CHAPTER X

RELIABILITY

10.1 INTRODUCTION

Reliability in psychological measurement has somewhat different connotation than that in everyday common sense use. A reliable measuring instrument is one which gives the same results trial after trial. For example, if one rates the intelligence of one of his colleagues as slightly below average at some particular point of time, and then subsequently also gives the same rating whenever required, his rating is said to be consistent or reliable. To the extent that there is a variation in the judgment, there is the element of unreliability. In common sense, however, to be called reliable, his judgment should not only be consistent but also conform to what is real. That is to say, if the person rated is in fact slightly below average in intelligence, then only

the judgment would be considered reliable. Substitute word for this is dependable. In psychological measurement, ⸱ this dependability of measuring the attribute in question accurately is called validity of a test. In validity, emphasis is on the test's agreement with some criterion measure; whereas in reliability emphasis is on the test's agreement with itself. Thus reliability in the present discussion means the consistency of measuring, irrespective of whatever is measured.

When any psychological attribute is required to be measured, there is no direct measurement possible. This is so, because, first of all, the attribute in question is less tangible than physical attributes which can be directly sensed and measured. It cannot be directly observed. It can only be inferred on the basis of some behaviour samples. Therefore, when one says that a person is slightly below average in intelligence, he does not actually see intelligence as such, which is a hypothetical concept but he only makes inference on the basis of

his observation of a sample of the person's behaviour. Secondly, the scope of behaviour defining any particular attribute is generally not clear. For example, what would an extravert do or how would he behave can only be stated in broad general terms. But the total behaviour that belongs under a particular concept or attribute is not clearly definable. Psychological measurement is, therefore, a kind of inferential judgment of attributes of human behaviour based on a sample and has similar limitations as the sample statistics. The reliability of test scores, therefore, depends upon the representativeness and adequacy of the behaviour sampled by the test. Reliability of some of the personality inventories is discussed in general manner in the following section to present a comparative picture.

10.2 RELIABILITY OF THE PERSONALITY INVENTORIES IN GENERAL

Brief reference to this point was made in the third chapter while discussing the evaluation of the personality inventories. It was stated that, in

general, the reliability of the personality inventories was less than that of the ability and achievement tests. But it was also observed that they had greater reliability than most of the other techniques of personality assessment. The reliability coefficients of various personality scales range from .75 to .85 in majority of the cases. Occasionally, they are greater than these. Here are the reliability values of some of the commonly used personality inventories:

<u>Colgate Mental Hygiene Test</u>

   B_1 Schedule (Psychoneurotic Tendencies)
         by split-half method - .79
         by test-retest method - .85

   C_1 Schedule (Introversion-Extraversion
         Scale)
         by split-half method - .45
         by test-retest method - .67

<u>The Allport's Ascendance-
Submission Reaction Study</u>

   Split-half reliability coefficient when corrected for length, is reported to be .74.

Bernreuter Personality Inventory

Average reliability coefficients by the split-half method as reported by Bernreuter are:[1]

	Scale	r
B 1	Neurotic Tendency	.87
B 2	Self-Sufficiency	.83
B 3	Introversion-Extraversion	.85
B 4	Dominance-Submission	.88

The Minnesota Personality Scales

Test-retest correlations on 13 scales are as follows:[2]

Scale	Men	Women
Morale	.65	.63
Inferiority	.61	.53
Attitude toward family	.64	.76
Attitude toward the legal system	.55	.57

1 R.G.Bernreuter, "The Theory and Construction of the Personality Inventory". J.soc.Psychol. IV: 387-405, 1933.

2 J.G.Barley, "Changes in Measured Attitudes and Adjustments".J.Soc.Psychol.,IX:189-199,1938.

Scale	Men	Women
Economic conservatism	.79	.59
Education	.46	.63
General adjustment	.61	.64
Home adjustment	.71	.82
Health adjustment	.72	.81
Social adjustment	.84	.78
Emotional adjustment	.68	.70
Social preferences	.73	.62
Social behaviour	.64	.69

Guilford and Guilford-Martin Inventories[3]

Social introversion-extraversion	.90
Thinking introversion-extraversion	.84
Depression	.94
Cycloid disposition	.88
Rhathymia	.90
General activity	.89
Ascendance-submission	.88

