CHAPTER 4

THE RELIABILITY AND VALIDITY OF MENTAL HEALTH INVENTORY AND DATA COLLECTION

4.1 Introduction

4.2 Standardization of mental health inventory

4.3 Reliability

4.3.1 Characteristics of Reliability

4.3.2 Types of Reliability

1. Test-Retest Method

2. Internal Consistency Reliability

A. Split-Half Method

A1. Spearman and Brown Method

A2. Rulon Method

A3. Flanagan Method

4.4 Factors influencing Reliability

4.5 Validity

4.5.1 Nature of Validity

4.5.2 Types of Validity

4.5.2.1 Construct Validity

4.5.2.1 Content Validity

4.5.2.2 Face Validity

4.5.2.3 Concurrence Validity

4.6 Factors affecting Validity

4.7 Relation between Validity and Reliability

4.8 Usability of the inventory

4.9 Manual for Mental Health Inventory

4.10 Final run of the Mental Health Inventory
4.11 Conclusion

4.1 Introduction

In previous chapter item analysis of the mental health inventory has been discussed.

The main purpose of standardizing any psychological tool is to establish its reliability and validity. To carry on with the data collection a good test is required. There are certain tools of evaluation required to obtain pertinent data. We can choose any type of tool which is best suited for a particular purpose. In order to ascertain that tool is good enough to provide accurate measurements, we should know the criteria of a good tool. As per the qualities and characteristics of a good test for the measurement of sample of desired behavior, the test should ensure that how well it measures the required qualities and to what extent it intends to measure. However the following are the characteristics of a good test:

1. Reliability
2. Validity
3. Norms
4. Usability or Practibility
5. Balance
6. Efficiency
7. Difficulty
8. Discrimination
9. Speed
10. Fairness

We can classify the most common and essential criteria of the inventory into validity, reliability, Norms and usability. We may note that the result of evaluation is based on information obtained from tools of evaluation. This means one has to prepare or select a proper tool to obtain accurate results of evaluation. Reliability, Validity, Norms and Usability are described below:

4.2 Standardization of mental health inventory

Following are the steps of standardization of a test:

1. Preparing the initial test
2. Pilot test
a. Assessing the difficulty value and the discrimination value
b. Assessing the effectiveness of the alternatives
c. Assessing the similarity of different forms
d. Assessing the instructions
e. Assessing the time limit

3. Preparing the final form of the test
4. Testing the reliability and validity
5. Sample selection
6. Applying the test
7. Analysis of the data
8. Establishing the norms
9. Evaluation

4.3 Reliability

The concept of reliability relates to the question of ‘accuracy’ with which we measure ‘what’. A test must be reliable as it must have the ability to consistently yield the same results when repeated measurements are taken of the same individual under the same conditions. A test is said to be reliable if it functions in a consistent manner. A reliable test gives stable and trustworthy results. In measuring reliability, the emphasis is upon the agreement of the test with itself.

Thus, reliability can be defined as the degree of consistency between two measurements of something that is, how consistent test scores are from one measurement to another. It means that when the same test is given twice at certain interval if time to the same sample, how consistent the scores remain.

According to Anastasi and Urbina (2002)

“Reliability refers to the consistency of scores obtained by the same persons when they are reexamined with the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions.”

Some points of further clarification is given by Gronlund(1981).

1. Usually, test scores are not reliable. This means an estimate of reliability always relates to a particular type of consistency like consistency of scores over a period of
time (stability) or consistency of scores over different samples of questions (equivalence).

2. Basically, reliability is statistical in nature, which means the scores obtained on two successive occasions are correlated with each other. This coefficient of correlation is known the ‘reliability coefficient’.

3. Reliability refers to the results obtained with an evaluation instrument and not to the instrument itself.

### 4.3.1 Characteristics of Reliability

Following are the characteristics of reliability:

1. Reliability is consistency of test scores.
2. It is the measure of variable error or chance error or measurement error.
3. It is the function of a test length.
4. It refers as the stability of the test for a certain population.
5. It is self- correlation.
6. It is the coefficient of stability and internal consistency.
7. It is reproducibility of the scores.
8. It is the important characteristics of measuring instrument.
9. It refers to the accuracy or precision of a measuring instrument.

