CHAPTER - 7

RELIABILITY AND VALIDITY OF THE INVENTORY
CHAPTER 7

RELIABILITY AND VALIDITY OF THE INVENTORY

Meaning of Reliability
A test or an inventory is said to be reliable if it gives the same result on different occasions of the abilities or adjustments of those to whom it is applied. This concept of reliability is true also when the inventory gives the same scores with similar sets of equivalent items or under other variable but similar examining conditions. It underlies the computation of the error of measurement of a single score whereby the range of fluctuation likely to occur in a single individual's score as a result of irrelevant chance of factors can be predicted [1]. The reliability of an inventory also means the self-consistency with which it works at all times. A perfectly reliable inventory gives the same scores of persons when they are again applied the same or an alternative inventory.

Methods of Determining Reliability
Several methods are employed to determine the reliability of a test or an inventory, but the reliabilities thus obtained are not of the same type.

Test-Retest Reliability
A very common method of determining reliability of an inventory derives its form from the definition of
reliability. The inventory is administered again to the same persons to whom it was given earlier and the two sets of scores thus obtained are correlated to find out the test-retest reliability. There are many difficulties arising in this procedure, however,

(a) If the inventory is repeated immediately or after a short interval, many pupils tend to recall their first answers and so finish all items very quickly. They tend to increase their scores on certain adjustments.

(b) The practice and confidence induced by familiarity with the material tend to increase their scores on certain adjustments.

(c) The above affects are not found uniform in all examinees. Clever pupils increase their scores considerably while others may not get that much advantage.

(d) If the inventory is administered after a long time, then the above defects can be eliminated, but the pupils might change their adjustments during the long interval of time and so the sets of scores thus obtained on the second occasion may vary considerably.

(e) This may be because the emotional condition of different individuals may be different on the two occasions.
Split-half Reliability

It is thus more advantageous, if the inventory is administered only once. The items of the inventory are divided into two equivalent halves and the correlation between the two halves is amended by Spearman-Brown prophesy formula, so that the inventory retains its length.

This method is regarded as the best method of calculating the reliability of an inventory. The greatest advantage in this method is that all emotional conditions under which the test is administered remain exactly the same and so chance errors are minimised. But the greatest drawback in this method is that the inventory can be divided into two parts by a variety of ways thus yielding different figures of reliability. In the tests, this factor is not very potent, because the items of the test are graded into a sequence and odd-even splitting is possible only.

Alternate or Parallel Form Method

Instead of giving the same inventory again, an alternate or parallel form of the test is used in the second trial. If the alternate form is used after a long interval, the practice or confidence effects are minimised but then the parallel form becomes a new inventory, which would require all labour of standardisation second time. Secondly, some inventories cannot be split into two or more parts as they are whole units.
Reliability of the Present Inventory-Test-Retest Reliability

The present adjustment inventory was used again with the same subjects after a short interval of time and the test-retest reliability was computed. The items of the inventory were also split into two halves to obtain the split-half reliability. The results of these two methods are given below here. The coefficient correlation was calculated by a calculator which directly yields Y between two sets of scores.

Table-7.1
Test-Re:est Reliability of the Inventory

(A) Composition of the Sample

<table>
<thead>
<tr>
<th>Grade</th>
<th>No. of Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>XI</td>
<td>32</td>
</tr>
<tr>
<td>XII</td>
<td>36</td>
</tr>
<tr>
<td>Total</td>
<td>68</td>
</tr>
</tbody>
</table>

(B) Results

Mean of first administration : 29.67
Mean of second administration : 30.88
Y between the test and retest : 0.51

Split-half Reliability

The other method applied to find out the reliability of the inventory was the split-half-method. The 52 items of the final version of the inventory was split into two parts made of odd items and even items numbering 26 each. The score
obtained by 109 boys and girls of the sample were computed and then correlated with each other. The sample is described in the following table:

Table-7.2

<table>
<thead>
<tr>
<th>Grade</th>
<th>Boys</th>
<th>Girls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>XII</td>
<td>42</td>
<td>12</td>
<td>54</td>
</tr>
<tr>
<td>XI</td>
<td>47</td>
<td>8</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>89</td>
<td>20</td>
<td>109</td>
</tr>
</tbody>
</table>

The correlation between the two halves of the inventory with 26 items each came to be 0.59.

