Chapter VII

RELIABILITY AND VALIDITY OF SCORES IN FINAL ADMINISTRATION

(A) RELIABILITY

Consistency is a great virtue for a man, and also for a test. But a person can be consistently wrong and a test can measure consistently something that is wrong. High reliability is no guarantee for the validity of a test, but low reliability is a definite proof of a poor test. A test should measure what it is meant to measure; and it should measure it consistently. To summarize, an ideal test tells the truth consistently.

(a) Nature of Reliability

A test is said to be reliable only if the correlation between the set of scores at the first time of

administration and the set of scores at the second time of administration is high. If a test is applied for the second time under similar conditions, and the testee's scores differ widely from those previously obtained, then it is obviously a poor test.

There are several meanings attached to reliability. It includes dependability, consistency, objectivity and stability. Each signifies something different as applied to measurements. Different types of reliability coefficients answer to different questions; they indicate different aspects of reliability, and hence permit different inferences regarding the evidence.

Reliability of a test refers to the consistency of scores obtained by the same individuals on different occasions or with different sets of equivalent items. The first type is called test-retest reliability, the second one is called parallel or alternate from reliability. The concept of reliability is based on the phenomenon of the error of measurement of a single score. With the help of reliability coefficient we can predict the range of fluctuation likely to occur in a single individual's score as a result of chance, irrelevant factors. The imperfectness of the reliability of statistical measures arises because of sampling errors, while

imperfect reliability of a test is due to errors of measurement. Thus reliability of a test is the measure of the consistency of results when observations are repeated on a group of individuals.

As told above, test reliability indicates the extent to which individual differences in test-and-retest scores are attributable to chance errors of measurement. Every measure of test reliability denotes what proportion of the total variance of the test scores is 'error variance'. Any condition, which is irrelevant to the purpose of the test, represents error variance. When the tester tries to maintain uniform testing conditions by controlling the testing environment, instructions, time-limits, rapport and such other things, he is reducing error variance and making the test scores more reliable. But no test can be a perfectly reliable instrument. Hence the characteristics of the sample and the type of reliability which was measured should be specifically stated.

The main techniques of measuring the reliability of a test are :

(i) Test-retest Method.

(ii) Parallel or Alternate Forms Method.

(iii) Split-half Method.

(iv) Analysis of Variance.

Important characteristics of these techniques are well-known and hence need not be discussed here.

(b) Technique of Reliability Measurement adopted in The Present Investigation

In the present investigation, the test-retest technique was adopted for the following reasons :

(i) Preparing two really equivalent test-forms is very difficult. Especially in aptitude and ability testing, parallel forms are difficult to prepare.

(ii) The test-retest procedure is very simple. It has been criticised for the fact that practice effect and memory effect viciate the results of the second-time testing, and hence it is less reliable than others. In spite of these limitations, this technique was adopted, because stability is the key concept in reliability and test-retest gives a good measure of stability over a period of time.

(iii) In the split-half technique, the correlation yielded gives the reliability measure, not of the full test, but only of one half as long. It shortens the test by one half and thus decreases its reliability.

Again, it is very difficult to frame two items which are equivalent in all regards, namely content validity, difficulty value, discriminative index etc.

(iv) The Kudar-Richardson formula measures the homogeneity of the test; but it rests on the assumption that all items are of equal difficulty value. Now, the present test consists of items having varying difficulty values. So the Kudar-Richardson technique will not be applicable to the present test.

(c) Reliability of The Present Test

The reliability coefficient of the present test was calculated according to the test-retest method. 100 students, out of the total sample of 1000, were retested after an interval of about 70 days (10 weeks). All other testing conditions were the same. The two sets of scores were then correlated. The coefficient of reliability is .82.

Reliability r = .82

N = 100

Time interval = 10 weeks.

Index of reliability can be measured by the following formula given by Garrett (1962) :

$$r_1 = \sqrt{r_{11}}$$

$$\text{Here } r_1 = \sqrt{.82}$$

$$= .9 \text{ approximately.}$$

Test-retest reliabilities of separate tests were also calculated. They are as follows :

Test I	=	.80
Test II	=	.82
Test III	=	.85
Test IV	=	.83

The scattergrams for these five retests are given in Tables 7.1 and 7.2.

