CHAPTER : 6

ESTIMATION OF RELIABILITY AND VALIDITY

6.1 RELIABILITY

6.1.1 TYPES OF RELIABILITY

6.1.1.1 TEST-RETEST RELIABILITY

6.1.1.2 ALTERNATE OR PARALLEL FORMS RELIABILITY

6.1.1.3 SPLIT-HALF RELIABILITY

6.1.1.4 METHOD OF RATIONAL EQUIVALENCE

(KUDE RICHARDSON FORMULAS)

6.1.2 ESTIMATING RELIABILITY OF VII

6.1.3 PREDICTION OF COMBINED RELIABILITY OF FORMS A AND B

6.1.4 DIFFERENT RELIABILITIES OF THE VII IN GUJARATI VERSION

6.2 VALIDITY

6.2.1 TYPES OF VALIDITY

6.2.1.1 OPERATIONAL AND PREDICTIVE VALIDITY

6.2.1.2 FACE VALIDITY

6.2.1.3 CONTENT VALIDITY

6.2.1.4 FACTORIAL VALIDITY

6.2.1.5 CONSTRUCT VALIDITY

6.2.1.6 CONCURRENT VALIDITY

6.2.2 VALIDITY OF THE VII
6.1 RELIABILITY

Reliability and validity represent two extremely essential psychometric characteristics of a sound measuring instrument. Tests can predict, diagnose or measure a trait accurately and efficiently only if they are dependable and trustworthy.

According to Anastasi:

Reliability refers to the consistency of scores obtained by the same individuals when re-examined with the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions.

This concept of reliability underlies the computation of coefficients of test-retest reliability as well as of two alternate forms of the same test or the standard error of measurement of a single score, whereby one can predict the range of fluctuations likely to occur in a single individual's score as a result of irrelevant chance factors.

Whenever anything is measured, whether in the physical, biological or behavioural sciences, there is some possibility of chance errors. This is true of psychological tests as well. Variations of results obtained with the same test administered more than once, using the same persons as subjects, or within the parts of a test given only once, are due not only to chance factors, which should be eliminated so far as possible, but also to actual differences among the individuals taking the test and to whatever defects may be inherent in the instrument itself.
In psychological testing, the term reliability has been used to denote two closely related though somewhat different aspects. In the first place, reliability refers to the consistency of scores obtained from a single administration of the test and is an index of internal consistency. In the second place, it refers to the extent to which the test yields the same results on several occasions. This is known as an index of temporal stability and is associated with the predictive value of the instrument.

6.1.1 Types of Reliability:

Four types that are commonly used to estimate reliability of a test are:

1. Test - retest method
2. Alternate - forms or parallel forms method
3. The split-half method
4. The method of Rational Equivalence

6.1.1.1 Test-Retest Reliability: The simplest way to find the reliability of the test is by means of a retest, or repetition of the identical test on a second occasion. The reliability coefficient which is also known as the coefficient of stability is nothing but correlation between the scores obtained by the same subjects on the two administrations of the test. This corresponds to the concept of temporal stability. The main advantage of this method is the inclusion of a new dimension - one that is external to the test and independent of it - time and that is perhaps
why Helim considers this method the most satisfactory means of assessing test consistency.

When retest reliability is reported in a test manual, the interval over which it was measured should always be specified, since retest correlations decrease progressively as this interval lengthens. There is not one, but an infinite number of retest reliability coefficients for any test. It is also desirable to give some indications of relevant intervening experiences of the subjects on whom reliability was measured.

"Test reliability (by test-retest method) is an extent to which the same group of people get (sic) the same scores (therefore, rank the same) on reported testings... It is important because if a test isn't consistent in the scores and ranking obtained from it, it can't be considered an accurate measuring device. A test on which persons scored very high (low) the first time, but very low (high) the second time, is like a broken ruler, it's not useful in measuring anything." 2

6.1.1.2 Alternate or Parallel Forms Reliability: In case of alternate or parallel forms reliability, retesting of the same group is done with a parallel or equivalent form of the test. The statistics thus obtained is known as the coefficient of equivalence of the two forms.

Like test-retest reliability, alternate forms reliability should always be accompanied by a statement of the length of the interval between two test administrations,
as well as a description of relevant intervening experiences. If the two forms are administered in immediate succession or at essentially the same time, the resulting correlation is designated as the coefficient of equivalence.

