CHAPTER IX

RELIABILITY

A test is said to be reliable when it always gives the same result if administered again and again. If the scores on the test are stable when administered to the same person at different times, it is said to be a reliable test. In short, by 'reliability' is meant the degree to which the test agrees with itself[1]. Reliability has to do with accuracy and precision of measurement procedure. In a word, reliability means consistency[2].

Methods of Determining Reliability

From the discussions of the various factors involving reliability and the methods of measuring it, three major methods are used. These are as follows:

(a) The Split Half Method

(b) The Test-Retest Method

(c) The Parallel Form Method

---

1 Thorndike/and Hagen/: Measurement and Evaluation in Psychology and Education. New York, John Wiley and Sons., 1955. p.129.

2 Ibid., p.129.

(a) The Split Half Method

In this method the test is usually divided into two equivalent parts and the correlation between the two parts shows the index of reliability. However, it must be noted that the computed correlation is between two half length tests. This value is not directly applicable to the full length tests which is the actual instrument prepared for use. So from the half test reliability the self correlation of the whole test is computed by the Spearman-Brown Formula. The division into two halves is usually effected in the following two ways:

(1) The first half versus the second half.

(2) The odd items versus the even items.

According to Thorndike[3], "This is usually a sensible procedure since items of similar form, content or difficulty are likely to be grouped together in a test. For a reasonably long test, say, of more than 60 items splitting the test up in this way will tend to balance out factors of item form, content covered and difficulty level.

According to Garret[4], the split-half method is generally regarded as the best of the methods for determining test reliability.

---

3 Thorndike and Hagen : Op.cit., p. 128

4 Garret, H.E. : Statistics in Psychology and Education. New York, Longmans Green and Company, 1953 p.334.

The following advantages are quite obvious :

(1) All the data are obtained on one single occasion.

(2) Same motivational factors operate for both the halves.

(3) It saves duplication.

The question of practice effect does not arise. As the administration is done under similar conditions in both the halves, the coefficient of correlation is more reliable.

(b) The Test-Retest Method

Repetition of a test is the simplest method of determining reliability. The test is administered and then readministered to the same group after a lapse of some interval. The correlation between the first and the second sets of scores gives an estimate of the reliability coefficient. It is used for two different purposes : "variation within the individual as well as variation due to the operation of measurement".[5]

If the period intervening the two administration is short, the scores on the test at the time of the second administration tend to be high because of familiarity. If the intervening period is long the scores on the test at the second administration will again be high on account of maturity.

---

[5] Thorndike and Hagen : Op.cit., p. 125.

It follows that only those tests which are not appreciably effected by repetition tend themselves to the retest technique. Because of this difficulty, other methods are preferred to this method.

In spite of the above disadvantages this method is used to test the reliability of intelligence tests by nearly all workers in the field of intelligence testing. Consequently the author decided to use it for this experiment.

(c) The Parallel Form Method

When parallel forms of a test are constructed the correlation between the two forms is taken as a measure of the self correlation of the test. "When the test is not available in two parallel forms it is but natural to get an estimate of the reliability by administering the same test twice"[6]. As exact alternative form is not feasible. So this method is discarded.

The Reliability of the Present Test

After considering the pros and cons of the methods discussed above, it was decided to check the reliability of the present test by the following two methods :

(i) Split-Half Method

(ii) Test-Retest Method

Reliability by the Split-Half Method

For calculating the reliability coefficient by the split-half method a sample of 897 booklets was drawn from the total number of booklets which was 5,372. Every sixth booklet was taken out.

---

6 Guilford, J.P. : Psychometric Methods. New York, McGraw-Hill Book Company, 1954.p.412.

Procedure. - The scores of these 897 pupils on the two equivalent parts of the test were calculated and the Pearson's produce moment coefficient of correlation between them was found out.

