERY AND ULTIMATE SELECTION OF TESTS

In order to construct the final test battery a choice had to be made between the items and tests of Forms A and B. Although the original intention was to select the best items from each form and to combine the best possible items, this proved to be unnecessary after item analysis had been carried out. Except for tests 4, 7, 8 and 11, tests could be used almost in their entirety for inclusion in the final test battery, since most of the items from the two forms complied with the set requirements (see Appendices B-1 and B-2). However, it was decided to change a few of the items and a few of the distractors.

In the light of the fact that it was found that tests could be retained almost in their entirety, it was decided to give attention to the more global aspects of the tests also in deciding on the composition of the final test battery. The aspects referred to here are the means and reliabilities of the tests. The theoretic means of the tests were borne in mind (namely 12,5) as well as the reliability coefficients of the tests. With regard to the means it was decided to use the means of the middle group (i.e. Standard 7) as a guide to the general suitability of the tests in the two forms of the battery. This basis would ensure that neither the Standard 6 nor the Standard 8 group would be hampered with regard to differentiation ability. If it were decided to select Standard 6 as the group whose means should closely resemble the theoretic mean for example, then the means of the Standard 8 group would be too high and differentiation for them thus harmed and vice versa. It must be emphasized at this stage, however, that it was extremely coincidental that the tests could be kept as a whole and that this outcome had not been expected. Had this not been the case it would not have been possible to consider global aspects (statistics) in the composition of the final battery also.

The final battery consists of 7 tests from From A, 2 tests from Forms A and B which were changed radically, and 1 test which was added (namely a social insight test). The manner in which this structure was obtained is as follows:

In the *first place* a selection was made between the two verbal reasoning tests, one from Form A (test 1) and one from Form B (test 1). On the basis of means and item statistics it was decided to retain test 1 of Form A (see Table 4.1 and Appendices B-1 and B-2).

Secondly a choice was made between the four non-verbal reasoning tests, two from Form A (tests 2 and 3) and two from Form B (tests 2 and 3). On the basis of means and reliability indices (see Tables 4.1 and 3.2.4.2), it was decided to retain one of the tests of Form A. Test 2 of Form A was finally chosen on the basis of the acquaintance of Indian pupils with this type of test in contrast with the Blocks and Circles Test (test 3). Test 2 of Form B might have been a suitable test for the Standard 8 pupils (the mean being 12,12), but it would have been fairly difficult for the Standard 6 pupils (the mean being 9,52).

In the *third place* attention was given to a computational (numerical) test (test 4). As can be gathered from Table 4.1, the tests of both forms were too easy even on a Standard 6 level. The computational test for the final battery was consequently *altered significantly* with regard to difficulty value. Since it was also reasoned that the numerical reasoning factor wasn't fully acknowledged in the original test and that the test would thus have limited predictive value, it was decided to include the *reasoning aspect* in the next application of this test also.

In the *fourth place* it was decided to include a two-dimensional spatial relations test in the final battery. A choice was made between test 5 of Form A and test 5 of Form B (see Table 4.1). Test 5 of Form A tended to be a bit easy for the Standard 8 group (the mean being 15,79), whilst test 5 of Form B tended to be a bit difficult for the Standard 6 group (the

TABLE 4.1 MEANS AND STANDARD DEVIATIONS OF TESTS IN FORMS A AND B

Test[1)] No.	Form A						Form B					
	Mean			Standard deviation			Mean			Standard deviation		
	St. 6	St. 7	St. 8	St. 6	St. 7	St. 8	St. 6	St. 7	St. 8	St. 6	St. 7	St. 8
1	10,93	14,09	15,44	4,34	4,35	4,17	13,43	15,50	17,29	4,29	3,92	3,50
2	9,77	12,07	14,92	4,82	5,24	4,97	9,52	11,15	12,12	3,88	4,61	4,94
3	10,20	12,51	15,15	4,92	5,28	4,71	13,33	15,28	16,65	3,53	3,30	3,28
4	18,25	19,97	21,54	3,42	3,24	3,18	18,49	20,35	21,82	3,64	3,04	2,83
5	11,78	13,36	15,79	6,70	7,13	6,93	10,04	11,69	13,85	6,02	6,24	6,58
6	7,98	9,81	13,25	4,37	5,40	5,74	8,00	9,75	12,66	4,45	4,82	5,68
7	7,55	8,46	9,48	3,08	3,46	4,27	7,59	8,85	10,08	3,50	3,55	3,81
8	20,13	21,42	22,24	4,83	3,94	3,82	19,65	22,19	22,57	5,48	3,58	3,54
9	15,29	17,35	18,03	4,93	4,81	5,93	16,21	18,75	19,89	5,14	3,73	3,43
10	10,28	10,85	12,00	2,97	2,99	3,00	10,49	11,17	12,62	2,70	2,68	2,96
11	7,69	9,46	10,73	3,55	3,24	4,50	7,82	10,10	11,38	3,56	3,25	3,25
12	7,29	10,22	11,71	3,60	3,87	4,05	6,95	9,50	10,57	3,51	3,15	3,44

1) For full names of tests see page 32

mean being 10,04). Test 5 of Form A was retained for consistency but a few items were altered.

In the *fifth place* it was decided to include a three-dimensional spatial relations test in the final battery. A choice was made between tests 6 and 7 of the two forms. It was decided to retain test 6 of Form A for consistency. However, it is clear that test 6 of Form A and test 6 of Form B produced almost identical results with regard to all statistics that were examined.

In the *sixth place* it was decided to eliminate the Memory Test (test 8), in the first instance because of the means of the tests (see Table 4.1) and in the second place because doubts arose about the predictive value of this type of test with regard to a specific vocational group. The conclusion was reached that this test would definitely have more theoretical than practical value. It could be used in examining the structure of the intellect for instance, since this variable appears as a factor in Guilford's model of the structure of the intellect. A very small part of the test was retained in the Social Insight Test which will be discussed later.

Next it was decided to retain the Clerical Speed Test (test 9) of Form A, because of the means of the tests and because of item statistics (see Table 4.1, and Appendices B-1 and B-2). It was further decided to impose a heavier time limit for this test as a result of the high means obtained in the different standards. In essence this test is actually a speed test, so that the time factor can influence performance considerably.

In the *eighth place* it was decided to retain the Art Test of Form A (test 10) in the final battery for the sake of consistency. Here again there was little choice between the tests of the two forms. A small alteration was made in the first section of the test since certain aspects of art could be related to periods in history and certain items could have been influenced by this fact.

In the *ninth place* it was decided to present the Word Building Test (test 11) in *altered form* in the final test battery. This decision was mainly due to the difficulty level of the test (see Table 4.1), but also since the author was of the opinion that it would be difficult to determine whether the test measured pure linguistic ability or whether the test measured a combination of linguistic and creative ability. For these reasons the test was changed to ensure that more of a linguistic ability would be measured. The vocabulary aspect of the test was retained, but it was decided to include other aspects of the language factor in the test also, for example correct use of verbs, etc.

Penultimately it was decided to retain the Mechanical Insight Test of Form A (test 12) owing to means and reliability coefficients (see Tables 3.2.4.2 and 4.1). It was clear that especially the Standard 6 pupils found this test very difficult. Since this test forms a very important part of aptitude testing, however, it was decided to retain it in the final test battery. More attention could probably be given to the instructions with the Standard 6 pupils.

Ultimately it was decided to add a Social Insight Test since a demand for a test of this type often arises when vocational guidance is given. The importance of measuring social ability has been expressed by several writers, amongst whom are Thorndike (as in Kaluger and Unkovic, 1969, p.123), Guilford (as in Whitla, 1968, p.220), Hahn and MacLean (1955, p.173) and Munn (1954, p.573).

Thorndike was one of the first to refer to the existence of a social intelligence in dividing intelligence into abstract, mechanical and social areas. He defined social intelligence in terms of understanding and managing people.

Hahn and MacLean (1955, p.174) also mentioned several important aspects of social intelligence amongst which are the following:

"(i) Social intelligence is not a unitary trait although it may be a generalized trait.

(ii) Social intelligence is not a synonym for personality.

(iii) Social intelligence appears to be more loaded with environmental factors and learning than with heredity.

(iv) Social intelligence is not composed of a number of responses which invariably occur when similar situational stimuli are presented".

More recently Guilford (as in Whitla, 1968, p.220) has also given acknowledgement to the social aspect in intelligence by including it in his SI-model. He emphasized its existence by referring specifically to the behavioural *content* which is social in nature (as in Whitla, 1968, p.220).

Since the rationale of the Social Insight Test has not been given yet, it will briefly be given now:

Rationale of the Social Insight Test

This test is a measure of the ability to make an adequate adjustment in a social situation, to form associations between names and faces readily, and to maintain good social relations in general.

The higher mental processes are involved to a certain extent in this test but the main aspect is the ability to act correctly in different social situations.

This test ought to have predictive value for any of the fields of a social assisting nature, for example for professions like social worker, teacher and public relations officer.

In summary it can be stated that the final battery consisted of 7 tests from Form A, 3 which could just as well have been chosen from Form B (tests 6, 7 and 8); 2 from Forms A and B of which the contents were changed considerably (tests 4 and 5); and one which was added. After the tests had been selected from the two forms a few of the distractors were also modified.

The composition of the final test battery is summarized below:

Final Test Battery

Test no.	Test name	Abbreviation
1	Verbal Reasoning	V.R.
2	Series Completion (Non-verbal reasoning)	S.C.
3	Social Insight	S.I.
4	Language Usage (Modified Word Building Test)	L.U.
5	Numerical Reasoning (Modified Computations Test)	N.R.
6	Spatial Perception (2-D)	2-D
7	Spatial Perception (3-D)	3-D
8	Visual Arts	V.A.
9	Clerical Speed and Accuracy	C.S.& A.
10	Mechanical Insight	M.I.

4.2 RELATION BETWEEN TESTS OF THE JATISA AND FACTORS OF GUILFORD'S SI-MODEL

4.2.1 Introduction

Whether one is a believer in factor analysis or not, it is always good to compare the contents of an existing test with a theoretical model and to use such a model as a basis for reference. By doing this one not only becomes aware of the field which is covered by a certain instrument in terms of systematic components, but one also becomes acquainted with the method of analyzing an instrument in detail.

With regard to the JATISA it was decided to use Guilford's Structure-of-Intellect model (as in Whitla, 1968, p.220) as a frame of reference. There are also other models available for reference; but none of them are as comprehensive as Guilford's. The following table is a guide to Guilford's Structure-of-Intellect designation of factors. The three columns given, correspond with the three axes of Guilford's cubic model (see page 11 of thesis).

TABLE 4.2.1 GUIDE TO THE STRUCTURE-OF-INTELLECT DESIGNATION OF FACTORS

First Letter (Operation)	Second letter (Content)	Third letter (Product)
C - Cognition M - Memory D - Divergent production N - Convergent production E - Evaluation	F - Figural S - Symbolic M - Semantic B - Behavioural	U - Unit C - Class R - Relation S - System T - Transformation I - Implication

4.2.2 Relation of tests of the JATISA to factors of the SI-Model

The factors which have been related tentatively to the tests of the JATISA are as follows:

TEST 1 - Verbal Reasoning

CMR - Cognition of semantic relations
CSR - Cognition of symbolic relations
CSC - Cognition of symbolic classes
CMC - Semantic classification
CMS - General reasoning
ESR - Evaluation of symbolic relations

TEST 2 - Series Completion

CFR - Figural relations (cognition)
NFR - Figural relations (convergent production)

TEST 3 - Social Insight

MFU - Visual memory
MSR - Rote memory
EBS - Evaluation of behavioural systems

TEST 4 - Language Usage

EMU - Evaluation of semantic units
EMR - Evaluation of semantic relations
EMS - Evaluation of semantic systems

TEST 5 - Numerical Reasoning

NSI - Convergent production of symbolic implications

MSI - Numerical facility

DSR - Symbolic relations

TEST 6 - Spatial Perception (2-D)

EFU - Perceptual speed

CFT - Spatial visualization

TEST 7 - Spatial Perception (3-D)

CFT - Spatial visualization

CFR - Figural relations

TEST 8 - Visual Arts

CFU - Visual cognition

NFI - Figural implication

TEST 9 - Clerical Speed and Accuracy

ESU - Symbolic identification

CMC - Semantic classification

TEST 10 - Mechanical Insight

NFS - Convergent production of figural systems

NFI - Convergent production of figural implications

From the above it can be concluded that approximately twenty-five factors[1] or dimensions were gauged by the JATISA, which clearly indicates the heterogeneity of the instrument. This matter was not discussed in detail, however, since the JATISA did not rely on a factorially pure structure, for its composition.

---

1) It must be mentioned that the above relation of factors to tests is based on logical reasoning and not on factor analytical results.

4.3 ADMINISTRATION AND SCORING

4.3.1 Sample tested

During the first term of 1970 the test battery was applied in Indian schools for the purpose of norm calculation. The strata that were taken into consideration for the drawing of the sample were the same as for the first application, namely province, area (urban or rural) and sex. As an illustration the actual number of pupils tested in Standards 6, 7 and 8 according to province, area and sex is indicated in Table 4.3.1.

TABLE 4.3.1 TOTAL NUMBER OF PUPILS TESTED FOR CALCULATION OF NORMS

Province and area	Std. 6		Std. 7		Std. 8	
	Boys	Girls	Boys	Girls	Boys	Girls
Natal-Urban	178	151	209	152	205	135
Natal-Rural	127	92	107	83	110	64
Transvaal-Urban	44	36	48	36	66	42
Transvaal-Rural	6	5	8	6	10	6
Sub-total	355	284	372	277	391	247
Total	639		649		638	

4.3.2 Testing times

The testing times were based on the average time it took 90% of the pupils (in the different standards) to complete the different tests. The testing times for the different tests were as follows:

TEST 1 : Verbal Reasoning	:	25 minutes
TEST 2 : Series Completion	:	19 minutes
TEST 3 : Social Insight	:	17 minutes
Section A & Section B	:	14 minutes
Section C	:	3 minutes
TEST 4 : Language Usage	:	17 minutes

INTERVAL OF 15 MINUTES

TEST 5 : Numerical Reasoning : 28 minutes
TEST 6 : Spatial Perception (2-D) : 13 minutes
TEST 7 : Spatial Perception (3-D) : 19 minutes
TEST 8 : Visual Arts : 12 minutes
Sections A,B,C,D : 3 minutes each

INTERVAL OF 10 MINUTES

TEST 9 : Clerical Speed and Accuracy : 7 minutes
Section A : 2 minutes
Section B : 3 minutes
Section C : 2 minutes
TEST 10 : Mechanical Insight : 25 minutes

TOTAL TIME : 3 hours 27 minutes

4.3.3 Scoring of answer sheets

Two NB 758 answer sheets were used by each pupil to answer the tests. The answer sheets could either be scored by hand in which case a set of scoring keys would be required, or by means of an IBM 1230 scoring-machine. In this particular case the answer sheets were scored by means of a computer. Although the answer sheets of the Group Intelligence Test were used in this instance, it can be foreseen that answer sheets designed specifically for the JATISA will be made available in the future.

4.4 TEST RESULTS

Item analysis was conducted again owing to the alteration of certain tests of the original forms and owing to the addition of the Social Insight Test. In general the results were very favourable. However it was decided to impose a heavier time limit for the Social Insight Test in future investigations. The remaining results of importance will now be discussed.

4.4.1 Reliability and standard error of measurement

4.4.1.1 Reliability

Reliability data for raw scores on the ten tests of the JATISA, for Standards 6 to 8 are provided in Tables 4.4.1(a), (b) and (c). The data include coefficients of reliability, standard errors of measurement, means and standard deviations. The reliability coefficients were calculated by means of the Kuder-Richardson formula 20 and are thus homogeneity indices. Reliability coefficients range from 0,539 to 0,918 for Standard 6, from 0,526 to 0,929 for Standard 7, and from 0,657 to 0,940 for Standard 8.

