CHAPTER VII

RELIABILITY AND VALIDITY OF THE TEST.

Introduction.

The basic facts taken into account while evaluating standardised instrument are (a) reliability, (b) validity and (c) practicability. Reliability deals with the accuracy and precision of measures. Validity deals with the extent to which the instrument measures what it proposes to measure. Practicability deals with factors like economy, convenience, and interpretability of scores. The test has been administered as per instructions in the manual. The norms have been established and have been presented in the form of a table from which the raw scores of the test can directly be converted into IQ measures. The test can be given in two class periods and can very easily be scored with the key. Thus the instrument is practicable.

Reliability.

A reliable instrument is that which gives the same measures of the quality of a thing, when measured by any person by following the instructions precisely and at any time until the thing does not change the quality that is being measured.

Intelligence is an innate ability which is almost constant.

Thus any reliable intelligence test should produce the same scores when it is taken by an individual at different times until he is in the same age group.

Definition of Reliability.

Reliability, as used in testing, " refers to the stability of a given measure on repeated application or as it is sometimes put, to the extent to which a test is consistent in measuring whatever it does measure".[1/]

When the test is prepared it is necessary to evaluate the accuracy or consistency of the obtained measures, or scores. It is practically not possible to get the same scores when the test is administered to the same individual at two different occassions. There will be deviations in the scores of the individuals. Reliability will tell about these deviations of the scores obtained by the same individual at two different occassions. A reliable test is precise, trustworthy, consistent and objective.

When the test is administered on two occassions, the measurement introduces some error due to chance. This error may be either large or small. If this difference is too much, the test is unreliable and if it is too small, it is reliable.

Importance of Reliability

"In any study of prediction and in any study of improvement resulting from training, some degree of reliability in the measure

1/Goodenough F.L., Mental Testing, Holt, Rinehart and Winston, New York 1961. pp. 564.

of the criterion being predicted or in the ability being trained is imperative if one is to achieve any predition on the one hand or any evidence of improvement on the other".[1/]

The information regarding reliability is crucial in the analytical study of the relationships among groups of tests.

<u>Factors Affecting Reliability</u>.

Reliability is the consistency of the achievement of individual or a group, but as said earlier, no two measures on the test on two different occassions, are identical. The reasons for this deviation, as classified by Thorndike[2/] are given below:-

1. Lasting and general characteristics of the individual (general skills of taking test, ability to comprehend instructions, etc.)
2. Lasting but specific characteristics of the individual. (knowledges and skills, specific to certain forms of test items)
3. Temporary but general characteristics of the individual. (health, fatigue, motivation etc.)
4. Temporary and specific characteristics of the individual. (comprehension of tasks, specific tricks, level of practice of skills involved etc.)
5. Systematic or chance factors affecting the administration.

1/Thorndike R.L., 'Reliability', Chapter 15, <u>Educational Measurement</u> Lindquist E.F., (Editor, American Council on Education, Washington D.C. 1966. pp. 563.

2/<u>Ibid</u>, pp. 568.

(conditions of testing, freedom from distructions, clarity of instructions etc.)

6. Variance not otherwise accounted for.
(luck in guessing).

Cronbach[1/] has also suggested similar type of analysis of factors affecting the reliability of a test.

Evaluation of Reliability.

The evaluation of reliability of measuring instrument involves two types of operations - experimental and statistical. The test is given to a defined group of individuals under specified conditions and the obtained scores are treated statistically to yield a value to represent the reliability characteristics of the test.

The problem of estimating the error between the two scores obtained at two different administrations, is attacked in two different ways. In the first, the actual magnitude of the error of measurement is found in the same units in which the scores are expressed. The deviation of the scores are expressed in terms of standard error (SE) of the measure. The SE of all statistics, considered so far, have already been found out.

The second approach is in terms of the consistency with which the individual maintains his position in the total group when the measurement is repeated.

The first approach deals with the reliability of the obtained

1/Cronbach L.J., Essentials of Psychological Testing. Harper & Row, New York, and John Weatherhill, Inc. Tokyo 1965. pp. 128.

statistics and the second approach deals with the reliability of the whole test.

There is a possibility of confusing reliability of a test with the reliability of statistical measure like mean, median, standard deviation, skewness, kurtosis, correlation or difference between means etc. which is expressed in terms of SE. The difference between these two types of reliability is made very clear by Anastasi in his statement "Sampling error pertains to the consistency of results obtained when observations are repeated on different individuals; error of measurement, to the consistency of results obtained when the observations are repeated on the same sample".[1/]

Reliability Coefficient.

The reliability is statistically expressed in the form of reliability coefficient. It is the correlation between the two sets of measurement obtained in the same manner. Technically the reliability coefficient gives information regarding proportions of true variance and error variance. The characteristics of reliability coefficient as stated by Cronbach are as given below:-

"A reliability coefficient tells what proportion of the test variance is nonerror variance.

The reliability coefficient depends on the length of the test.

The reliability coefficient depends on the spread of scores in the group studied.

1/Anastasi Anne, Psychological Testing, The Macmillan Company, New York 1965. pp. 105.

A test may measure reliably at one level of ability and unreliably at another level.

The validity coefficient cannot exceed the square root of the reliability coefficient".[1/]

Types of Reliability.

The word reliability is used to cover several aspects of score consistency. No one type of reliability is universally preferred. The choice depends on the use for which the test is put.

The various types of determiners of reliability coefficients, some times increase the reliability coefficient and some times decrease it. Moreover each type gives rise to different reliability coefficients. So to distinguish between the reliabilities of a test obtained by different methods, they are named differently. The three types of reliability coefficients, that are generally used in expressing the consistency of measurement of a psychological test are as given below:-

(a) Coefficient of stability

(b) Coefficient of equivalence

(c) Coefficient of internal consistency.

Coefficient of stability.

This tells us how stable this particular performance is over a given period of time.

1/Cronbach L.J., Op.cit., pp. 129.

The test is given to the same group of individuals under identical conditions after a certain period of time and the co-rrelation between the two sets of scores is computed. This method is called the 'test-retest' method and the obtained correlation is called the coefficient of stability. This is a simple method of computing reliability of a test. Moreover it is very easy to apply.

If the test is given, immediately, it is possible that the subjects would recall the previous answers and will have more time at their disposal for dealing with the new items which they had not attempted during the first trial. "Besides the memory effect, practice and the confidence induced by familiarity with the material will almost certainly affect scores when one takes the test for the second time".[1/] To minimise this error, if the time interval between the two trials is increased, the factors like growth, and maturity of the individual will affect the coefficient of stability. There is no definite experimental evidence to decide about the time interval between the two trials, some suggest that it should be some weeks.

Moreover the factors like the moods of the individuals, extreme climatic conditions etc., which are beyond the control of the administrator and the testee, are likely to affect the performance of the group at two different times.

1/Garrett H.E., Statistics in Psychology and Education. Longmans, Green and Co., New York, London, Toronto 1954. pp. 333.

"Because of the difficulty in controlling the conditions which influence scores on different administrations of a test, the test-retest method is used less generally than are the other two methods".[1]

Coefficient of Equivalence.

To avoid the errors introduced in the measurement of reliability due to too short or too long time intervals between the two administrations (as done in test-retest method), equivalent forms of tests are constructed. These two tests are similar to each other in the kind of content, mental processes required, number of items, difficulty levels, discriminating indices etc. Statistically, they have equal means, equal variances, and very high correlation with each other.

One form is given first and the second one is given as early as possible. The coefficient of correlation between the two sets of scores is computed. This measure of reliability is called coefficient of equivalence. If the obtained 'r' is high, both the forms measure, what they propose to measure, with equal accuracy.

There are certain problems in applying this procedure. The two equivalent forms may have some specific variance in both which may under-estimate the reliability. If they overlap to a great extent, it will introduce not merely chance error but some systematic error too.