Table : 7.1

TEST-RETEST RELIABILITY COEFFICIENT

Scattergram showing the Correlation between the First Testing and Retesting of the Present Test

(Time Interval : 10 Weeks)

Second Testing	181-195	196-210	211-225	226-240	241-255	256-270	271-285	286-300	301-315	Total of Second Testing
316-330								1	2	3
301-315						3	3	2		8
286-300						3	4	4		11
271-285				3	4	5	3	2		17
256-270			2	7	6	4	3			22
241-255			2	3	7	4	1			17
226-240		3	5	2	3					13
211-225	1	2	2	1						6
196-210	1	2								3
Total of First Testing	2	7	11	16	20	19	14	9	2	100

FIRST TESTING

r = .82

N = 100

Time Interval = 10 Weeks (70 days)

Table : 7.2

S C A T T E R G R A M S

showing
The Test-retest Reliabilities
of
Separate Tests

TEST : I

RE - TEST

FIRST TESTING	111-115	116-120	121-125	126-130	131-135	136-140	141-145	146-150	Total
141-145						2	3	2	7
136-140					1	3	4	1	9
131-135				2	5	5	4		16
126-130			2	3	9	7	2		23
121-125		1	3	6	5	3			18
116-120	1	4	4	1	1				11
111-115	2	5	2	1					10
106-110	3	2	1						6
Total	6	12	12	13	21	20	13	3	100

r = .80

Table : 7.2
(Contd.)

TEST : II

RE-TEST

FIRST TESTING	26-30	31-35	36-40	41-45	46-50	51-55	56-60	Total
51-55					2	3	4	9
46-50				2	3	4	3	12
41-45			2	4	6	3	1	16
36-40		1	5	9	8	2		25
31-35		1	8	5	4			18
26-30	2	8	1					11
21-25	6	2	1					9
Total	8	12	17	20	23	12	8	100

r = .82

TEST : III

RE-TEST

FIRST TESTING	46-50	51-55	56-60	61-65	66-70	71-75	76-80	81-85	Total
76-80						1	2	3	6
71-75					2	3	4	2	11
66-70				3	5	9	4	1	22
61-65			1	5	5	6			17
56-60			4	3	4	4			15
51-55	1	3	3	4	1				12
46-50	3	4	2						9
41-45	4	3	1						8
Total	8	10	11	15	17	23	10	6	100

r = .85

Table : 7.2
(Contd.)

TEST : IV

RE-TEST

FIRST TESTING	13-15	16-18	19-21	22-24	25-27	28-30	Total
28-30				1	3	7	11
25-27			2	3	10	4	19
22-24		1	6	11	5	2	25
19-21		4	10	9	2		25
16-18	1	11	1				13
13-15	5	2					7
Total	6	18	19	24	20	13	100

r = .83

(d) Interpreting the Test Reliability

There is no set formula to judge the satisfactoriness of a given reliability coefficient. The best than can be done is to compare the reliability coefficient of this test with those of other ability and aptitude tests.

Not many tests measuring language ability are available in Gujarati. The language ability test prepared by Dr. Urvashi Desai has the reliability coefficient of .60 (retest),

.75 (Rulon formula), .81 (Guttman formula) and .82 (analysis of variance). Compared to that criterion, the reliability of the present test is quite satisfactory.

(B) VALIDITY

(a) Nature of Validity

In addition to reliability, a test must have validity also. It is not enough that a test is sufficiently reliable; it should be valid also.

Validity means that test must be suitable to the purpose. It must yield the kind of results which are needed. Test validity concerns what a test measures, and also how well it does so. What a test measures can be determined only by an examination of the procedures employed in finding the validity of the test, and especially by the nature of the criterion. A test is valid to the extent that it measures what it purports to measure.

Ross and Stanley (1941), defining validity, say,

> One kind of validity concerns the degree to which the test or other measuring instrument measures what it claims to. In a word, validity means truthfulness.