TABLE 16

The Reliability of the Test by Split-Half Method

A. The Composition of the Sample

Age	No. of Boys	No. of Girls	Total
12	120	30	150
13	160	40	200
14	220	55	275
15	117	30	147
16	100	25	125
Total	717	180	897

B. Scatter Diagram Between

Odd Items and Even Items

Scores on the Even Items	Scores on the Odd Items											
	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	111-120	121-130	Total
121-130									2	1	1	4
111-120							3	2	3	2	1	11
101-110						9	8	3	1	1		22
91-100					6	7	24	21	11	2		71
81-90					4	26	33	16	1	1		81
71-80			2	6	40	64	15	7	3			137
61-70			6	35	57	33	2	2				135
51-60			29	60	41	11	4					145
41-50		3	31	57	32	4	3					130
31-40		28	45	37	7	1						118
21-30		20	23									43
Total		51	99	131	140	153	153	89	51	21	7	2	897

r = 0.89 S.E. = .02

$$r = \frac{\Sigma x' y' - \left(\frac{\Sigma fx'}{N} \times \frac{\Sigma fy'}{N} \right)}{\sqrt{\left(\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N} \right)^2 \right) \left(\frac{\Sigma fy'^2}{N} - \left(\frac{\Sigma fy'}{N} \right)^2 \right)}}$$

$$= \frac{4.29 \text{ -- } 1.32 \times 0.30}{\sqrt{\left\{ 16.01 - (1.32)^2 \right\} \left\{ 4.48 - (0.30)^2 \right\}}}$$

$$= \frac{3.894}{\sqrt{4.2676 \times 4.39}}$$

$$= \frac{3.8940}{\sqrt{18.73}}$$

$$= \frac{3.894}{4.33}$$

$$= .89$$

The product moment coefficient of correlation as calculated from the above table is .89.

This coefficient of correlation does not give the reliability of the whole test. It merely gives the reliability of the comparable halfs of the test. Spearman-Brown formula[7] is applied to get the reliability of the whole test.

$$r_1 \, I = \frac{2 \, r_{\frac{1}{2}} \frac{I}{II}}{1 + r_{\frac{1}{2}} \frac{I}{II}}$$

in which

$r_1 I$ = the reliability coefficient of the whole test;

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

Substituting the obtained value.

$$r_1 I = \frac{2 \times .89}{1 + .89}$$

$$= \frac{1.78}{1.89}$$

$$= .941$$

---

7 Garret, H.E. : Statistics in Psychology and Education. New York, Longmans Green and Co. p. 343.

The reliability coefficient of the test as calculated by the Split-Half Method is .941.

Reliability by Test-Retest Method

In the present work three hundred eighty four students were administered the test on the 20th May 1962 and they were again tested on 15th December i.e. after seven months. The second test given after an interval of seven months can be considered to give same results as the first one because test items were almost forgotten by that time. Cattel[8] opines, " In many instances one wishes to retest a child's intelligence after a lapse of some months, it is quite safe to use the same test".

The mean score on the second administration was found to be high as compared with the mean score on the first administration. The two sets of scores were then arranged in the form of scatter diagram and the coefficient of correlation was computed, which came out to be .90.

The following table gives the analysis of the sample and the calculations of the product moment 'r' between the two sets of scores.

---

8 Cattell, R.B. : A Guide to Mental Testing. London, University of London Press, 1953. p.13.

TABLE 17

A. Composition of the Sample

Age	No. of boys	No. of Girls	Total
12	48	12	60
13	55	15	70
14	107	22	129
15	65	20	85
16	32	8	40
Total	307	77	384

B. Scatter Diagram Between Test and Retest Scores

I / II	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	111-120	121-130	Total
121-130										1	1	2
111-120								1	2	2		5
101-110						1	3	2	3			9
91-100					2	7	6	11	4			30
81-90				1	3	17	9	5				35
71-80			1	6	6	25	20	2				60
61-70			7	13	27	9		1				57
51-60			15	25	18	4						62
41-50	4	8	19	14	9	2						56
31-40	16	18	14	1	1							50
21-30	2	16										18
f_2	22	42	56	60	66	65	38	22	9	3	1	384