In the development of alternate forms, care should, of course, be exercised to ensure that they are truly parallel. Fundamentally, parallel forms of a test should be independently constructed tests designed to meet same specifications. The test should contain the same number of items, and items should be expressed in the same form and should cover the same type of content. The range and level of difficulty of the items should also be equal. Instructions, time-limits, illustrative examples, format and all other aspects of the parallel form must likewise be checked for comparability.

6.1.1.3 Split-half Reliability: From a single administration of one form of a test, it is possible to arrive at a measure of reliability by various split half procedures. In such a way, two scores are obtained for each individual by dividing the test into comparable halves. It is apparent that split half reliability provides a measure of consistency with regard to content sampling. Temporal stability of the scores does not enter into such reliability, because only one test session is involved. This type of reliability coefficient is sometimes called a coefficient of internal consistency, since only a single administration of a single form is required.
To find split half reliability, the first problem is how to split the test in order to obtain the most nearly comparable halves. Any test can be divided in many different ways. In most tests, the first half and the second half would not be comparable, owing to differences in nature and difficulty level of items, as well as of the cumulative effects of warming up, practice, fatigue, boredom and any other factors varying progressively from the beginning to the end of the test. A procedure that is adequate for most purposes is to find the scores on the odd and even items of the test. If the items were originally arranged in an approximate order of difficulty, such a division yields very nearly equivalent half scores.

Once the two half scores have been obtained for each subject, they may be correlated by the usual method and the reliability of the full test can be computed by using Spearman-Brown prophecy formula.

6.1.1.4 Method of Rational Equivalence (Kuder-Richardson Formulas): A fourth method for finding reliability, also utilizing a single administration of a single form is based on the consistency of subjects' responses to all items in the test. This inter-item consistency is influenced by two sources of error variance. (1) content sampling (as in alternate form and split half reliability); and (2) heterogeneity of the behaviour domain sampled. The more homogeneous the domain, the higher the inter-item consistency.
The most common procedure for finding inter-item consistency is that developed by Kuder and Richardson. As in the split half method, inter-item consistency is found from a single administration of a single test. Rather than requiring two half scores, however, such a technique is based on an examination of performance on each item.

The method of rational equivalence stresses the inter correlations of the items in the test and the correlations of the items with the test as a whole. Four formulas for determining test reliability have been derived of which the one given below is perhaps the most useful 3.

\[ \gamma_{11} = \frac{n}{n-1} \times \frac{\sigma_t^2 - \sum pq}{\sigma_t^2} \]

in which

\[ \gamma_{11} = \] reliability coefficient of the whole test  
\[ n = \] number of items in the test  
\[ \sigma_t = \] 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

The product of p and q is computed for each item, and these products are then added for all items, to give \( \sum pq \). Since in the process of test construction p is often routinely recorded in order to find the difficulty level of each item, this method determining reliability involves little additional computation.
A simple approximation to the above formula is often useful to teachers and others who want to determine the reliability quickly.

The formula is:

\[ Y_{1II} = \frac{\eta \sigma^2_t - \bar{M} (\eta - \bar{M})}{\sigma^2_t (\eta-1)} \]

Where:

\[ Y_{1II} \] = reliability of the whole test  
\[ \eta \] = number of items in the test  
\[ \sigma^2_t \] = SD of the test scores  
\[ \bar{M} \] = the mean of the test scores

As H.E. Garrett remarks:

"The rational equivalence formulas are not strictly comparable to the three methods already outlined. Like the spilt-half technique, these formulas provide an estimate of the internal consistency of the test and thus of the dependability of test scores. Rational equivalence is superior to the spilt-half technique in certain theoretical aspects, but the actual difference in reliability coefficients found by the two methods is never large and is often negligible."

6.1.2 Estimating Reliability of VII:

In the present study, the problem of reliability of VII has been approached from the following different angles:

(a) Coefficient of Temporal Stablility by test-retest method
(b) Coefficient of Equivalence by parallel forms method
(c) Coefficient of Internal Consistency by split-half method

(d) Coefficient of Inter-item Consistency by method of rational equivalence

(a) Coefficient of Temporal Stability by test-retest method: The coefficient of temporal stability was determined by test retest method. If the time interval between the two administrations of the test is short, the immediate memory effect, practice and the confidence created by familiarity with the test material may vitiate the reliability of the test. On the other hand, if the time interval is pretty long, the real changes in behaviour in terms of growth may again vitiate the reliability of the test.