Gullicksen (1950) defines validity as follows :

> Validity of a test is the correlation of the test with some criterion.

Validity thus refers to the truthfulness of a test. It can be determined experimentally by finding the correlation between the test and some independent criterion. A criterion may be an objective measure of performance, or a qualitative measure such as judgment of the character or excellence of work done.

(b) Validity of The Present Test

(i) Predictive Validity

The purpose of the present test is to measure the ability to use language (Gujarati language) effectively in the areas of reading comprehension and written expression. Indirectly it also aims at predicting whether the testee will be successful in tasks demanding language ability. The direct and primary purpose is to measure the testee's language ability as it is presently revealed; only indirectly it aims at predicting success in future tasks that involve effective use of language. So the present investigation does not attempt to establish predictive validity of the instrument. Even from the viewpoint of practicability, it is not possible to establish predictive validity within the short span of 2 or 3 years.

Even then, the present test can be useful to some extent in predicting future success in linguistic tasks. As has been discussed at length in Chapter I present proficiency serves to some extent as an indicator of future success. Ability includes both present proficiency and an aptitude indicating future success in an area. The scores of an ability test can be made to serve in the prediction of future performance, and to that extent it functions as an aptitude test. The present investigator believes that the present ability test does have some measure of predictive value. But beyond that, nothing precise or statistical can be claimed at this stage. If some further investigation establishes expectancy tables, showing correlation between this test and future linguistic performances, that will contribute to the predictive validation.

(ii) Content Validity

The content validity of the present test is a matter of rational judgment, not of statistical, experimental verification. The investigator believes that the content validity of the test was established in the following manner:

(a) Judgments and weightages were secured from eminent teachers of Gujarati language and eminent linguists regarding the components of language ability in Gujarati.

The present test was constructed in accordance with their judgment. The component abilities listed by them were included in the present test.

(b) Each of the items in the subtests and all the subtests were discussed with experts, keeping in view whether the language skills expected of an S.S.C.-pass student have been duly covered by the test or not, and whether the level is maintained or not. The experts judged their fitness for the population for which the test is intended, namely the college-entrants. They also screened the test-items to see whether they would measure what they are meant to measure. Items were included in the test after careful scrutiny.

(iii) Construct Validity

As discussed in chapter IV on page 99, the construct of language ability was analysed threadbare with a view to identifying the possible components of the construct. The investigator has taken a position that the present test to measure "language ability" should reflect the analysis made earlier with maximum fidelity. It is a truism that establishing construct validity amounts to bringing the construction, planning and preparation of the test as near as possible to the theoretical position taken by the investigator earlier.

Most of the items in the present test are

multiple-choice items; only a few are matching items. Thus all the items are objective-type items. Their difficulty value and discrimination value are precisely scrutinized through item analysis. So it can be said that the manner in which the items and the test are constructed reflect with precision the behaviours and mental performances indicative of the criterion behaviour. To that extent, the present test can be said to possess construct validity.

(iv) Congruent Validity

Ideally, congruent validity is based on the comparison of the test scores with those of a similar valid measure of the same function. But there is no existing language ability test standardized for the college-entrants, consequently, the nearest to the ideal was accepted as a compromise. There is a language ability test for high school students, standardized by Dr. Urvashi Desai. High School leavers comprise a population very similar to the college-entrants. So the investigator decided to use that test as the external criterion for establishing congruent validity. That test was administered to 60 students out of the total sample of 1000 for the present test, and the correlation between the two was computed. The scattergram is given in Table No. 7.3. The results can be summarized as follows:

N	=	60
Congruent Validity r	=	.69
The Standard Error of r	=	.0675

Table : 7.3

CONGRUENT VALIDITY

Scattergram showing the Correlation between The Present Test and Dr. Urvashi Desai's Language Ability Test (Std. XI)

PRESENT TEST (A) : (TOTAL SCORE 400)