TABLE 4.4.1(a) RELIABILITY COEFFICIENTS (K-R 20), ERRORS OF MEASUREMENT, MEANS AND STANDARD DEVIATIONS (STD. 6)

Test no.	Test name[1]	$R^{2)}$	Error of measurement: Raw Scores	Error of measurement: Stanines	$\bar{X}^{3)}$	$s^{4)}$	
1	V.R.	0,734	2,121	1,414	10,703	4,115	
2	S.C.	0,821	2,112	1,419	9,594	4,992	
3	S.I.	0,757	2,128	1,428	12,734	4,318	
4	L.U.	0,630	2,229	1,425	12,726	3,665	
5	N.R.	0,689	2,165	1,429	9,381	3,697	(N=639)
6	2-D	<u>0,918</u>	2,023	1,424	13,844	<u>7,055</u>	
7	3-D	0,751	2,152	1,416	8,672	4,538	
8	V.A.	<u>0,539</u>	2,235	1,429	10,940	<u>3,267</u>	
9	C.S.&A.	0,796	2,014	1,410	12,215	4,459	
10	M.I.	0,582	2,235	1,436	9,402	3,445	

1) For full names see page 81
2) = reliability coefficient
3) = mean
4) = standard deviation

TABLE 4.4.1(b) RELIABILITY COEFFICIENTS (K-R 20), ERRORS OF MEASUREMENT, MEANS AND STANDARD DEVIATIONS (STD. 7)

Test no.	Test[1] name	$R^{2)}$	Error of measurement		$\bar{X}^{3)}$	$S^{4)}$	
			Raw Scores	Sta-nines			
1	V.R.	0,680	2,154	1,425	14,556	3,790	
2	S.C.	0,838	2,160	1,427	12,779	5,368	
3	S.I.	0,721	2,070	1,432	15,759	3,919	
4	L.U.	0,642	2,132	1,413	15,832	3,562	
5	N.R.	0,721	2,118	1,433	12,940	3,702	(N=649)
6	2-D	0,929	1,926	1,430	15,466	7,231	
7	3-D	0,811	2,161	1,432	11,277	5,252	
8	V.A.	0,526	2,197	1,426	12,824	3,169	
9	C.S.&A.	0,771	2,038	1,441	14,752	4,260	
10	M.I.	0,606	2,229	1,429	11,699	3,537	

1) For full names see page 81

2) = reliability coefficient

3) = mean

4) = standard deviation

TABLE 4.4.1(c) RELIABILITY COEFFICIENTS (K-R 20), ERRORS OF MEASUREMENT, MEANS AND STANDARD DEVIATIONS (STD. 8)

Test no.	Test[1)] name	$R^{2)}$	Error of measurement		$\bar{X}^{3)}$	$s^{4)}$	
			Raw Scores	Sta-nines			
1	V.R.	0,699	2,112	1,422	15,777	3,849	
2	S.C.	0,856	2,073	1,419	15,195	5,462	
3	S.I.	<u>0,657</u>	1,903	1,407	18,722	<u>3,174</u>	
4	L.U.	0,626	2,044	1,414	17,604	3,340	
5	N.R.	0,745	1,940	1,426	15,841	3,576	(N=638)
6	2-D	<u>0,940</u>	1,714	1,436	18,187	<u>7,015</u>	
7	3-D	0,853	2,085	1,422	14,671	5,681	
8	V.A.	0,647	2,117	1,423	13,739	3,495	
9	C.S.&A.	0,750	1,948	1,398	17,267	3,897	
10	M.I.	0,683	2,181	1,424	13,593	3,816	

1) For full names see page 81

2) = reliability coefficient

3) = mean

4) = standard deviation

In the light of test homogeneity some of these indices, are rather low. However, as is known, item homogeneity does not ensure high retest reliability and furthermore an assumption with the K-R formula 20 is that the items are at the same difficulty level. The latter aspect can be rectified to a certain extent by making use of Horst's formula, Loevinger's formula or Tucker's formula, but the first aspect is unchangeable. Tucker's formula (Psychometrika, 1949, p.117) was computed for the data of Tables 4.4.1(a), (b) and (c) for the sake of interest, but in none of the cases did the reliability coefficient increase by more than 0,05.

On studying the formula for the K-R 20 reliability index it will be noticed that this index is mainly a function of test variance. From the tables it will also be seen that for each standard the test with the lowest variance has the lowest reliability (except Standard 8, test 3) and the test with the highest variance has the highest reliability. This implies that the variance of a test can be used to a certain extent to make a prediction of the reliability of a test (if the other factors are held constant). Thus it can be reasoned that the nature of the group in contrast to the nature of the test might determine the size of the K-R reliability index. Unfortunately this implies that the test could have a higher reliability for a general group than for a specific group. Generally speaking, however, this problem should not cause serious concern since the difficulty level of the test will also be different for the two groups and since certain corrections can be made for restriction of range.

Although test-retest reliability coefficients cannot be reported at this stage, these coefficients will have to be determined for tests 6 and 9 in any event, since they are in essence speed tests.

4.4.1.2 Standard error of measurement

In interpreting individual scores, the standard error of measurement is more helpful than a reliability coefficient. Amongst other things it has the advantage that it can be interpreted in terms of probability statements. In this connection Seibel (as in Whitla, 1968, p.274) makes the following statement: "The outstanding advantage of using the standard error of measurement is that it enables us to make certain probability statements about an individual's "true score", given the score he actually received". The standard error of measurement for a test is an estimate of the amount of error (in test score units) associated with the score a person receives on a test, or, stated in another way, it is an indication of the variation in an individual's score which may be expected by chance. As an example, if the standard error for a given test is 3, and an individual obtains a score of 20 on the same test then one can conclude with about 68 per cent confidence, that the person's true score lies within the band 17-23 (i.e. between one standard error above and one standard error below the obtained score).

The formula used for obtaining the standard error of measurement is as follows (Whitla, 1968, p.275):

$$S_e = S_x \sqrt{1-r} \text{ , where}$$

S_e = standard error of measurement

S_x = standard deviation of group

r = reliability coefficient of test.

The alternative method of obtaining the standard error of measurement is to examine the differences between the scores on two performance samples, and to obtain the standard deviation of these differences (Whitla, 1965, p.273). This is related to the test-retest method of reliability determination but is usually considered as impractical and uneconomical. However, this is actually the proper method for determining the standard error of measurement, whereas the method described previously yields an estimate of the standard error of measurement only.

4.4.2 Intercorrelations

Tables 4.4.2(a), (b) and (c) list the intercorrelations for the ten tests of the test battery for Standards 6, 7 and 8 separately. The formula used to compute these intercorrelations is the familiar Product-moment correlation formula.

Although many correlations are significant at the 0,05 and the 0,01 level, significance has only been attached to correlations bordering on 0,5 and larger than 0,5. The motivation for this cut-off point is that a correlation coefficient of 0,5 accounts for 25 per cent common variance, which seems a reasonable figure for a lower limit. In the light of this cut-off point the following correlations can be considered (see Tables 4.4.2(a), (b) and (c)).

TABLE 4.4.2(a) INTERCORRELATIONS OF TESTS FOR STD. 6

Test[1)] No.	1	2	3	4	5	6	7	8	9	10
1	1,000									
2	0,520	1,000								
3	0,453	0,307	1,000							
4	0,428	0,255	0,412	1,000						
5	0,525	0,480	0,317	0,295	1,000					(N=639)
6	0,260	0,248	0,158	0,124	0,310	1,000				
7	0,290	0,356	0,270	0,207	0,321	0,230	1,000			
8	0,297	0,360	0,291	0,262	0,305	0,277	0,307	1,000		
9	0,324	0,323	0,371	0,318	0,348	0,256	0,269	0,378	1,000	
10	0,357	0,357	0,293	0,321	0,422	0,211	0,321	0,347	0,290	1,000

1) Test 1 - Verbal Reasoning
Test 2 - Series Completion
Test 3 - Social Insight
Test 4 - Language Usage
Test 5 - Numerical Reasoning
Test 6 - Spatial Perception (2-D)
Test 7 - Spatial Perception (3-D)
Test 8 - Visual Arts
Test 9 - Clerical Speed and Accuracy
Test 10 - Mechanical Insight

TABLE 4.4.2(b) INTERCORRELATIONS OF TESTS FOR STD. 7

Test[1)] No.	1	2	3	4	5	6	7	8	9	10
1	1,000									
2	0,509	1,000								
3	0,437	0,346	1,000							
4	0,373	0,300	0,456	1,000						
5	0,481	0,472	0,352	0,317	1,000					
6	0,198	0,303	0,112	0,141	0,254	1,000				(N=649)
7	0,364	0,410	0,195	0,160	0,350	0,208	1,000			
8	0,268	0,395	0,198	0,143	0,313	0,280	0,360	1,000		
9	0,316	0,355	0,239	0,223	0,358	0,208	0,339	0,308	1,000	
10	0,385	0,435	0,250	0,277	0,411	0,305	0,357	0,373	0,254	1,000

1) Test 1 - Verbal Reasoning
Test 2 - Series Completion
Test 3 - Social Insight
Test 4 - Language Usage
Test 5 - Numerical Reasoning
Test 6 - Spatial Perception (2-D)
Test 7 - Spatial Perception (3-D)
Test 8 - Visual Arts
Test 9 - Clerical Speed and Accuracy
Test 10 - Mechanical Insight

TABLE 4.4.2(c) INTERCORRELATIONS OF TESTS FOR STD. 8

Test No. [1]	1	2	3	4	5	6	7	8	9	10
1	1,000									
2	0,583	1,000								
3	0,379	0,300	1,000							
4	0,370	0,296	0,368	1,000						
5	0,546	0,510	0,296	0,355	1,000					(N=638)
6	0,356	0,391	0,133	0,105	0,370	1,000				
7	0,480	0,514	0,265	0,187	0,471	0,378	1,000			
8	0,329	0,405	0,220	0,128	0,305	0,362	0,446	1,000		
9	0,411	0,414	0,280	0,236	0,369	0,290	0,382	0,341	1,000	
10	0,425	0,526	0,188	0,216	0,417	0,365	0,504	0,443	0,394	1,000

1) Test 1 - Verbal Reasoning
Test 2 - Series Completion
Test 3 - Social Insight
Test 4 - Language Usage
Test 5 - Numerical Reasoning
Test 6 - Spatial Perception (2-D)
Test 7 - Spatial Perception (3-D)
Test 8 - Visual Arts
Test 9 - Clerical Speed and Accuracy
Test 10 - Mechanical Insight

(i) For the Standard 6 group the correlations between variables 1 and 2, 1 and 5, and 2 and 5 can be regarded as significant. Since the variables 1, 2 and 5 are all related to a reasoning factor, these correlations are quite explicable and conform with the constructs which were hypothesized when composing the battery.

(ii) For Standard 7 the correlations to which importance can be attached are exactly the same as for Standard 6, with the result that the explanations for the correlations remain the same.

(iii) For the Standard 8 group there are eight correlations to which importance can be attached. They are the following:

In the second column between variables 1 and 2, 1 and 5, 1 and 7;
In the third column between variables 2 and 5, 2 and 7, 2 and 10;
In the sixth column between variables 5 and 7;
In the eigth column between variables 7 and 10.

The correlations of importance in the second column can easily be ascribed to a common reasoning factor and those of the third, sixth and eighth columns mainly to a common spatial perception factor. The correlations between variables 2 and 10 (0,526), and 7 and 10 (0,504) for Standard 8 confirm the theory that mechanical insight is composed of a reasoning factor and a spatial factor.

4.4.3 Factor Analysis

4.4.3.1 Introduction

Factor analysis was done on the ten tests of the test battery. The "Principal Factor Analysis" method was used, and the axes were rotated according to the "Varimax" criterion. Significance was attached to factor loadings in the vicinity of, and larger than, 0,4.

According to Fruchter (1958, p.151) factor loadings of 0,2 and less can be regarded as insignificant, of 0,3 to 0,5 as reasonable, 0,5 to 0,7 as high, and higher than 0,7 as very high. This is in agreement with the viewpoint of Du Bois. He laid down 0,3 as a lower cut-off point (1965, p.250). The above criteria all refer to large samples and are thus applicable to this study.

The results obtained in the factor analyses are given in Tables 4.4.3(a), (b) and (c).

4.4.3.2 Description of results

On studying the results the points mentioned in Chapter 3 should be borne in mind again. A description of the results obtained now follows:

(i) For Standard 6, factor I was identified as a *perceptual factor* owing to the loadings of variables 7 (Spatial Perception), 8 (Visual Arts), and 9 (Clerical Speed) on this factor. Factor II was identified as a *language factor* owing

TABLE 4.4.3(a) FACTOR ANALYSIS OF TESTS FOR STD. 6

Test no.	Test name[1)]	Commu-nality	Relia-bility	Factor loading			
				I	II	III	IV
1	V.R.	0,616	0,734	0,215	-0,386	-0,615	-0,204
2	S.C.	0,495	0,821	0,202	-0,235	-0,348	0,527
3	S.I.	0,426	0,692	0,218	-0,566	-0,172	0,166
4	L.U.	0,381	0,630	0,176	-0,551	-0,176	0,125
5	N.R.	0,501	0,689	0,258	-0,213	-0,454	0,427
6	2-D	0,225	0,918	0,254	-0,093	-0,671	0,383
7	3-D	0,282	0,751	0,398	-0,127	-0,221	0,241
8	V.A.	0,372	0,539	0,506	-0,210	-0,110	0,245
9	C.S.&A.	0,366	0,804	0,464	-0,322	-0,156	0,149
10	M.I.	0,331	0,582	0,340	-0,257	-0,237	0,306
	Latent root			3,288	0,330	0,272	0,103
	Difference			2,958	0,058	0,169	
	Percentage of factor variance			10,37	11,08	9,09	9,21

1) For full names see page 81

TABLE 4.4.3(b) FACTOR ANALYSIS OF TESTS FOR STD. 7

Test no.	Test name[1)]	Commu-nality	Relia-bility	Factor loading				
				I	II	III	IV	V
1	V.R.	0,520	0,680	0,235	-0,466	-0,310	-0,388	-0,022
2	S.C.	0,515	0,838	0,434	-0,307	-0,386	-0,286	-0,043
3	S.I.	0,455	0,618	0,107	-0,624	-0,183	-0,071	-0,128
4	L.U.	0,414	0,642	0,149	-0,608	-0,075	-0,072	0,105
5	N.R.	0,451	0,720	0,343	-0,368	-0,325	-0,300	0,052
6	2-D	0,225	0,929	0,439	-0,089	-0,149	-0,049	0,017
7	3-D	0,377	0,811	0,314	-0,109	-0,475	-0,194	0,049
8	V.A.	0,372	0,526	0,447	-0,114	-0,365	-0,052	0,156
9	C.S.&A	0,317	0,727	0,231	-0,227	-0,450	-0,052	-0,084
10	M.I.	0,441	0,606	0,540	-0,228	-0,219	-0,173	-0,138
	Latent root			3,249	0,509	0,142	0,106	0,081
	Difference			2,740	0,367	0,036	0,025	
	Percentage of factor variance			15,17	13,36	10,15	4,03	0,82

1) For full names see page 81

TABLE 4.4.3(c) FACTOR ANALYSIS OF TESTS FOR STD. 8

Test no.	Test[1)] name	Commu-nality	Relia-bility	Factor loadings			
				I	II	III	IV
1	V.R.	0,583	0,699	0,394	0,444	-0,458	-0,146
2	S.C.	0,568	0,856	0,554	0,319	-0,390	-0,084
3	S.I.	0,367	0,519	0,135	0,406	-0,145	-0,403
4	L.U.	0,366	0,626	0,108	0,579	-0,121	-0,069
5	N.R.	0,521	0,745	0,373	0,310	-0,505	-0,175
6	2-D	0,326	0,940	0,505	0,075	-0,255	-0,019
7	3-D	0,523	0,853	0,643	0,249	-0,219	-0,010
8	V.A.	0,406	0,647	0,607	0,121	-0,052	-0,140
9	C.S.&A.	0,341	0,740	0,452	0,310	-0,147	-0,136
10	M.I.	0,502	0,683	0,608	0,085	-0,281	-0,214
	Latent root			3,728	0,502	0,156	0,117
	Difference			3,226	0,346	0,039	
	Percentage of factor variance			18,70	10,77	8,64	3,06

1) For full names see page 81

to the loadings of variables 3 (Social Insight) and 4 (Language Usage) on this factor. As far as factors III and IV are concerned it can be postulated that, if fewer factors had been extracted, these factors would probably have converged and formed a single *reasoning factor*. The variables of relevance here are variables 1 (Verbal Reasoning), 2 (Series Completion) and 5 (Numerical Reasoning).