To avoid these difficulties, the present investigator administered the retest after 4 weeks. Test-retest method was applied for both the groups: (1) Secondary school pupils and (2) Higher Secondary school pupils. Sanghavi High School from the urban area and Sheth C.K. High School from the semi-urban area were selected at random. The proportionate number of pupils from each grade (IX, X, and XI & XII) was selected by systematic random sampling. All the pupils of
different grades were administered the same inventory again, in one group, as a retesting. The total number of boys, thus, selected were 86 while the total girls were 81. The test-retest reliability for all the ten areas of interest was found out using product-moment 'γ' for boys and girls separately and, the average γ were found out using Fisher's Z.

Table 6.1 presents γ of test-retest reliability.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Area of Interest</th>
<th>Coeff. of temporal stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Boys</td>
</tr>
<tr>
<td>1</td>
<td>Outdoor</td>
<td>.67</td>
</tr>
<tr>
<td>2</td>
<td>Mechanical</td>
<td>.68</td>
</tr>
<tr>
<td>3</td>
<td>Computational</td>
<td>.66</td>
</tr>
<tr>
<td>4</td>
<td>Scientific</td>
<td>.72</td>
</tr>
<tr>
<td>5</td>
<td>Persuasive</td>
<td>.75</td>
</tr>
<tr>
<td>6</td>
<td>Artistic</td>
<td>.74</td>
</tr>
<tr>
<td>7</td>
<td>Literary</td>
<td>.56</td>
</tr>
<tr>
<td>8</td>
<td>Musical</td>
<td>.80</td>
</tr>
<tr>
<td>9</td>
<td>Clerical</td>
<td>.75</td>
</tr>
<tr>
<td>10</td>
<td>Social Service</td>
<td>.61</td>
</tr>
</tbody>
</table>

It can be read from Table 6.1 that throughout all areas of interest, boys are having higher coefficients of temporal stability ranging from .56 (for literary interest) to .80 (for musical interest). The girls' results of test-reliability
are inferior, the minimum being .51 (for mechanical interest) and the maximum, .66 (for musical interest). Would it be that girls would be less serious than boys in taking this inventory; at least in the second administration? The average of both boys and girls together were computed by converting $Y$ into Fisher's $Z$, coefficients and reconverting them again into $Y$. The average $Y$s have been shown in the last column; they range from .55 to .68. One should not forget the fact that one cannot expect very high coefficients of reliability as those found with intelligence and aptitude tests. Personality traits and interests are not comparatively so stable and hence generally the later has low reliability.

(b) Coefficient of Equivalence by Parallel Forms Method: Parallel forms reliability or coefficient of equivalence was determined for the 10 vocational interest variables.

The sample of 100 subjects were selected out of 1237 subjects by a systematic proportionate (gradewise) method of sampling and included pupils of all the grades IX, X, XI, and XII. The total number of pupils selected from these four grades was 23, 24, 28 and 25 respectively. These students were already administered both the Forms A and B.
The areawise coefficients of equivalence, thus, obtained have been presented in Table 6.2.

TABLE 6.2
AREA OF INTERESTWISE COEFFICIENTS OF EQUIVALENCE
(N=100)

<table>
<thead>
<tr>
<th>Sr. NO.</th>
<th>AREA OF INTEREST</th>
<th>COEFFICIENT OF EQUIVALENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Outdoor</td>
<td>.68</td>
</tr>
<tr>
<td>2</td>
<td>Mechanical</td>
<td>.72</td>
</tr>
<tr>
<td>3</td>
<td>Computational</td>
<td>.68</td>
</tr>
<tr>
<td>4</td>
<td>Scientific</td>
<td>.96</td>
</tr>
<tr>
<td>5</td>
<td>Persuasive</td>
<td>.71</td>
</tr>
<tr>
<td>6</td>
<td>Artistic</td>
<td>.72</td>
</tr>
<tr>
<td>7</td>
<td>Literary</td>
<td>.60</td>
</tr>
<tr>
<td>8</td>
<td>Musical</td>
<td>.65</td>
</tr>
<tr>
<td>9</td>
<td>Clerical</td>
<td>.62</td>
</tr>
<tr>
<td>10</td>
<td>Social service</td>
<td>.70</td>
</tr>
</tbody>
</table>

It can be read from Table 6.2, the coefficients of equivalence by parallel forms method range from 0.60 to 0.96; the least value being for 'Literary' interest while the highest value being for 'Scientific' interest. Thus, all the coefficients of equivalence are 0.60 and above.