1/Ibid, pp. 333.

It is also very tiresome to prepare two forms of a test merely to obtain an estimate of the reliability of a test. It also requires time for two administrations, which introduces the errors like moods, climatic conditions etc.

In order to avoid this difficulty, the test is artificially divided into two half lengths and the correlation between the performances on the two parts is computed. This is the reliability of the half test. Then by using the Spearman-Brown formula the reliability of the whole test is found out. This method of finding the reliability is known as split-half method.

There are different ways of splitting the test into two parts. "The more usual procedures include: (a) selecting sets of items for the two half tests which appear equivalent in content and difficulty, (b) putting alternate items or trials in each half test, (c) putting alternate groups of items or trials in each half test, (d) using the first half of the items or trials as one half-test and second half as the other".[1/] The most commonly used procedure for splitting the test is putting alternate items in each half test.

The two parts are not separately timed but the performances on the two parts are adjacent due to which the fluctuations in conditions and minute-to-minute variance in performance are equated for both sets of scores.

1/Thorndike R.L., Op.cit., pp. 579-580.

Coefficient of Internal Consistency.

It is the term used to indicate the extent to which separate items or parts of a test are correlated with each other. It is a type of reliabilitỵ coefficient obtained when either split-halves or Kuder-Richardson formulas are used for computing it".[1/]

The split-half method has already been described above. Kuder-Richardson method doesnot require splitting the test into two halves. It also does not require the rescoring of the test and calculation of the correlation coefficient. The data required for simpler method are the number of items in the test, standard deviation of the scores and their arithmetical mean. The formula used is

$$r_{1I} = \frac{n\,\sigma_t^2 - M\,(n - M)}{\sigma_t^2\,(n - 1)}$$

in which

r_{1I} = reliability of the whole test

n = number of items in the test

σ_t = SD of the test scores

M = the mean of the test scores.

Another formula used for estimating test reliability coefficient is Kuder-Richardson formula 20 which reads as

$$r_{1I} = \frac{n}{(n - 1)} \times \frac{\sigma_t^2 - pq}{\sigma_t^2}$$

1/Remmers H.H., Gage N.L., Rummel J.F., A Practical Introduction to Measurement and Evaluation. Harper & Row, New York and John Weatherhill Inc. Tokyo 1966. pp. 371.

in which

r_{1I} = reliability coefficient of the whole test

n = number of items in the test

σ_t = the SD of the test scores

p = the proportion of the group answering a test item correctly

q = (1 - p) = the proportion of the group answering a test item incorrectly.

Another method of estimating the coefficient of internal consistency is suggested by Cyril Hoyt. As described by Shah, "He assumes that the score of an individual on a test may be divided into four independent components, as follows:-

(i) A component common to all individuals and to all items

(ii) A component associated with item

(iii) A component associated with the individual

(iv) An error component that is independent of (i), (ii) and (iii)".[1/]

Reliability coefficient is computed by using the formula,

$$\text{Reliability} = 1 - \frac{\text{Error variance}}{\text{Variance among individuals}}.$$

Reliability of the Present Test.

The methods used for computing the reliability of the present

1/Shah M.M., An aptitude Test for Secondary School Teachers. The Maharaja Sayajirao University of Baroda, Baroda 1965. pp. 174.

test are,

1. Test-Retest method

2. Split-half method

3. Kuder-Richardson formula (approxim-ation to formula 20)

Test-Retest Method

Retesting was done in the following schools after a period of five weeks:-

1. Zilla Parishad Boys High School, Nilanga.

2. Bharat Vidyalaya, Omerga.

3. Kamdhenu Vidyalay, Makegaon.

4. Zilla Parishad Boys High School, Kallam.

5. Shri Krishna Vidyalay, Gunjoti.

6. Shri Paramhansa Vidyalay, Yeneguru.

Only 361 students tested in these schools were available for retesting. The performance of 9 pupils out of these 361 had been discarded in the first trial, as these were either above 17 or below 13 years of age. The performance of two more pupils selected at random was discarded to make N a round figure 350.

The answer sheets were evaluated and the scores on two trials were tabulated. The correlation coefficient of the two sets of scores was computed by product-moment method with the obtained scattergram given below:-

Table 39. Scattergram of Scores Used in Test-Retest Method.

Test Scores (y variables)	Re-test Scores (X variable) 30-39	40-49	50-59	60-69	70-79	80-89	90-99	100-109	110-119	120-129	fy
120-129	-	-	-	-	-	-	-	-	1	2	3
110-119	-	-	-	-	-	-	-	1	4	-	5
100-109	-	-	-	-	-	2	2	6	3	-	13
90-99	-	-	-	-	1	2	20	5	-	-	28
80-89	-	-	4	10	11	21	10	5	-	-	61
70-79	-	-	3	13	71	15	6	1	-	-	109
60-69	-	2	9	35	10	12	5	-	-	-	73
50-59	2	3	14	7	4	1	-	-	-	-	31
40-49	1	8	6	3	2	-	-	-	-	-	20
30-39	4	2	1	-	-	-	-	-	-	-	7
fx	7	15	37	68	99	53	43	18	8	2	350

-246-

	Test	Retest
Mean	73.59	75.07
SD	16.31	17.12
SE_{Mean}	0.3094	0.3247
SE_{SD}	0.2197	0.2305

From this it can be seen that the mean of the sample used for retesting is almost the same as the mean of the whole group. (M = 71.987 and SD = 15.42)

The coefficient of correlation computed by product moment method is 0.81.

SE of r was computed by using the formula

$$SE_r = \frac{(1-r^2)}{\sqrt{N}} = \frac{1 - .81^2}{\sqrt{350}} = .01839$$

As N = 350, df = 350-2 = 348.

Table No. 25 [1/] shows that when df = 348, r greater than 0.106 and .138 is significant at 0.05 and 0.01 levels of significance respectively. The obtained r .81 is larger than .138 and hence it is significant at .01 level of significance.

$$P.E \text{ of } r = 0.6745 \times \frac{1 - r^2}{\sqrt{350}}$$

$$= 0.6745 \times \frac{.3439}{\sqrt{350}}$$

$$= .01240$$

1/Garrett H.E., _Op.cit_. pp. 200.

The coefficient of stability of the test is .81 and PE being 0.01240.

It is seen that the increase in the mean is 1.48 (75.07 - 73.59).

Split-half Method.

A sample of 1300 out of the total sample of 7745, was selected for applying the split-half method. The test was divided into two halves, odds and evens. The answer sheets with serial numbers ending with 0 and 5 were selected.

The sample selected should be very similar to that from which it has been drawn.

Table 40. Data Grouped for the Calculation of Mean & Standard Deviation of the sample selected for split half method.

Class interval scores.	f	x'	fx'	fx'^2	cum f
1	2	3	4	5	6
120-129	2	5	10	50	1300
110-119	22	4	88	352	1298
100-109	67	3	201	603	1276
90-99	137	2	274	548	1209
80-89	212	1	212	212	1072
70-79	299	0	000	000	860
60-69	302	-1	-302	302	561
50-59	221	-2	-442	884	259

(concluded on next page)

Table 40. (concluded)

1	2	3	4	5	6
40-49	34	-3	-102	306	38
30-39	4	-4	- 16	64	4

$N = 1300 \quad \Sigma fx' = -77 \quad \Sigma fx'^2 = 3321$

$$C = \frac{fx'}{N} = \frac{-77}{1300} = -0.05924$$

$$c^2 = (-0.05924)^2 = 0.003510$$

$$ci = -0.05924 \times 10 = -0.5924 = .59$$

$$\text{Mean} = \text{Assumed Mean} + ci = 74.5 - 0.59 = 73.91$$

$$\text{Median} = l + \left(\frac{\frac{N}{2} - F}{fm} \right) = 69.5 + \frac{650 - 561}{299} \times 10 = 72\ .477$$

$$SD = i \sqrt{\frac{fx'^2}{N} - c^2} = 10 \sqrt{\frac{3321}{1300} - .003510} = 10 \sqrt{2.555 - .003510} = 10 \times 1.597 = 15.97$$

Table 41. Mean, Median and SD of the whole sample and sample selected for split-half method.