According to Hopkin reliability means, the consistency with which a test measures whatever it measures. Reliability can be measured by different methods:

### 4.3.2 Types of Reliability

It is not possible to calculate reliability exactly. Instead, the estimate of reliability is calculated. There are four general types of reliability estimates. Different methods to find the reliability of the test are as under:

- **Inter- Rater or Inter- Observer Reliability**

In this method the researcher checks how consistent the result is obtained when the test is given to different people. It is seen that whenever the humans are used as a part of measurement procedure, the researcher is worried whether the results obtained are reliable or consistent. This is because the individuals are distractible, tired of repeatedly doing the same
work, misinterpret the concept due to misunderstanding. Hence, by this method the researcher does not get the correct results so he manipulates the results to establish inter rater reliability. If the study goes for a long time, the researcher needs to re-establish inter-rater reliability from time to time to assure that the response of the sample aren’t changing. The correlation between these evaluations would give a rough calculation of the reliability or consistency between the individuals who are assessed. This type of reliability can be used to arrange the participants according to the given standards.

- **Parallel-Form Reliability**

This method is used to find the correspondence or uniformity between two tests constructed same in terms of content, form, range and level of difficulty.

In the test-retest method, the health, emotions, motivations, mental health, and mental set of the sample subject do not remain same in both the administrations. Hence it is possible that it affects the test-retest scores. The equivalent or parallel form method is an improvement over the test-retest method which overcomes the problems of memory, practice, carry over and recall factors. In this type of reliability, two parallel forms are created and thus two instruments are created. Both the instruments are administered to the same sample. In this method a large set of questions are divided into two sets. Thus the estimate of reliability is obtained by correlating the two parallel forms is the estimate of reliability. The main problem is to frame many items that agree to the same concept. The parallel forms approach is similar to the split-half reliability. The major difference is that parallel form are constructed so that the two forms can be used independent of each other and considered equivalent measures. While in split-half reliability the test is divided into two-odd half and even half.

In this method, two equivalent or parallel forms are administered to the same group of subjects and the two sets of test scores are correlated to obtain the reliability of the test. Further, it requires that the items in the two forms are not too similar. When the parallel forms are identical, reliability is too high. When parallel forms are not identical, reliability is too low. Pearson’s method of correlation is used for calculating the coefficient of correlation between two sets of scores obtained by administering the two forms of the test.

Assumptions of the Parallel Form Method

1. The memory practice and carry over factors do not affect the scores.
2. Two forms of a test are virtually alike with reference to difficulty and discrimination indexes, content and type of items, size, number of items, and time of administration.
3. Two forms of the test are equivalent.

- **Internal Consistency Reliability**

This method is used to find the consistency of results of the items within the test. There are various ways to calculate internal consistency of the test.

(A) Split Half Reliability
   1. Spearman and Brown Formula
   2. Rulon Formula
   3. Flanagan Formula

(B) Cronbach’s Alpha

(C) Method of rational equivalence
   1. Kunder Richardson – KR20
   2. Kunder Richardson – KR2

In the present study the researcher has checked the reliability of the test by the following methods:

1. **Test- Retest Reliability**

In test-retest reliability method the same test is given to the same group of people at two different times. The scores thus obtained are then correlated and compared.

The reliability coefficient in this case is simply the correlation (r) between the scores obtained by the same sample on the two administration of the test. This approach supposes that there is no firm change in the concept being measured between the two occasions. The amount of time allowed between measures is extremely important. The shorter the time gap, the higher the correlation; the longer the time gap, the lower the correlation. This is because the two observations are related over time, the lesser time the more similar are the factors that contribute to error.

It is used to assess the consistency of a measure from one time to another. Test-retest reliability is estimated when the same test is administered to the same sample on two different occasions. Since, here the same thing is measured twice, the correlation is
calculated. The correlation between the two observations depends on how much time passes between the two measurement occasions. It is the simplest method of estimating reliability index of test scores. In test-retest method, with a reasonable time gap, the single form of a test is administered twice on the same sample. In this way, two sets of scores are obtained by administering a test twice. The correlation coefficient is calculated between the two set of scores as the reliability which indicates to what extend the individuals retain their relative position as measured in terms of test score over a gap of time. Time gap of retest score should not be more than six months. This correlation coefficient is also referred as a correlation of stability.