3 Manuals of the above inventories,
(Sheridan Supply Co.).

Masculinity-femininity	.85
Inferiority feelings	.91
Nervousness	.89
Objectivity	.83
Agreeableness	.80
Co-operativeness	.91

Bell's Adjustment Inventory[4]

Home adjustment	.89
Health adjustment	.80
Social adjustment	.89
Emotional adjustment	.85
Total adjustment	.93

Minnesota Multiphasic Personality Inventory

"Typical coefficients (split-half estimates of reliability) for total scores are in the range of .7 to .9. Typical retest reliabilities seem to be

---

4 H.M.Bell, The Theory and Practice of Psychological Counselling. Stanford University Press, 1939, As cited by Ferguson, Personality Measurement, (McGraw Hill Book Co., 1952), p.231.

in the range of .6 to .8."[5]

Edwards Personal Preference Schedule[6]

Scale	Split-half	Test-Retest
Achievement	.74	.74
Deference	.60	.78
Order	.74	.87
Exhibition	.61	.74
Autonomy	.76	.83
Affiliation	.70	.77
Intraception	.79	.86
Succorance	.76	.78
Dominance	.81	.87
Abasement	.84	.88
Nurturance	.78	.79
Change	.79	.83
Endurance	.81	.86
Heterosexuality	.87	.85
Aggression	.84	.78
Consistency score		.78

5 J.P.Guilford, Personality (New York: McGraw Hill Book Co., 1959), p.180.

6 EPPS Manual (Psychological Co-operation, 1959).

Gordon Personal Profile[7]

Reliability coefficients by different methods for different scales range from .74 to .95 with majority around .85.

Maudsley Personality Inventory[8]

Neuroticism scale – between .85 and .90

Extraversion scale – between .75 and .85

The extensive illustrations above show that the range of reliability values for personality inventories is large, i.e. from .45 to .95. Majority of the scales have reliability values of about .75 to .85 depending upon the method of determining the reliability.

10.3 RELIABILITY OF THE PRESENT TEST

The present test consisted of two scales and, therefore, it yielded two indices of reliability,

---

7 Manual (World Book Co., 1953).

8 Manual (University of London Press,1959).

one for each scale. There are different types of
reliability values of a test. They are:[9]

1. Alternate forms reliability,

2. Split-half reliability,

3. Test-retest reliability,

4. Kuder-Richardson reliability, and

5. Analysis of variance reliability.

The present test had only one form and
therefore, the first type is not reported here. The
Kuder-Richardson formula 20 gives identical results[10]
as the analysis of variance method, therefore,
analysis of variance method was also not applied.
Reliability values by all the rest of the methods
were calculated.

Reliability by Split-half Method

A sample of one hundred students from the

___

9 J.P.Guilford, Psychometric Methods (New
York: McGraw Hill Book Co., 1954), Ch. XIV.

10 Ibid. p.385.

Arts College was selected for determining the relia-
bility values for the two scales of the Inventory.
Scores on the odd items of the scale were correlated
with the even items of the same scale. The data used
is tabulated in the scatter diagrams below:

TABLE X-1

Scatter Diagram of Scores on Two Halves of
the Introversion-Extraversion Scale

Odd series of items

Even series of items	2	3	4	5	6	7	8	9	fy
8								1	1
7						3	2		5
6				1	8	2	3		14
5		2	2	13	14	5			36
4		3	9	11	6				29
3	2	1	8	1					12
2	1	2							3
fx =	3	8	19	26	28	10	5	1	100

TABLE X-2

Scatter Diagram of Scores on Two Halves
of the Normal-Neuroticism Scale

Odd series of items

		3	4	5	6	7	8	fy
Even series of items	7			1	1	2	1	5
	6		2	2	5	5	1	14
	5		2	14	12	3	1	32
	4		5	13	15			33
	3	1	5	4	1			11
	2	2	2	1				5
fx =		3	16	35	33	10	3	100

The reliability coefficients calculated from the above data were:

Introversion-extraversion Scale   .70

Normal-Neuroticism Scale   .59

But these values were for only half the
-ed
lengths of the scales and need/correction for length.
This was done by applying the Spearman - Brown

formula.[11] After correction, the reliability
coefficients were:

  Introversion-Extraversion Scale   .83

  Normal-Neuroticism Scale   .74

Reliability by Test-Retest Method

Another group of about 120 students from the
Arts College was used for calculating the coefficient
of stability or reliability by the test-retest
method. The Inventory was administered to the sample
twice at the interval of about four weeks. Correct
data for 106 students was available at both the
administrations. Six cases were dropped randomly
to facilitate calculations. The data is tabulated
in the following scatter-diagrams for the two scales.