Sample	Mean	Median	SD
Whole sample	71.987	71.690	15.42
Sample for split-half method	73.91	72.477	15.97

From the table above it may be observed that the statistics of the whole sample and sample selected for split-half method are almost the same.

The selected answer-sheets were re-assessed. The scores on the odd and even items were found out separately and the scattergram of the scores on odd and even items was prepared for the computation of correlation coefficient by product-moment method.

Table 42. Scattergram of scores used in split-half method.

Scores on Odd Itǝms (x variable)

Scores on Even Items (y variable)	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	fy
66-70	-	-	-	-	-	-	-	-	-	-	-	0
61-65	-	-	-	-	-	-	-	-	-	1	1	2
56-60	-	-	-	-	-	-	-	2	14	8	-	24
51-55	-	-	-	-	-	-	21	45	15	-	-	81
46-50	-	-	-	-	-	57	75	20	5	-	-	157
41-45	-	-	-	-	33	78	37	17	4	-	-	169
36-40	-	-	-	62	164	83	39	1	-	-	-	349
31-35	-	-	21	85	77	16	3	-	-	-	-	202
26-30	-	27	71	58	20	-	-	-	-	-	-	176
21-25	4	36	57	20	2	-	-	-	-	-	-	119
16-20	4	13	4	-	-	-	-	-	-	-	-	21
	8	76	153	225	296	234	175	85	38	9	2	1300

-251-

The coefficient of correlation computed from this scattergram is 0.9067. As the test has been divided into two parts, this is the reliability coefficient of half the test. By using Spearman-Brown formula the self correlation of the whole test was computed.

The Spearman-Brown formula[1/] for estimating the reliability from two comparable halves of a test is as given below:-

$$r_{1I} = \frac{2r_{\frac{1}{2} \frac{I}{II}}}{1 + r_{\frac{1}{2} \frac{I}{II}}}$$

in which

r_{1I} = reliability coefficient of the whole test

$r_{\frac{1}{2} \frac{I}{II}}$ = reliability coefficient of one half of the test found experimentally.

$r_{\frac{1}{2} \frac{I}{II}}$ in the present case is o.9067. Substituting this value in the above formula, we get

$$r_{1I} = \frac{2 \times 0.9067}{1 + 0.9067}$$

$$= \frac{1.8134}{1.9067}$$

$$= 0.9508 = 0.95$$

$$\text{The PE of } r = 0.6745 \times \frac{1 - r^2}{\sqrt{1300}}$$

$$= \frac{0.6745 \times (1-(0.9508)^2\ )}{\sqrt{1300}}$$

1/Ibid, pp. 341.

= 0.002261

Kuder-Richardson Method - Rational Equivalance:

The simple approximation formula is used to determine the reliability of the test. The formula used is

$$r_{1I} = \frac{n\,\sigma_t^2 - M\,(n-M)}{\sigma_t^2\,(n-1)}$$

$n = 154$

$\sigma_t = 15.42$

$M = 71.98$

Therefore $r_{1I} = \frac{154 \times (15.42)^2 - 71.98(154 - 71.98)}{(15.42)^2\,(154-1)}$

$= 0.8443$

$= 0.84$

Table 43. Reliability Coefficient Obtained by Different Methods.

Method	Obtained Reliability Coefficient
Test-Retest Method	0.81
Split-Half Method	0.95
Kuder-Richardson Method	0.84

The split-half method gives a little higher value. Shaha suggests that the discripancy between the reliability coefficients obtained by split-half method and by Kuder-Richardson Method "might be attributed to overestimation of reliability coefficient by the split-half method or to under-estimation by the use of K-R formula ......". The reliability coefficient estimated by test-retest method may be either high or low depending upon the nature of the test and the difficulties in controlling conditions which influence scores on retest. So the test retest method is generally less useful. In this case it is observed that the reliability coefficient obtained by this method is less than that obtained by other methods.

So it can safely be said from the experimentally results, that the reliability coefficient of the test will not be less than 0.81. For interpretation and other uses of reliability coefficient, the obtained minimum value namely 0.81 has been treated as the reliability of the test.

Reliability Coefficient as a Measure of True Variance.

The variance of the test score consists of two parts namely variance of true scores and variance of chance errors.

The relation between them is expressed mathematically as giv en below:-[1/]

$$1 = \frac{\sigma_{\infty}^2}{\sigma_x^2} + \frac{\sigma_e^2}{\sigma_x^2}$$

where $\frac{\sigma^2}{\sigma_x^2}$ = true score variance

1/Garrett H.E., Statistics in Psychology and Education. Vakils, Feffer and Simons Private Ltd., Bombay 1. 1971 pp. 346.

$$\frac{\sigma_e^2}{\sigma_x^2} = \text{error variance}$$

Under the reasonable assumption, that true scores and errors are independent, it may be said that reliability coefficient is the true score variance.

So, $$r_{1I} = \frac{\sigma_\infty^2}{\sigma_x^2}$$

The above equation changes to

$$1 = r_{1I} + \frac{\sigma_e^2}{\sigma_x^2}$$

$$\therefore \quad r_{1I} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$$

So if the variance of chance error is small, the reliability of the test is high.

Estimating True Scores Using Regression Equation and Reliability Coefficient.

True score can be estimated from the reliability coefficient by using the regression equation given below:-

$$\bar{X}_\infty = r_{1I} X_1 + (1 - r_{1I})M_1$$

where

$\bar{X}_\infty$ = estimated true score on the test

X_1 = obtained score on test 1.

M_1 = mean of test 1 distribution (71.98)

r_{1I} = reliability coefficient of test 1 (0.81)

$$\bar{X} = 0.81\,X + (1 - 0.81) \times 71.98$$

$$= 0.81\,X + 0.19 \times 71.98$$

$$= 0.81\,X + 13.68$$

The standard error (SE) of an estimated true score is computed by using the formula

$$SE_{\infty} = \sigma\sqrt{r_{1I} - r_{1I}^2}$$

where

SE_{∞} = Standard error of estimated true score

σ = standard deviation (15.42)

r_{1I} = reliability coefficient (0.81)

$$\therefore SE_{\infty} = 15.42\sqrt{.81 - .81^2}$$

$$= 15.42\sqrt{.81 - .6561}$$

$$= 15.42\sqrt{.1539}$$

$$= 6.091$$

= 6 (nearest whole number).

0.95 interval is $\bar{X} \pm 1.96 \times 6$

$= \bar{X} \pm 11.76$

$= \bar{X} \pm 12$ (nearest whole number).

Index of Reliability.

The correlation between the obtained scores and their corresponding true scores is given by the formula,

$$r_{1\infty} = \sqrt{r_{1I}}$$

where

$r_{1\infty}$ = the index of reliability or the correlation between obtained and true scores

r_{1I} = reliability coefficient (0.81)

$$\therefore r_{1\infty} = \sqrt{0.81}$$
$$= 0.9$$

So 0.9 is the maximum correlation which the test is capable of yielding in the present form.

Validity.

As Cureton points out, "The essential question of test validity is how well a test does the job it is employed to do".[1/] The validity may be high, moderate or low according to the purpose for which the test is put.

A highly reliable instrument may not necessarily be a valid one. A false balance, a balance with unequal arms, may be highly reliable as it gives the same weight of a body when weighing is repeated under the same conditions. But the obtained weight is is not a valid one because if the weight of the body is found by using a balance of known validity there is a significant variation

1/Cureton E.E., "Validity". chapter 16. Educational Measurement. Lindquist E.F., (Editor), American Council on Education, Washington D.C. 1966. pp. 621

in the weight. Similarly a reliable test, i.e. a test which is capable of giving consistent scores when repeatedly administered on a particular group, may not necessarily be a valid one.