(ii) For Standard 7, it can be postulated that factors 1 and 3 would probably have combined into one factor namely a *perceptual factor*, if fewer factors had been extracted. The variables concerned here are variables 2 (Series Completion), 6 (Spatial Perception), 7 (Spatial Perception), 8 (Visual Arts), 9 (Clerical Speed and Accuracy) and 10 (Mechanical Insight). Further, factor II was identified as a *language factor*, owing to the loadings of variables 1 (Verbal Comprehension), 3 (Social Insight) and 4 (Language Usage) on this factor. Finally factor 4 was identified as a *reasoning factor* because of the loading of variable 1 on this factor. Factor V was not interpreted since the loadings on this factor were very low.

(iii) For Standard 8, factor 1 was identified as a *perceptual* factor owing to the loadings of variables 2 (Series Completion), 6 (Spatial Perception 2-D), 7 (Spatial Perception 3-D), 8 (Visual Arts), 9 (Clerical Speed and Accuracy) and 10 (Mechanical Insight) on this factor. Factor II was identified as a *language factor* because of the loadings of variables I (Verbal Comprehension), 3 (Social Insight), and 4 (Language Usage) on this factor. Factor III was identified as a *reasoning factor* owing to the loadings of variables 1, [illegible] and 5 on this factor. Factor IV was identified as a *social factor* owing to the loading of variable 3 (Social Insight) on this factor.

Further it can be postulated that a numerical factor would probably also have emerged if use had been of a marker test, and if the Numerical Reasoning Test had made less use of reasoning. It can also be noted that the factors which were identified here are typical factors. This fact becomes evident when, for instance, one inspects the Holzinger Crowder Uni-factor Tests and the Multi Aptitude Test Battery (Bauernfeind, 1963, p.202).

Finally it can be mentioned that approximately forty per cent of the variance of the tests was accounted for by the factors with regard to each standard group. It is thus clear that a large amount of variance was not accounted for, as can be expected owing to the lack of marker variables.

4.4.4 Validity

The types of validity which received attention during the second application of the JATISA are as follows:

4.4.4.1 Factorial Validity

Factorial validity was determined by factorising the intercorrelations of the various tests. The results of the factor analyses are given in Tables 4.4.3(a), (b) and (c). In general it can be stated that the factorial validities of the tests are reasonably high. The factor loadings of the tests are reasonable but the communalities of the tests are not very impressive. As was mentioned earlier, however, this battery was not based upon factorial pureness for its composition.

4.4.4.2 Construct Validity

This type of validity has received considerable attention during the past few years. Since this construct (aptitude) involves the higher mental processes there should be a *reasonable* correlation between scores obtained in the aptitude tests and school marks for instance. This is confirmed by the correlations indicated in Tables 4.4.4(a), (b) and (c). If the battery had been constructed to determine scholastic aptitude the intercorrelations should have been much higher.

4.4.4.3 Predictive Validity

Although much attention has not been given to the predictive validity of the test battery yet, the need for this is very urgently felt. It is planned to determine in the near future, the relation between the original predictions for a number of pupils, and their eventual vocational choices. In addition the predictive value of aptitude tests in relation to school achievement can also be determined. The predictive aspect mentioned here actually refers to differential prediction (Cronbach, 1970, p.371). Aptitude test results and school marks can be correlated and different profiles drawn up. In addition multiple regression equations can also be used for prediction of school success on the basis of the aptitude test results. A small study in this connection has already been carried out and the results are quite favourable (see Tables 4.4.4(a), (b) and (c)). This study refers to the Transvaal only and was conducted during 1970. A description of some of the results obtained now follows:

TABLE 4.4.4(a) CORRELATIONS BETWEEN TESTS AND SCHOOL SUBJECTS - STD. 6 (N = 89, EXCEPT FOR ART N = 40)

Test name	School subjects and average						
	English	Arithmetic	Mathematics	Art	General Science	Book keeping	Average
V.R.	0,336[1]	-	0,368	0,375	0,286	-	0,468
S.C.	0,329	-	0,415	0,300	0,403	-	0,409
S.I.	0,279	-	0,180	0,112	0,208	-	0,342
L.U.	0,269	-	0,415	0,209	0,344	-	0,469
N.R.	0,190	-	0,353	0,256	0,289	-	0,361
2-D	0,072	-	0,172	0,276	0,144	-	0,127
3-D	-0,006	-	0,170	0,242	0,218	-	0,157
V.A.	0,118	-	0,365	0,366	0,405	-	0,336
C.S.&A	0,346	-	0,330	0,099	0,464	-	0,419
M.I.	0,317	-	0,450	0,195	0,541	-	0,530

1) A few of the largest correlation indices are underlined in each column.

TABLE 4.4.4(b) CORRELATIONS BETWEEN TESTS AND SCHOOL SUBJECTS - STD. 7 (N = 96, EXCEPT FOR ART N = 40)

Test name	School subject and average						
	English	Arithmetic	Mathematics	Art	General science	Bookkeeping	Average
V.R.	0,263	0,322	0,231	0,273	0,436	0,250	0,347
S.C.	0,316	0,219	0,303	0,557	0,469	0,368	0,396
S.I.	0,266	0,689	0,130	-0,026	0,267	0,124	0,245
L.U.	0,437 1)	0,217	0,265	0,147	0,293	0,329	0,398
N.R.	0,205	0,480	0,476	0,602	0,547	0,508	0,554
2-D	0,305	0,202	0,163	0,193	0,309	0,181	0,252
3-D	0,323	0,539	0,501	0,513	0,370	0,464	0,537
V.A.	0,099	0,373	0,110	-0,161	0,323	0,106	0,236
C.S.&A	0,210	0,353	0,189	0,062	0,426	0,207	0,308
M.I.	0,365	0,271	0,344	0,473	0,340	0,438	0,428

1) A few of the largest correlation indices are underlined in each column.

TABLE 4.4.4(c) CORRELATIONS BETWEEN TESTS AND SCHOOL SUBJECTS - STD. 8 (N = 119)

Test name	School subject and average						
	English	Arithmetic	Mathematics	Art	General science	Bookkeeping	Average
V.R.	0,342	0,414	0,353	-	0,336	0,307	0,504
S.C.	0,278	0,375	0,296	-	0,330	0,298	0,456
S.I.	0,319	0,065	0,155	-	0,324	0,244	0,316
L.U.	0,484 1)	0,109	0,291	-	0,363	0,309	0,457
N.R.	0,354	0,486	0,443	-	0,388	0,464	0,596
2-D	0,143	0,325	0,250	-	0,242	0,276	0,352
3-D	0,246	0,207	0,502	-	0,275	0,270	0,361
V.A.	0,210	0,368	0,304	-	0,208	0,156	0,318
C.S.&A.	0,203	0,359	0,330	-	0,292	0,308	0,994
M.I.	0,139	0,295	0,330	-	0,230	0,133	0,288

1) A few of the largest correlation indices are underlined in each column.

(i) For the Standard 6 group the correlations of significance are the following:

(a) In column 2 between English and Verbal Reasoning (0,336), and between English and Clerical Speed and Accuracy (0,346). These correlations are quite acceptable on the basis of logical reasoning although a higher correlation could be expected between English and Language Usage (0,269). Several explanations can be offered for the latter finding, one of which is that the Standard 6 pupils might encounter language problems and might find language as a school subject closely associated with the reasoning process.

(b) In column 4 between Mathematics and Series Completion (0,415), Mathematics and Language Usage (0,415) and Mathematics and Mechanical Insight (0,450). Two of these correlations have logical foundations. However, the third, namely the correlation between Mathematics and Language Usage once again could point to language difficulties being encountered.

(c) In column 5 between Art and Verbal Reasoning (0,375) and between Art and Visual Arts (0,366). Logical bases can be found for these correlations but it must be mentioned that Art as a school subject is not a very general subject.

(d) In column 6 between General Science and Clerical Speed (0,464), and between General Science and Mechanical Insight (0,541). It was suspected that the factor responsible for the first correlation mentioned was the perceptive factor which could be common to scientific word and clerical work. However, when the effect of spatial perception (test 7) was "partialled" out, the correlation coefficient only decreased by 0,02.

(e) In column 8 the correlations of importance are those between Average and Verbal Reasoning (0,468), Average and Language Usage (0,469), and Average and Mechanical Insight (0,530). The second correlation can once again be ascribed to language difficulties, and the third correlation to the common reasoning factor, which is one aspect of mechanical insight (the other being spatial visualization), as well as to the language factor. When the influence of the language factor was "partialled" out in the latter correlation, the coefficient decreased by 0,06.

(ii) For the Standard 7 group the correlations of significance are the following:

(a) In column two between English and Language Usage (0,437) and between English and Mechanical Insight (0,365). The latter correlation can be ascribed to the language aspect of the Mechanical Insight Test.

(b) In column three between Arithmetic and Numerical Reasoning (0,480) and between Arithmetic and Spatial Perception (0,539). Both of these correlations have logical basis.

(c) In column four between Mathematics and Numerical Reasoning (0,476) and between Mathematics and Spatial Perception (0,501). These correlations have a very logical basis also.

(d) In column five between Art and Series Completion, (0,557) and between Art and Numerical Reasoning (0,602). These correlations are quite explicable but a correlation which is difficult to explain is the correlation between Art and Visual Arts (-0,167), which are two related spheres. It must be remembered, however, that Art as it is taught in most schools is not an examination subject.

(e) In column six between General Science and Series Completion (0,469), and between General Science and Numerical Reasoning (0,547). These correlations are also logically based.

(f) In column seven between Bookkeeping and Numerical Reasoning (0,508), and Bookkeeping and Spatial Perception (0,464). The latter correlation is perhaps a bit difficult to explain.

(g) In column eight between Average and Numerical Reasoning (0,554) and between Average and Spatial Perception (0,537). These correlations can also be explained on a logical basis.

(iii) For the Standard 8 group the correlations of significance are the following:

(a) In column two between English and Language Usage (0,484). This is a logical correlation.

(b) In column three between Arithmetic and Verbal Reasoning (0,414), and Arithmetic and Numerical Reasoning (0,486). These correlations also have very logical bases.

(c) In column four between Mathematics and Numerical Reasoning (0,443), and Mathematics and Spatial Perception (0,502). The latter correlation can probably be ascribed to the acquaintance with geometry, woodwork, etc. in the higher standards.

(d) In column six between General Science and Language Usage (0,363), and General Science and Numerical Reasoning (0,388). The latter correlation is quite reasonable but the former is a bit difficult to explain.

(e) In column seven between Bookkeeping and Numerical Reasoning (0,464). This correlation is very logically based.

(f) In column eight between Average and Verbal Reasoning (0,504), and Average and Numerical Reasoning (0,596). These correlations can be described as typical correlations with regard to aptitude tests and achievement tests.

In summary it can be stated that some of the correlations obtained are quite remarkable in view of the fact that the aptitude test battery is a general aptitude test battery, and not a scholastic aptitude test battery.

Further it must be mentioned that the value of the correlations obtained can also be increased substantially by making use of multiple correlations. The following tables serve as an illustration of this statement:

TABLE 4.4.4(d) MULTIPLE CORRELATIONS OF APTITUDE TESTS WITH SCHOOL AVERAGE (STD. 7)

R[1)]	N	VA.NR.[2)] & (Beta weight)
0,660	96	5(0,412) 7(0,386)
0,682	96	5(0,356) 7(0,362) 4(0,186)
0,687	96	5(0,388) 7(0,389) 4(0,217) 1(-0,108)
0,693	96	5(0,386) 7(0,421) 4(0,176) 1(-0,171) 3(0,120)
0,696	96	5(0,411) 7(0,435) 4(0,184) 1(-0,171) 3(0,125) 9(0,072)

1) R = Multiple correlation coefficient
2) VA.NR. = Variable Number
VA.NR.5 = Numerical Reasoning
VA.NR.7 = Spatial Perception (3-D)
VA.NR.4 = Language Usage
VA.NR.1 = Verbal Reasoning
VA.NR.3 = Social Insight
VA.NR.9 = Clerical Speed

TABLE 4.4.4(e) MULTIPLE CORRELATIONS OF APTITUDE TESTS WITH ENGLISH (STD. 7)

R[1)]	N	VA.NR.[2)] & (Beta weight)
0,705	119	1(0,392) 5(0,380)
0,739	119	1(0,296) 5(0,322) 10(0,264)
0,767	119	1(0,256) 5(0,247) 10(0,260) 4(0,231)
0,778	119	1(0,181) 5(0,204) 10(0,246) 4(0,218) 2(0,182)
0,782	119	1(0,174) 5(0,194) 10(0,243) 4(0,178) 2(0,169) 3(0,096)

1) R = Multiple correlation coefficient
2) VA.NR. = Variable Number
VA.NR.1 = Verbal Reasoning
VA.NR.5 = Numerical Reasoning
VA.NR.10= Mechanical Insight
VA.NR.4 = Language Usage
VA.NR.2 = Series Completion
VA.NR.3 = Social Insight

In Table 4.4.4(d) it can be seen that the multiple correlation coefficient (for Standard 7) can be raised from 0,660 when using two variables to 0,696 when using six variables. From a *practical* viewpoint, however, the best combination would probably be the use of variables 5,7 and 4 which provides a multiple correlation coefficient of 0,682. Regression equations will in future be supplied for predicting school average. In Table 4.4.4(e) a substantial difference is again observed in the correlation coefficient when using two variables and when using six variables. The most *practical* combination would probably be combination 3 which provides a correlation coefficient of 0,767. Regression equations will in future be supplied for the most efficient combinations (of aptitude tests) and this will serve as a basis for prediction of success in school subjects.

4.4.5 Comparative prediction

This aspect was actually tentatively referred to in the previous paragraph, but since it is of such vital importance in aptitude testing it will be discussed in slightly more detail now.

One approach to the problem of comparative prediction relates to the use of the multiple regression technique, by means of which it is possible to predict either differences between criterion scores or the separate criterion scores directly. When the prediction relates to differences between criterion scores, one obtains an indication of which of two occupational groups an individual belongs to, and when the prediction is of the criterion scores separately one obtains an estimate of the probability of success in a particular occupation.

In either of the above cases the form of the prediction equation will be:

$$Y_i = a_i + b_{i1}X_1 + b_{i2}X_2 + \dots\dots\dots + b_{ij}X_j,$$

where Y_i represents a criterion of success in occupation i (or difference between criterion scores), the a_i, b_{i1}, ... b_{ij} represent the regression weights, and X_1,

X_2, ... X_j represent the aptitude test scores. By considering the above equation one can also determine some of the limitations of the method. The first assumption made is that a linear relationship exists between test scores and criterion. Violation of this assumption will only have the effect of less accurate prediction. The second assumption made is that exceptionally high ability in one area can compensate for low ability in a second. Failure to meet this assumption will have more serious consequences. Helmstadter (1964, pp.218-219) gives the following example in this connection:

Suppose the prediction equation for a particular criterion is

$$Y = 5X_1 + 10X_2 + 7X_3, \text{ and}$$

two individuals receive the following set of scores:

	X_1	X_2	X_3
Person A	10	5	0
Person B	7	3	5 .

For both persons Y = 100, so that the same degree of success will be predicted for person A and person B. However, person A does not even possess a minimum of the ability measured by X_3 and surely the aptitude X_3 is important for the particular vocation or it would have a zero weight.

The above error is actually the kind of error for which the multiple cut-off approach was specifically designed to avoid. However the latter method also has certain disadvantages.

Since each approach has certain disadvantages it was decided to investigate one of the approaches which seemed reasonable. A number of multiple correlation coefficients relating achievement in the aptitude tests to achievement in school subjects, were computed (regression equations will be determined later). The results obtained in this connection have already been discussed in the previous paragraph. It is planned to make use of other approaches like the Multiple Discriminant Function and the Multiple Cut-off method in future also, since each approach has certain distinct advantages.