(c) Coefficient of Internal Consistency by Split-half method:

Split half reliability coefficients or coefficients of internal consistency were determined for the 10 vocational interest areas.
The sample of 100 pupils selected at random was the same as that used in Parallel Forms method. The internal consistency coefficients, corrected by the Spearman-Brown formula are presented in Table 6.3.

**TABLE 6.3**
**AREA OF INTERESTWISE COEFFICIENTS ON INTERNAL CONSISTENCY (N=100)**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>AREA OF INTEREST</th>
<th>COEFFICIENT OF INTERNAL CONSISTENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Outdoor</td>
<td>.62</td>
</tr>
<tr>
<td>2</td>
<td>Mechanical</td>
<td>.88</td>
</tr>
<tr>
<td>3</td>
<td>Computational</td>
<td>.74</td>
</tr>
<tr>
<td>4</td>
<td>Scientific</td>
<td>.79</td>
</tr>
<tr>
<td>5</td>
<td>Persuasive</td>
<td>.70</td>
</tr>
<tr>
<td>6</td>
<td>Artistic</td>
<td>.82</td>
</tr>
<tr>
<td>7</td>
<td>Literary</td>
<td>.78</td>
</tr>
<tr>
<td>8</td>
<td>Musical</td>
<td>.72</td>
</tr>
<tr>
<td>9</td>
<td>Clerical</td>
<td>.79</td>
</tr>
<tr>
<td>10</td>
<td>Social service</td>
<td>.79</td>
</tr>
</tbody>
</table>

Table 6.3 reveals that the coefficients of internal consistency vary from 0.62 to 0.88; only one coefficient of internal consistency is less than 0.70 (for 'Outdoor' interest).

The internal consistency of the VII developed in Gujarati version seems to be quite moderate. Yet, one point should not be forgotten, and that is, the format of the schedule is totally different wherein one has to show one's preference for each and every statement on a five point scale; while in the original schedule out of the triad one has to select one activity for liking and the other for disliking.

260
The Method of Rational Equivalence (Kuder-Richardson Coefficient):

Besides the test-retest reliability, parallel forms reliability and split-half reliability, the rational equivalence method was also used to estimate reliability of VII.

The Kuder-Richardson formulas 20 and 21 which are very commonly used in estimating reliability by the method of rational equivalence are not applicable for VII. The K.R. formulas are "applicable only to those tests whose items are scored right or wrong, or according to some other all-or-none system."

Some tests, however, may have multiple scored items. In this VII, the testee receives a different numerical score on an item from 4 to 0, depending on whether he or she checks 'likes the most', 'likes it', 'indifferent', 'dislikes it' and 'dislikes the most', respectively. In such a case, a generalised formula has been derived, known as coefficient alpha (Cronbach 1951....) It is as follows:

$$\gamma_{t} = \left( \frac{n}{n-1} \right) \frac{SD_t^2 - \sum (SD_t^2)}{SD_t^2}$$

Wherein, $$\sum pq$$ is replaced by $$\sum (SD_t^2)$$, the sum of the variance of item scores". In this study, the above Cronbach alpha formula has been applied to estimate reliability by the method of rational equivalence.
Table 6.4 presents the data about coefficients of inter-item consistency.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>AREA OF INTEREST</th>
<th>Cronbach alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Outdoor</td>
<td>.77</td>
</tr>
<tr>
<td>2</td>
<td>Mechanical</td>
<td>.84</td>
</tr>
<tr>
<td>3</td>
<td>Computational</td>
<td>.77</td>
</tr>
<tr>
<td>4</td>
<td>Scientific</td>
<td>.89</td>
</tr>
<tr>
<td>5</td>
<td>Persuasive</td>
<td>.78</td>
</tr>
<tr>
<td>6</td>
<td>Artistic</td>
<td>.75</td>
</tr>
<tr>
<td>7</td>
<td>Literary</td>
<td>.71</td>
</tr>
<tr>
<td>8</td>
<td>Musical</td>
<td>.73</td>
</tr>
<tr>
<td>9</td>
<td>Clerical</td>
<td>.75</td>
</tr>
<tr>
<td>10</td>
<td>Social service</td>
<td>.78</td>
</tr>
</tbody>
</table>