Mean of odd items = 14.73
Mean of even items = 16.06

Spearman-Brown [2] prophesy formula was applied to this result, since it was a correlation between the two halves of the inventory.

\[ r = \frac{2}{11 + r} = .75 \]

This reliability can be considered satisfactory as this is an inventory and not a test.

**Split-half Reliability by Rulon Formula**

Rulon gave a simple formula to estimate reliability by split-half method. The formula given in Ebel’s book [3] is-
\[ \gamma^2 = 1 - \frac{\sigma_d^2}{\sigma_t^2} \]

where \( \sigma_d^2 \) is the variance of the differences between the half test scores and \( \sigma_t^2 \) is the variance of the sum of the half test scores.

\[ \gamma = 1 - \frac{2.686^2}{9.7428} = 1 - \frac{6.96}{9.7428} = 1 - 0.717 = 0.283 \]

This value of split-half reliability by Rulon's formula is quite high.

Stanley has suggested a further simplification, using the following formula [4].

\[ \gamma = 1 - \frac{D_d^2}{D_s^2} \]

where \( D_d^2 \) represents the squared difference between the sum of the difference scores on the 27 per cent of papers having smallest half-test difference scores. \( D_s^2 \) represents the squared difference between the sums of total scores on the 27% of papers having largest total scores and the sum of total scores on the 27 per cent of papers having smallest total scores.

In the present inventory, \( D_d^2 = \) and \( D_s^2 = \frac{328}{1478} \).

So, \[ \gamma = 1 - \frac{D_d^2}{D_s^2} = 1 - \frac{328}{1478} = 1 - 0.22 = 0.78 \]

This result is very near the one obtained by Spearman-Brown formula.
Reliability by Kunder-Richardson Formula or the Method of Rational Equivalence

Kunder-Richardson formula reads as follows: [5]

\[
\gamma_{11} = \frac{\eta \sigma_e^2 - M(n-M)}{\sigma_e^2 (n-1)}
\]

where \( \sigma_e^2 \) is the variance of the inventory

\( M \) is the mean of inventory scores

\( n \) is the number of items in the inventory

and \( \gamma_{11} \) is the reliability of the whole test.

So

\[
\gamma_{11} = \frac{52 \times 50 - 31 (52-31)}{50(51)}
\]

\[
= \frac{2600 - 651}{2550} = .76
\]

The next question pertains to the inventory score. How reliable is the inventory score?

The Index of Reliability

An individual's true score on an inventory is defined as the mean of a very large number of determinations made of the same person on the same inventory or on its parallel form. The correlation between a set of obtained scores and their corresponding true counterparts is given by the formula [6]

\[
\rho = \sqrt{\gamma_{11}}
\]

where \( \rho \) is the correlation between obtained and true score and \( \gamma_{11} \) is the reliability coefficient.
Since the reliability coefficient of the inventory is about .75

\[ r_{100} = \sqrt{\frac{\sigma_i}{\sigma_{II}}} = \sqrt{.75} = .87 \]

\[ = \sqrt{.75} = .87 \]

The Standard Error of the Obtained Score

The effects of variable or chance errors in producing divergencies of scores from their true values is given by the formula [7].

\[ SE = \sigma_I \sqrt{1 - r_{II}} \]

where \( SE \) is the SE of an obtained score,

\( \sigma_I \) is the SD of the scores on the inventory

and \( r_{II} \) is the reliability co-efficient.

\[ SE = 7.07 \times .75 = 5.25 \times .77 \]

\[ = 7.07 \times .25 = 1.77 \]

Thus the SE of the obtained score on the inventory is 1.77.

The standard error of an estimated true score is given by the formula [8]

\[ SE = \sigma_I \sqrt{\sigma_{II} - r_{II}^2} \]

\[ = 7.07 \times .75 - .75 \]

\[ = 7.07 \times .75 - .56 \]

\[ = 7.07 \times .44 \]

\[ = 3.08 \]

Validity

"The validity of a test concerns what the test measures and
how well it does so."[9] If a watch tells right time, i.e. the standard time, then we may call the watch valid, an 'inventory is valid if it measures', what it purports to measure. There are several types of validities.