4.5 NORMS

4.5.1 Basis for norm groups

As a basis for internal norm differentiation use was made of the significance of differences between means for different groups. The significance of differences was obtained for the separate standards with regard to the qualities sex, province and region (urban or rural). To determine these differences a homogeneity of variance analysis was first conducted, and thereafter the appropriate t-formulas were used for establishing the differences. The t-formulas that were used are the following (Popham, 1967, p.148):

(i) $n_1 \neq n_2$ and $s_1^2 = s_2^2$; pooled variance formula with $df = n_1 + n_2 - 2$:

$$t/z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{\Sigma x_1^2 + \Sigma x_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where, $\bar{X}_1$, $\bar{X}_2$ = means

x_1^2, x_2^2 = squares of deviations from means

n_1, n_2 = sizes of samples

(ii) $n_1 \neq n_2$ and $s_1^2 \neq s_2^2$; separate variance formula with tabled t value determined by averaging t values for (a) degrees of freedom equal to n_1-1 and (b) degrees of freedom equal to n_2-1:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where, $\bar{X}_1$, $\bar{X}_2$ = means

s_1^2, s_2^2 = variances

n_1, n_2 = sizes of samples.

Very few differences were found with regard to province and region. However, a substantial number of differences were found with regard to sex (see Tables 4.5.1(a), (b) and (c)). Of the thirty differences which were examined for the three groups, 19 were significant at the 5% or 1% levels. In addition tests 5, 6 8, and 10 indicated significant differences for all three standards. These tests are the Numerical Reasoning Test, the Spatial Perception Test (2-D), the Art Test and the Mechanical Insight Test. In all these cases the results of the boys eclipsed those of the girls. These findings are logically based, except for the Art Test where one would expect the girls to perform better. As a result of these findings it was decided to establish separate norms for boys and girls.

TABLE 4.5.1(a) MEANS AND STANDARD DEVIATIONS AND T-VALUES FOR STD. 6 ACCORDING TO SEX

Test no.	Males			Females			Type of test
	X_1	S_1	N_1	X_2	S_2	N_2	t/z(two sided)
1	10,788	4,110	353	10,599	4,109	284	0,576
2	9,818	5,075	353	9,317	4,873	284	1,257
3	16,249	3,574	353	17,102	3,177	284	3,140 2)
4	12,510	3,703	353	12,993	3,601	284	1,652
5	9,771	3,836	353	8,898	3,456	284	2,979 2)
6	14,510	7,054	353	13,021	6,969	284	2,655 2)
7	8,471	4,603	353	8,922	4,444	284	1,241
8	11,239	3,305	353	10,569	3,180	284	2,579 2)
9	15,649	4,197	353	15,865	4,484	284	0,623
10	9,724	3,560	353	9,000	3,252	284	2,643 2)

1) $p < 0,05$ 2) $p < 0,01$

TABLE 4.5.1(b) MEANS AND STANDARD DEVIATIONS AND T-VALUES FOR STD. 7 ACCORDING TO SEX

Test no.	Males			Females			Type of test
	$\bar{X}_1$	S_1	N_1	$\bar{X}_2$	S_2	N_2	t/z(two sided)
1	14,757	3,727	370	14,274	3,857	227	1,515
2	13,305	5,405	370	12,065	5,242	227	2,748 2)
3	18,762	2,940	370	19,134	2,684	227	1,549
4	15,500	3,540	370	16,264	3,546	227	2,554 1)
5	13,432	3,613	370	12,264	3,712	227	3,790 2)
6	16,632	7,153	370	13,884	7,036	227	3,578 2)
7	11,618	5,424	370	10,809	4,981	227	1,820
8	13,184	3,188	370	12,326	3,072	227	3,227 2)
9	18,027	3,657	370	18,762	3,129	227	2,600 2)
10	12,476	3,808	370	10,652	2,824	227	6,683 2)

1) $p < 0{,}05$
2) $p < 0{,}01$

TABLE 4.5.1(c) MEANS AND STANDARD DEVIATIONS AND T-VALUES FOR STD. 8 ACCORDING TO SEX

Test no.	Males			Females			Type of test
	$\bar{X}_1$	S_1	N_1	$\bar{X}_2$	S_2	N_2	t/z(two sided)
1	16,137	3,790	394	15,206	3,873	248	3,000 2)
2	16,228	5,144	394	13,552	5,549	248	6,215 2)
3	19,997	2,468	394	20,125	2,262	240	0,660
4	17,515	3,233	394	17,746	3,499	248	0,853
5	16,419	3,278	394	14,923	3,829	248	5,089 2)
6	19,261	6,596	394	16,480	7,315	248	4,978 2)
7	15,417	5,692	394	13,722	5,506	248	3,712 2)
8	13,995	3,571	394	13,332	3,331	248	2,341 1)
9	19,565	3,151	394	19,544	3,377	248	0,080
10	14,737	3,757	394	11,777	3,145	248	10,733 2)

1) $p < 0{,}05$
2) $p < 0{,}01$

4.5.2 The normal curve and normality

Frequency distributions can take on an infinite variety of shapes. However, there is one shape that is of particular importance to educators, psychologists, and statisticians and this shape is represented by a normal curve.

By definition a normal curve is a graphical representation of any one of the group of mathematical functions (density functions) defined by the equation:

$$y = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}$$

where y is the height of the curve at any particular x score, e and π are mathematical constants (2,7183 and 3,1416 respectively), and μ (mean) and σ (standard deviation) are known as the parameters of the equation. Two important facts concerning a normal curve must be borne in mind. First, a normal curve is not itself a frequency distribution; as has been stated, it is a graphical representation of a mathematical function. Second, there is not one, but a whole family of normal curves. With regard to the latter aspect the error is often made in practice that the idea of a normal curve is linked with a specific (single) bell-shaped curve. Stilson stresses this point when he remarks (1966, p.312): "The specific meaning of normal probability density function is given by formula (X.19), not by a verbal description of the characteristics of the normal curve ("symmetrical, bell-shaped", etc.). Many other formulas define symmetrical, bell-shaped curves that lack some of the important properties of the normal density function. Actually there are as many normal curves as there are different pairs of numerical values for μ and σ". As an illustration of the above statement two specific normal curves (μ = 20, σ = 5 and μ = 25, σ = 10) are presented in Figure 4.5.2.

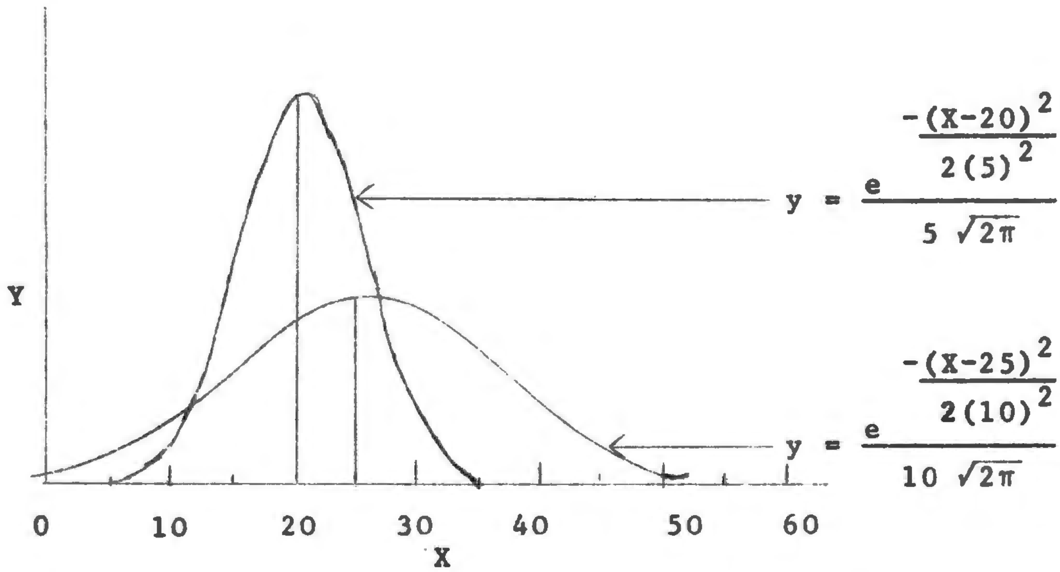


FIGURE 4.5.2 TWO NORMAL CURVES WITH DIFFERENT VALUES FOR THE PARAMETERS μ AND σ.

Although these curves appear to have different shapes they are both normal curves, because the functions they represent are variations of the general normal-curve function.

Stilson (1966, p.312) remarks that it is important to notice that IQ's are only approximately normally distributed. The reason for the approximation is first, that intelligence is not measured continuously, and second that although the range of a normally distributed variable extends from $-\infty$ to $+\infty$, an intelligence test does not yield IQ's beyond certain finite limits. The above-mentioned criticisms naturally apply to aptitude also. Stilson summarizes by saying: "...... the normal curve is used as a convenient approximation of some empirical distributions, but no empirically defined variable can have an exactly normal density function. Thus the normal distribution is a theoretical density function which may describe an observation variable more or less accurately".

The best statistical method that can be employed for investigating the normality of a distribution is the Kolmogorov-Smirnov technique. This technique was applied to the tests of the JATISA and it was found that most tests were normally distributed on the 0,01 level. When distributions are not normal, non-parametric statistics should be employed or the distributions normalized. Certain problems are encountered in this respect however, some of which are the following:

(i) the scores may be non-normally distributed owing to the nature of the test *or* owing to the nature of the group;

(ii) non-parametric statistics are less powerful than parametric statistics (Popham, 1967, p.273);

(iii) scores based on samples may not be distributed normally whereas those based on the population may be distributed normally, and vice versa;

(iv) when normalizing scores the original nature of the distribution is changed.

The advantage of non-parametric statistics is, of course, that far fewer assumptions of population data have to be made for the use of these techniques.

For the purpose of establishing norms non-normal distributions are usually normalized. Examples of normalized scores are normalized t-scores, stanines and deviation IQ's.

4.5.3 Use of norm tables

Two types of norms are available for this test battery, namely percentile norms (which are non-normalized scores) and stanines (which are normalized scores).

4.5.3.1 Percentile norms

Percentile norms have been developed for boys and girls separately. Norm tables are given in the manual. The percentile ranks are found on the left-hand side of the stanine scores in the manual. If a boy (in Standard 8) obtains a raw score of 18 in the Mechanical Insight Field his percentile rank will be 80, e.g. (he will be on the 80th percentile level). This means that 80 percent of the boys in his group obtained mechanical scores lower than 18.

Although percentile norms can be criticised in many ways, they also have certain distinct advantages, for example their ease of interpretation. The main disadvantage of percentile norms is that the percentile ranks of different pupils cannot be compared directly. Thus, if pupil A has a percentile rank of 45 and pupil B has a percentile rank of 90 we cannot say that B's results are twice as good as A's. The reason for this is that percentile scales are not evenly divided as far as scores are concerned, but are based on percentage of individuals obtaining a certain score - in other words we have a nonlinear relationship.

In interpreting percentile scores it is important to understand the concept of statistical average. Roughly one quarter of any heterogeneous group is above average and one quarter is below average, leaving a balance of 50 per cent to comprise the average group. Therefore a percentile rank of 75 or higher indicates an above average interest, a percentile rank of 26 through 74 indicates "average" interest and a percentile rank of 25 or less indicates "below average" interest. Most attention is usually given to percentiles above the 75th and below the 25th levels.

4.5.3.2 Stanines (Standard nine point scales)

The stanine scale consists of 9 scale divisions (intervals) and is based on a distribution with a mean of 5 and a standard deviation of 1,96. The advantange of this type of score over the percentile score is that it has been normalized and thus makes direct comparisons possible. Except for the extremes the scale units all have the same length. Each stanine represents a certain portion of the norm group as shown in Table 4.5.3.2.

The main advantages of stanines are that they are easy to interpret and that they will fit in a single column of a punched card. The main disadvantage of stanines is the lack of discrimination at the extremes.

Because there is no upper limit to stanine 9 for example, individuals with a stanine score of 9 may differ widely from one another with regard to the variable being measured. However, for most practical purposes the scale discriminates adequately among those individuals who fall between stanines 1 and 9.

As an example of how to use this norm, assume that a girl (Standard 6) obtained a raw score of 16 in the Language Usage field. If one looks at the norm table for girls (which is given in the manual) one will notice that this corresponds with a stanine of 8. Thus one immediately knows that she is above average and also how far above average. Her results can also be directly compared with those of another pupil (except perhaps at the extremes).

To aid the teacher or counsellor with the interpretation of stanines the following table is supplied:

TABLE 4.5.3.2 INTERPRETATION OF STANINES

Stanine level	Percentage of cases	Interpretation
1	4	Low (4%)
2 3	7) 12)	Below average (19%)
4 5 6	17) 20) 17)	Average (54%)
7 8	12) 7)	Above average (19%)
9	4	High (4%)

4.6 INTERPRETATION OF RESULTS

Scores are obtained in the following ten tests:

Test 1 Verbal Reasoning

Test 2 Series Completion

Test 3 Social Insight

Test 4 Language Usage

Test 5 Numerical Reasoning

Test 6	Spatial Perception (2-D)
Test 7	Spatial Perception (3-D)
Test 8	Visual Arts
Test 9	Clerical Speed and Accuracy
Test 10	Mechanical Insight

After the profile of a pupil has been determined, occupations which are related to the sections in which peaks are obtained (see Appendix C and Appendix D) should be recommended. For certain occupations only one or two peaks will be relevant, whilst for others several peaks will have to be taken into consideration. Table 4.6 has been drawn up to assist the counsellor with such recommendations although a word of caution will be issued later. The following method could be of further use for the specific circumstances:

(i) <u>One high score is obtained</u> (high scores can be regarded as scores above the 75th percentile level)

In this case the relevant area should be found in Table 4.6 and the list of occupations should be carefully studied for recommendation to the subject. For example, if a pupil obtains a percentile rank of 80 in test 5, the occupations listed under 5 in the table should be studied for making recommendations. Since this list only indicates a small number of the occupations specific to the aptitude fields included in the test battery, any additional sources on the classification of occupations can also be used.

(ii) <u>More than one high score is obtained</u>

The lists for the separate numbers should be consulted in this case, and the best possible combination should be sought. By doing further research lists with combined numbers (e.g. 3-4) can also be drawn up. Such combined lists should actually have more value than single lists since most vocations require several qualities for being successful.

(iii) <u>No high scores are obtained</u>

In such an event scores below the 75th percentile level and above the 60th percentile level should be inspected in decreasing order. Scores in this range will also have some significance although they cannot be evaluated on the same level as the higher scores.

(iv) <u>All scores are low and at the same level</u>

This is probably a rare situation. It could mean that the pupil's aptitudes have not yet fully developed. Examples of this are found among Standard 6 pupils. Such findings will mainly have educational value. Findings of this nature could also mean that the answer sheets were not completed correctly. However, the observant test administrator should be able to detect such a situation readily.

Occasionally it also occurs that all the scores of a pupil are high. This may be owing to a wide general aptitude which will probably be associated with high intelligence or it may be ascribed to the structure of the test battery.

Lastly it must be emphasized that in all the above cases *aptitude and achievement test results should be used in conjunction with interest questionnaire (and personality questionnaire) results.* Thus if a person shows high aptitude in five fields but only shows high interest in two of these five, the discussion should be limited to these two fields. The same would apply the other way round. It must also be mentioned that care must be taken in explaining the difference between interests and ability to the pupils.

Although the above rules can be used as a guide, the counsellor must always interpret scores realistically as is illustrated by the following example: a student obtains a percentile rank of 86 in the Social Insight category, a percentile rank of 76 in the Mechanical Insight category and a percentile rank of 66 in the Language Usage category. An unskilled counsellor would only discuss the scores above the 75th percentile level i.e. Social and Mechanical scores. However, the combined area Social-Mechanical can be limited to only a few vocations like lecturer in engineering courses, workshop teacher, and trade school teacher. On the other hand, if the Social Insight and Language Usage categories are combined, a much wider field will be open to the pupil, for example personnel officer, hotel manager, clothing dealer, public relations officer, canvasser, guide and sales demonstrator. This illustrates that one must always be practical in the interpretation of test results. For this reason Table 4.6 should be regarded as only a very rough guide to the vast number of vocations which pertain to the relevant aptitude areas. Special caution should also be emphasized with regard to areas 1 & 2, since these are more indicative of *level* of ability than of *area* of ability.

At a later stage an attempt will also be made to provide profiles for specific vocations or specific vocational groups, so that comparisons can be drawn between the profiles of scores for individuals and the profiles of scores for different vocational groups.