Dr. Urvashi Desai's Test (B) (Total Score : 100)	181-195	196-210	211-225	226-240	241-255	256-270	271-285	286-300	301-315	Total of Test (B)
81-85							1	1		2
76-80						1	1		1	3
71-75					1	1	2	1		5
66-70					2		1	2	1	6
61-65			1	1	2	1	3	1		9
56-60			1	2	4	4	1	1		13
51-55		1	1	2	1	3				8
46-50		1	2	2	1	1				7
41-45		1	1	1	1					4
36-40			1	1						2
31-35	1									1
Total of Test (A)	1	3	7	9	12	11	9	6	2	60

r = .69

(v) Concurrent Validity

Concurrent validity of the present test has been established by correlating the scores on the present test with marks in Gujarati (subject) at the S.S.C. examination (secondary school certificate examination). The S.S.C. Gujarati marks of 100 students from the sample were collected. They were correlated with the scores on the present test. The scattergram is given in Table No. 7.4.

The result can be summarized as follows :

N	=	100
Concurrent Validity r	=	.75
Standard Error of r	=	.0438

Table : 7.4

CONCURRENT VALIDITY

Scattergram showing the Correlation between The Present Test and Scores in Gujarati at the S.S.C. Examination

PRESENT TEST (A) : (TOTAL SCORE 400)

S.S.C. Examination Marks in the Subject of Gujarati (Total Score : 100)	181-195	196-210	211-225	226-240	241-255	256-270	271-285	286-300	301-315	Total of S.S.C. Scores
63-65								1	1	2
60-62							1	1	1	3
57-59						2	2	1		5
54-56					3	4	2	2		11
51-53				1	3	4	4	4		16
48-50			1	8	8	5	4			26
45-47		2	5	3	3	3	1			17
42-44		3	4	3	3	1				14
39-41	2	2	1	1						6
Total of Test (A)	2	7	11	16	20	19	14	9	2	100

r = .75

(C) CORRELATIONS AMONG THE TESTS OF THE PRESENT BATTERY

The first step in the direction of factor analysis is to calculate intercorrelations among the tests and to prepare a correlation matrix. The investigator prepared scattergrams for Tests I and II, I and III, I and IV, II and III, II and IV, and III and IV. Then he calculated correlations among the tests according to the following formula :

$$r = \frac{n.\Sigma fxy - (\Sigma fx)(\Sigma fy)}{\sqrt{n.\Sigma fx^2 - (\Sigma fx)^2} \cdot \sqrt{n.\Sigma fy^2 - (\Sigma fy)^2}}$$

The scattergrams are given in Table : 7.5 and the correlation matrix so formed is given in Table : 7.6.

Table : 7.5

SCATTERGRAMS SHOWING CORRELATIONS

Correlations between
Test I (Vocabulary) and Test II (Structure)

TEST II : (STRUCTURE): SCORES

Test I:(Vocabulary):Scores	21-25	26-30	31-35	36-40	41-45	46-50	51-55	Total
141-145			14	21	10	8	16	69
136-140			5	11	22	23	30	91
131-135				42	43	37	45	167
126-130	21	26	48	61	45	28		229
121-125	20	29	65	53	20			187
116-120		21	17	27	23	22		110
111-115	25	20	23	26				94
106-110	23	20	10					53
Total	89	116	182	241	163	118	91	1000

r = .58

Table : 7.5
(Contd.)

Correlation between
Test I (Vocabulary) and Test III (Spelling)

TEST III : (SPELLING) : SCORES

Test I:(Vocabulary)Scores	41-45	46-50	51-55	56-60	61-65	66-70	71-75	76-80	Total
141-145				10	19	20	12	8	69
136-140			7	12	26	24	16	6	91
131-135	5		19	17	28	42	33	23	167
126-130	21		20	27	30	87	27	17	229
121-125	28	19	21	28	36	25	21	9	187
116-120	10	31	23	23	16	7			110
111-115	8	27	12	20	13	14			94
106-110	6	10	17	12	8				53
Total	78	87	119	149	176	219	109	63	1000

r = .47

Correlation between
Test I (Vocabulary) and Test IV (Comprehension)

TEST IV : (COMPREHENSION) : SCORES

Test I:(Vocabulary) Scores	13-15	16-18	19-21	22-24	25-27	28-30	Total
141-145			10	23	21	15	69
136-140			16	25	30	20	91
131-135			34	48	40	45	167
126-130		19	58	84	47	21	229
121-125	10	32	51	40	37	17	187
116-120	21	42	27	8	12		110
111-115	27	20	27	20			94
106-110	14	18	21				53
	72	131	244	248	187	118	1000

r = .58

Table : 7.5
(Contd.)

