CHAPTER 6

VALIDITY

** MEANING OF VALIDITY

** TYPES OF VALIDITY

- Face Validity

- Construct Validity

- Concurrent Validity
  a) Contrasted Group
  b) Ratings by Teachers
  c) Interview
  d) Correlation with other tests

- Inter-correlations
The validation of a test is a long process rather than a single event. In the beginning of the testing movement, the widely accepted procedure was to define first what was intended to measure and then to collect evidences to demonstrate how successfully it was being measured. But today, one must analyse the content of the test and examine many correlations with different criteria in various groups to have an adequate information to understand just what it is measuring (104).

Understanding what is involved in the validation of tests carries three clear implications for test users as under:

1. One is that if a test is to be employed in making decisions about individuals or groups, all the available evidence should be studied before any attempt is made to interpret the scores.

2. Another implication is that whenever possible a test to be used for prediction or selection should be validated in the specific situation in which it is to be employed.

3. In light of measuring theoretical knowledge of individual differences the ideas about what the traits are as well as what the test measures must change as new evidence comes in.

TYPES OF VALIDITY

There are several kinds of validity and procedures for determining each validity. The type of validity necessary for a particular test depends upon the purpose(s) for which it is constructed.
According to Adams (1), Psychologists, Educationists, Counsellors and Personnel Managers carry out testing for several purposes:

a) to determine how an individual would perform at present in a given universe of situations of which the test situation constitutes a sample,

b) to estimate an individual's present status on some variable external to the test,

c) to predict an individual's future performance, and

d) to infer the degree to which the individual possesses some trait or quality, presumed to be reflected in the test scores, so Adam talked about four types of validity:

1) content validity, 2) concurrent validity, 3) predictive validity and 4) construct validity — one for each purpose respectively.

Freeman (41) has broadly indicated two types of validity — operational validity and predictive validity. Operational validity means that the task required by the test are adequate for the measurement of psychological characteristics or activities under consideration. While predictive validity means the extent to which it is efficient in forecasting and differentiating behaviour or performance in a specified area under consideration.

The APA Technical Recommendations, as noted by Anastasi (4) classified various validity procedures under four categories designated as content, predictive, concurrent and construct validity.
As the purpose of the present inventory is to assess the degree of adjustment in general and in specific life areas of an individual, so the question of establishing predictive validity was not attempted by the present author. The section that follows discusses the various validity procedures used in determining different types of validity of this test.

Face Validity

Face validity refers to what it appears superficially to measure. It pertains to whether the test 'looks valid' to the subjects who take it and to the administrative personnel who decide upon its use, and other untrained observers. Basically the question of face validity concerns rapport and public relations. As Anastasi (4) says, if the test content appears irrelevant, inappropriate, silly or childish, the result will be poor cooperation, regardless of the actual validity of the test.

Further, according to her, it is not sufficient for a test to be objectively valid. It also needs face validity to function effectively in all practical situations.

This type of validity 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. It was claimed most often when tests of educational achievement and of personality, and to a lesser extent with tests of specific aptitudes (41).

According to this type of validity, the content of the test seems to be relevant to its stated purpose, and no further effort is made to confirm the assumption objectively. This concept of an operational validity is based upon specific judgement of specialists making the diagnosis.
The main purpose of the present inventory is to determine the degree of general and specific adjustment of an individual. Therefore the items collected from various sources were given to a group of five specialists for examining its relevance. The five specialists used were two clinical psychologists, a psychiatrist, and two counsellors. The items on which more than two experts differed were eliminated. Thus, the face validity of the inventory was checked by experts.

Content Validity

Content validity involves mainly the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured. Each item should be a sampling of the knowledge or performance which the test purports to measure. Taken collectively, the items should constitute a representative sample of the variable to be tested.

Content validity is most appropriately applied to tests of proficiency and of educational achievement, although such validity may be supplemented by several types of statistical analysis. As Freeman (41) says, validity of content should be based upon careful analysis by several specialists of the matter under consideration.

Content validity should not be confused with face validity. The face validity is not validity in the technical sense; it refers not to what the test actually measures, but to what it appears superficially to measure (4).

This validation procedure consisted of two parts - a) Experts' analysis of the content to be included in the test, and b) the use of various statistical procedures to refine the original selection of items (41).
This type of validity is inadequate and may be some time misleading for aptitude and personality tests as they are not based on a specified course of instruction or uniform set of prior experiences from which the test content is drawn. Therefore, the selection of representative items measuring sample of behaviour is difficult.