TABLE 4.6 CLASSIFICATION OF OCCUPATIONS ACCORDING TO MAJOR APTITUDES

1 & 2

VERBAL REASONING & SERIES COMPLETION

Architect	Psychologist
Chemist	Sociologist
Doctor	Teacher
Engineer	Aero Mechanic
Lawyer	Electronics Technician
Mathematician	Radiotrician
Physicist	X-ray technician

3

SOCIAL INSIGHT

Air Hostess	Policeman
Adviser	Public Relations Officer
Diplomat	Radio Announcer
Doctor	Teacher
Employment Officer	Vocational Counsellor
Nurse	Welfare Officer
Occupational Therapist	
Personnel Officer	

4

LANGUAGE USAGE

Author	Literary Critic
Editor	Magazine Writer
Fiction Writer	Novelist
Historian	Proofreader
Interpreter	Librarian
Journalist	Printer
Language Teacher	

5

NUMERICAL REASONING

Accountant	Physicist
Astronomer	Programmer (Computing)
Auditor	Salaries Clerk
Bookkeeper	Statistician
Cost Clerk	Surveyor
Calculating-machine Operator	
Engineer	
Income Tax Clerk	
Mathematician	
Physicist	

6 & 7

SPATIAL PERCEPTION (2-D & 3-D)

Architect	Plumber
Bricklayer	Sheet-metal Worker
Civil Engineer	Taylor
Draughtsman	Toolmaker
Dressmaker	Watchmaker
Fashion Designer	
Patternmaker	

8

VISUAL ARTS

Artist	Hairdresser
Barber	Interior Decorator
Builder	Jeweller
Carpenter	Landscape Artist
Chemical Engineer	Milliner
Commercial Artist	Photographer
Florist	Photo-engraver
Furniture Designer	Sculptor

9

CLERICAL SPEED AND ACCURACY

Bank Teller	Mail Clerk
Correspondence Clerk	Postman
Factory Worker	Secretary
Filing Clerk	Sorter
Insurance Clerk	Stenographer
Income-tax Clerk	Storeman
	Typist

10

MECHANICAL INSIGHT

Aeroplane Mechanic	Mechanic
Civil Engineer	Mechanical Engineer
Crane Operator	Millwright
Electrician	Miner
Fitter and Turner	Pilot
Machinist	Plumber
	Welder

4.7 TEST MATERIALS NEEDED WHEN APPLYING THE JATISA

The complete set of instruments required for the application of the JATISA, the scoring of the answer sheets and the interpretation of the results are as follows:

(i) A test booklet of the JATISA (H.S.R.C., 1971)
(ii) Two answer sheets, scribbling-paper, a pencil, and a rubber

(iii) A manual for the JATISA (H.S.R.C., 1971)
(iv) A set of scoring-keys for the JATISA (hand-scoring).

The contents of the test booklet have already been described in detail and will therefore not be discussed again. It can just be mentioned that the final test battery consists of ten tests, which can be administered in 3 to 3½ hours.

A set of scoring keys has to be used if the answer sheets are not scored by a computer. The instructions for scoring are presented in the manual.

The manual for the test battery contains certain background information concerning the tests. It also contains the instructions for the application and scoring of the tests, and the norms for interpreting the test-scores.

4.8 SUMMARY

In this chapter attention is first given to the construction of the final test battery. It is indicated that most of the tests of Form A will be retained in the final test battery and that a need is felt for certain alterations to two of the existing tests, namely, the Computations Test and the Word Building Test. The need for an additional test, a Social Insight test, is also indicated and the rationale of the test is given. The composition of the final test battery is then set forth.

The final application and the ensuing results are discussed thereafter. The reliability and validity of the tests in the JATISA are dealt with among other things, and reference is made to the intercorrelations of the tests.

Finally the types of norms which are available for the interpretation of the test results are discussed. The standardization of the test battery has now been finalized and the test booklet is ready for distribution to authorized persons. A large number of follow-up studies will still be conducted especially with regard to the validity of the tests. Reference is also made to the other instruments which will be used when applying the JATISA, e.g. the manual and answer sheets.

CHAPTER 5

A FEW OF THE PROBLEMS ENCOUNTERED BY THE COUNSELLOR, AND THE INTEGRATION OF THE JATISA WITH A VOCATIONAL GUIDANCE PROGRAM

5.1 CERTAIN PROBLEMS WHICH CONFRONT THE COUNSELLOR

In applying psychological tests for guidance purposes the counsellor is often confronted with a number of problems, some of which are readily soluble and others less so. Attention will now be given to five of these problems.

5.1.1 Selection of tests

The counsellor usually has at his disposal a wide variety of tests from which he must make a selection, and from which he sometimes makes an incorrect selection owing to lack of background information on the tests. As an example, it might happen that the Lee Thorpe Interest Questionnaire is applied to a group of Standard 6 pupils or that the Strong Vocational Interest Blank is applied to a group of Standard 8 pupils. However, Forbes and Cottle (1953, p.188) estimated that the average grade level of readability of the Lee Thorpe Questionnaire is 9,4, and of the Strong Questionnaire is 11,4. It is thus clear that these pupils could encounter language difficulties with some of the questions included in these questionnaires and this could influence the validity of the test.

As another example, it may happen that the "Holzinger Crowder Uni Factor Tests (Bauernfeind, 1963, p.202) or the "Multiple Aptitude Tests (Manual, p.2) are applied to a group of pupils for guidance purposes. Although neither of these test batteries can be criticised with regard to their psychological value, it must be borne in mind that these tests relied fairly heavily on factor analysis as a basis for their composition. This becomes quite clear when the number of variables which can be regarded as marker variables are observed. Unfortunately this implies more tests of the same nature with a smaller range of content sampling. This does not imply that some of the tests included in such a battery may not have considerable practical value. It only implies that the utility of such a battery in a practical set-up is influenced by the procedure laid down for the construction of the battery.

The above examples indicate that great care should be taken when selecting a test or a group of tests for guidance purposes. With regard to the guidance program for Indian pupils it can be mentioned that the problems referred to in the examples above were attended to by furnishing the meanings of difficult words in the *Interest Questionnaire for Indian South Africans* (H.S.R.C., 1969) and by using *practical value* as a basis for the construction of the aptitude test battery. With regard to the latter aspect less importance was attached to factor analysis (a theoretical contribution), and more emphasis was placed on the integrability of the aptitude tests with a guidance program. Certain job classifications were also studied in order to determine which vocational fields have to be represented by the test battery.

5.1.2 Availability of tests and norms

In contrast with the position sketched in 5.1.1 where sufficient tests are usually available, the situation sometimes arises that appropriate instruments are not available. This is often the position when we are dealing with foreign cultural groups. So, for instance, it would not be correct to use the same tests for a group of Indian pupils (whose second language is English) and a group of English-speaking pupils since this would be a highly biased situation. Unfortunately the problem is often that there are not suitable tests available for such foreign groups and the next best alternative is then taken.

It also happens quite frequently that the group on which the norms are based and the group which is being investigated, although related, differ markedly from each other with regard to certain strata. In such a case it would also be difficult to justify the use of the available norms.

With regard to the Indian South Africans the abovementioned problem is very pertinent, since only a limited number of tests have been standardised for Indian pupils, in South African schools. The present position has been sketched in Chapter 1 and it can be concluded that it is only during the past five or six years that different types of tests have been standardized specifically for Indian South Africans. The organization responsible for this task is the Human Sciences Research Council (Previously called the National Bureau of Educational and Social Research).

The standardization of the JATISA can be regarded as one of the attempts to overcome the problem pictured above. At present several other psychological instruments are also being standardized for Indian pupils in South African schools. Amongst these are a personality questionnaire and an adjustment questionnaire.

5.1.3 <u>Inadequacy of certain sections of existing tests</u>

Owing to the contents (structures) of existing psychological instruments, the counsellor is often compelled to use the results of only one type of measurement (e.g. an interest questionnaire) to make general predictions for a certain area. The situation may arise, for instance, that a counsellor must make a prediction of a pupil's mechanical or artistic ability purely on the basis of results obtained on an interest questionnaire. It will be realised that this is not a sound state of affairs. This matter has already been referred to in Chapter 1.

It may be mentioned that prior to constructing the JATISA the contents of the other instruments which would be used in the guidance program as well as a large number of existing aptitude test batteries were thoroughly studied.

5.1.4 Interpretation of test results

The interpretation and comparison of test results is a much disputed subject. Since it is often not clearly put to the counsellor that ample use must be made of a concept like "standard error of measurement" when interpreting test scores, he is inclined to regard test scores as invariable entities and does not provide for any chance fluctuations. If he is not fully aware of the restrictions of percentile ranks, to mention another example, direct comparisons may be drawn between test scores, in terms of percentile ranks; or if the counsellor is not fully aware of the rationale of the tests of a certain test battery, predictions may be made for specific vocations instead of for vacational groups (compare e.g. the Strong and Kuder Interest Questionnaires). The writer is of the opinion that the latter situation often arises in practice. Reference will again be made to the above situation in paragraph 5.2.

5.1.5 The influence of the dependence of the counsellor upon other professionals

Since the counsellor is dependent on the test constructor for his supply of testing materials, he is usually not aware of the obstacles which are encountered by the test constructor when constructing these tests. Consequently the counsellor often accepts the validity of certain statistical techniques and/or psychological instruments, upon which certain doubts should actually be cast.

A typical example here is the use of forced-choice (ipsative) interest questionnaires or personality questionnaires. An example of the former is the C.V. Interest Questionnaire and an example of the latter is the Edwards Personal Preference Schedule. Although the forced-choice instruments undoubtedly have certain valuable qualities, few test constructors and even fewer counsellors and pedagogues are apparently aware of the ipsative nature of these questionnaires and the *problems which are associated with ipsative measures*. As an illustration, Guilford (as in Jackson and Messick, 1967, p.312) made the following comments concerning ipsative measurements: "...... it represents a kind of measurement that is not suited to correlations of the normal kinds where normative measurements are needed," and "It is improper to use normative scores in a Q-technique analysis and to use ipsative scores in an R-technique analysis. Unfortunately both of these mistakes are sometimes made". Bauernfeind (1963, p.217) also regarded this matter in a serious light, as can be concluded from the following remark: "It should be clear from this discussion that ipsative test scores are extremely difficult to interpret. Yet, at the time of this writing only one test manual accompanying an ipsative forced-choice instrument discusses this frankly and openly. We refer here to the following caution cited in the manual for the Allport-Vernon Study of Values ....".

Another problem which the author would like to mention relates to the function or purpose of a specific instrument. It is well-known that certain instruments have more value in a theoretical connection than in a practical connection. Unfortunately, however, the counsellor is not always acquainted with the background of the instruments amongst which he must make a selection. Thus it may happen that an instrument, which was originally developed for investigating the nature of certain constructs and not for predicting success in a school subject or vocation, is selected. This applies to aptitude tests as well as other types of tests and questionnaires. With regard to aptitude tests one would give preference to the Differential Aptitude Tests in a practical relation and to the Holzinger Crowder Uni-Factor Tests in a theoretical relation, as an illustration.

On the basis of the previous discussion it seems a good idea to suggest closer communication between the counsellor and the test constructor in future, so that the dependence relation does not only function in one direction. In this way closer attention can be given to the needs of the counsellor also. The counsellor or test-conductor furthermore can supply valuable information to the test-constructor in connection with follow-up studies.

5.1.6 Conclusive remarks

To obviate the above obstacles, considerable care was taken when constructing the psychological instruments which will presumably be used in the guidance program. The author was also in the fortunate position that he had gained experience of counselling, test construction and the techniques employed by mathematical statisticians prior to the construction of the aptitude test battery.

At this stage it may seem appropriate to mention a few of the problems encountered by the test constructor also. However, some of the problems encountered by the counsellor are also problems which confront the test constructor. In addition certain problems which confront the test constructor have already been referred to (Chapter 2, pp.20-21).

5.2 THE INTEGRATION OF THE JATISA WITH A VOCATIONAL GUIDANCE PROGRAM

5.2.1 Aids towards integration

In adjusting this battery for integration with a vocational guidance program, attention was given to the following aspects, amongst others:

5.2.1.1 Basis for selection of tests in battery

In constructing the test battery the view was taken that intellectual ability does not consist of a single factor but is composed of several totally different factors. This is in accordance with the findings of factor analysts like Thurstone (as in Kaluger p.122) and Thorndike (as in Kaluger, p.122). *The other instruments which will be used in the guidance program likewise consist of a number of different fields and assist the counsellor in the differentiation and integration process.*

Several job classifications and descriptions, test batteries and factor analyses were consulted to determine which tests to include in the test battery so that the theoretical and practical views could be combined as far as possible. The writer was mainly interested in *practical aspects* and for this reason sometimes criticised certain factor analytic approaches which have not thus far proved to have greater practical value than instruments not having factor analysis as a basis, although they may have certain theoretical advantages. In this respect the writer shares the view of (1) Mc Nemar (as in Whitla, 1968, p.167) who states: "...... the apparent fact that after twenty years of factoring there is altogether too little acceptance of the method and the results obtained thereby," and (2) Mischel (1968, p.51) who states: "Factor analysis is a very useful tool for reducing a large set of correlated measures to fewer unrelated measures and as such it is a powerful method in psychological research. But there is no reason to assume that the procedure does more".

The author is convinced that factor analysis can serve a valuable function if correctly applied. In the case of aptitude tests where the time factor plays an important role, however, it is perhaps not so easy to construct a factorially pure test, *especially if a large range of vocational areas is represented.* To construct a final battery consisting of ten tests, for instance, the original battery would have to consist of twenty to thirty tests to provide for the necessary "marker" variables. In the case of personality questionnaires and interest questionnaires where the time taken for the application is negligible, it is much easier to include a large number of variables and to aim at factorial pureness. As was stated previously, however, a factorial basis does not guarantee a higher practical value, which is actually what is required in guidance.

5.2.1.2 Co-ordination of instruments used in the guidance program

The instruments which are normally used in a guidance program are an aptitude test battery, an intelligence test battery, an interest questionnaire and a personality questionnaire. Sometimes an instrument for measuring attitudes or one for measuring motives or one for measuring values is also included. Most value is however, attached to the first-mentioned instruments. Since these instruments have supplementary functions, a careful study was made of the *existing instruments in these categories* and an attempt was made to relate the newly constructed instrument (JATISA) to the *existing instruments* in the relevant categories. In other words, the *different psychological tests and questionnaires were not regarded as detached entities*. Thus when constructing the JATISA the areas (or sections) of the existing intelligence test, the Group Test for Indian South Africans (Manual, 1968) and the areas of the existing interest questionnaire, the Interest Questionnaire for Indian South Africans (Manual, 1969) were taken into consideration for purposes of co-ordination. This is also the main reason why areas like Visual Arts, Mechanical Insight, and Social Insight were included in this battery. These aspects were all properly gauged by the relevant interest questionnaire and it seemed unfair to have to make predictions on the basis of these results alone. This view is probably also in accordance with the principle of using all the information available in the counselling process. The importance of these areas for the specific cultural group being dealt with was also pointed out previously.

Unfortunately a standardized personality questionnaire was not available when the JATISA was constructed, so that the areas of the JATISA could not be related to those of any existing personality questionnaire.

5.2.1.3 Interpretation of test results

Special attention was given to the interpretation and use of test results by explaining in detail concepts like "error of measurement" and "stanines" in the test manual and by supplying tables for relating performance in certain areas (e.g. Arts, Language) to possible vocational choices. The latter aspect can be further investigated by administering tests to certain vocational and technical groups. With regard to the use of the statistic "error of measurement" in the "band" approach, the following important remark of Goldman (1961, p.285) can be cited: "...... the specific applications to particular tests have been used little by counsellors. Few seem to make the relatively simple translation from reliability coefficient or standard error of measurement to the *band of scores* used above. A promising sign, however, is the use of the *band approach* for score description in some recently published tests". The bands which are referred to here have the length of one or two errors of measurement to the right and to the left of the score obtained, according to the degree of accuracy of prediction which is required. The following is an illustration of test scores in terms of the bands referred to above (the thicker lines represent one standard error of measurement and the thinner lines two errors of measurement).