Correlation between :
Test II (Structure) and Test III (Spelling)

TEST II : (STRUCTURE): SCORES

Test III (Spelling) Scores	21-25	26-30	31-35	36-40	41-45	46-50	51-55	Total
76-80		4	8	9	10	13	19	63
71-75			3	37	25	24	20	109
66-70	7	6	7	80	45	46	28	219
61-65	8	22	43	35	32	23	13	176
56-60	13	18	55	24	20	8	11	149
51-55	15	23	32	30	19			119
46-50	26	32	19	10				87
41-45	20	11	15	16	12	4		78
Total	89	116	182	241	163	118	91	1000

r = .49

Correlation between :
Test II (Structure) and Test IV (Comprehension)

TEST II : (STRUCTURE) : SCORES

Test IV (Compr.) Scores	21-25	26-30	31-35	36- 40	41-45	46-50	51-55	Total
28-30			16	26	25	31	20	118
25-27		10	22	55	35	41	24	187
22-24	14	17	45	66	49	30	27	248
19-21	34	36	51	48	54	11	10	244
16-18	29	28	27	32		5	10	131
13-15	12	25	21	14				72
Total	89	116	182	241	163	118	91	1000

r = .48

Table : 7.5
(Contd.)

Correlation between
Test III (Spelling) and Test IV (Comprehension)

Test III (Spelling) Scores	TEST IV : (COMPREHENSION): SCORES						
	13-15	16-18	19-21	22-24	25-27	28-30	Total
76-80		5	22	21	9	6	63
71-75		7	30	39	21	12	109
66-70		19	25	54	69	52	219
61-65		20	31	54	46	25	176
56-60	10	21	43	35	24	16	149
51-55	24	22	36	20	10	7	119
46-50	20	17	32	10	8		87
41-45	18	20	25	15			78
Total	72	131	244	248	187	118	1000

r = .42

Table : 7.6

CORRELATION MATRIX SHOWING
INTERCORRELATIONS AMONG THE FOUR SUBTESTS

	$Test_1$	$Test_2$	$Test_3$	$Test_4$
Test 1 (Vocabulary)	-	.58	.47	.58
Test 2 (Structure)		-	.49	.48
Test 3 (Spelling)			-	.43
Test 4 (Comprehension)				-

(D) FACTORIAL VALIDITY

(a) Theoretical Nature

Guilford (1936), explaining the relevance of factor theory, says,

> Factor theory has highlighted a very serious fault in psychological tests. This is the fact that any test measures more than one common factor to a substantial degree, yields scores that are psychologically ambiguous and very difficult to interpret. What is worse, almost all tests have a complexity greater than one, that is, they measure more than one common factor.

To meet this situation - to fulfill the search for the unitary traits of personality - a statistical approach such as factor analysis is necessary. The shortcomings of commonly used single-score tests have been revealed by the statistical procedure known as factor analysis.

This procedure was first developed and applied to mental ability tests by Spearman and his co-workers in England. In the United States, pioneer work in this field was conducted by Thurstone. Following the work of Spearman and Thurstone, many other investigators

have used factor analysis to study the nature of human abilities. Some of these studies have been used merely to gain a greater understanding of the organization and components of such abilities. Others have been used as the foundation for the construction of multifactor tests. Through factor analysis, psychologists have contributed much to the understanding of tests of human behaviours.