It can be read from Table 6.4 that the range of Cronbach alpha coefficients is from 0.71 to 0.89 which can be considered quite satisfactory looking to the number of items (that is, 12 items), in each area of interest. It is a very open fact that reliability of a test depends upon its length. Simply by increasing the length of a test (adding more items into it), the reliability of a test can be boosted.

6.1.3 Prediction of Combined Reliability of Forms A and B:

It was mentioned in Chapter 5 that the standardisation
of the Form B would be taken up immediately as a post doctoral work. However, coefficient of equivalence by parallel forms A and B each containing 12 items has already been estimated.

Taking these data as a split half reliability and applying the very well known Spearman-Brown Prophecy formula, one can predict the reliability of both the forms A and B together. "As we increase the length of the test, chance plays a less and less important role."

Table 6.5 presents the data about coefficient of equivalence of parallel forms which is now split-half reliability and the predicted reliability of the full inventory (Forms A and B).

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>AREA OF INTEREST</th>
<th>Split half Reliability (From Table 6.2)</th>
<th>Predicted Reliability of the full VII (Forms A and B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Outdoor</td>
<td>.68</td>
<td>.81</td>
</tr>
<tr>
<td>2</td>
<td>Mechanical</td>
<td>.72</td>
<td>.84</td>
</tr>
<tr>
<td>3</td>
<td>Computational</td>
<td>.68</td>
<td>.81</td>
</tr>
<tr>
<td>4</td>
<td>Scientific</td>
<td>.96</td>
<td>.98</td>
</tr>
<tr>
<td>5</td>
<td>Persuasive</td>
<td>.71</td>
<td>.83</td>
</tr>
<tr>
<td>6</td>
<td>Artistic</td>
<td>.72</td>
<td>.84</td>
</tr>
<tr>
<td>7</td>
<td>Literary</td>
<td>.60</td>
<td>.75</td>
</tr>
<tr>
<td>8</td>
<td>Musical</td>
<td>.65</td>
<td>.79</td>
</tr>
<tr>
<td>9</td>
<td>Clerical</td>
<td>.62</td>
<td>.77</td>
</tr>
<tr>
<td>10</td>
<td>Social service</td>
<td>.70</td>
<td>.82</td>
</tr>
</tbody>
</table>

It is naturally visualised from Table 6.5 that predicted reliability of the whole VII (Forms A and B) is
boosted up, area of interestwise. The range from 0.60 to 0.96 has been increased to from 0.75 to 0.98. It can, therefore, be concluded that whenever time permits, both the forms A and B developed by the present investigator should be administered to the subjects, after the Form B is standardised.

6.1.4 Different Reliabilities of the VII in Gujarati Version

To have the gestalt of different types of reliability, the present investigator has added one more Table 6.6 wherein all types of coefficients of reliability are shown together to have a clear-cut picture, at sight, of the reliability of the VII, in Gujarati Version.
### TABLE 6.6
DIFFERENT TYPES OF RELIABILITIES OF GUJARATI VII