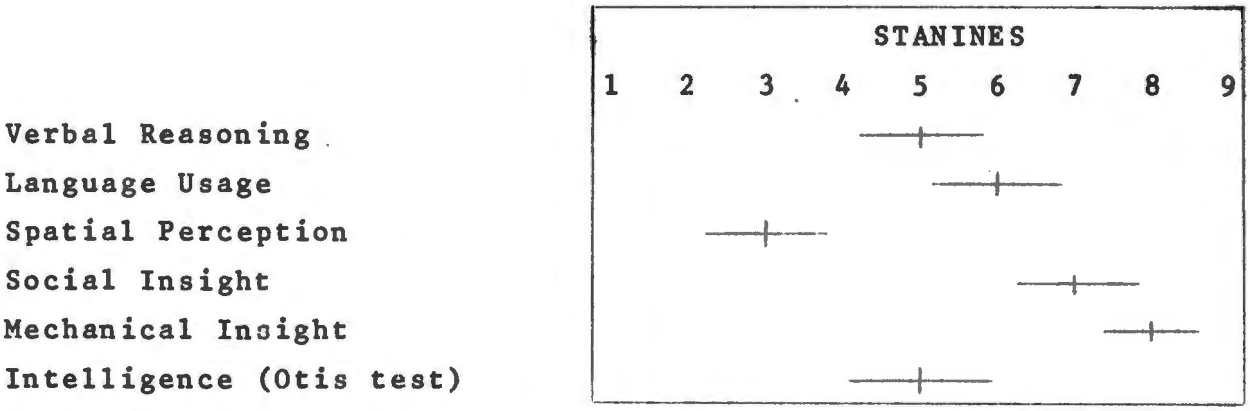


FIGURE 5.2.3 STANINE SCORES WITH ACCOMPANYING ERRORS OF MEASUREMENT OF PUPIL NO. 19 ON A NUMBER OF TESTS

The same types of norms are available for interpreting the test scores obtained in the different instruments, with the result that the integration of the different scores to form a total picture of an individual for the purpose of guidance should not be much of a problem.

5.2.2 The actual integration of the JATISA with a vocational-guidance program

The main instruments which will probably be used in the vocational guidance program are an aptitude test battery (the JATISA), an intelligence test, an interest questionnaire and later probably a personality questionnaire. With the exception of the personality questionnaire these tests have all been standardized. The Junior Aptitude Tests for Indian South Africans (HSRC, 1971) will be used to differentiate between the abilities of the pupils, and the Group Test for Indian South Africans (Inter. Form, S.A. Dept. of Higher Educ., 1968) will be used to indicate the level of ability. Finally the Interest Questionnaire for Indian South Africans (HSRC, 1969) will be used to indicate the preference of the pupils for certain activities which are representative of the more general vocational groups. The contents of the three instruments referred to will now be given.

CONTENTS OF THE JUNIOR APTITUDE TESTS (JATISA)

Test 1 : Verbal Reasoning
Test 2 : Series Completion
Test 3 : Social Insight
Test 4 : Language Usage
Test 5 : Numerical Reasoning
Test 6 : Spatial Perception (2-D)
Test 7 : Spatial Perception (3-D)
Test 8 : Visual Arts
Test 9 : Clerical Speed and Accuracy
Test 10 : Mechanical Insight

CONTENTS OF THE GROUP INTELLIGENCE TEST

Non-verbal Reasoning

Test 1 :	Figure Analogies	Compare Test 2 of the JATISA
Test 3 :	Number Series	
Test 5 :	Pattern Completion	

Verbal Reasoning

Test 2 :	Classification of Pairs of Words	Compare Test 1 of the JATISA
Test 4 :	Verbal Reasoning	
Test 6 :	Analogies of Words	

CONTENTS OF THE INTEREST QUESTIONNAIRE

Field 1 : Language (compare Test 4 of the JATISA)
Field 2 : Arts (compare Test 8 of the JATISA)
Field 3 : Social Service (compare Test 3 of the JATISA)
Field 4 : Science (compare Tests 5 and 10 of the JATISA)
Field 5 : Mechanics (compare Test 10 of the JATISA)
Field 6 : Business (compare Test 5 of the JATISA)
Field 7 : Office Work (compare Test 9 of the JATISA)

The use of the above instruments corresponds with the position sketched by Joubert (Fiat Lux, 1966, p.61) of the Department of Indian Education when he remarked that standardized tests and questionnaires would be used to determine the abilities, aptitudes and interests of pupils, and that these results would be used to guide the pupils towards choosing the most suitable vocation.

It must be mentioned that normally a fixed set of instruments is not used in a vocational guidance program, although certain types of instruments are almost always included. The absolute minimum of instruments required for a vocational guidance program would be an aptitude test battery and an interest questionnaire. In this case the verbal and non-verbal reasoning tests of the aptitude test battery would serve to indicate the level of intelligence. According to need, certain additional tests or questionnaires can be included in the guidance program, or substitions made (or both). So, for instance, a personality questionnaire can be added to the two instruments mentioned or an intelligence test substituted for the verbal and non-verbal reasoning tests of the JATISA, or both modifications can be made. With regard to the personality questionnaires there will presently be two types available, one for determination of the structural (more static) aspects of personality and one for the determination of the more *dynamic* (adjustmental) aspects of personality. The factors which will be evaluated by the adjustmental questionnaire are: personal relations, formal relations (morality, etc.), and vocational relations. The factors which will be evaluated by the *structural* questionnaire are fourteen to sixteen in number, but will not be listed here since final decisions have not been reached in this respect. It should be clear that any of the two personality questionnaires referred to can supply valuable additional information in a guidance program. These questionnaires are being standardized at present and should be available during 1973.

An aspect of importance with regard to *integration* is the *co-ordination of the fields of the different instruments* employed in the vocational guidance program. That special attention was given to the integrational or comparative aspect can be deduced from the previous outline of the contents of the JATISA, the Group Test for Indian South Africans and the Interest Questionnaire for Indian South Africans. It can be seen that the fields of all three instruments bear relation to each other. Further research will still be done to relate the Science Field of the Interest Questionnaire to fields of the JATISA, but is is suspected that future research on creativity will also cest further light on this matter.

With regard to the *selection of tests* for the guidance program the counsellor will be dependent on the tests and questionnaires that have been standardized by the Human Sciences Research Council, most of which consist of one form only. It can be foreseen, however, that in future different forms of the same type of instrument will be available, as well as a wider selection of tests.

Since no aptitude tests had previously been standardized for Indian South Africans, great care had to be taken when selecting the tests of the JATISA.

Another aspect of importance in the integration process is the *time* required for the application of the different tests which are used in the vocational guidance program. The time allowed for the completion of the JATISA is approximately 3½ hours, for the completion of the Group Test approximately 1½ hours and for the completion of the Interest Questionnaire approximately 1½ hours. It is thus clear that most of the time is absorbed by the completion of the JATISA. It must be remembered, however, that the largest area of abilities is tapped by this instrument and that a section of the JATISA can also serve as a substitute for an intelligence test under certain circumstances. Apart from the 6½ hours mentioned above for the completion of the different tests, at least one hour will also have to be devoted to the interview of the counsellor with the counsellee.

From the above description and from previous descriptions it should be clear that the JATISA fulfils a very specific need in the educational program for Indian pupils, not only with regard to vocational guidance but also with regard to scholastic guidance. Under vocational guidance is mainly included the prediction of the most appropriate vocations at the most appropriate levels. Thus a Standard 8 pupil very much interested in mechanical engineering aspects cannot take an engineering degree, but he can enroll for fitting and turning or motor mechanics. Under scholastic guidance is mainly included the prediction of the most suitable school subjects. Aptitude and interest profiles for the different school subjects will be used, as well as regression equations relating the test variables to school subjects.

Whereas some of the other instruments can still be excluded from a guidance program it is hardly conceivable that the same can be said about an aptitude test battery. Nevertheless the intention is not that the JATISA should be used independently of other psychological instruments.

The JATISA was standardized for Indian pupils in Standards 6 to 8 at the request of the Department of Indian Education, as was indicated in Chapter 1 (p.7). The greatest need for an aptitude test battery arose in this range since most Indian pupils leave school after completion of these standards. The question of availability of vocations for Indian pupils leaving school does not cause concern either as was indicated in Chapter 2 (p.25). Rapid progress is being made towards selfdevelopment by the Indian community as is indicated by the all-Indian commercial bank recently opened in Natal, and most well-known professions are already open to the Indians.

5.3 SUMMARY

In the first place a few of the problems encountered by the counsellor prior to and during the guidance session are mentioned in this chapter. Amongst the problems that are discussed, are the selection of tests and the inadequacy of certain sections of existing tests. A problem that is also discussed is the dependence of the counsellor upon other professionals for the tests employed by him in the vocational guidance setting.

In the second place the integration of the JATISA with a vocational guidance program is discussed. First, certain aids towards integration are discussed and then the actual integration process is discussed. The following topics are referred to, amongst other things: the basis for the selection of the tests in the JATISA; the co-ordination of the fields of the instruments used in the guidance program; the interpretation of test results.

CHAPTER 6

SUMMARY, PERSONAL REMARKS AND PROPOSED FUTURE RESEARCH

6.1 SUMMARY

The contents of this thesis can briefly be summarized under the following headings;

6.1.1 Introduction

It may be stated that the aims which were set forth in Chapter 1 were largely achieved with the completion of this project. This project mainly included the production of an objective, reliable, and valid aptitude test battery that would be used for a cultural group which differs markedly from other cultural groups in South Africa. The test battery forms part of a guidance program for Indian pupils in Standard 6 to 8, with the emphasis on vocational guidance. In this respect one step forward has also been made towards the achievement of the aims set forth by the Department of Indian Education in 1966 (Joubert, 1966, p.91). Apart from vocational guidance the JATISA can also be used for scholastic guidance (guidance with regard to choice of school subjects) and for selection purposes, but in these cases a few modifications may have to be made. The present position with regard to psychological testing of Indian South Africans was also described in Chapter 1 and the need for aptitude tests and other psychological tests was clearly indicated.

In Chapter 2 the nature of the concepts "aptitude" and "aptitude testing" was first described. On discussing the concept "aptitude", different views were presented ranging from the earliest (those of Bingham, etc.) to the latest (those of Guilford, etc.). Several definitions were provided and the viewpoint of the author was also cited. Next a distinction was made between special aptitude tests and multiple aptitude tests, and also between aptitude tests and other types of psychological instruments like performance tests and interest questionnaires. Finally the construction of aptitude tests in general and of the JATISA in particular was discussed. It was also motivated why it was decided to construct a completely new test battery and not to make use of an existing test battery. In Chapters 1 and 2 certain introductory information was thus provided.

6.1.2 The standardization process

In order to provide the final test battery the standardization process had to be covered in all its phases. The application for item analysis was presented in Chapter 3 and the application for norm calculation in Chapter 4. For the purpose of item analysis a test battery consisting of two forms with twelve tests in each form was constructed. These forms were applied to a representative sample consisting of 1828 pupils during 1969. The results of the first application were discussed in Chapter 3 but before this was done a few methods of item analysis were first presented. Reference was also made to aspects like reliability, validity, intercorrelations and factor analysis. Both the theoretical and practical aspects of these concepts were presented.

In Chapter 4 attention was first given to the construction of the final test battery. It was indicated that most of the tests of Form A would be retained in the final test battery and that a need was felt for certain alterations to the Computations Test and the Word Building Test. The need for a Social Insight Test was also indicated and the rationale of this type of test was given. The composition of the final test battery was then presented. For the purpose of norm calculation the final test battery was applied to a representative sample of 1926 pupils during 1970. When presenting the results the reliability and validity of the tests in the JATISA were discussed amongst other things. Afterwards the interpretation of test results and the use of norm tables were discussed. To finalize the standardization process, however, a manual and scoring keys had to be constructed. These were successfully completed towards the end of 1971.

In discussing the standardization process an attempt was made to present the different phases of the process eclectically, i.e. different views were presented and an attempt was made to motivate the choice of certain techniques above others. To mention just a few aspects of importance:

(i) different methods of item analysis were presented;
(ii) some of the problems associated with factor analysis were indicated;
(iii) the concepts reliability and validity were discussed in perspective. Modern views concerning these concepts were presented and also certain problems associated with the determination of these test characteristics. With regard to reliability the influence of restriction of range on a K-R reliability coefficient was indicated, and with regard to validity the differential predictive aspect was discussed amongst other things.

In the light of the above-mentioned facts this thesis should serve as an aid to the prospective test constructor even if he does not favour the specific techniques employed by the author.

6.1.3 Certain problems which confront the counsellor and the integration of the JATISA with a vocational guidance program

In Chapter 5 certain problems which confront the counsellor were first presented, because these problems have an important bearing on the guidance set-up. Amongst the problems discussed, were the selection of tests, and the inadequacy of certain sections of existing tests. Another problem which was discussed and to which the author attaches considerable value, is the dependency of the counsellor upon other professionals for the tests employed by him in the guidance process. It was pointed out that the test constructor as well as the counsellor are confronted by certain statistical problems. The nature of this dependency effect can be indicated as follows:

Statistician $\longrightarrow$ Test Constructor $\longrightarrow$ Counsellor

The above implies that the counsellor and pedagogue are dependent on the test constructor for their supply of tests (intelligence tests, aptitude tests, etc.), whilst the test constructor again is dependent on the statistician and programmer for the application of the correct statistical techniques. In the latter respect it may be necessary for the test constructor to employ fairly complicated statistical techniques like maximum likelihood methods, factor analysis or multiple correlational analysis and he may have to approach a mathematician or statistician for the correct application of these techniques or the interpretation of the results. It is only in very recent years that mathematics and mathematical statistics have been included in degree courses which relate to psychometrics with a view to lightening the dependency effect. It is hoped that more attention will also be given to test construction in courses related to counselling, in future.

The second aspect which was discussed in Chapter 5 was the integration of the JATISA with a vocational guidance program. Certain aids towards the integrational process were first discussed, and then the actual process of integration. Aspects which were discussed in the latter section are: the basis for the selection of the tests in the JATISA, the co-ordination of the instruments used in the guidance program and the interpretation of test results.

6.2 PERSONAL REMARKS

It cannot be claimed at this stage that all the aims which were proposed before the construction of this battery have been achieved. However, it does seem, from the results obtained during the standardization process, as if the battery will be able to function satisfactorily. Favourable results with regard to the relation between achievement on the aptitude tests and achievement in school subjects have already been obtained (see paragraph 4.3.4.3).

The author would like to suggest at this stage that more attention is given to integration and specialization in test construction in future. The integration aspect has already been referred to repeatedly in the previous chapters, and the writer is convinced that the integration problem is common to most counselling situations. The most important aspect in the integration process is the co-ordination of the fields of the instruments employed in the counselling situation.

With regard to the specialization aspect, the author's view is that the only real way of improving the predictive validity of existing instruments is to make greater use of specialization and not to change the original basis or rationale of the constructs. So, for instance, the JATISA can serve the *supplementary* function of prediction of suitable school subjects, but the *main function* is prediction of suitable *vocational choices*. In order to improve the predictive ability with regard to choice of appropriate school subjects, further specialization is required, and a few additional tests might have to be added to the test battery. Another example is the differentiation between a Vocations-Interest Questionnaire, a School Subjects-Interest Questionnaire and perhaps a University Courses-Interest Questionnaire. Owing to the different purposes that are served, the setup changes each time, the same as a saloon-car and a sports coupé differ in appearance owing to the different functions they have to fulfill. It is with regard to the specialization aspect also that concepts like "moderator variable", "suppressor variable" and "partial correlation" should play a more prominent part in test construction. Finally the writer is of the opinion that it is at the specialization stage that factor analysis can be usefully employed to examine whether the groups for which specialization facilities are provided do actually form homogeneous clusters.

6.2 PROPOSED FUTURE RESEARCH

Amongst other things it is planned to do research on the following aspects in the future:

6.3.1 <u>Test-retest reliability</u>

It is planned to determine this type of reliability especially for the tests which are to a certain extent speed tests (tests 6 and 9).

6.3.2 <u>Predictive value</u>

To determine the predictive value of the test battery, the relation between the original predictions for a group of subjects and their eventual vocational choices, and perhaps degree of success in these vocations, will be determined.

6.3.3 <u>Construct validity</u>

It is planned to apply the JATISA as well as certain other psychological instruments like intelligence tests and personality questionnaires to a small group of pupils in order to determine the uniqueness of these constructs. In addition the JATISA will be applied to certain external groups (e.g. vocational groups). The first proposed investigation of the latter kind will be related to a technical college, where a technical group, a commercial group and a domestic group will be tested.

6.3.4 <u>Prediction of school achievement</u>

Further research will be done on the best predictors for scientific achievement and for average school achievement. Use will be made of existing intelligence and aptitude models if necessary.