Definition.

"(1) In mental measurement the term is defined as the degree to which a test measures that which it purports to measure (Otis); (2) in more general sense a conclusion is said to be valid if it is a logical deduction from the premisis assumed".[1/]

The performance on a test is measured in terms of scored. But the mere sdores are meaningless unless they are related with magnitude of certain ability which the test proposes to measure. As Thorndike and Hagen propose, "We must find some way of establishing the extent to which the performance on the test actually corresponds to the quality of behaviour in which we are directly interested".[2/]

They also propose, "A test may be thought of as corresponding to some aspect of human behaviour in any of the three senses. For these three senses we shall use the terms (1) represent, (2) predict and (3) signify".[3/]

Validity as Representing.

By undergoing certain education or training, the individual

1/Goodenough F.L., Op.cit., pp. 569

2/Thorndike R.L., Hegen E., Measurement and Evaluation in Psychology and Education, John Wiley & Sons, Inc. New York, London 1961. pp. 161.

3/Ibid, pp. 161.

is expected to achieve certain goals, in the form of knowledge, comprehension, skill etc. If the performance on the test calls for knowledge, skill etc., the performance on the test represents the achievement on these goals. Since analysis of the items of the test is largely in terms of the content of the test, the term content validity is also used for this purpose. As it deals with the achievement of certain goals, it is important for estimating the validity of achievement test.

Validity as Predicting.

Tests are also used to predict some specific future outcomes. The procedure used is to give the test to a group of persons who are entering some job or training, follow them up, and then measure their success in the particular field or training. Find the correlation between the scores on the test and success in the course (criterion measure). Higher the correlation, better is the predictive validity of the test.

The four qualities that are expected to be possessed by criterion measures, as suggested by Thorndike and Hagen,[1/] are (a) relevance (b) freedom from bias (c) reliability and (d) availability.

Validity as Signifying.

This type of validity tries to answer the question "How well does this test mean or signify?". It tells what the scores tell

1/Ibid, pp. 166

about the individual. So it is sometimes called the construct validity.

Validity is a relative term. A test valid for a particular purpose may not be valid for certain other purpose. So several types of validity may be thought of, depending upon the purpose for which the test is to be used. The different types of validity that have been used in the testing programme may be classified as follows:-

Validity

- Rational Validity
  - Content validity
  - Concept or construct validity
- Statistical Validity
  - Congruent validity
  - Concorrent validity
  - Predictive validity

To this may be added factor validity.

Content Validity

It is concerned with the content of curriculum and that of the test. "Content validity is evaluated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are to be drawn".[1] So it is used in case of achievement tests.

Construct Validity.

"Construct validity is evaluated by investigating what

1/Remmers H.H. etc., *Op.cit*., pp. 120

psychological qualities a test measures, or in other words, by demonstrating that certain explanatory constructs account for performance on the test".[1/] It is used when the tester has no definite criterion measure.

Congruent Validity.

Congruent validity refers to the correlation of the test with an existing similar measure of the same function. This validity coefficient is valuable only if the validity of the criterion test is testified.

Concurrent Validity.

"Concurrent validity is evaluated by showing how well the test scores correspond to already accepted measures of performance or status made at the same time".[2/]

The individual's performance in the school subjects is greatly influenced by the intelligence he possesses. So to evaluate concurrent validity, the scores on the test are correlated with the teachers report on the abilities of the tested individuals. If the correlation between these two estimates is high then the concurrent validity is high.

Predictive Validity.

"Predictive validity is evaluated by showing how well predictions made from the test are confirmed by evidence gathered at

---

1/*Ibid*, pp. 121

2/*Ibid*, pp. 120

some subsequent time".[1/] It is similar to concurrent validity but evidence in this case is collected after some time. It will tell how well the individual will do in his college courses, or in any profession or vocation he proposes to undertake.

Factorial Validity.

In the process of factor analysis, the intercorrelations of tests are examined and they are accounted for in terms of smaller number of factors. By applying this process, the validity of a test is defined in terms of factor loadings.

Validity of the Present Test.

(1) The purpose of the present test is to measure the general mental ability - intelligence. On the basis of the performance of this test, the child is to be given educational and vocational guidance.

To give him educational guidance, estimation of achievement in the secondary school certificate examination conducted by the Maharashtra State Board of Secondary Examination is to be found out. The criterion measure in this case is the score obtained in the S.S.C.Examination. For this purpose, computation of concurrent validity and Predictive validity of the test is necessary.

(2) There are number of tests prepared so far, though no one of them has been standardised for the children in this area.

1/Ibid, pp. 120.

However the results of this test can be compared with those of another test with known validity. Thus the congruent validity is to be estimated.

(3) As the test is based on the hierarchical model, it should indicate the factors that are measured by the test. So there is a need to estimate the factorial validity of the test. This factorial validity will also indirectly speak about the construct validity.

(4) As it is not an achievement test, there is no need to find out the content validity of this test. The content validity is found out by judgement.

In chapter III the abilities involved in the performances of an individual on this test have been described in details. The inspection of the test shows that the universe from which the items have been selected is reasonably wide. It may be also observed that the sampling of the abilities and the fields of experiences from which items are drawn, are reasonably adequate when they are compared with other tests of intelligence. This is enough to prove the content validity of the test.

(5) As the criterion measure, in terms of scholastic achievement is available, there is no need of giving the concept or construct validity. Moreover the factorial validity is giving the nature of concept of intelligence as measured by the test.

(6) There are many studies done so far to find the relationship between the IQ of the individual and his proficiency in a

particular field of life. If the IQs obtained by this test match with the IQs obtained from other recognised tests, then the predictive value of this test is the same as that of the recognised test. Moreover it requires time and the follow up of the success of the individuals in the various fields of life. Thus the predictive validity of the test is not calculated in this case.

So the following three validities have been computed.

1. The concurrent validity in terms of correlation between the scores on the test and school marks.

2. The congruent validity in terms of correlation with another test of intelligence.

3. Factorial validity in terms of factor loadings and the correlation of each sub-test with the whole test.

Concurrent Validity.

The test was given to 400 students in four schools. The students were from standards VIII and IX. Their answer sheets were scored and their IQs corresponding to these scores and their ages were found out from the tables of norms.

As the students were from different schools and from different standards instead of using the absolute total marks obtained in their annual examination the percentage of marks were used for preparing the scattergram of the IQs and the scores on the scholastic achievement.

Table 44. Scattergram of IQs of the individuals and the percentage of marks in the annual examination.

IQs of individuals 'y' variable	Percentage of Marks in Annual examination 'X' variable										
	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	71-75	fy
130 & above	-	-	-	1	1	-	-	-	2	1	5
120-129	-	3	3	2	-	-	15	12	10	-	45
110-119	4	5	6	4	46	20	20	21	-	-	126
100-109	1	6	1	3	59	41	12	-	-	-	123
90-99	7	3	7	8	11	5	7	5	-	-	53
80-89	5	9	13	3	2	-	-	-	-	-	32
70-7 9	2	4	8	2	-	-	-	-	-	-	16
	19	30	38	23	119	66	54	38	12	1	400

-265-

The coefficient of correlation obtained from the above scattergram by product moment method works out to be 0.5552 and PE = 0.02333.

The validity computed by this method is usually low because the assessment of the papers of the annual examinations is more liberal as the schools do not wish to detain more pupils in the class. The liberal assessment is observed especially in case of students getting 25 to 35 percent of marks. Moreover the pupils are likely to adopt foul means to pass the annual examination. If they fail in the examination they lose one year and also the economical benifits like freeships etc.

Congruent Validity.

