VALIDITY
CHAPTER VI
VALIDITY

- What is validity
- Factors influencing empirical validity
  - Test variables
  - Criteria used
  - Variability differ
  - Group differences
  - Additional information
- Types of Validity
  - Content validity
  - Face validity
  - Construct validity
  - Concurrent validity

Contrasted group
Correlation with other groups

- References
Validity is one of the most important attributes of standardized tests. Nothing will be inferred from testing unless the test has some validity for the use we wish to make of it. In brief validity of a test concerns what the test measures and how will it does so. The statement of validity reveals to us what a test actually measures what it intends to measures (6.2, 6.3, 6.21).

Generally, test names provide short convenient labels for identifying the purposes. Sometimes those names are far too broad and vague to provide meaningful insight to the behavioral areas covered. The trait measured by a particular test can be identified or determined only through an examination of the objective sources of information and empirical procedures employed in establishing its validity. Further, the validity of the test cannot be reported in general or vague terms, because no test can be said to be of high or low validity in abstract. Its validity must be established with regard to the specific use for which it is being considered. (6.1, 6.5, 6.8, 6.19)
Secondly, a test which has high validity for one purpose may have moderate validity for another, and will not be valid for a third one. There are no fixed rules for deciding what is meant by high validity, moderate validity, low validity, etc. Necessary skills in making such decisions can be acquired through training and experience (6.2, 6.8, 6.19).

Broadly speaking, all procedures for determining validity are primarily concerned with the relationship between performance on the test and other independently observable criteria or measures about the behavioral characteristics under consideration. The specific techniques and procedures employed for investigating those relationships are numerous and have been prevalent by different names. All these procedures have been classified into three principal classes: content, criterion related, and construct validity (6.2, 6.19).

The validation of a test is a very long process. In this, one must examine and analyse the content of the test and various correlations with different
criteria, in several groups, to have an adequate and appropriate information just to understand what it measures. In this sense, the correlations between the test score and criterion measures determine the validity of the test.

There are two main approaches to measure validity: logical approach and empirical approach. The logical approach consists of logical analysis of the test score to decide what it measures. The empirical approach consists of experimental studies in which test scores are correlated to various criterion measures (6.19).

Different types of validity and its procedures have been suggested by psychometricians depending upon the purpose of the test. Freeman has stated operational and predictive validity; Guilford has advocated intrinsic and face validity; Anastasi defines face validity, criterion related validity, construct and factorial validity, while Carret suggests concurrent, content, and construct validity, and Byrom suggests face validity, content validity and empirical validity (6.12).
Factors influencing Empirical Validity of a Test:

As discussed earlier, considerable skill and experience are required in interpreting test scores and their validity coefficient. In general, the higher the coefficient of correlation between test scores and criterion measures, the better; however, many other factors will have to be taken into account.

1. Test Variables Differ:

Some tests easily and naturally lend themselves validation studies than do others. E.g., examination marks are normal criterion to be employed in validating scholastic tests. While on the other hand, it is rather difficult to find out the good criterion for validating anxiety tests, where good criteria are hard to find, usually one can not expect high validity coefficients and in some cases the test itself is a better measure of the characteristic than criterion is.

2. Criteria Used:

The criteria used in validating one test may be more appropriate or relevant to the purpose than the criteria used. So criteria also differ.
3. **Variability** differs: 

Validity coefficients are likely to be higher when the variability of the group employed is higher. In much the same way, a crude test can discriminate nicely if there are gross differences among the tested, but very good test may not discriminate adequately if the group is highly homogeneous.

4. **Group differences:**

For several reasons, the test which works well with one group need not do so with different groups; e.g. the test which discriminates between the group having high intelligence than the group intellectually retarded drawn from secondary schools may be ineffective among the groups drawn from higher secondary schools.