Factor analysis consists of analysing tests in order to ascertain their factorial composition, and of analyzing the criterion in order to determine the nature and weight of the factors which enter into it. The former step makes possible the refinement of tests, cutting down the number of tests needed to predict success by eliminating overlapping and making each test do a maximum of work. The second step, namely the analysis of the criterion, indicates what types of tests should be stressed in order to improve prediction. In this investigation, only the former procedure has been adopted, while the criterion behaviour is analyzed through expert rating.

The use of factor analysis implies that tests can be statistically analyzed into a limited number of independent traits or aptitudes. The application of

Thurstone's centroid method of factor analysis with rotation of axes to a battery of tests has given us three types of component variances: (1) several common or group factors, that is components that appear in several tests, (2) possible specific factors, appearing in only one test and (3) error variance, arizing from the unreliability of measurement.

Defining factor analysis Fruchter (1967) writes,

> Factor analysis is essentially a statistical tool. In factor analysis a series of test scores or other measures are intercorrelated to determine the number of dimensions the test space occupies, and to identify those dimensions in terms of traits or other general concepts.

Factor analysis starts with inter-correlations among factors. These correlations indicate whether tests possess a common element. A formal factor analysis goes beyond inspection and calculates how much each test is influenced by various factors.

Three types of factors are commonly distinguished: General, Group and Specific. A specific factor is present in one test but not in any of the others. A group factor is present in more than one test. A general factor is a factor found in all tests. If all the correlations among

a set of tests are positive, one can find a general factor. The mathematical methods of factor analysis determine the correlation between each test, and hence each factor. These correlations provide a table of factor loading. The square of the factor loading tells how much each factor contributes to the variance of the test.

One important step in factor analysis is 'rotation'. Rotation is a procedure for placing factors so that results will be most meaningful. The analyst goes on eliminating more general factors in order to arrive at more specific factors.

Thurstone introduced the principle of simple structure. His aim was to describe complex performances as composites of simpler performances, that is, to break test scores into more fundamental elements. He planned his factor analysis to find group factors having small loadings in some tasks and large loadings in others. A 'simple structure' is one in which a large number of factor loadings are near zero, so that each test is described in terms of just a few factors. Thurstone first aimed to track down group factors, which would have zero loadings in some tests. Second, he aimed to discover 'pure' tests each of which would have a high loading on just one factor.

Thus verbal factor measures ability that is demanded by many tests and criteria, but is almost independent of numerical or mechanical ability.

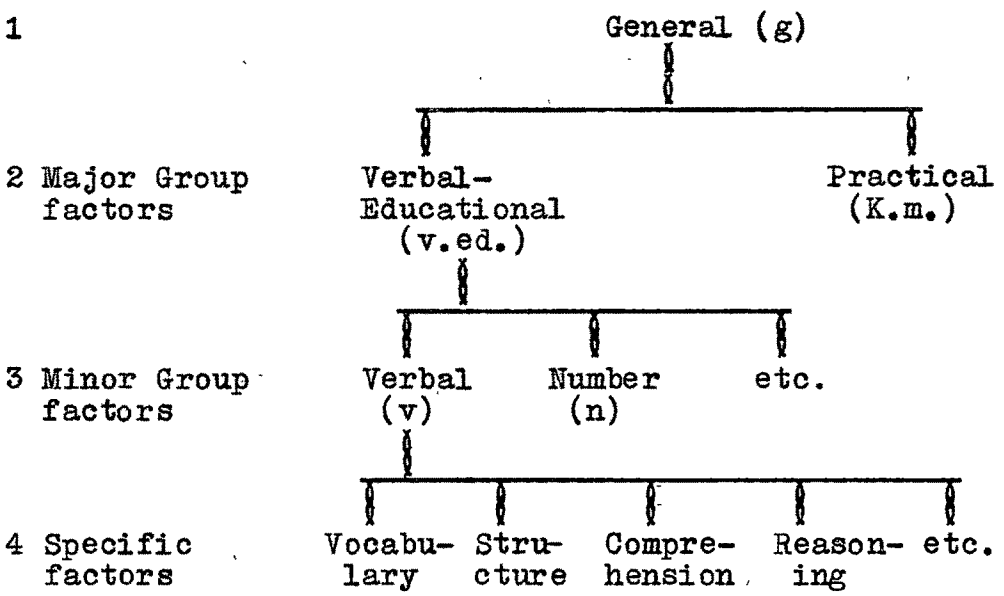
British investigators like Burt and Vernon are much less interested in pure measures of simple abilities; they, instead, rotate factors so as to identify broad factors present in a large number of tasks, for example V.ed. (Verbal-educational). British psychologists stress the existence of general ability. Burt, as quoted by Cronbach (1960), says,