The congruent validity of a test is found out in terms of the coefficient of correlation between the scores obtained by the same group on two tests doing the same function.

Since there was no verbal test suitable in Marathi, a non-verbal test of intelligence (N.V.T.I.) by Dr. Nafade has been used for finding the congruent validity. The same test has been used for validating the items. The other details about the test have already been given on page 63, and the test is used by the Guidance Bureau of the Department of Education, Government of Maharashtra.

This test was given to 220 pupils of standards VIII, IX and X in four high schools and on the very next day the N.V.T.I. was administered to the same group. The IQs of the individuals on both the tests were computed separately and the scattergram was prepared for computing the coefficient of correlation.

Table 45. Scattergram of IQ scores Obtained by N.V.T.I. and the Present Test.

IQ scores on Present Test (Y axis)	IQ scores on N.V.T.I. (X axis)							
	70-79	80-89	90-99	100-109	110-119	120-129	130-139	fy
130-139	-	-	-	-	-	-	2	2
120-129	-	-	-	1	8	9	-	18
110-119	-	-	-	12	27	7	-	46
100-109	-	-	12	27	9	-	-	48
90-99	-	2	35	21	3	-	-	61
80-89	-	28	12	-	-	-	-	40
70-79	5	-	-	-	-	-	-	5
	5	30	59	61	47	16	2	220

-267-

The coefficient of correlation by product moment method is 0.8586 and PE = 0.01060.

Is the obtained validity coefficient of this test adequate ? It is very difficult to answer this question. As indicated earlier, validity is not general. It is specific. Cronbach points out, "The ultimate judgement as to the validity of the test must be made by the user, who alone can decide whether the evidence indicates that the test is suitable for his unique purposes and situation".[1/]

As the test constructor cannot anticipate the various purposes of the user, the former may only state the different validity coefficients in the manual.

Another way is to compare the obtained validities and reliabilities with those of the other tests.

The statistics of some of the tests are given below:-

Name of the Test.	Reliability		Validity	
	Test Retest	Split Half	School Marks	With other Test.
1	2	3	4	5
1. Desai's Grouptest of intelligence	.77	.94	.53	.
2. Pathak's Test of intelligence.		.89		.74

(continued on next page)

1/Cronbach L.J., "Validity", Encyclopedia of Educational Research. Harris C.W. (Editor), The Macmillan Company, New York 1960 pp. 1555.

1	2	3	4	5
3. Prayag Mehta's Test			.44	
4. N. Samarth's adaptation of Northumberland mental Tests No. 2		.70	.56	
5. Group Test of Intelligence by Lele and others	.653 to .885		.41 to .58	0.55 to .85
6. C.L.Bhat's Test	.86	.96	.45 to .57	0.68 to .88
Present Test	.81	.95	.555	0.8586

(All figures are adopted from the First Mental Measurement Hand Book for India)

From the above table it can be seen that the coefficients of reliability and validity of this test are comparable with some of the standardised intelligence tests prepared in Bilingual Bombay State or Maharashtra. So it can be said with a certain degree of confidence that this test is considerably good for measuring intelligence of pupils in this area.

Factor Validity.

The test constructors are interested in knowing whether the test scores are due to a single source of variation or are due to the combined functioning of different mental traits. They also are interested in knowing whether the various abilities that have been named so far, are really different from each other or represent the combinations of the some basic ones. A statistical approach namely factor analysis tries to answer these questions.

An ability or capacity, as pointed out by Vernon, "....... implies the existence of a group of category of performances which correlate highly with one another, and which are relatively distinct from (............) other performances".1/

The correlation between two tests of the mental ability is due to the general factor like intelligence, which enters all abilities to some extent. It may be to some extent due to a group factor which occurs in a group of performances of a restricted type.

"The statistical investigations of Spearman (1927) and others have shown that it is possible to account for practically the whole of a set of test inter-correlations by postulating appropriate common factors".2/ The statistical approach used for such accounting, is known as factor analysis.

As Fruchter describes it, "It is a method of analyzing this set of observations from their inter-correlations to determine whether the variations represented can be accounted for by a number of basic categories smaller than that with which the investigation was started".3/

1/Vernon P.E., The Measurement of Abilities. University of London Press Ltd., London, E.C. 4 1961. pp. 131.

2/Ibid, pp. 137.

3/Fruchter B., Introduction to Factor Analysis, D.Van Nostrand Company, Inc. Princeton, Afficiated East-West Press Ltd., New Delhi 1967. pp. 1.

As Anastasy points out " The principal object of factor analysis is to simplify the description of data by reducing the number of necessary variables or dimensions".[1/] The factor analysis starts with correlation matrix and ends with a factor matrix. The first (correlation matrix) is a table showing the correlation of each test with each other test and the factor matrix is the loading of each of the factorsin each test.

Factor.

Factors are not entities of mind. They primarily consist of categories for classifying mental tests. Factor is "one of the elements or qualities which enter into a product determined by factor analysis".[2/] These are not casual factors but descriptive categories. They are not psychological entities but functional unities, or aggregates of elementary components.

Assumptions.

(1) "The basic assumption of factor analysis is that a battery of intercorrelated variables has common factors running through it and that the scores of an individual can be represented more economically in terms of these reference factors".[3/] The score of an individual on a test depends on (a) the particular abilities assessed by the test and (b) the particular abilities possessed by

1/Anastasy Anne, Op.cit. pp. 338.

2/Goodnough F.L., Op.Cit. pp. 551.

3/Fructer B., Op.cit. pp. 44.

the individuals.

Variance is the index of extent to which a test discriminates individual differences. The variance of a variable can be subdivided into three parts namely common variance, specific variance and error variance. The portion of the variance which correlates with other variables is the common variance and the one which does not correlate with any other variable is the specific variance and the part due to chance error is the error variance. The reliable variance is the sum of common variance and specific variance and total variance is the sum of the reliable variance and the error variance.

The values of the square roots of the common variances are called the factor loadings. The sum of the independent common variance is called the communality and is represented by the symbol h^2.

(2) "A second assumption of factor analysis is that the correlation between two variables j and k can be accounted for the nature and extent of their common factor loadings. For orthogonal factors this can be represented by the equation:

$$r_{jk} = a_{j1} a_{k1} + a_{j2} a_{k2} + \dots\dots\dots + a_{jr} a_{kr} \text{ ."}$$ 1/

The factor validity of a given test is defined in terms of its factor loadings and are given by its correlation with each factor.

1/*Ibid*, pp. 47-48.

Methods of Factor Analysis.

"Since Spearman proposed his criterion of the tetrad difference, a number of procedures for factor analysis have been proposed".[1/] The chief among them are "the method of principal components" by Hotelling, "the method of principal axes by Kelley, " the method of summation" by Burt, and "the centroid method" by Thurstone. The first two have much in common. Similarly the second two also have much in common. The methods of Hotelling and Kelley are mathematically more regorous but the factors are difficult to be interpreted psychologically. Burt and Holzinger methods impose some arbitrary restrictions, one of which is the requirement of g as a factor. In England mostly Burt's method is used and in America, Thurston's method is used. The main purpose of factor analysis is to reduce the number of variables to explain the obtained data. For this purpose, as Guilford points out, "almost any method of factor analysis will do, with or without rotation of axes".[2/], The centroid method of Thurston has been used for the factor analysis in this case, as it is computationally less laborious.

Sample For Inter Correlations.

The first step in carrying out factor analysis is to compute the inter-correlations of each test with other tests. For this a sample of 1,000 pupils has been selected. The sample of 1300 used

1/Guilford J.P., Psychometric Methods, McGraw-Hill Book Company, New York, Kogakusha Company Ltd., Tokyo. 1959. pp. 477

2/Ibid, pp. 522.

for split half method has been used for this purpose. But every third answer sheet has been discarded until 1000 answer sheets were left. The statistics of this sample are as given below:-

Table 46. Data Grouped for finding Mean, SD, Median of the Sample selected for finding Inter-correlations.