Assumption of Test-Retest Method

1. The memory, practice and carry over will not affect the retest scores because the numbers of test items are large.
2. The growth and maturity will not affect the retest score since the innate ability of an individual remains constant.
3. Time gap of retest score should not be more than six months.
4. It is useful for speed test.

The value of “r” will always fall within the range -1 to +1. Guilford (1956) gave the informal interpretation of the value of r, as shown in the table 4.1.

<table>
<thead>
<tr>
<th>Value of r</th>
<th>Informal interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 0.2</td>
<td>Slight, almost no relation</td>
</tr>
<tr>
<td>0.2- 0.4</td>
<td>Low correlation; definite but small relation</td>
</tr>
<tr>
<td>0.4- 0.7</td>
<td>Moderate correlation; substantial relationship</td>
</tr>
<tr>
<td>0.7- 0.9</td>
<td>High correlation; strong relationship</td>
</tr>
<tr>
<td>0.9- 1.0</td>
<td>Very high correlation; very dependable relation</td>
</tr>
</tbody>
</table>

Under this method the researcher has administered the test on 100 students of standard IX of ASIA English School. Calculated the scores of the same and after the interval of 15 days the same test was administered on the same 100 students and the scores were calculated. The
scores obtained in the first administration of the test and the scores obtained when the test was administered again were correlated with each other to find an estimate of reliability index. The correlation coefficient provides a measure of stability and this index is called as coefficient of stability.

The table 4.2 below shows the reliability index.

**Table 4.2**

**Test-Retest Reliability**

<table>
<thead>
<tr>
<th>Test-Retest Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>Rtt</td>
</tr>
<tr>
<td>( S_{Er} )</td>
</tr>
</tbody>
</table>

In the above table 4.2, it is shown that reliability index is 0.98. Considering the values of \( r \) in the table 4.1, it can be interpreted that the correlation between the test and retest of the mental health inventory is high, indicating a strong relationship. Hence the inventory is reliable.

2. **Internal Consistency Reliability**

Test- retest reliability method has the disadvantage that it is time consuming. The researcher will want to find the reliability from single administration of a test. Hence, the other method i.e. internal consistency reliability is used. Internal consistency reliability represents the consistency with which items on an instrument provide similar scores. Internal consistency reliability defines the consistency or uniformity of the results expressed in a test, assuring that the different statements measuring the different concepts gives compatible scores.

In the present study to estimate the internal consistency reliability of the mental health inventory Split- Half technique is used.

The statistical analysis for Split- half reliability is done using the following formulas:

- Spearman and Brown Formula
- Rulon formula
- Flanagan formula
A. **Split-Half Method**

In this method, the test is divided into two equal halves, and the scores of half the items are correlated with the scores of the other half. The items of the test can be divided into two sets in a variety of ways. From the reliability of the half test, the self-correlation of the whole test is estimated by the Spearman-Brown formula. This method measures the internal reliability of the test. If the two halves do not correlate highly, it suggests that they are not measuring the same thing. Since, it can be divided into two halves in a number of ways, the correlation between the scores on the two halves may not have a unique value which is a point of criticism. It least bothers about memory, carry over, practices, skills, maturity factors i.e. it has the advantage of controlling fatigue, practice effects etc.

When the parallel forms of a test are not advisable and repletion is not possible, this method is employed to compute reliability of test. This method is useful for attitude scale, personality, inventories, questionnaire, intelligence, performance test, e.g. picture completion, and puzzle solving Rorschach test. Further, since all data for computing reliability are obtained on one occasion, the variations brought about by difference between the two testing situations are eliminated. This method is limited to the chance errors which may affect scores on the two halves of the test and tends to make the reliability index too high. Further, the test can be divided into two parts in a number of ways, so that the reliability coefficient is not a unique value. Again, it is not possible to divide the test into two halves in such a way that both halves are equivalent. This method cannot be used in power tests and heterogeneous tests. Also, it is not possible to split the test items in two equivalent forms because items of a test measure the same trait or ability.