> In nearly every factorial study of cognitive ability, the general factor commonly accounts for quite 50 % of the variance, while each of the minor factors accounts for only 10 % or less.

Thurstone and his students discard the view that factors are irreducible. While verbal tests have enough in common to define 'a verbal factor', they can be divided into several subgroups, thus establishing narrower factors within the verbal domain. The present investigation also tries to do the same.

According to Vernon (1956), abilities are most clearly described by a hierarchy ranging from the very broad factors to those present only in very specific

tests. He has suggested the following hierarchy :



Application of factor analysis techniques to results of test performance provides an insight about test validity. Authors of tests are becoming increasingly concerned about the extent to which various parts of their test, as well as the test as a whole, isolate and measure relatively independent factors. As a result of the wide-spread use of the factor analysis technique, now the authors of tests are able to provide the test-users with more exact information concerning their component elements.

Regarding the present position in factor theory, Cronbach (1960) writes,

From some points of view factor analysis has been a great success. It provides methods for handling large numbers of variables and for reducing them to a much smaller number of scores with little loss of information. It is a highly important statistical method. Secondly, factor analysis has cut through a large amount of nonsensical interpretation which results from assuming that every test with a different name measures a different ability. Thirdly, factor analysis helps to describe what a test measures. It is gradually establishing a reference system that all psychologists can use to describe tests.

(b) Different Methods of Factor Analysis

Various methods of factor analysis have been evolved. The chief among them are :

1. The Principal Component Method of Hotelling.
2. The Centroid Method of Thurstone.
3. The Principal Axes Method of Kelley.
4. The Summation Method of Cyril Burt.

The third is very similar to the first, and the second to the fourth. So only the first two are discussed.

1. The Hotelling Interative Procedure

The primary purpose of factor analysis for the present work was to provide some inference regarding the construct validity of the test. To achieve this Hotelling Interative procedure of latent roots or trace roots can be much useful.This procedure is mathematically more rigorous than others and can be applied completely objectively.

The data processing can be done on a computor. The programme reads a parameter card for

(a) the size of the matrix, and

(b) the test criterion for testing the difference between trial sectors.

The programme then reads and stores the entire correlation matrix and calculates the trace of the matrix as it is stored. The first set of factor loadings is then computed and punched. The sum of squares of the factor loading is computed and this is used in determining the percentage of total variance accounted for by this factor. The percentage of total variance accounted for is calculated from a cumulative sum of squares from factor to factor. The programme pauses after each factor to allow the operator to make a decision about extracting another factor. If another

extraction is desired, pushing start causes the residual matrix to be confronted and a branch to extract another factor. The self-correlations of the subtests are taken as 1. on the diagnonal axis.

This facilitates giving a complete account of all the factors in all the subtests neglecting the error variance.

To establish the construct validity of present test as a verbal ability test, it can be administered along with some tests purporting some other abilities such as Numerical Ability, or Mechanical Comprehension to the same sample. The inter-correlations among them may be arranged in a matrix. If the Hotelling Iterative Procedure is applied to such a matrix, it would show the following results :

(i) the high correlation among all the three tests - indicating the 'G' factor - general intelligence or general ability.

(ii) low or negative correlations of Verbal Ability with other abilities and a very high loading of the verbal factor with the verbal test.

This can be treated as a sufficient evidence for establishing the construct validity of the present test as a language ability test.

Test	N	L	M	Language Ability Criterion C
N	-	low	low	low
L		-	low	very high
M			-	low
Language Ability Criterion C				-

2. Thurstone's Centroid Method

This method has been developed on the basis of matrix algebra.