Class interval scores	f	x'	fx'	fx'^2	cum.f.
120-129	1	5	5	25	1000
110-119	14	4	56	224	999
100-109	43	3	129	387	985
90-99	91	2	182	364	942
80-89	152	1	152	152	851
70-79	229	0	000	000	699
60-69	284	-1	-284	284	470
50-59	161	-2	-322	644	186
40-49	22	-3	- 66	198	25
30-39	3	-4	- 12	48	3

N = 1000 $\quad \Sigma fx' = -160 \quad \Sigma fx^2 = 2326$

$c = \frac{-160}{1000} = -.160$

$ci = .160 \times 10 = -1.60$

$c^2 = (-.16)^2 = .0256$

$$AM = 74.5$$
$$ci = -1.6$$
$$Mean = 72.9$$

$$Median = 69.5 + \frac{500 - 470}{229} \times 10$$
$$= 70.81$$

$$SD = 10\sqrt{\frac{2326}{1000} - .0256}$$
$$= 15.17$$

Table 47. Statistics of the Sample and the whole group.

Sample	Mean	Median	SD
Sample selected for finding inter correlations	72.9	70.81	15.17
Total sample	71.98	71.69	15.42

From the table given above, it may be seen that the statistics of the selected sample are almost the same as those of the whole sample.

Centroid Method.

Centroid is the centre of gravity. Statistically it is the mean.

Number of Expected Factors.

The first decision to be made before extracting the factors is the number of factors to be expected from the number of given tests. The formula used to make this decision is as given below:

$$r = \frac{2n + 1 - \sqrt{8n + 1}}{2}.$$

where

r = number of factors

n = number of variables (tests)

In the present case n = 8

$$\therefore \quad r = \frac{2 \times 8 + 1 - \sqrt{8 \times 8 + 1}}{2}$$

$$= \frac{17 - \sqrt{65}}{2}$$

$$= \frac{17 - 8.071}{2}$$

$$= \frac{8.929}{2}$$

$$= 4.464 \text{ i.e. } 4$$

This indicates that there is a possibility of four centroid factors.

<u>Criteria for Significant Factors</u>.

There are no exact criteria for stopping extraction of factors. Number of empirical criteria have been developed. Vernon has listed as many as twentyfive criteria. Some of them are Tucker's Phi, Humphrey's Rule, Coomb's Criterion etc.

Humphrey's Rule has been applied in the present case. This procedure takes into account the size of the sample and the two highest factor loadings rather than the entire matrix. The rule is if the product of the two highest loadings is more than twice the standard error of a correlation coefficient of zero the obtained factor is significant.

The actual factor analysis of this data is given in the following tables.

Table 48. Extraction of the First Centroid Factor from the Correlation Matrix.

Test No.	1	2	3	4	5	6	7	8	E
1	(.6486)	.6486	.5577	.5852	.5314	.2823	.2993	.3500	3.9031
2	.6486	(.6486)	.5514	.5509	.4681	.3402	.3311	.3714	3.9103
3	.5577	.5514	(.5577)	.5542	.4719	.3096	.3271	.3926	3.7222
4	.5852	.5509	.5542	(.5852)	.4663	.2441	.2714	.4139	3.6712
5	.5314	.4681	.4719	.4663	(.5314)	.3119	.4871	.3221	3.5902
6	.2823	.3402	.3096	.2441	.3119	(.4630)	.2354	.4630	2.6495
7	.2993	.3311	.3271	.2714	.4871	.2354	(.5184)	.5184	2.9882
8	.3500	.3714	.3926	.4139	.3221	.4630	.5184	(.5184)	3.3498
E	3.9031	3.9103	3.7222	3.6712	3.5902	2.6495	2.9882	3.3498	$T = 27.7845$ $\sqrt{T} = 5.271$ $m = \frac{1}{\sqrt{T}} = 0.1897$
$mE=a_1$	.7404	.7418	.7060	.6965	.6811	.5027	.5669	.6354	5.2714

-278-

Table 49. First Factor Matrix.

First Factor loadings (a_1)	Test No.	.7404	.7418	.7060	.6965	.6811	.5027	.5669	.6354
		1	2	3	4	5	6	7	8
.7404	1	.5483	.5493	.5228	.5157	.5043	.3722	.4198	.4706
.7418	2	.5493	.5503	.5237	.5166	.5053	.3729	.4206	.4714
.7060	3	.5228	.5237	.4984	.4917	.4808	.3549	.4002	.4486
.6965	4	.5157	.5166	.4917	.4851	.4743	.3502	.3949	.4426
.6811	5	.5043	.5053	.4808	.4743	.4638	.3424	.3861	.4328
.5027	6	.3722	.3729	.3549	.3502	.3424	.2527	.2849	.3195
.5669	7	.4198	.4206	.4002	.3949	.3861	.2849	.3214	.3602
.6354	8	.4706	.4714	.4486	.4426	.4328	.3195	.3602	.4038

-279-

Table 50. First Residual Correlation Matrix

Test No.	1	2	3	4	5	6	7	8	
1	(.1003)	.0993	.0349	.0695	.0271	-.0899	-.1205	-.1206	.0001
2	.0993	(.0083)	.0277	.0343	-.0372	-.0327	-.0895	-.1000	.0002
3	.0349	.0277	(.0593)	.0625	-.0089	-.0453	-.0731	-.0560	-.0011
4	.0695	.0343	.0625	(.1001)	-.0080	-.1061	-.1235	-.0287	.0001
5	.0271	-.0372	-.0089	-.0080	(.0676)	-.0305	.1010	-.1107	.0004
6	-.0899	-.0327	-.0453	-.1061	-.03[illegible]5	(.2103)	-.0495	.1435	-.0002
7	-.1205	-.0895	-.0731	-.1235	.1010	-.0495	(.1970)	.1582	.0001
8	-.1206	-.1000	-.0560	-.0287	-.1107	.1435	.1582	(.1146)	.0003
	.0001	.0002	-.0011	.0001	.0004	-.0002	.0001	.0003	-.0001

-280-

Table 51. Extraction of the Second Centroid Factor from the First Residual Correlation Matrix in which variables 1, 2, 3 and 4 are Reflected.

Test No.	1	2	3	4	5	6	7	8	
1	(.1206)	.0993	.0349	.0695	-.0271	.0899	.1205	.1206	0.6282
2	.0993	(.1000)	.0277	.0343	.0372	.0327	.0895	.1000	0.5207
3	.0349	.0277	(.0731)	.0625	.0089	.4053	.0731	.0560	0.3815
4	.0695	.0343	.0625	(.1235).	.0080	.1061	.1235	.0287	0.5561
5	-.0271	.0372	.0089	.0080	(.1107)	-.0305	.1010	-.1107	0.0975
6	.0899	.0327	.0453	.1060	-.0305	(.1435)	-.0495	.1435	0.4810
7	.1205	.0895	.0731	.1235	.1010	-.0495	(.1582)	.1582	0.7745
8	.1206	.1000	.0560	.0287	-.1107	.1435	.1582	(.1582)	0.6545
E	.6282	.5207	.3815	.5561	.0975	.4810	.7745	.6545	T = 4.0940
									$\sqrt{T} = 2.023$
									$m = \frac{1}{\sqrt{T}} = .4943$
$mE = a_2$	.3106	.2574	.1886	.2749	.0482	.2378	.3828	.3235	2.0238