5. **Additional information:**

Lastly, the validity coefficient of a particular test is also evaluated in term of how much additional information it will provide to the user. Because some tests correlate high with criteria, but not help us beyond that, the test will not be of high value unless it gives something new to the understanding
of the tested (6.19)

Above discussed are some of the illustrative types of factors influencing empirical validity. And that is why one can not say light away "the higher the validity coefficient, the better". The given statement regarding validity coefficient will be true provided other things are equal.

In view of the previous discussion, the present author has attempted to determine different validity coefficients of this scale. The author describes them in the following section in detail.

A Content validity:

The evaluation of the content validity of the test involves examination of its content. Mainly, it consists in ascertaining whether it covers a representative sample of the behaviour domain it claims to measure. This procedure is specifically employed in evaluating achievement test. Some times it is also used in illustrating the likely domain of tests. Here, to some extent it may be evaluated on the basis of sources from where items have been
As Freeman maintains, that the validity of content should be based upon careful analysis by several specialists of the behaviour under consideration (6.2, 6.3, 6.19).

Further, it should not be confused with face validity. The face validity is not a validity in the technical sense; it indicates not what the test actually measures, but to what it appears superficially to measure. The content validity of the present scale could be reasonably ascertained as is done in the case of the items included in the scale. The final scale was drawn from several sources such as experts, other tests and current literature.

3. Face Validity:

Face validity is not a validity in the technical sense. It indicates what the test appears superficially to measure. Mainly it indicates whether the test "looks valid" to the testees, the administrative personnel, and other technically untrained observers. Basically, the problem of face validity concerns
rappor and public relations. The general usages of
the word validity in this regard create more confusion.
If face validity itself appears irrelevant, inappropriate,
silly and childish to the person concerned the results
will be poor cooperation, regardless of the actual
quality of the test. So it is not necessary that
the test should be valid only but it also requires to have
face validity to become effective in actual practice.
According to this type of validity the test content
should appear to be relevant to its purpose. No further
test attempt is necessary in this regard. Further, face
validity should never be considered as a substitute
for objectively determined validity. The face validity
of the test be determined by objective examination
of the actual test content. (6.2, 6.9)

This type of face validity has been disapproved
as more sophisticated procedures have come into existence.
As a matter of fact, face validity in the earlier days
of test development was the criterion used by many
experts. It was frequently ascertained with tests of
educational achievement and of personality and to a
As the purpose of this scale is to determine the degree of neuroticism in an individual or a group of individuals, the items collected from various sources were given to experts in the field for examining their relevance to the purpose. The experts for checking were clinicians, clinical psychologists, and psychiatrists. The items on which more than half the experts agreed in regard to their relevance were retained for further analysis. Thus the face validity was checked by experts.

6. **Construct Validity**

The construct validity of a test is the extent to which it may be said to measure or assess a theoretical construct or trait. The construct is developed to explain and systematize observed response consistencies. The construct validity derives from established inter-relationships among several behaviour measures. Mainly it requires the gradual accumulation of information from a variety of sources.
Any data throwing light on the nature of the trait under consideration, conditions affecting its development, and manifestations are materials for determining this validity.

The present scale was correlated with existing tests—Eysenck n-scale and scale for extroversion and mental health analysis questionnaire. The data regarding this are presented in table numbers 10, 11, and 12.

The table reveal that it is significantly and positively correlated with n-scale and while it is negatively correlated with extroversion test (6.2, 6.16)

Further the construct validity was determined by maintaining internal consistency. In this item scores were correlated with the total scores. The items showing significant correlations
were retained for the final form of the scale.

(See appendix).

----------------------------------
Read Table
Number 13 Hero
----------------------------------

Lastly, the inter correlation matrix between sub scales including total scale given in table number 13 indicates that each sub-scale is positively and significantly correlated with total scale.

All these procedures suggest a reasonable construct validity of the present scale.

D. Concurrent Validity:

For the diagnosis and assessment of personality, the concurrent validity is inevitable. The concurrent validity is different from the predictive validity. They both differ not only on time, but on the objective of testing. The concurrent validation is relevant to tests employed for diagnosis and assessment of the present status, rather than prediction of future outcomes.
The concurrent validity of the present scale was determined on the basis of the following criteria:

a. Contrasted groups

b. Correlation with other similar tests.