6.3.5 <u>The process of integration</u>

The value of integration in guidance and test validation will be further investigated.

6.3.6 <u>The inclusion of additional tests in the battery and/or exclusion of existing tests in the battery</u>

Amongst other things the suitability of the Social Insight Test for Standard 8 and the Mechanical Insight

Test for Standard 6 will be further investigated. A small investigation has already been conducted in 1971 with regard to the Social Insight Test, and the indications are that it will be possible to improve this test still further.

6.3.7 <u>The introduction of a detailed differentiation program</u>

This aspect will be further investigated for all areas of psychological testing. The author already has certain fixed views regarding this topic.

APPENDICES

APPENDIX A. SAMPLE FOR THE FIRST APPLICATION OF THE JUNIOR APTITUDE TESTS.

NATAL	Std. 6				Std. 7				Std. 8			
	B 1)		G 2)		B		G		B		G	
URBAN HIGH SCHOOLS	A	B	A	B	A	B	A	B	A	B	A	B
Avoca No. 1	9	9	8	8	10	10	7	7	11	11	7	7
Chatsworth					10	10	7	7	11	11	7	7
Clairwood					10	10	7	7	11	11	7	7
Glenover					10	10	7	7	11	11	7	7
Lakehaven	9	9	8	8	10	10	7	7	11	11	7	7
Mayville					10	10	7	7	11	11	7	7
Burnwood High					10	10	7	7	11	11	7	7
Raisethorpe					10	10	7	7	11	11	7	7
Tagore	9	9	8	8								
Westcliffe					10	10	7	7				
Woodlands	9	9	8	8	10	10	7	7	11	11	7	7
	36	36	32	32	100	100	70	70	99	99	63	63
	136				340				324			
RURAL HIGH SCHOOLS												
Greytown					10	10	7	7	10	10	6	6
M.L. Sultan (Glencoe)	8	8	7	7								
Mt. Edgecombe					10	10	7	7	10	10	6	6
St. Oswald					10	10	7	7	10	10	6	6
Umzinto	8	8	7	7	10	10	7	7	10	10	6	6
Windsor					10	10	7	7	10	10	6	6
	16	16	14	14	50	50	35	35	50	50	30	30
	60				170				160			
URBAN PRIMARY SCHOOLS.												
Essendene	10	10	8	8								
Fairhaven	10	10	8	8								
Overport	10	10	8	8								
Springfield Model	10	10	8	8								
Umhlatuzana	10	10	8	8								
	50	50	40	40								
	180											

1) B = Boys

2) G = Girls

	Std. 6				Std. 7				Std. 8			
	B[1)]		G[2)]		B		G		B		G	
	A	B	A	B	A	B	A	B	A	B	A	B
RURAL PRIMARY SCHOOLS												
Ashram	10	10	8	8								
Harry Bodasing	10	10	8	8								
Mount Partridge	10	10	8	8								
Shakaskraal	10	10	8	8								
	40	40	32	32								
		144										
TOTALS: NATAL.	142	142	118	118	150	150	105	105	149	149	93	93
		520				510				484		
TRANSVAAL:												
URBAN HIGH SCHOOLS:												
Laudium	11	11	9	9	8	8	6	6				
Lenasia	11	11	9	9								
Pretoria High									11	11	7	7
Roodepoort					8	8	6	6	11	11	7	7
William Hills					8	8	6	6	11	11	7	7
	22	22	18	18	24	24	18	18	33	33	21	21
RURAL PRIMARY SCHOOLS												
Standerton	3	3	3	3	4	4	3	3	5	5	3	3
TOTALS TRANSVAAL:	25	25	21	21	28	28	21	21	38	38	24	24
GRAND TOTALS:	167	167	139	139	178	178	126	126	187	187	117	117
	334		278		356		252		374		234	
		612				608				608		

1) B = Boys

2) G = Girls

APPENDIX B-1 ITEMS FROM FORM A WHICH DID NOT COMPLY WITH THE REQUIRED CRITERIA

Test number.	Test name.	Std. 6			Std. 7			Std. 8		
		Cr.1*	Cr.2	Cr.3	Cr.1	Cr.2	Cr.3	Cr.1	Cr.2	Cr.3
1	(V.C.)		Items: 23,24			Items: 23,24			Item: 32	Item: 17
2	(S.C.-1)		Item 23			Item 17			Item 23	
3	(S.C.-2)		Item 21							
4	(Calc.)			Items: 4,6,8 10,12 16,21			Items: 2,4,6 8,9, 12,13, 17,21			Items: 1,2,4, 6,7,8, 9,11, 12,13, 16,17, 18,21
5	(F.P. 2D)		Items: 10,18			Item 18				
6	(F.P. $3D_1$)		Items: 2,8							
7	(F.P. $3D_2$)		Items: 5,7,15 19,22			Items: 5,15, 22,24			Items: 5,22, 24	
8	(Mem.)			Items: 5,7 8,9,			Items: 2,4,5, 7,8,9, 12,17			Items: 1,2,3, 4,5,6, 7,8,9, 12,14, 15,17, 18
9	(Cl.Sp)		Item: 13						Items: 6,20, 24	
10	(V.A.)		Items: 6,7			Items: 5,7,23			Items: 5,6, 23,25	
11	(W.B.)		Items: 6,13 20,21, 23			Items: 2,4, 10,23			Items: 10,23	
12	(M.I.)		Items: 8,19, 21			Items: 19,21			Item: 19	

* Cr. 1 = Correlation with external field.
Cr. 2 = Correlation with component field.
Cr. 3 = Proportion answering item correct, more than 90% or less than 10%.

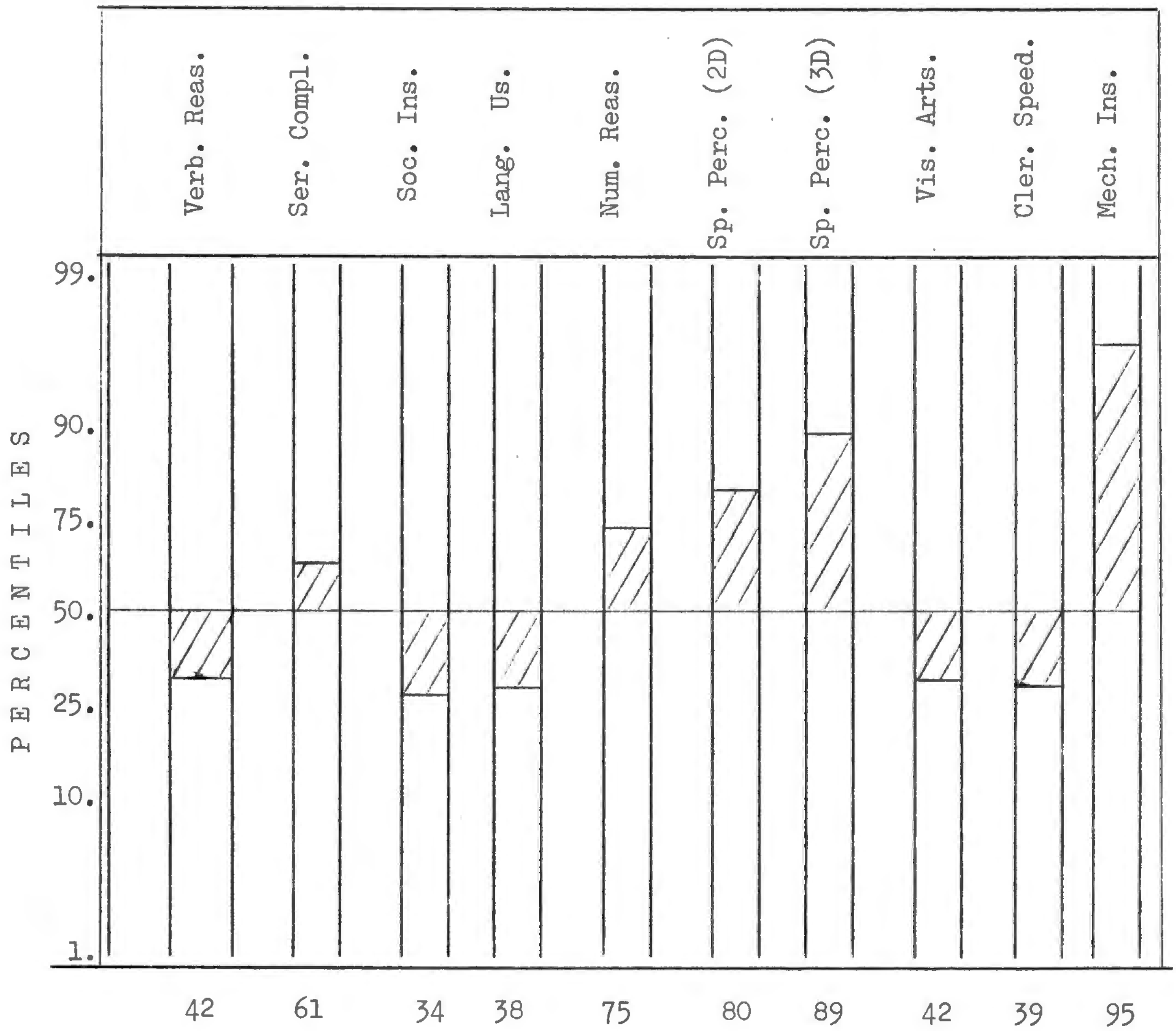
APPENDIX B-2 ITEMS FROM FORM B WHICH DID NOT COMPLY WITH THE REQUIRED CRITERIA

Test number.	Test name.	Std. 6			Std. 7			Std. 8		
		*Cr.1	Cr.2	Cr.3	Cr.1	Cr.2	Cr.3	Cr.1	Cr.2	Cr.3
1	(V.C.)		Items: 9,14,24			Items: 9,14			Item 9	Items: 2,5,6, 11
2	(S.C.- 1)		Items: 12,18, 23,25						Item 23	
3	(S.C.- 2)		Item 17							
4	(Calc.)			Items: 1,6,7, 16,17, 20			Items: 1,2,7 9,12, 13,16, 17,18, 21,23			Items: 1,2,6, 7,9,11, 12,16, 17,21, 23,24
5	(F.P.2D)		Item 25			Items: 23,25			Item 23	
6	(F.P. $3D_1$)		Item 11			Item 2				
7	(F.P. $3D_2$)		Items: 15,18 23	Item 1		Items: 15,18, 25	Item 1		Items: 15,18, 20	Item 1
8	(Mem.)			Items: 13,14			Items: 1,2,3, 4,5,7, 9,13, 14,15, 17,18 35			Items: 1,2,3, 4,5,7, 9,10, 13,14, 15,16, 17,18, 25
9	(Cl.Sp)		Item 14			Item 20	Items: 4,8, 19,25,			Items: 2,4,5, 8,17, 19,25
10	(V.A.)		Items: 5,7,23			Items: 5,7,23			Items: 1,5, 20,23	
11	(W.B.)		Items: 5,15,17			Items: 10,17, 24,25			Items: 1,10,24	
12	(M.I.)		Items: 18,19			Items: 8,13,18			Item 8	

* Cr.1 = Correlation with external field.
Cr.2 = Correlation with component field.
Cr.3 = Proportion answering item correct, more than 90% or less than 10%.

APPENDIX C

PROFILE OF A. NAIDOO ON THE TESTS OF THE JATISA IN TERMS OF PERCENTILES

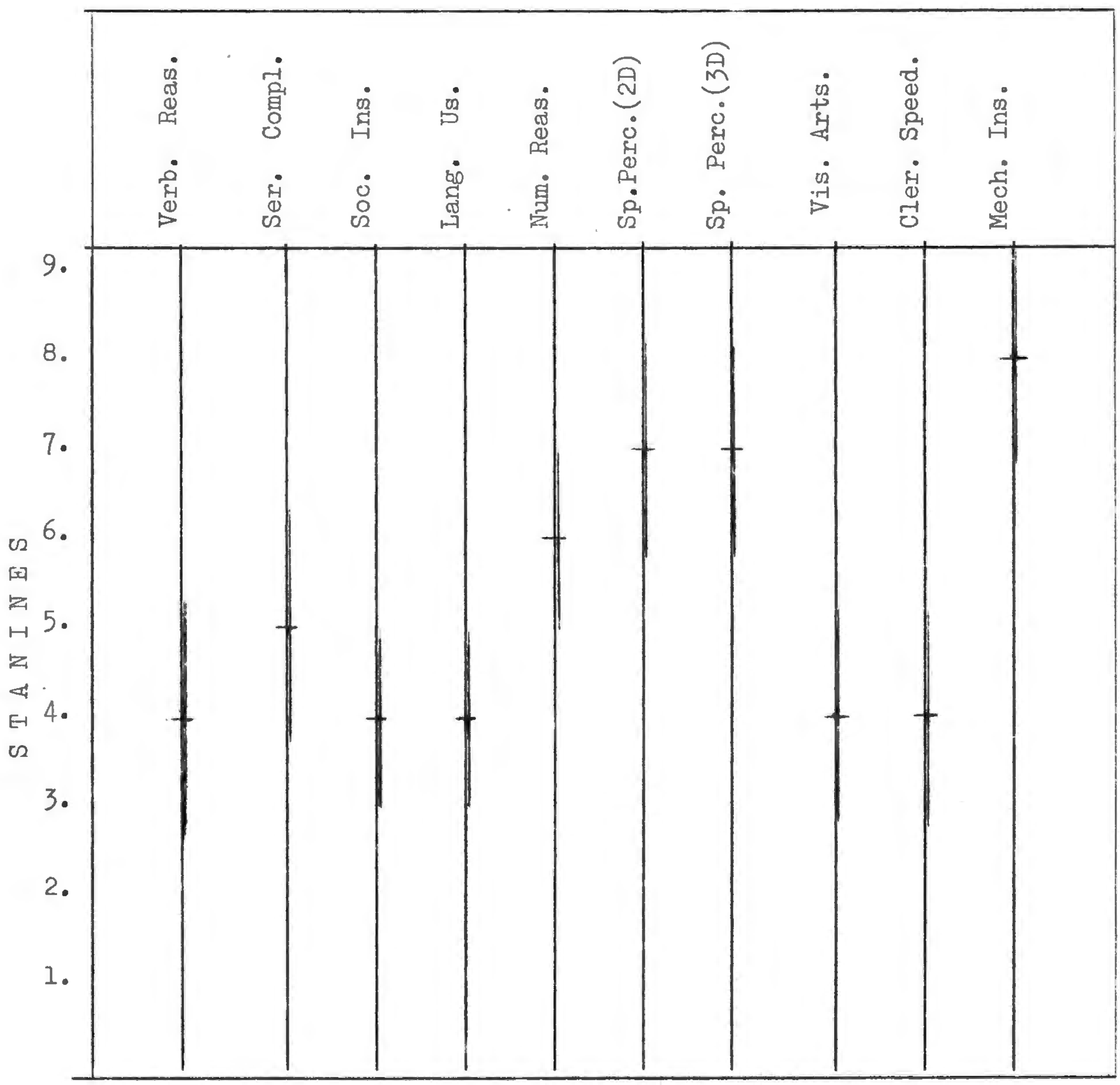


NAME:	SEX	STANDARD
A. NAIDOO*	B	6

* This is an imaginary name.

Note: Use was not made of percentile bands in the above profile but it is advisable to do so. The use of percentile scores can be recommended when the scores of an individual are compared with those of the group.

APPENDIX D. PROFILE OF A. NAIDOO ON THE TESTS OF THE JATISA IN TERMS OF STANINES



NAME:	SEX	STANDARD
A. NAIDOO*	B	6

* This is an imaginary name.

Note: Preference may be given to stanine scores when comparing the results of different individuals.

OPSOMMING

TITEL: DIE STANDAARDISERING VAN 'N AANLEGTOETSBATTERY VIR INDIËRLEERLINGE IN STANDERDS 6 TOT 8 VIR INTEGRASIE BY 'N BEROEPSVOORLIGTINGPROGRAM

SKRYWER: S. OOSTHUIZEN, M.A.

PROMOTOR: PROF. L.A. GOUWS

DEPARTEMENT: SIELKUNDE

GRAAD: D.Phil.