-281-

Table 52. Second Factor Matrix

Second Factor loadings (a_2)		.3106	.2574	.1886	.2749	.0482	.2378	.3828	.3235
	Test No.	1	2	3	4	5	6	7	8
.3106	1	.0964	.0799	.0586	.0854	.0150	.0738	.1189	.1005
.2574	2	.0799	.0663	.0485	.0707	.0124	.0612	.0985	.0833
.1886	3	.0586	.0485	.0356	.0518	.0091	.0448	.0722	.0610
.2749	4	.0854	.0707	.0518	.0755	.0132	.0653	.1052	.0889
.0482b	5	.0150	.0124	.0091	.0132	.0023	.0115	.0185	.0156
.2378	6	.0738	.0612	.0448	.0653	.0115	.0565	.0910	.0769
.3828	7	.1189	.0985	.0722	.1052	.0185	.0910	.1466	.1239
.3235	8	.1005	.0833	.0610	.0889	.0156	.0769	.1239	.1047

-282-

Table 53. Second Residual Correlation Matrix

Test No.	1	2	3	4	5	6	7	8	
1	(.0242)	.0194	-.0237	-.0159	-.0421	.0161	.0016	.0201	-.0003
2	.0194	(.0337)	-.0208	-.0364	.0248	-.0285	-.0090	.0167	-.0001
3	-.0237	-.0208	(.0375)	.0107	-.0002	.0005	.0009	-.0050	-.0001
4	-.0159	-.0364	.0107	(.0480)	-.0052	.0408	.0183	-.0602	.0001
5	-.0421	.0248	-.0002	-.0052	(.1084)	-.0420	.0825	-.1263	-.0001
6	.0161	-.0285	.0005	.0408	-.0420	(.0870)	-.1405	.0666	.0000
7	.0016	-.0090	.0009	.0183	.0825	-.1405	(.0116)	.0343	-.0003
8	.0201	.0167	-.0050	.0602	-.1263	.0666	.0343	(.0535)	-.0003
	-.0003	-.0001	-.0001	.0001	-.0001	.0000	-.0003	-.0003	-.0011

-283-

Table 54. Extraction of Third Centroid Factor from the Second Residual Correlation Matrix in which Variables 1, 3, 4, 6 and 8 are Reflected.

Test No.										
1	(.0421)	-.0194	-.0237	-.0159	.0421	.0161	-.0016	.0201	.0598	
2	-.0194	(.0364)	.0208	.0364	.0248	.0285	-.0090	-.0167	.1018	
3	-.0237	.0208	(.0237)	.0107	.0002	.0005	-.0009	-.0050	.0263	
4	-.0159	.0364	.0107	(.0602)	.0052	.0408	-.0183	-.0602	.0589	
5	.0421	.0248	.0002	.0052	(.1263)	.0420	.0825	.1263	.4494	
6	.0161	.0285	.0005	.0408	.0420	(.1405)	.1405	.0666	.4755	
7	-.0016	-.0090	-.0009	-.0183	.0825	.1405	(.1405)	-.0343	.2994	
8	.0201	-.0167	-.0050	-.0602	.1263	.0666	-.0343	(.1263)	.2231	
E	.0598	.1018	.0263	.0589	.4494	.4755	.2994	.2231		$T = 1.6942$
										$\sqrt{T} = 1.302$
										$m = \frac{1}{\sqrt{T}} = .7683$
$mE = a_3$	.0460	.0782	.0202	.0452	.3453	.3654	.2301	.1714	1.3018	

-284-

Table 55. Third Factor Matrix

Third Factor Loadings a_3		.0460	.0782	.0202	.0452	.3453	.3654	.2301	.1714
	Test No.	1	2	3	4	5	6	7	8
.0460	1	.0021	.0036	.0009	.0021	.0159	.0168	.0106	.0079
.0782	2	.0036	.0061	.0016	.0035	.0270	.0286	.0180	.0134
.0202	3	.0009	.0016	.0004	.0009	.0070	.0074	.0047	.0035
.0452	4	.0021	.0035	.0009	.0021	.0156	.0165	.0104	.0078
.3453	5	.0159	.0270	.0070	.0156	.1192	.1262	.0795	.0592
.3654	6	.0168	.0286	.0074	.0165	.1262	.1336	.0841	.0626
.2301	7	.0106	.0180	.0047	.0104	.0795	.0841	.0529	.0394
.1714	8	.0079	.0134	.0035	.0078	.0592	.0626	.0394	.0294

-285-

Table 56. Third Residual Correlation Matrix

Test No.	1	2	3	4	5	6	7	8	
1	.0400	-.0230	-.0246	-.0180	.0262	-.0007	-.0122	.0122	-.0001
2	-.0230	.0303	.0192	.0329	-.0022	-.0001	-.0270	-.0301	.0000
3	-.0246	.0192	.0233	.0098	-.0068	-.0069	-.0056	-.0085	-.0001
4	-.0180	.0329	.0098	.0581	-.0104	.0243	-.0287	-.0680	.0000
5	.0262	-.0022	-.0068	-.0104	.0071	-.0842	.0030	.0671	-.0002
6	-.0007	-.0001	-.0069	.0243	-.0842	.0069	.0564	.0040	-.0003
7	-.0122	-.0270	-.0056	-.0287	.0030	.0564	.0876	-.0737	-.0002
8	.0122	-.03 01	-.0085	-.0680	.0671	.0040	-.0737	.0969	-.0001
	-.0001	.0000	-.0001	.0000	-.0002	-.0003	-.0002	-.0001	.0010

-286-

Table 57. Extraction of Fourth Centroid from the Third Residual Correlation Matrix in which Variables 2, 3 and 4 have been Reflected.

Test No.	1	2	3	4	5	6	7	8	
1	(.0262)	.0230	.0246	.0180	.0262	-.0007	-.0122	.0122	0.1173
2	.0230	(.0329)	.0192	.0329	.0022	.0001	.0270	.0301	0.1674
3	.0246	.0192	(.0246)	.0098	.0068	.0069	.0056	.0085	0.1060
4	.0180	.0329	.0098	(.0680)	.0104	-.0243	.0287	.0680	0.2115
5	.0262	.0022	.0068	.0104	(.0842)	-.0842	.0030	.0671	0.1157
6	-.0007	.0001	.0069	-.0243	-.0842	(.0842)	.0564	.0040	0.0424
7	-.0122	.0270	.0056	.0287	.0030	.0564	(.0737)	-.0737	0.1085
8	.0122	.0301	.0085	.0680	.0671	.0040	-.0737	(.0737)	0.1899
E	.1173	.1674	.1060	.2115	.1157	.0424	.1085	.1899	$T = 1.0587$ $\sqrt{T} = 1.029$ $m = \frac{1}{\sqrt{T}} = .9716$
$mE=a_4$	.1140	.1627	.1030	.2060	.1127	.0412	.1054	.1845	1.0295

-287-

In table no. 57 it is observed that the two highest factor loadings are 0.2060 and 0.1845 in columns 4 and 8 respectively. Their product is 0.03801 (.2060 x .1845)

The size of the sample selected is 1000.

$$\therefore SE = \frac{1}{\sqrt{1000}} = 0.03162$$

$$\therefore 2 \times SE = 2 \times .03162$$

$$= .06324$$

As the product of the two highest loadings of the fourth factor is less than twide the SE, the obtained fourth factor is not significant. So there are only three factors obtained.

Table 58. Centroid Factor Matrix with Proportions of Variances contributed by the Centroid Factors, obtained and guessed Communalities.

Test	Factor Loading			Variances			$h^2_{obt.}$	h^2_{guess}
	I	II	III	I^2	II^2	III^2		
1	.7404	-.3106	.0460	.5483	.0964	.0021	.6468	.6486
2	.7418	-.2574	-.0782	.5503	.0663	.0061	.6227	.6486
3	.7060	-.1886	.0202	.4984	.0356	.0004	.5344	.5577
4	.6965	-.2745	.0452	.4851	.0755	.0021	.5627	.5852
5	.6811	.0482	.3543	.4638	.0023	.1192	.5853	.5314
6	.5027	.2378	-.3654	.2527	.0565	.1336	.4428	.4630
7	.5669	.3828	.2301	.3214	.1466	.0529	.5209	.5184
8	.6354	.3235	-.1714	.4038	.1047	.0294	.5379	.5184
Total				3.5238	0.5839	0.3458	4.4535	
				79 %	13 %	8 %	100 %	

-289-

The sums of the squared loadings show that the first factor takes out about 79 % of the total common factor variance, the second factor takes out about 13 % of the common factor and the third factor takes out about 8 % of the common factor.