I. Contrasted groups:

The technique of contrasted groups is used quite commonly in the validation of personality tests. For determining this validity the scale was given to the group of prisoners, neurotics, and individuals awaiting trials or final depositions either at remand home or at Vikas Griha. The total scores of each group were compared with normals. This comparison was undertaken on the assumption that the evidence of neurotic behaviour among the prisoners, neurotics and individuals selected from the remand home and Vikas Griha are likely to be more frequent and intense than the normals (6.2, 6.8, 6.19).

For the purpose of comparison a group of criminals were drawn from the prisons at Chavnaigri and Sabarmati. By and large, the prisoners were from
the group of post adolescent and adults. The number table 14 shows the comparison of mean scores on a total scale between criminals and normals.

It reveals that the obtained difference between the two groups is highly significant. The average score of the criminal group is higher than the average score of the normal (6.2, 6.5).

For the purpose of comparison of neurotics and normals the required sample of neurotics was obtained from the private clinics. The table 15 shows the comparison.

The comparison of average scores of neurotics and normals shows a significant difference between the two. The average score of the neurotic is higher than the average score of the normals.
For the purpose of comparing the individuals awaiting trials or those whose trials have not been completed, were drawn from remand home and Vikas Griha at Ahmedabad. The table number 16 shows this comparison.

The comparison reveals that the average score of the group awaiting trials or those whose trials have not been finally completed is significantly different from that of the normals. In this regard, the average score of the group drawn from remand home and Vikas Griha is higher than the normals (6.2, 6.8, 6.19).

Further the comparison of averages high and low scoring groups on this scale, data was obtained from two hundred subjects. The comparison data is given in table number 17.
It revealed that the difference between the two averages is significant at .001 level.

II Correlations with other tests:

Frequently attempts have been made to establish concurrent validity by correlating the test scores under consideration with the scores of other previously standardized tests. When the new test is an abbreviated or modified form of a currently used test then it can reasonably be accepted as a criterion measure.

In this context the present scale was correlated with Eysenck's n-scale and extroversion scale separately and Mental Health Analysis questionnaire. Table numbers 10 and 11 reveal data concerning this validity.

-------------------------------------
Read Table
Numbers 10 and 11 More
-------------------------------------

Eysenck has expressed the opinion that n-scale is negatively correlated with extroversion scale. So the author has attempted to find correlation between extroversion scale and the present scale.
The data revealed that Eysenck's scale is significantly and positively related while extroversion scale is significantly and negatively related with the present scale. This demonstrates the concurrent validity of this scale (6.4, 6.16).

Further, under the assumption that the mental health of the neurotics would be poorer than the normals, this scale was correlated with NMAQ. The table number 12 shows these results.

The validity coefficient calculated is found statistically significant. It shows considerable positive correlation between the two.

The various validity coefficients are found fairly satisfactory. They throw considerable light on the validity of the test, and they also indicate the various purposes for which this scale may be employed.


"Construction And Validation of an Adjustment Inventory for the study of Maladjustment particularly Among Adolescent" *Center of Advance Study in Education Research Monogram*

C.S. University of Baroda.

6.7 Bhatt K.K. (1971) Validation of The 188 for The Use of Gujarati Speaking Adolescent

*Vidya J. of Gujarat University* Vol. 14, 1

87-89.


"CONSTRUCT VALIDITY IN PSYCHOLOGICAL TESTS". Psychological Bulletin Vol. 52

201-302.


Statistics in Psychology and Education—Bombay: Vakils, Peffer and Simons

Private Ltd.


6.17 Patel A.S. (1965) "A Rational Construction And Tryout of Test Scale Measuring some Personality Traits (In Gujarati) Extroversion Introversion Scale". Advances in Education. M.S. University of Baroda. 1, 1, 3-17.


6.20 Sinha D. (1965) "Validation of an Anxiety Scale". J. Psychological Researches 9.18-26