**Face Validity**

If from its appearance an inventory looks what it is designed to measure, it has face validity. Face validity pertains to whether the inventory "looks valid" to examinees who take it, the administrative personnel who decide on its use and other technically untrained observers.

Face validity can be easily improved by merely reformulating the items of the inventory in terms of relevance and plausibility in the particular setting in which they will be used. Fortunately, inventories are meant for pupils of later adolescence or adults, so very little change is needed to improve their items.

The present inventory is compiled to include items checking adjustments of various types and on the face of it it looks to be a proper type of inventory or in other words, the inventory has good face validity.

**Content Validity**

Content validation involves the systematic examination of the contents of the inventory to determine whether it covers a representative sample of behaviour domain to be measured.
It is also important to guard against the temptation to over-generalize regarding the domain sampled by the items of the inventory.

The present inventory was based on the six factors emphasised by Mary Jahoda to measure adjustment and maladjustment viz.

(1) Attitude towards the self
(2) Growth, development and self-actualization
(3) Integration
(4) Autonomy
(5) Perception of reality
and (6) Environmental mastery

The items of the inventory reflect these six criteria very well. Thus it has a good content validity.

**Construct Validity**

The construct validity of an inventory is the extent to which the inventory may be said to measure a theoretical construct or trait. In the present inventory it is adjustment for adolescents and as seen in the discussion of content validity, adjustment is properly viewed from different angles. In cases like these, the construct validity is the same as the content validity of the inventory. Thus it can be safely assumed that the construct validity of the present inventory is quite good.
**Concurrent Validity**

Concurrent Validity is measured by the coefficient of correlation between the new inventory and some older inventory which has proved to be efficient through use. In the present example, the new inventory has been correlated with K.G. Desai's well-known inventory of adjustment which has been used on a big scale in Gujarat. To estimate the correlations, both the inventories were administered to boys and girls of Grade XI and Grade XII of two schools in Ahmedabad. The following table shows the description of the sample and the next one the coefficient of correlation.

**Table-7.3**

*Sample for Concurrent Validity*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Boys</th>
<th>Girls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>XI</td>
<td>30</td>
<td>34</td>
<td>64</td>
</tr>
<tr>
<td>XII</td>
<td>38</td>
<td>32</td>
<td>70</td>
</tr>
<tr>
<td>Total</td>
<td>68</td>
<td>66</td>
<td>134</td>
</tr>
</tbody>
</table>

The coefficient of correlation between the two inventories is quite good which shows the concurrent validity of the present inventory.
Table-7.4
Correlation Matrix for Concurrent Validity

<table>
<thead>
<tr>
<th>Scores on Desai inventory</th>
<th>0-9</th>
<th>10-19</th>
<th>20-29</th>
<th>30-39</th>
<th>40-49</th>
<th>50 and above</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>90-99</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80-89</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70-79</td>
<td>11</td>
<td>12</td>
<td>7</td>
<td>1</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-69</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-59</td>
<td>1</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td></td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>30-39</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td></td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>10-19</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>0-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

\[ r = .68 \]

Factorial Validity

As the means of identifying psychological traits, factor analysis is relevant to construct validity. It is a refined statistical technique for analysing the inter-relationship of behaviour data. Factor analysis reduces to the underlying factors the large number of variables or items the inventory is compared of.
The present inventory has 52 items. On a sample of 100, inter-item correlations were computed which were then subjected to Hotteling's iterative factor analysis and the following five factors were extracted to represent the 52 items classified into seven variables.

First the 52 items were classified into 7 variables as described in chapter-6.

Table 7.5

<table>
<thead>
<tr>
<th>Variables</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Positive attitude towards self (self-confidence and self acceptance)</td>
<td>2, 8, 13, 21, 28, 31, 37, 42, 51</td>
</tr>
<tr>
<td>2. Adequate perception of reality</td>
<td>1, 7, 10, 22, 26, 27, 32, 34, 38, 45</td>
</tr>
<tr>
<td>3. Personal autonomy</td>
<td>9, 15, 19, 23, 30</td>
</tr>
<tr>
<td>4. Environmental mastery</td>
<td>11, 14, 20, 36, 41, 44</td>
</tr>
<tr>
<td>5. Integration</td>
<td>5, 16, 18, 25, 29, 43, 48, 50</td>
</tr>
<tr>
<td>6. Initiative</td>
<td>4, 17, 35, 39, 47, 52</td>
</tr>
<tr>
<td>7. Social and emotional competence</td>
<td>3, 6, 12, 24, 33, 40, 46, 49</td>
</tr>
</tbody>
</table>