The first factor represents the 'g' factor. Nearly half the variance of the second factor is shared by the test numbers 7 and 8 which involve dominantly the operations with the numbers. Thus this variance is mainly due to the numerical ability.

Nearly two third of the variance of the third factor is shared by tests 5 and 6 which require Language Comprehension. Thus this variance is due to verbal ability.

Table 59. The Factors and The Proportions of their Variances.

Factor	Variance Percent
'g'	79
Verbal	8
Numerical	13
	100

From the table given above it is seen that 79 % of the performance is accounted for by the 'g' factor, 8 % by the verbal factor and 13 % by the numerical factor.

These statistics are very much in agreement with the requirements of the hierarchical model used for preparing this intelligence test.

Sub-Tests in a Battery.

"The term "battery" is conventionally applied to a set of separate tests to be administered to the same group of individuals in order to meet a single measurement objective, or a closely interrelated set of such objectives".[1/]

The present test consists of eight subtests to measure a single ability intelligence. The hierarchical model of 'g' has been used for preparing the test. The factor analysis has shown that 'g' factor contributes nearly 79 % of the variance and the other two factors being verbal and numerical contributing nearly 8 and 13 percent of total variance respectively. Thus the major factor measured is 'g'. The other factors functioning are in traces.

The test items are classified in eight groups according to the particular common way of solving the items included in that group. Each group of test items is called a sut-test.

The subtests should be so selected that they should have high correlation with the scores of the whole test but low correlation with each other. This avoids duplication and each test used contributes maximally to the forecast.

1/Moiser C.I., "Batteries and Frofiles" chapter 18, Educational Measurement, Lindquist E.F. (Editor), American Council of Education, Washington D.C. 1966. pp. 764.

Table 60. Correlations of the tests with each other and with the whole Test.

Test	1	2	3	4	5	6	7	8	Whole Test
1		.6486	.5577	.5852	.5314	.2823	.2993	.3500	.7777
2	.6486		.5514	.5509	.4681	.3402	.3311	.3714	.7478
3	.5517	.5514		.5542	.4719	.3096	.3271	.3926	.7446
4	.5852	.5509	.5542		.4663	.2441	.2714	.4319	.7487
5	.5314	.4681	.4719	.4663		.3119	.4871	.3221	.6960
6	.2823	.3402	.3096	.2441	.3119		.2354	.4630	.5231
7	.2992	.3311	.3271	.2714	.4871	.2354		.5184	.5283
8	.3500	.3714	.3926	.4139	.3221	.4630	.5184		.5848

-292-

It may be observed from the above table that the correlation of any sub-test with the whole test is more than its correlation with any other subject.

IQ's of the Whole Sample.

The distribution of population according to IQ is given below.[1/] (Table no. 15 has been reproduced).

Table 61. Distribution of Standardization Sample in Composite Stanford-Binet IQ on forms L and M.

IQ	Percentage of cases.	Classification.
160-169	0.03	Very superior
150-159	0.2	
140-149	1.1	
130-139	3.1	Superior
120-129	8.2	
110-119	18.1	High average
100-109	23.5	Normal or average
90-99	23.0	
80-89	14.5	Low average.

(concluded on next page)

1/Annastasi Anne, Op.cit. pp. 208

Table 61. (concluded)

1	2	3
70-79	5.6	Borderline defective
60-69	2.0	Mentally defective
50-59	0.4	
40-49	0.2	
30-39	0.03	

In the light of this distribution the study of the sample used for the standardization of the present test has been done.

By using the tables of morms established, the IQs of all the 7745 pupils were calculated and the frequency table was prepared for calculating the mean, median and SD.

Table 62. Data grouped for finding Mean, Median and SD of IQs of the Sample selected for Standardization of the Test.

Class interval scores 1	f 2	x' 3	fx' 4	fx'^2 5	cum.f 6
130-139	329	+3	+ 987	2961	7745
120-129	754	+2	+1508	3016	7416

(concluded on next page)

Table 62. (concluded)

1	2	3	4	5	6
110-119	1171	+1	+1171	1171	6612
100-109	1470	0	0000	0000	5491
90-99	1760	-1	-1760	1760	4121
80-89	1434	-2	-2868	5736	2261
70-79	827	-3	-2481	7441	827
	N =7745		$\sum fx' = -3443$	$\sum fx'^2 = 22,085$	

$$c = \frac{-3443}{7745} = -0.4446 \qquad ci = -0.4446 \times 10 = -4.446 \qquad c^2 = (-0.4446)^2 = 0.2489$$

$$\therefore AM = 104.500$$
$$ci = -4.446$$
$$\text{Mean} = 100.054$$

$$\text{Median} = 89.5 + \frac{3872.5 - 2261}{1760} \times 10$$
$$= 89.5 + 9.160$$
$$= 98.660$$

$$SD = i\sqrt{\frac{fx'^2}{N} - c^2}$$
$$= 10\sqrt{\frac{22085}{7745} - 0.2489}$$
$$= 16.11$$

The mean (100.05) and SD (16.11) of the sample used for standardization agree very closely with the parameters (Mean = 100 and SD = 16.4) of the population for whom the test is prepared.

The test aims at measuring IQs between 70 and 130 children. The classification of pupils in the sample used for standardization and the parameter are given below.

Table 63. Distribution of Standardization Sample.

IQ	Percentage of Cases	
	In composite Stanford-Binet IQ on L & M forms	In the present Test
130 onwards	4.43	4.25
120-129	8.2	9.73
110-119	18.1	15.12
100-109	23.5	18.98
90-99	23.0	22.73
80-89	14.5	18.51
70-79	8.23	10.68
Total	99.96	100.00

It can be seen that the distribution of individuals according to IQ in the sample used for standardization of the present test fairly agrees with those used for standardization of L and M forms of Stanford-Binet 1937 Scale. Thus the present test may be treated as a fairly reliable and valid test for measuring the IQs of pupils attending standards VIII to X in the secondary schools in Marathwada region.

References.

1. Anastasi Anne, Psychological Testing. The Macmillan Company New York 1965.
2. Cronbach L.J., Essentials of Psychological Testing. Harper & Row New York and John Weatherhill, Inc. Tokyo. 1965.
3. Fruchter B., Introduction to Factor Analysis. D.Van Nostrand Company, Inc. Princeton, Affiliated East-West Press Ltd., New Delhi. 1967.
4. Garrett H.E., Statistics in Psychology and Education. Longmans, Green and Co. New York, London, Toronto. 1954.
5. Garrett H.E., Statistics in Psychology and Education. Vakils, Feffer and Simons Private Ltd., Bombay 1. 1971.
6. Goodnough F.L., Mental Testing. Holt Rinehart and Winston, New York. 1961.
7. Guilford J.P., Psychometric Methods. Mc Graw-Hill Book Company, New York. Kogakusha Company Ltd., Tokyo. 1959.
8. Harris C.W., Encyclopedia of Educational Research. The Macmillan Company, New York. 1960.
9. Remmers H.H., Gage N.L., Rummel J.F., A Practical Introduction to Measurement and Evaluation. Harper & Row, New York and JohnWeatherhill Inc. Tokyo. 1966.
10. Shah M.M., An Aptitude Test for Secondary School Teachers. The Maharaja Sayajirao University of Baroda, Baroda. 1965.
11. Vernon P.E., The Measurement of Abilities. University of London Press Ltd., London E.C. 4. 1961.