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Area of Interest</th>
<th>Method</th>
<th>Test Retest N=127</th>
<th>Parallel Forms N=100</th>
<th>Split Half Method N=100</th>
<th>Cronbach alpha method N=100</th>
<th>Split half method (N=100) Corrected by S.B. Prophecy Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Outdoor</td>
<td></td>
<td>0.61</td>
<td>0.68</td>
<td>0.62</td>
<td>0.77</td>
<td>0.81</td>
</tr>
<tr>
<td>2</td>
<td>Mechanical</td>
<td></td>
<td>0.60</td>
<td>0.72</td>
<td>0.88</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>3</td>
<td>Computational</td>
<td></td>
<td>0.61</td>
<td>0.68</td>
<td>0.74</td>
<td>0.77</td>
<td>0.81</td>
</tr>
<tr>
<td>4</td>
<td>Scientific</td>
<td></td>
<td>0.64</td>
<td>0.96</td>
<td>0.79</td>
<td>0.89</td>
<td>0.98</td>
</tr>
<tr>
<td>5</td>
<td>Persuasive</td>
<td></td>
<td>0.68</td>
<td>0.71</td>
<td>0.70</td>
<td>0.78</td>
<td>0.83</td>
</tr>
<tr>
<td>6</td>
<td>Artistic</td>
<td></td>
<td>0.67</td>
<td>0.72</td>
<td>0.82</td>
<td>0.75</td>
<td>0.84</td>
</tr>
<tr>
<td>7</td>
<td>Literary</td>
<td></td>
<td>0.55</td>
<td>0.60</td>
<td>0.78</td>
<td>0.71</td>
<td>0.75</td>
</tr>
<tr>
<td>8</td>
<td>Musical</td>
<td></td>
<td>0.74</td>
<td>0.65</td>
<td>0.72</td>
<td>0.73</td>
<td>0.79</td>
</tr>
<tr>
<td>9</td>
<td>Clerical</td>
<td></td>
<td>0.65</td>
<td>0.62</td>
<td>0.79</td>
<td>0.75</td>
<td>0.77</td>
</tr>
<tr>
<td>10</td>
<td>Social Service</td>
<td></td>
<td>0.57</td>
<td>0.70</td>
<td>0.79</td>
<td>0.78</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Note: The table includes different types of reliabilities for Gujаратi VII, including social service, with varying correlation and coefficient values.
Table 6.6 is self explanatory and needs no discussion. Reliabilities estimated by four different methods have different specifications and they can, in no way, be averaged, a well-known fact for the psychometricians.

6.2 VALIDITY:

The validity of a test is related to what the test purports to measure and how well it does so. The validity of a test or of an inventory is frequently defined as, "The extent to which the test or inventory actually measures what it purports to measure." The determination of the validity of an inventory would involve the correlation between scores on the inventory and some pure criterion measure of what the inventory purports to measure.

The construction and use of a test imply that instrument has been evaluated against criteria regarded by experts as the best evidence of the traits to be measured by the test.

The first essential quality of a valid test is that it should be highly reliable. If a reliability coefficient of a test is very low, it will not be correlated with anything. A test that yields inconsistent results (low reliability) cannot correlate well with a measure of another variable; in this instance, a criterion.
Glenn and others observes:

Test validity is an extent to which scores on a test correspond to the actual behaviour the test is designed to measure. If it turns out that the behaviour the test was designed to measure has little or no relation to the test scores, the test is useless.

6.2.1 Types of Validity:

6.2.1.1 OPERATIONAL AND PREDICTIVE VALIDITY: Operational validity means that the tasks required by the test are adequate for the measurement and evaluation of certain defined psychological traits. The predictive validity of a test is the extent to which it is efficient in forecasting and differentiating behaviour or performance in a specified area under actual working and living conditions.

Predictive validity is dependent, at least in part, upon the operational validity of the test. The reason is that the psychological operations required by the test were included because they were found essential for testing, in certain actual situations. Hence, if the psychological operations, or the information, or the specific skills are not measured truly, prediction or later performance will be adversely affected.

6.2.1.2 Face validity: This form of validity now, more or less neglected, has been disparaged since more sophisticated procedures have been devised. As a matter of fact, however, face validity in the earlier days of test development was the criterion used by many competent psychologists as a
first step. Validation of their test-content at face value was not as capricious, haphazard, or casual as some critics have said. On the contrary, their content was based upon whatever psychological knowledge and insight could, then, be utilised. Face validity was claimed most often with tests of educational achievement and of personality and to a lesser extent with tests of specific aptitudes.

6.2.1.3 Content Validity: It is the most appropriately applied only to tests of proficiency and of educational achievement, although such validity may be and should be supplemented by several types of statistical analysis; validity of content, however, should not depend upon the subjective judgement of only one specialist. It should be based upon careful analysis, by several specialists, of instructional objectives and of the actual subject-matter studied.