Scores on each variable were correlated with those on other variables. The following table shows these inter-correlations.
Table-7.6

Intercorrelations of Seven Variables

<table>
<thead>
<tr>
<th>Serial No.</th>
<th>Correlation with other variables of variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>-.35</td>
<td>.21</td>
<td>.41</td>
<td>.39</td>
<td>.21</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>-.28</td>
<td>.32</td>
<td>.37</td>
<td>.29</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>-.40</td>
<td>.38</td>
<td>.29</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>-.24</td>
<td>.34</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>-.29</td>
<td>.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>-.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These correlations were then subjected to Hotteling's Principal Axis Method of factor analysis and the following five factors were extracted:

Table-7.7

Factor Loading

<table>
<thead>
<tr>
<th>Variables</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.68</td>
<td>-.25</td>
<td>-.03</td>
<td>.51</td>
<td>-.22</td>
</tr>
<tr>
<td>2</td>
<td>.32</td>
<td>.29</td>
<td>-.04</td>
<td>.32</td>
<td>-.12</td>
</tr>
<tr>
<td>3</td>
<td>.29</td>
<td>.04</td>
<td>-.01</td>
<td>-.09</td>
<td>-.09</td>
</tr>
<tr>
<td>4</td>
<td>.54</td>
<td>-.12</td>
<td>.23</td>
<td>-.02</td>
<td>.31</td>
</tr>
<tr>
<td>5</td>
<td>.32</td>
<td>.30</td>
<td>.32</td>
<td>.11</td>
<td>.21</td>
</tr>
<tr>
<td>6</td>
<td>.49</td>
<td>-.10</td>
<td>.41</td>
<td>.03</td>
<td>.18</td>
</tr>
<tr>
<td>7</td>
<td>.52</td>
<td>.02</td>
<td>-.12</td>
<td>-.01</td>
<td>-.05</td>
</tr>
</tbody>
</table>

168
After varimax rotation, these loadings were converted as follows:

Table 7.8

<table>
<thead>
<tr>
<th>Variables</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.51</td>
<td>.02</td>
<td>.02</td>
<td>.32</td>
<td>-.03</td>
</tr>
<tr>
<td>2</td>
<td>.42</td>
<td>.31</td>
<td>.10</td>
<td>.22</td>
<td>-.10</td>
</tr>
<tr>
<td>3</td>
<td>.52</td>
<td>.14</td>
<td>.01</td>
<td>-.01</td>
<td>.02</td>
</tr>
<tr>
<td>4</td>
<td>.31</td>
<td>-.05</td>
<td>.18</td>
<td>-.03</td>
<td>.21</td>
</tr>
<tr>
<td>5</td>
<td>.55</td>
<td>-.12</td>
<td>.24</td>
<td>-.10</td>
<td>.38</td>
</tr>
<tr>
<td>6</td>
<td>.40</td>
<td>.30</td>
<td>.55</td>
<td>.20</td>
<td>-.05</td>
</tr>
<tr>
<td>7</td>
<td>.30</td>
<td>.10</td>
<td>-.09</td>
<td>.01</td>
<td>-.02</td>
</tr>
</tbody>
</table>

The first factor shows positive and appreciable loadings with the seven variables. This can therefore, be called the factor of 'general adjustment'.

The second factor can be called 'perception of reality' as it has appreciable correlation with variable 2 and 6.

The third factor shows good correlation with the variable 6 and appreciable with variable 5 and so can be called 'initiative inadjustment'. The fourth factor shows appreciable correlation with variable 1 and a little less with variable 2 and 6. It can be called 'self-acceptance'.
The fifth factor shows appreciable correlation with variable 5 and a little less with variable 6 and so it may be called 'integration of personality'.

Thus the factor analysis indicates the factorial validity of the adjustment inventory.

In this way, the present adjustment inventory shows validity of all different types to a large extent.

References

1. Anastasi, Anne, op cit., p.102.


5. Garret, H.E. op cit., p.341.