**Assumptions**

1. It provides the internal consistency of a test score.
2. The test is homogenous and all the items of the test measure the same trait or ability.
3. All the items of the test bear same difficulty value.
4. The two halves are equivalent.
5. The assumptions of Spearman-Brown prophecy formula also apply to this method.
6. Linearity, the assumptions of Pearson’s method is also applied.
In this method the statements of the inventory were divided into two halves. The odd- even split was made. According to that the odd numbered items of the test constituted one half of the test and the even numbered items formed the second half. After the two halves were formed, a score for each half was determined taking 100 students. The scores obtained were used for estimating reliability coefficient.

The coefficient was obtained by correlating the scores of both the halves using Spearman and Brown, Rulon and Flanagan formulas.

A1. Spearman and Brown Formula

The Spearman and Brown formula was designed to estimate the reliability of a test n times as long as the one which we know a self-correlation. From the reliability of the half test, the self-correlation coefficient of the whole test is estimated by applying Spearman and Brown formula:

\[
 r_{tt} = \frac{2r_{hh}}{1 + r_{hh}}
\]

Where, \( r_{tt} \) = Reliability of a test estimated from reliability of one of its halves

(Reliability coefficient of the whole test)

\( r_{hh} \) = Self correlation of a half test (Reliability coefficient of the half test)

The calculation was done using SPSS software and the reliability coefficient was 0.93.

The table 4.3 below shows the reliability index:

### Table 4.3

<table>
<thead>
<tr>
<th>Split-Half Reliability</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>
In the above table 4.3 it is shown that reliability index is 0.93. Considering the values of \( r \) in the table 4.1, it can be interpreted that the correlation between both the half tests of the mental health inventory is high, indicating a strong relationship. Hence the inventory is reliable.

**A2. Rulon Method**

It is an alternate method for finding Split- Half reliability and it was developed by Rulon. It requires only the variance of differences between each student's scores on the two half-test (SD\(_d^2\)) and variance of total scores (SD\(_x^2\)); These two values are substituted in the following formula:

\[
\hat{r}_{tt} = 1 - \frac{SD_d^2}{SD_x^2}
\]

Where, \( \hat{r}_{tt} \) = Reliability of the test

\[SD_d = SD \text{ of difference of the scores}\]

\[SD_x = SD \text{ of the scores of the whole test}\]

The calculation was done using SPSS software and the reliability coefficient was 0.93.

The table 4.4 below shows the reliability index:

<table>
<thead>
<tr>
<th>Table 4.4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rulon Method</strong></td>
</tr>
<tr>
<td>( r_{tt} )</td>
</tr>
<tr>
<td>( S_{Er} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rulon Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>SDd</td>
</tr>
</tbody>
</table>
In the above table it is shown that reliability index is 0.93. Considering the values of $r$ in the table 4.1, it can be interpreted that the correlation between both the half tests of the mental health inventory is high, indicating a strong relationship. Hence the inventory is reliable.

**A3. Flanagan Method**

Flanagan also gave a parallel formula for finding reliability using split half method. He gave the following formula:

$$r_{tt} = 2 \times \left( 1 - \frac{SD_1^2 + SD_2^2}{SD_t^2} \right)$$

Where, $r_{tt} =$ Reliability of the test

$SD_1 = $ SD of the scores on $1^{st}$ half

$SD_2 = $ SD of the scores on $2^{nd}$ half

$SD_t = $ SD of the scores of whole test

The calculation was done using SPSS software and the reliability coefficient was 0.94.

The table 4.5 shows the reliability index:

<table>
<thead>
<tr>
<th>Flanagan Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>SD1</td>
</tr>
<tr>
<td>SD2</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>
In the above table 4.5 it is shown that reliability index is 0.94. Considering the values of r in the table 4.1, it can be interpreted that the correlation between both the half tests of the mental health inventory is high, indicating a strong relationship. Hence the inventory is reliable.

### 4.4 Factors influencing the reliability of test scores:

There are some intrinsic and extrinsic factors which affect the reliability of the test scores:

a. **Length of the test:** The reliability of the test increases with its length.

b. **Speed:** In a speed test, reliability will be problematic. This is because every student cannot complete all of the items in a speed test. In contrast, a power test is a test in which every student is able to complete all the items.

c. **Group Homogeneity:** The test is more reliable if the group of students on which the test is administered is more heterogeneous.

d. **Item Difficulty:** The test items should have certain difficulty level so as to maintain the reliability of the test i.e the items of the test should not be very easy or very hard.

e. **Objectivity:** Objective test will have higher reliability compared to subjective test.

f. **Variation with the testing situation:** Deviation during the administration of the test such as noise level and distraction can cause test scores to vary, which may affect the reliability of the test.