6.2.1.4 Factorial validity: In the statistical method called factor analysis, the items of a large number of tests are examined and if possible, accounted for in terms of a much smaller number of more general 'factors' or 'trait categories'. The factors presumably run through the often complex abilities measured by the individual tests. It is sometimes found, for example, that 3 or 4 factors will account for the inter correlations obtained among fifteen or more tests. The validity of a given test is defined by its factor loadings and these are given by the correlations of
the test with each factor. Factor analysis is a specialized mathematical technique widely used and highly important in modern test construction.

6.2.1.5 Construct Validity: It is concerned with the meaning and interpretation of the test scores obtained in terms of psychological or theoretical constructs. Construct validity is concerned not only with the test itself but also with the theory which seeks to explain, or to account for the results which are obtained when the test is used.

Construct validity is established through a long continued experimentation based on imagination, reasoning and observation. First imagination suggests that a construct X accounts for the test performance. The researcher, then reasons, "If that is so, individuals with a high score should have characteristic or attribute Y." An experiment is performed and observations are made, and the expectations are confirmed; the interpretation is supported. But as various deductions are tested, some of them prove to be inaccurate. The proposed interpretation is, then, altered by involving a different concept, by introducing an additional concept, or by altering the theory of the concept itself. The process of construct validation is the same as that by which scientific theories are developed.
6.2.1.6 Concurrent Validity: Originally, psychologists spoke to validation with other tests, validation with a proficiency rating, validation with school grades, or, in the case of a personality test, validation with a recent diagnosis. At present, psychologists prefer the term, "concurrent validity" to indicate the process of validating the new test by correlating it, or otherwise comparing it for agreement, with some present source of information. This source of information might have been obtained shortly before or after the new test has been administered.

Dougless and Choen rightly remark:

Two methods that demonstrate validity by correlating tests with independent criteria are predictive and concurrent validity. It is assumed that psychological tests predict independent, external criteria that might include specific behaviour (e.g. suicide attempts), general traits (e.g. scores on other psychological tests), or certain future outcomes (e.g. graduation from college).... the higher the correlation, the greater the test's predictive power, and more valid the test is in placement. This is the essence of predictive validity: current test scores foretell some independent criterion in future.

Concurrent validity is nearly identical validity, except that the outside criterion is assessed at the same time the test is administered. Often a pre-existing or well established test is used as a criterion with which a new test is correlated .... Obviously, concurrent validation depends on having identified a suitable external criterion, such as the Wechsler scales.
6.2.2 VALIDITY OF THE VII

Face Validity: The definition of vocational interests accepted for this study was that of Kuder - "Vocational Interests, in general, tell you what kind of duties you like to perform." (p.111). To be more specific, the investigator would like to give an operational definition as follows:

"A Vocational Interest is an expressed preference for engaging in a given occupational activity. The degree of a person's apparent occupational interests is reflected in the extent to which the person expresses a greater preference for one activity (or collection of activities) than for another activity (or collection of activities)."

As the present inventory assesses students' "expressed interest", it can be claimed that it possesses "face validity".

Content Validity: The rationale for the selection of activities that are represented in the items on the inventory was three-fold.

a day-to-day / seasonal activities
b curricular / co-curricular activities run in the schools, and
c certain duties or activities carried out in different occupations commonly found in this region.

The activities under (a) and (b) were such that they were bases for future complicated or sophisticated types of
occupational duties or activities; they might be considered as miniature activities for future occupations.

The newly coined items presented on pages 139-146 were also devised on this rationale.

On page 158, the investigator has enlisted the newly coined items actually selected in the present inventory by giving serial numbers that can be found in Appendix H (English version - pp. 412-16).

These thirty seven items are divided into three parts as per the rationale.

a  18,20,21,46,53,58,64,103 -> 8 items
b  1,27,28,39,44,54,57,74,112,117,119 -> 11 items

Thus, the selection of activities in the present inventory represents viable options available in the state of Gujarat. The present inventory, therefore, samples in a reasonable way the 'content' of the labour market available to the students, in this region. It can be concluded that the present inventory possesses content validity.