The term 'centroid' is borrowed from mechanics. It is a point in a mass where the centre of gravity is located. In factor analysis, the centroid of the end points of the test vectors might be considered the locations of the centre of gravity of equal weights at the points. A centroid is then the centre of gravity. Statistically regarded, it is a mean.

The purpose of factoring a correlation matrix is to account for the inter-correlations with fewer factors than there are tests. This factoring should be done so as to minimize the residuals after each factor has been determined. The main centroid axis is regarded as an approximation to the major principal axis of the

factor configuration. This main centroid axis is so placed that it has zero projections on all the remaining coordinate axes. This fact leads to the rule that "the sum of the coefficients in the correlation matrix is equal to the square of the sum of the first centroid loadings". The rule permits factoring through simple summational procedure after appropriate reflection. By 'reflecting' it is meant that each test vector retains its same length, but it extends in the opposite direction. The general policy is to reflect one test at a time and note the results; then to reflect a second one, and so on.

The extraction of each factor loading reduces the residuals in the correlation matrix. The factoring process is stopped when the standard deviation of the residuals is less than the standard error of a zero correlation.

With regard to sample size in Thurstone technique, Guilford advises to have minimum N of 200 when Pearson's 'r's are used. Factor loadings from samples near 200 have been fairly consistent with loadings in the same factors and tests - from samples above 1000.

(c) Factorial Validity of This Test

The investigator decided to adopt the Hotelling Principal Component technique for factor analysis. He took the correlation matrix to the Physical Research Laboratory, Ahmedabad, and got the cards punched for preparing the program. The data was fed to the computor and was analysed on it. The following is the factor matrix provided by the computor :

Table : 7·7 FACTOR MATRIX

	F_1	F_2	F_3	F_4
$Test_1$ (Voc.)	.84	- .26	- .10	.47
$Test_2$ (Stru.)	.81	- .04	- .49	- .33
$Test_3$ (Spel.)	.74	.66	.12	.07
$Test_4$ (Compr.)	.78	- .30	.49	- .22
% variance	63	14	13	10
Cummulative variance	63	77	90	100

The above factor matrix can be interpreted as follows :

(1) There is a group factor that covers 63 % of total variance and which can be called the Group Verbal factor.

(2) This verbal factor is positively correlated with all the tests.

(3) All the tests are highly loaded with this verbal factor, but especially Test I and Test II, that is Tests of Vocabulary and Structure.

(4) The second factor extracted is loaded by some ability requiring the command over spelling. This second factor accounts for 14 % of total variance.

(5) The third factor is related to comprehension which accounts for 13 % of total variance.

(6) The fourth factor accounts for 10 % of total variance. It is related to the knowledge and correct use of vocabulary.

(7) The first factor bears the following relation with the four tests separately (sum of squares) :

Vocabulary	+ .84
Structure	+ .81
Spelling	+ .74
Comprehension	+ .78

This indicates that there is a common verbal factor that enters all the four tests with high loading. It specially manifests itself through structure.

(8) Other factors related to spelling, comprehension and vocabulary have their separate identity.

These findings confirm the hypothesis with which the investigator started this research. It establishes the factorial validity of this test battery.

REFERENCES

1. Cronbach L. J., Essentials of Psychological Testing, New York, Harper, 1960. (Pp. 258, 262).

2. Fruchter Benjamin, Factor Analysis (East-West Student Edition), New York, Van Nostrand, 1967. (P. 2)

3. Garrett H. E., Statistics in Psychology and Education, Bombay, Allied Pacific Pvt. Ltd., 1962. (Pp. 347-348).

4. Guilford J. P., Psychometric Methods, New York, McGraw Hill, 1936. (P. 366)

5. Gullicksen H., Theory of Mental Tests, New York, Wiley, 1950. (P. 88).

6. Ross C. C. and Stanley J. C., Measurement in To-day's Schools, New Jersey, Prentice-Hall, 1941. (P. 107).

7. Vernon Philip E., The Measurement of Abilities (Second Edition), London, University of London Press, 1956.