For obtaining content validity of the present inventory, an attempt was made to collect items from various available sources - a) subjects free writing, b) expert opinion, c) group discussion with subjects, d) personal interview of the subjects, and e) relevant studies and other interventions. Items collected from various sources were given to a panel of five experts for examining its relevance to the purpose of the inventory. The items for which more than two experts differed about its relevance were eliminated. Items approved by the experts were subjected to statistical analysis. It gave the indices of internal consistency and discrimination power. The criterion of applicability was also used. Items having satisfactory internal consistency, discrimination power and criterion of applicability were selected. These procedures demonstrated fairly high content validity.

Construct Validity

The construct validity of a test is the extent to which the test may be said to measure or assess a theoretical construct or a trait. In comparison to previously discussed types of validity, construct validity requires the gradual accumulation of information from a variety of sources. Any information throwing light on the nature of the trait under consideration are grist for this type of validity mill. Construct validity depends upon
the degree to which the test items individually and collectively sample
the range or class of activities or traits, as defined by the mental
process or the personality trait being tested.

This procedure includes various techniques, the selection of which
depends upon the trait being measured by the test under consideration.

The construct validity of the present inventory was determined on
the basis of utilization of previous researches, and on the basis of theoretical construct, which was drawn or accepted from psychological literature. Those concepts are utilized in the inventory defined in terms of operational and theoretical definition.

The present inventory was correlated with similar earlier inventories - Saxena's "Vyaktitva Parakh Prashnavali" (Appendix d), Patel's "Vyaktitva Mapasuchi" (Appendix e), Badami's Vyaktitva Mapasuchi" (Appendix f) and Gujarati version of incomplete sentence blank (Appendix h). The relevant statistics are presented in table 6.1.

Further its construct validity was determined by internal consistency. In this item - test correlation was obtained and items yielding significant correlation were selected for the final form of the inventory.

These procedures have provided satisfactory evidences about the construct validity of the present inventory.

Concurrent Validity

Concurrent validity is one of the newer terms employed by psychologists. Formerly, majority of psychologists spoke of validation
### TABLE 6.1.

**COEFFICIENT OF CORRELATION BETWEEN THE PRESENT INVENTORY AND OTHER INVENTORIES**

<table>
<thead>
<tr>
<th>Inventories</th>
<th>Coefficient of Correlation</th>
<th>N</th>
<th>Level of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saxenas' Vyaktitva</td>
<td>0.75</td>
<td>0.80</td>
<td>.001</td>
</tr>
<tr>
<td>Parakh Prashnavali</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patels' Vyaktitva</td>
<td>0.72</td>
<td>0.75</td>
<td>.001</td>
</tr>
<tr>
<td>Map Suchi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Badamis' Vyaktitva</td>
<td>0.67</td>
<td>120</td>
<td>.001</td>
</tr>
<tr>
<td>Map Suchi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incomplete Sentence</td>
<td>0.61</td>
<td>0.50</td>
<td>.001</td>
</tr>
<tr>
<td>Blank</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
with other tests, validation with a proficiency rating, validation with school grades, and for personality inventory (41) validation with a recent diagnosis. At present, the term 'concurrent validity' is preferred 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 information might have been obtained either before or after the administration of the new test (41).

The concurrent validity indicates the relation between test scores and indices of criterion status obtained at approximately the same time (4). Concurrent validity is attempted when the test is proposed as a substitute for some other information; this information is then the criterion. Designers of new tests frequently establish it for their instruments by comparing them with established tests (37). Concurrent validity differs from predictive validity. The logical distinction between concurrent validity and predictive validity is based on the objective of testing. Former is relevant to tests used for diagnosis of existing status while the later is relevant to tests used in predicting some future outcome.

This type of validation procedure utilizes several criteria. Generally, criteria information are available at the time of testing. Among the most commonly employed criteria for concurrent validation are contrasted groups, ratings, interviews, and other available similar tests.

As the main purpose of the present attempt was to develop a tool for the diagnosis of existing status, rather than prediction of future outcome, no effort was made to assess predictive validity of this inventory.
For the purpose of determining concurrent validity following criteria were employed:

a) contrasted groups,
b) ratings by teachers,
c) interviews, and
d) correlation with other similar tests.

a) Contrasted Groups

The method of contrasted groups is used quite commonly in the validation of personality tests. Validation by this method generally involves a composite criterion that reflects the cumulative and uncontrolled selective influences of everyday life. This criterion is ultimately based upon survival within particular groups versus elimination therefrom (4).

Life other personality tests, the present inventory is validated on the contrasted groups - prisoners awaiting trials and normals, and psychiatric patients and normals.