**Construct validity**: Generally, "Vocational Choice" is taken up as one of the 'constructs' for validation purposes. The higher scores on areas of interest are pitted against the actual vocational choice of a person. In India, there is a very acute
problem of unemployability; even graduates and some post-graduates are either unemployed or grasp any vocation available without taking into account the interest profile. There is a question of one's survival. Again the "social status" accompanied with an occupation is being given undue importance. Most of the unemployed persons crave for "White Collar Jobs" only, though there are availability of blue collar jobs. Under these circumstances, it would be futile to take vocational choice as a Construct.

Howsoever, the investigator has taken up "teachers' rating" as a construct and has estimated construct validity.

Teachers' Rating : The comparison of three highest scores achieved by 100 pupils of Pravin Vidya Vihar of Ahmedabad city was carried out with three highest ratings of teachers out of the list of ten areas of interest. Teachers were instructed to give first three ranks 1, 2 and 3 to those areas of interest which teachers thought, the pupil has shown such interest in different activities carried out in the school. Pravin Vidya Vihar was selected purposefully as some guidance activities like career talks, celebration of Guidance Week, keeping cumulative record cards, Parent - Teachers' association, etc. were implemented in that school. In short, the school was guidance-oriented that has been rarely found in Gujarat, except in those missionary schools.
Then, the number of frequencies was counted: (a) Common interest in three areas (pupils' scores and teachers' rating) (b) common interest in any two areas of interest (c) common interest in only one area and (d) no common area of interest between the two. Equal probability hypothesis was applied to the data and chi-square was found out.

To find out coefficient of contingency from chi-square, the following formula was applied:

$$ z = \sqrt{ \frac{\chi^2}{N+\chi^2} } $$

Table: 6.8 presents the necessary data.
TABLE 6.8
PERCENTILE NORMS AND TEACHERS' RATING

Congruence of Frequencies in

<table>
<thead>
<tr>
<th>ALL THE AREAS</th>
<th>ANY TWO AREAS</th>
<th>ANY ONE AREA</th>
<th>NONE AREA</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>14</td>
<td>63</td>
<td>12</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>100</td>
</tr>
</tbody>
</table>

$\chi^2 = 77.2$  \hspace{1cm} df = 3  \hspace{1cm} P < 0.001  
(Table Value 16.266)

Coefficient of Contingency = $C = 0.66$

It can be read from Table 6.8 that chi-square value is 77.2 which is significant beyond 0.001 level. The null hypothesis of equal probability is, therefore, not accepted. This chi-square value is, then, converted into contingency coefficient by the formula mentioned earlier. This coefficient clearly indicates there is a significant relationship between first three percentile scores of pupils and three areas of interest in which pupils show their interest as per rating of teachers.

Importance of Measurement Of Interest in Vocational Counselling:

The measurement of interests is not merely justified and desirable but is a necessity in a career planning as well as in vocational counselling. Two very effective ways in which a student engaged in a career planning can use his interest profile are as a means of uncovering some new and promising vocational leads within a vast network of existing vocations and as a device for narrowing the wide range of job opportunities to those most worthy of his intensive and serious investigation. The assumption underlying the importance of the survey of students' or people's interests is that expressed interest of the VII is one significant factor determining.
the types of occupations people seek to enter,

b the extent to which they remain in those occupations, and
c the level of success they enjoy in those occupations.

These three aspects can be taken up as 'predictive validity' only when the pupils selected in a sample would finish their study up to higher secondary level, certificate or diploma courses, graduation or post graduation.

Of course, one has to be cautious not to give undue weightage to 'interest' only; the other equally important factors that should be taken into consideration while doing counselling are academic achievement, mental ability (intelligence), aptitudes of pupils, personality traits, socio-economic level of the parents, job-opportunities available, the experiences a person undergoes, etc.

Two important aspects of validity, namely (1) what the VII truly measures and (2) why it is important to measure interest in vocational counselling have been discussed.

In conclusion, it can be said that a new inventory has sufficiently satisfactory reliability as well as validity.

The present investigator is quite aware that she need not get inflated in any way. Mursell very rightly remarks, "The ultimate validation of any test is to be found only in its wide and serviceable use". The present inventory must stand on its own feet in its day-to-day trial and evaluation by one's peers would be the final arbitrator.
RÉFÉRENCES


4. Ibid., p. 342.


6. Ibid., p. 117.


11. Ibid., p. 390.