---

11 J.P.Guilford, 1954, Op.Cit. p. 353.

TABLE X-3

Scatter Diagram of Scores on Introversion-Extraversion Scale of the Two Administrations of the Inventory

1st Administration

2nd Administration	4-5	6-7	8-9	10-11	12-13	14-15	16-17	fy
16-17							1	1
14-15					1	5	1	8
12-13			1	1	12	1		15
10-11		1	4	27	2			34
8-9	1	2	19	2	1			25
6-7	2	9	1					12
4-5	4	1						5
fx =	7	13	25	31	16	6	2	100

TABLE X-4

Scatter Diagram of Scores on Normal-Neuroticism Scale on
the Two Administrations of the Inventory

		1st Administration						
2nd Administration	4-5	6-7	8-9	10-11	12-13	14-15	16-17	fy
16-17								0
14-15				1	1	2		4
12-13				1	11	2	1	15
10-11		1	1	44	3	2		51
8-9		1	15	2	1			19
6-7	1	5	2	1				9
4-5	1	1						2
fx =	2	8	18	49	16	6	1	100

Reliability coefficient calculated from the above data were:

Introversion-Extraversion Scale - .91

Normal-Neuroticism Scale - .81

Reliability by K-R Method

The K-R Formula[12] which is commonly used is,

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2} \right)$$

where   n = number of items in the scale,

p = proportion of subjects responding in the keyed manner,

q = 1 - p.

The p and q values for each item on the two scales were calculated on the basis of the item analysis data obtained on the 370 subjects. The pq values for items on the two scales are tabulated below separately.
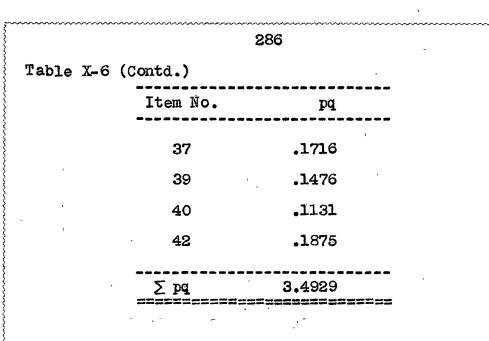
---

12 J.P.Guilford, 1954, Op.Cit. p.380.

TABLE X-5

pq Values for Items on Introversion-Extraversion Scale

Item No.	pq
1	.2139
3	.1924
5	.2211
7	.1824
9	.0900
11	.1275
13	.1476
15	.1771
17	.1131
19	.2059
21	.1204
23	.1530
25	.1131
27	.2176
29	.0900
31	.1204
33	.2304
35	.1924
38	.1411
41	.1875
Σpq =	3.1469

TABLE X-6

pq Values for Items on Normal-Neuroticism Scale

Item No.	pq
2	.1659
4	.1924
6	.1716
8	.1204
10	.1771
12	.2139
14	.1659
16	.1204
18	.1076
20	.1204
22	.1131
24	.2016
26	.1875
28	.1076
30	.2059
32	.1771
34	.1476
36	.1771

Table X-6 (Contd.)

Item No.	pq
37	.1716
39	.1476
40	.1131
42	.1875
Σ pq	3.4929

Reliability coefficients calculated from the above data were:

Introversion-Extraversion Scale - .60

Normal-Neuroticism Scale   - .55

This method is said to underestimate the reliability of the tests.[13] Even the split-half method underestimates the reliability and therefore, correction is applied through Spearman-Brown formula. In K-R method, test is split into n parts of one item each,[14] instead of two equal halves. When

---

13 J.P.Guilford, 1954, Op.Cit. p.385.

14 Ibid. p. 380.

length of the test is reduced reliability decrease.
That is why split-half method underestimates the
reliability. And by the same logic, when length of
the test part is reduced to one item each, it under-
estimates the reliability of the test still more.
Guilford says that reliability cannot be lower than
that obtained by K-R method or the analysis of
variance method.[15]