Inleiding

Hierdie projek het in die eerste plek die standaardisering van 'n objektiewe, betroubare en geldige aanlegtoetsbattery vir Indiër-Suid-Afrikaners in standerds 6 tot 8 behels. In die tweede plek was die doel om die belangrikheid van die integreerbaarheid van die verskillende instrumente wat in 'n beroepsvoorligtingprogram gebruik word, te ondersoek en aan te toon.

Inleidend word daar in die proefskrif eers sekere agtergrondgegewens ten opsigte van die Indiërkultuur, asook ten opsigte van aanleg en verwante begrippe verskaf. Ten opsigte van eersgenoemde aspek word die heterogeniteit van die Indiërras onder andere aangetoon asook hulle snelle tempo van ontwikkeling op opvoedkundige gebied. Ten opsigte van laasgenoemde aspek word daar sekere definisies van aanleg verskaf en die uitgangspunt van die skrywer word ook gestel. Die huidige stand van sielkundige toetsing ten opsigte van Indiërleerlinge in Suid-Afrikaanse skole word ook kortliks geskets en daar word aangetoon waarom dit raadsaam sou wees om 'n toetsbattery van meet af aan op te stel.

Eerste en tweede toepassings

Nadat sekere agtergrond-gegewens eers bespreek is, word die eerste en tweede toepassings van die aanlegtoetsbattery beskrywe. Die eerste toepassing van die battery is in 1969 op 'n groep van 1828 leerlinge gedoen, en die toepassing vir normberekening in 1970 op 'n groep van 1926 leerlinge. Dieselfde toetse is toegepas op die drie standerdgroepe en dieselfde toetstye is ook gebruik vir die drie groepe. Waar daar

aanvanklik twee vorms met twaalf toetse in elke vorm opgestel is, bestaan die finale toetsbattery uit een vorm met tien toetse. Die samestelling van die finale toetsbattery is soos volg:

Toets 1 - Verbale Redenering
Toets 2 - Reeksvoltooiing
Toets 3 - Sosiale Insig
Toets 4 - Taalgebruik
Toets 5 - Numeriese Redenering
Toets 6 - Ruimtelike Waarneming (2D)
Toets 7 - Ruimtelike Waarneming (3D)
Toets 8 - Visuele Kuns
Toets 9 - Klerklike Spoed en Akkuraatheid
Toets 10 - Meganiese Insig

Vir die beantwoording van die toetsbattery is 'n toetsboekie, twee IBM-antwoordblaaie, 'n potlood, 'n uitveër en 'n vel werkpapier nodig. Verder het die toetsafnemer ook nog 'n handleiding vir die afneem van die toetse en 'n sleutel vir die nasien daarvan nodig. Die totale tydsduur vir die toepassing van die finale toetsbattery, met pouses ingesluit, is 3 uur 27 minute.

As die steekproef ten opsigte van die eerste toepassing klaar bespreek is, word al die belangrike aspekte van toetskonstruksie bespreek, van itemontleding af tot by normberekening. Aspekte soos korrelasie, faktorontleding, betroubaarheid en geldigheid word onder andere hier bespreek, in 'n teoretiese sowel as praktiese verband. In die bespreking van die resultate van die tweede toepassing word die itemontledingsproses egter nie weer volledig uiteengesit nie.

<u>Probleme van die voorligter en die integrasie van die toetsbattery by 'n beroepsvoorligtingprogram</u>

Aangesien die aanlegtoetsbattery hoofsaaklik in 'n beroepsvoorligtingverband gebruik sal word, is daar besluit om ook kortliks aan bogemelde aangeleenthede aandag te gee. In die eerste deel van die aanbieding word probleme soos die keuse en die beskikbaarheid van toetse bespreek. In die tweede deel word aangeleenthede soos die uitgangspunt by die seleksie van die toetse van die toetsbattery, en die koördinasie van die instrumente wat in 'n beroepsvoorligtingprogram gebruik word, bespreek. Verskeie persoonlike menings word ook hier uitgespreek.

Ten slotte word sekere voorstelle van die skrywer aangebied, en 'n aanduiding word gegee van verdere navorsing wat ten opsigte van die Junior Aanlegtoetse vir Indiër-Suid-Afrikaners (JATISA) beplan word.

In eersgenoemde verband word daar onder andere voorgestel dat daar in die toekoms meer van differensiasie en spesialisasie by sielkundige meting gebruik gemaak word. Dit impliseer kortliks dat daar aanvullend tot elke "kerntoets" of "kernveld" van 'n instrument ook 'n groep gedifferensieerde toetse of velde beskikbaar moet wees, wat verband hou met die betrokke "kerntoetse" of "kernvelde". So kan 'n algemene ruimtelike waarnemingstoets byvoorbeeld 'n "kerntoets" van 'n aanlegtoetsbattery uitmaak terwyl die aanvullende gedifferensieerde velde mag bestaan uit

(i) 'n ruimtelike waarnemingstoets waarin gebruik gemaak word van *rotasie* van plat figure;

(ii) 'n ruimtelike waarnemingstoets waarin gebruik gemaak word van *samestelling* van plat figure;

(iii) 'n drie-dimensionele ruimtelike waarnemingstoets wat gebruik maak van planne en modelle;

(iv) 'n drie-dimensionele ruimtelike waarnemingstoets wat gebruik maak van rotasie van drie-dimensionele figure, ensovoorts.

Die gedagte is dan dat die "kerntoetse" algemene aanwysers sal wees terwyl kombinasies van die gedifferensieerde toetse gebruik sal word om voorspellings te maak vir spesifieke beroepe of vakrigtings.

Wat betref verdere navorsing ten opsigte van die JATISA word daar hoofsaaklik melding gemaak van geldigheidsondersoeke. Op hierdie terrein lê die veld natuurlik nog braak.

REFERENCES

BOOKS AND UNPUBLISHED WORKS

ANASTASI, A. *Psychological testing.* New York: The Macmillan Co. Ltd., 1969.

ANSTEY, E. *Psychological tests.* London: Nelson, 1966.

BAER, F. & ROEBER, E.C. *Occupational information: The dynamics of its nature and use.* Chicago: Science Research Associates Inc., 1964.

BAUERNFEIND, R.H. *Building a school testing program.* Boston: Houghton Mifflin Co., 1963.

BENGE, E.J. *The right career for you.* New York: Funk & Wagnalls Co., 1950.

BINGHAM, W.D. *Aptitudes and aptitude testing.* New York: Harper Bros., 1937.

BRUNING, J.L. & KINTZ, B.L. *Computational handbook of statistics.* Glenview, Illinois: Scott, Foresman & Co., 1968.

CATTELL, R.B. *A guide to mental testing.* London: University of London Press, 1953.

CLARKE, R.B., COLADARCI, A.P. & CAFFREY, J. *Statistical reasoning and procedures.* Columbus, Ohio: Charles E. Merrill Books Inc., 1965.

COOK, H. *Freehand technical sketching.* London: Methuen and Co. Ltd., 1965.

CRONBACH, L.J. *Essentials of psychological testing.* New York: Harper & Row Publishers, 1970.

DU BOIS, P.H. *An introduction to psychological statistics.* New York: Harper & Row Publishers, 1965.

ENGLISH, H.B. & ENGLISH, A.C. *A comprehensive dictionary of psychological and psychoanalytical terms.* London: Longmans Green Co., 1958.

FREEMAN, F.S. *Theory and practice of psychological testing.* New York: Henry Holt Inc., 1955.

FRUCHTER, B. *An introduction to factor analysis.* Princeton, New Jersey: D. van Nostrand Co., 1954.

FURST, E.J. *Constructing evaluation instruments.* New York: Longmans, Green & Co., 1958.

GARRETT, H.E. & WOODWORTH, R.S. *Statistics in psychology and education.* Toronto: Longmans, Green & Co., 1964.

GEKOSKI, N. *Psychological testing.* Springfield, Illinois: Charles, C. Thomas, 1964.

GOLDMAN, L. *Using tests in counselling.* New York: Appleton Century Crofts Inc., 1961.

GRONLUND, N.E. *Measurement and evaluation in teaching.* New York: The Macmillan Co., 1965.

GRONLUND, N.E. *Readings in measurement and evaluation.* New York: The Macmillan Co., 1968.

GUILFORD, J.P. *Psychometric methods.* Tokyo: Tokyo Insatsu Printing Co. Ltd., 1954.

GUILFORD, J.P. *Personality.* New York: Mc Graw-Hill Book Co., Inc., 1959.

GUILFORD, J.P. *Fundamental statistics in psychology and education.* New York: Mc Graw-Hill Book Co., 1965.

GUILFORD, J.P. When not to factor analyze. In: *Problems in human assessment,* edited by Jackson, D.N. & Messick, S. New York: Mc Graw-Hill Book Co., 1967, pp. 309-317.

GULLIKSEN, H. *Theory of mental tests.* New York: John Wiley & Sons Inc., 1962.

HAHN, M.E. & MACLEAN, M.S. *Counselling psychology.* New York: Mc Graw-Hill Book Co., Inc., 1953.

HARDAWAY, M. *Testing and evaluation in business and education.* Ohio: South Western Publishing Co., 1966.

HELMSTADTER, G.C. *Principles of psychological measurement.* London: Methuen and Co. Ltd., 1964.

HOGG, R.V. & CRAIG, A.T. *Introduction to mathematical statistics.* New York: The Macmillan Co., 1966.

HORST, P. *Psychological measurement and prediction.* Belmont, California: Wadsworth Publishing Co. Inc., 1966.

KALUGER, G. & UNKOVIC, C.M. *Psychology and sociology - an integrated approach to understanding human behaviour.* Saint Louis: C.V. Mosby Co., 1969.

KLEINMUNTZ, B. *Personality measurement.* Homewood, Illinois: The Dorsey Press, 1967.

LINDQUIST, E.F. *Educational measurement.* Washington: American Council on Education, 1963.

LOGUE, G.D. *An investigation into the suitability of the South African Group Test for Indian Children in Durban.* Unpublished M.Ed. dissertation, University of Natal, 1954.

LYMAN, H.B. *Test scores and what they mean.* Englewood Cliffs, New York: Prentice Hall Inc., 1963.

MEHRENS, W.A. & EBEL, R.L. *Principles of educational and psychological measurement.* Chicago: Rand Mc Nally & Co., 1967.

MORTENSEN, D.G. & SCHMULLER, A.M. *Guidance in today's schools.* New York: John Wiley & Sons Inc., 1964.

MISCHEL, W. *Personality and assessment.* New York: John Wiley & Sons Inc., 1968.

MURDOCH, D.C. *Linear algebra for undergraduates.* New York: John Wiley & Sons Inc., 1966.

NAIDOO, U.S. *Religion among Indians in South Africa.* Cape Town: O.U.P., 1947.

NUNNALLY, J.C. *Introduction to psychological measurement.* New York: Mc Graw-Hill Book Co., 1967.

OHLSEN, M. *Guidance services in the modern school.* New York: Harcourt, Brace & World Inc., 1964.

PEATMAN, J.G. *Introduction to applied statistics.* New York: Harper & Row Publishers Inc., 1963.

POPHAM, W.J. *Educational statistics.* New York: Harper & Row Publishers Inc., 1967.

RAMPHAL, C. *A study of three current problems of Indian Education.* Unpublished D. Phil dissertation, University of Natal, 1960.

REMMERS, H.H. & GAGE, N.L. *Educational measurement and evaluation.* New York: Harper & Row Publishers, 1955.

SHERTZER, B. & PETERS, H.J. *Guidance: Techniques for individual appraisal and development.* New York: The Macmillan Co., 1965.

SMITH, F.M. & ADAMS, S. *Educational measurement for the classroom teacher.* New York: Harper & Row Publishers, 1966.

STILSON, D.W. *Probability and statistics in psychological research and theory.* San Francisco, California: Holden-Day Inc., 1966.

SUPER, D.E. & CRITES, J.O. *Appraising vocational fitness.* New York: Harper & Row Publishers, 1965.

TYLER, L.E. *Tests and measurements.* Englewood Cliffs, New Jersey: Prentice Hall, 1963.

WHITLA, D.K. *Handbook of measurement and assessment in behavorial sciences.* Reading, Massachusetts: Addison Wesley Publishing Co., 1968.

PERIODICALS AND REPORTS

CATTELL, R.B. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology.* 54, 1-22 (1963).

CRONBACH, L.J. & WARRINGTON, W.G. Efficiency of multiple-choice tests as a function of the spread of item difficulties. *Psychometrika.* 17, 127-147 (1952).

DUFF, F.L. Item subtlity in personality inventory scales. *Journal of Consulting Psychology.* 29, 565-570 (1965).

FORBES, F.W. & COTTLE, W.C. A new method for determining the readability of standardized tests. *Journal of Applied Psychology.* 37, 185-190 (1953).

HORN, J.L. A rationale and test for the number of factors in factor analysis. *Psychometrika.* 30, No. 2, 179-185 (June 1965).

JOUBERT, P.R. Psychological services in Indian schools. *Fiat Lux.* 1, No. 3, 90-91 (Aug. 1966).

LISTER, J.L. & McKENZIE, D.H. A framework for the improvement of test interpretation in counseling. *Personnel and Guidance Journal.* 45, 61-65 (1965).

NAUDE, P.S. — *Indiërs en Indiëronderwys in Natal: 'n historiese en kritiese studie;* 1860-1960.

OOSTHUIZEN, S. — *Office Report of the Human Sciences Research Council.* No. 2, 6-7(1971).

OWEN, K. — *Technical Report of the Human Sciences Research Council,* No. 3, p.32 (1971).

THE EDITOR — Graduation ceremony. *Fiat Lux.* 3, No. 5, 146-157(June 1968).

THE EDITOR — Garage for Laudium. *Fiat Lux.* 5, No. 2, 10-11 (March 1970).

THE EDITOR — Many more posts for Indians. *Fiat Lux.* 5, No. 4, 10-11 (May 1970).

THE EDITOR — Profile of the Indian community. *Fiat Lux.* 5, No. 4, p.9 (May 1970).

THE EDITOR — Extracts. *Fiat Lux.* 5, No. 6, p.33 (Aug. 1970).

THE EDITOR — Selection and balance - the key to good composition. *Practical Photography.* 25-31 (March 1968).

TUCKER, L.R. — A note on the estimation of test reliability by the Kuder-Richardson formula 20. *Psychometrika.* 14, No. 2, 117-119 (1949).

VAN DER SPUY, D.T. — Die arbeidspatroon van Indiërs in Natal. *Navorsingsreeks nr. 62 van die Nasionale Buro vir Opvoedkundige en Maatskaplike Navorsing,* 1968.

YAMAMOTO, K. — "Creativity" - A blind man's report on the elephant. *Journal of Counseling Psychology.* 12, 428-434(1965).

<u>MANUALS</u>

BENNETT, G.K., SEASHORE, H.G. & WESMAN, A.G. — *Manual for the Differential Aptitude Tests.* New York: The Psychological Corporation, 1966.

CURETON, E.E. & CURETON, L.W. et.al. *Manual for the Multi-Aptitude Test.* New York: The Psychological Corporation, 1955.

HOLZINGER, K.J. & CROWDER, N.A. *Manual for the Holzinger-Crowder Uni-Factor tests.* New York: Harcourt, Brace and World, 1955.

HUMAN SCIENCES RESEARCH COUNCIL (South Africa) *Manual for the Interest Questionnaire for Indian South Africans.* Pretoria: Institute for Psychometric Research, 1969.

HUMAN SCIENCES RESEARCH COUNCIL (South Africa) *Manual for the Junior Aptitude Tests for Indian South Africans.* Pretoria: Institute for Psychometric Research, 1971.

RAAD VIR GEESTES-WETENSKAPLIKE NA-VORSING (Suid-Afrika) *Handleiding vir die Akademiese Tegniese Aanlegtoetse.* Pretoria: Instituut vir Psigometriese Navorsing, 1970.

S.A. DEPARTEMENT OF HIGHER EDUCATION *Manual for the Group Test for Indian South Africans.* Pretoria: National Buro of Educational and Social Research, 1968.

SEGEL, D. & RASKIN, E. *Manual for the Multiple Aptitude Tests.* Los Angeles, California: California Test Bureau, 1955.

STRONG, E.K. (Jr.) *Manual for the Strong Vocational Interest Blank for men.* Stanford: Stanford University Press, 1969.

UNITED STATES DEPARTMENT OF LABOR *Manual for the General Aptitude Test Battery.* Washington: United States Department of Labor, 1967.