Retest Scores

r = 0.90 (Approximate)

$$r = \frac{\Sigma x'y' - \left(\frac{\Sigma fx'}{N} \times \frac{\Sigma fy'}{N}\right)}{\sqrt{\left(\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2\right)\left(\frac{\Sigma fy'^2}{N} - \left(\frac{\Sigma fy'}{N}\right)^2\right)}}$$

$$= \frac{5.55 - (1.27 \times 1.32)}{\sqrt{4.71 \times 4.28}}$$

$$= \frac{3.90}{4.47}$$

$$r = 0.872$$

It can be seen that the different methods do not give exactly the same results although the reliability coefficient of the test by the Split-Half method and the Test-Retest method are very close to one another. The question arises when is the reliability coefficient satisfactory or how high the reliability coefficient we should require depends upon the nature of the group and the purpose for which the test was given. In order to differentiate between the means of two school grades of relatively narrow range, a reliability coefficient need be higher than 50 or 60. If the test is to be used to make individual diagnoses (i.e. to separate pupil from pupil) its reliability coefficient for a single grade should be .90 or higher.

One can satisfy oneself by comparing the reliability of the present test with some other known intelligence test.

TABLE 18

Reliability Coefficient of some well Known Tests*

Sr. No.	Test	Reliability Coefficients
1	Stanford-Binet (1916)	.90 to .95
2	Otis Self Administering Test	.95
3	Otis Group Intelligence Scale : Primary	.967
4	Terman Group Test of Mental Ability	.89
5	Thorndike Intelligence Examination	.85
6	Thurstone Psychological Examination	.959
7	McCall Multi Mental Test	.94

It is not known which methods of testing the reliability is used in many of the above mentioned tests. Any how, the reliability coefficient of the present tests compares favourably with those of all these well known tests.

* 1 to 7 taken from Kelley : Interpretation of Educational Measurement.

The table below gives the reliability coefficient of the present test by different methods.

TABLE 19

Reliability of the Test

Sr. No:	Methods	Reliability Coefficient
1	Split-Half Method	.94
2	Test-Retest Method	.90

Thus the reliability of the test varies from .90 to .94 which is fairly high reliability.

The reliability coefficient as found by test retest method actually shows that Chronbach calls, " The coefficient of stability". It indicates whether a sample of behaviour taken at one time is typical of the behaviour at other times or not. With the above coefficient of reliability we can say that the present test is satisfactory. Some standard is necessary against which reliability coefficient of the present test should be judged as satisfactory or not.

Kelley[9] suggests that a test with a reliability as low as 0.5 is useful for determining the status of a group in some subjects whereas reliability of more than .9 is useful for differentiating the status of an individual in a group.

9
Kelley : Interpretation of Educational Measurement.

The reliability coefficient of the present test is nowhere less than .90. Therefore it can be concluded that the test is sufficiently reliable.

Selected References

1. Cronbach, L.J. : Essentials of Psychological Testing, Ch. 3. New York, Harper and Brothers, 1949.

2. Deasi, K.G. : The Construction and Standardization of a Battery of Group Tests of Intelligence in Gujarati, Ahmedabad, Bharat Prakashan, 1954.

3. Garret, H.E. : Statistics in Psychology and Education, New York, Longmand Green and Company, 1951.

4. Guilford, J.P. : Psychometric Methods, New York, McGraw-Hill Book Company, 1954.

5. Lindquist, E.F. Educational Measurement, Ch.15, Washington, American Council of Education.

6. Ross, C.C. : Measurement in Today's Schools, Ch. 4, New York, Prentice Hall, Inc., 1956.

7. Thorndike Hagen : Measurement and Evaluation in Psychology and Education, New York, John Wiley and Sons, 1955.

8. Thorndike, R.L. : Personal Selection, Ch.4, New York, John Wiley & Sons Inc.

9. Thurstone, L.L. : The Principles of Vocational Guidance, Br. Jr. of Psychology, 1928, 14.Ch.