TABLE X-7

Summary of the Reliability Values

Reliability by	IE Scale	NN Scale
1. Split-half method*	.83	.74
2. Test-retest method	.91	.81
3. K-R method	.60	.55

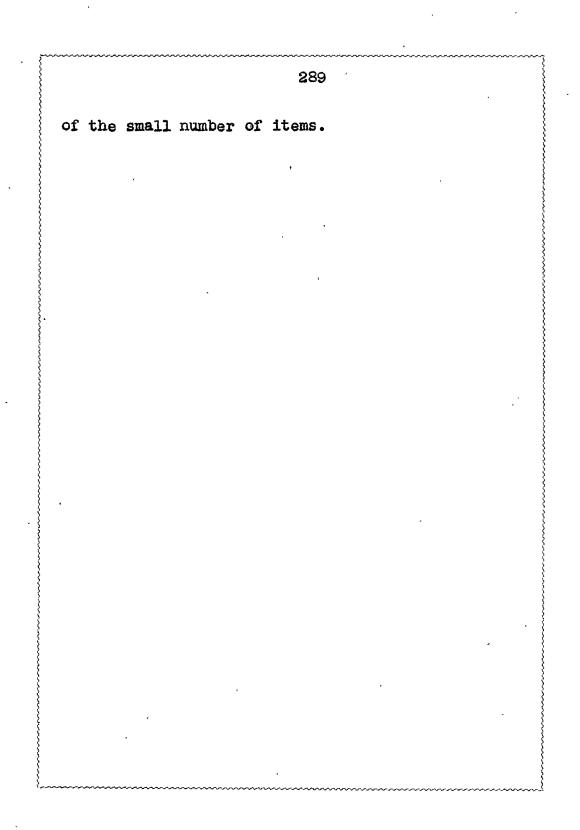* As corrected by Spearman-Brown Formula

The above table shows that the reliability
values are good for both the scales when compared
with most of the standardized inventories. They are

15 Ibid. p.385.

higher for the IE scale than for the NN scale. ˜

10.4 SUMMARY

Reliability of a test is the consistency with which it measures a thing from time to time. Measurement in psychology is not as reliable as that in the physical sciences. The ability and achievement tests are the most reliable ones, while the reliability of the personality tests varies from one method to another. In general, their reliability coefficients are lower than those of ability or achievement tests. Of the different methods of personality measurement, the personality inventories are the most reliable ones, with reliability values around .75 and .85. The present test was also subjected to reliability studies by three methods, viz., split-half method, test-retest method and Kuder-Richardson method. The values obtained by the first two methods were sufficiently high, while those by the third method were low. This was natural because the third method underestimates the reliability of the tests. To conclude, the test has a good reliability, in spite

of the small number of items.

REFERENCES

1. Barley, J.G., "Changes in Measured Attitudes and Adjustment." <u>J.soc. Psychol.</u> IX, 189-199, 1938.

2. Bell, H.M., <u>The Theory and Practice of Psychological Counselling.</u> Stanford University Press, As cited by Ferguson, Personality Measurement, McGraw Hill Book Co., 1939.

3. Bernreuter, R.G., "The Theory and Construction of the Personality Inventory". <u>J.soc.Psychol.</u> IV, 387-405, 1933.

4. Ferguson, L.W., <u>Personality Measurement.</u> New York: McGraw Hill Book Co., 1952.

5. Garrett, H.E., <u>Statistics in Psychology and Education.</u> New York: Longmans, Green & Co., Ch. XIII, 1958.

6. Guilford, J.P., <u>Personality</u>. New York: McGraw Hill Book Co., 1959.

7. Guilford, J.P., <u>Psychometric Methods.</u> New York: McGraw Hill Book Co., Ch. XIV, 1954.

8. Lindquist, E.F. (Ed.), <u>Educational Measurement.</u> Washington D.C.: American Council on Education, Ch. XV, 1950.

9. Micheel, W.J. and M.R.Karnes, <u>Measuring Educational Achievement.</u> New York: McGraw Hill Book Co., Ch. XVI, 1950.

10. Thorndike, R.L., <u>Personnel Selection.</u> New York: John Wiley and Sons, Ch. IV, 1949.

11. Vernon, P.E., <u>Personality Tests and Assessments.</u> New York: Henry Holt and Co., 1953.