For the purpose of validation of this procedure, approximately one hundred male prisoners awaiting trial were drawn from the Central Sabarmati Prison, Ahmedabad and fifty psychiatric patients of both the sexes were obtained from private clinics in the city of Ahmedabad. The required number of normals were drawn from the post-graduate departments of the Gujarat University.

Table indicate the comparison of mean scores of prisoners awaiting trials and normals. It is observed that the average adjustment of the normal group was significantly higher than the average adjustment of the prisoners. The obtained 't' is significant at .01 level.
### Table 6.2

Comparison of Personality Mean Score of Prisoners Awaiting Trials and Normals

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>X</th>
<th>SD</th>
<th>t</th>
<th>Level of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prisoners</td>
<td>100</td>
<td>34.30</td>
<td>14.60</td>
<td>6.25</td>
<td>.01</td>
</tr>
<tr>
<td>Normals</td>
<td>100</td>
<td>21.55</td>
<td>14.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6.3 indicates the comparison of average adjustment of normals and psychiatric patients.

The significance of difference between the two means is tested by 't' test. It indicates that the average adjustment of the normals is significantly higher than the psychiatric patients.

b) Ratings by Teachers

Generally, ratings have been employed in the validation of almost every type of test. They are especially useful in providing criteria where objective criteria are much more difficult to find, e.g., personality tests. The usefulness of ratings increases when they are obtained from more than one trained raters under carefully controlled conditions (4). The accuracy of ratings can be greatly increased by the use of well constructed rating scales with clearly defined, unambiguous units and with adequate safeguards against common rating errors. While selecting the raters, care must be taken to see that the selected raters have enough 'trait acquaintance' with the individual in the traits they are rating.

The rating scale prepared by Badami (9) was used for this validation procedure. (A copy of the rating scale is given in Appendix 6.9.) Ratings for about fifty college subjects were obtained from their teachers. The mean of two ratings for each subject was obtained. Thus fifty mean scores were correlated with fifty inventory scores obtained from the same group. The relevant data is presented in Table 6.4. It indicates a positive and a significant correlation between the two.
TABLE 6.3.

COMPARISON OF PERSONALITY MEAN SCORE OF PSYCHIATRIC PATIENT AND NORMALS

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>X</th>
<th>SD</th>
<th>'t'</th>
<th>Level of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychiatric patients</td>
<td>50</td>
<td>36.80</td>
<td>15.80</td>
<td>6.91</td>
<td>.01</td>
</tr>
<tr>
<td>Normals</td>
<td>100</td>
<td>21.55</td>
<td>14.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score</td>
<td>Coefficient of Correlation</td>
<td>N</td>
<td>Level of Significance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------</td>
<td>---</td>
<td>-----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers' rating and inventory score</td>
<td>.50</td>
<td>50</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
c) Interview

Frequently personality inventory utilizes interview data for establishing concurrent validity. For the purpose of this validation procedure, about twenty-five college subjects were personally interviewed for assessing their assessment by the present author. The subjects who were willing to spare time for personal interviews were included in the sample. Semi-structured interview method was adopted for this. (A copy of the interview schedule is given in Appendix C.) The content analysis was carried out and adjustment score for each individual was calculated. Score on this inventory was obtained for each subject and was correlated with score obtained from interview data. The coefficient of correlation was calculated using rank order method. Formula for the same is given in Appendix F. Necessary statistics are given in Table 6.5. It indicates significant relationship between the two.

Correlation with other tests

Frequently, attempts are made by calculating correlations between a new test and previously available similar test to demonstrate this type of validity. When the new test is an abbreviated or simplified form of a currently available test, the latter can reasonably be considered as a criterion measure (4).

The present inventory is correlated with "Vyaktitva Parakh Prashnavali" (80) "Vyaktitva Mapsuchi" (78) and Gujarati adaptation of Incomplete Sentence Blank (24). The Pearson Product Moment Method was used for calculating coefficient of correlation between the scores of two inventories. The relevant statistics are presented in Table 6.7. They indicate that
### TABLE 6.5

**COEFFICIENT OF CORRELATION BETWEEN THE PRESENT INVENTORY AND INTERVIEW**

<table>
<thead>
<tr>
<th>Score</th>
<th>Coefficient of Correlation</th>
<th>N</th>
<th>Level of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview and</td>
<td>0.60</td>
<td>25</td>
<td>0.001</td>
</tr>
<tr>
<td>Inventory Score</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the present inventory is significantly related with these inventories.
The coefficient of correlation is comparatively low in case of Incomplete Sentence Blank and high in case of "Vyaktitva Parakh Prashnavali."

The various validation procedures employed have revealed fairly satisfactory validity of this inventory.