### Table 4.6

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Method of Reliability</th>
<th>Sample (N)</th>
<th>r tt</th>
<th>S Er</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test- Retest method</td>
<td>100</td>
<td>0.98</td>
<td>0.004</td>
</tr>
<tr>
<td>2</td>
<td>Split Half method</td>
<td>100</td>
<td>0.93</td>
<td>0.014</td>
</tr>
<tr>
<td>3</td>
<td>Rulon method</td>
<td>100</td>
<td>0.93</td>
<td>0.014</td>
</tr>
<tr>
<td>4</td>
<td>Flanagan method</td>
<td>100</td>
<td>0.94</td>
<td>0.012</td>
</tr>
</tbody>
</table>
From the above table it shows that the reliability of the mental health inventory is from 0.93 to 0.98, which shows that mental health inventory has high reliability. The same is shown graphically in graph 4.1

**Graph showing Reliability Coefficients of Mental health Inventory**

<table>
<thead>
<tr>
<th></th>
<th>0.98</th>
<th>0.97</th>
<th>0.96</th>
<th>0.95</th>
<th>0.94</th>
<th>0.93</th>
<th>0.92</th>
<th>0.91</th>
<th>0.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Graph 4.1**

### 4.5 Validity

In education and psychology, no test is perfectly valid because measurement is indirect. It is never possible to ascertain that a test measures the precise ability or trait for which it has been designed. Validity is a matter of degree considered in terms of categories such as higher, moderate, and low validity. Validity pertains to the results of a test or evaluation instrument and not to the instrument itself. It is actually the validity of the test results or more specifically the validity of the interpretation to be made from the results and not the validity of a test. The validity of the test is obtained in relation to the specific use of the test. The validity of a test cannot be described in usual manner. The validity provides a direct check on how well the test fulfills its functions. Mostly, it connotes the extent to which the test satisfies its purpose. The test user wishes to introduce the intensity to which the individual possesses some hypothetical trait or capacity or quality (construct) presumed to be shown in the test achievement. It must be noted that each tool of evaluation has some specific purpose. Thus
while choosing a tool; we must look what is the purpose for which it is being used. This means any test used to measure any information should be valid. Validity is therefore an indispensable characteristic of the test. A valid test means it must measure, what it purports to measure. It can also be referred as the truthfulness of the test.

According to Freeman (2006)

“An index of validity shows the degree to which a test measures what it purports to measure, when compared with accepted criteria.”

Therefore, the concept of validity of a test is mainly a concern for the ‘basic honesty’ of the test- honesty in the sense of doing what one promises to do. In short, validity refers to how well a tool measures what it intends to measure in a test.

### 4.5.1 Characteristics of Validity

Following are the characteristics of validity:

1. It is an important characteristic of a measuring instrument or a test.
2. It is a measure of constant error while reliability is the measure of variable error.
3. It is an index of external correlates. The test scores are correlated with external criterion scores.
4. The criterion may be a set of operations or success in the job or as a predictor for future course of test scores.
5. It relates the objective of a test score.
6. It connotes the psychological construct of a variable which is indirectly measured with the help of behaviours.
7. It ensures the reliability of a test. Thus, if a test is valid, it must be reliable.
8. It refers the truthfulness of a test score.
9. It is the function of a length.

It determines how an individual performs, in different situations.

### 4.5.2 Nature of Validity:
There are certain cautions that one has to keep in mind while using the term validity in evaluation:

1. Validity is a matter of degree and hence does not exist on an all- or- none basis.
2. Validity refers to the results of a test or evaluation tool for a given group of individuals, not the tool itself.
3. Validity being a relative term, a tool would be valid for a particular situation. This means a particular tool is not valid in every situation.

Validity can be measured by different methods.

### 4.5.3 Types of validity

The basic types of validity which have been commonly used in educational and psychological measurement are as under:

1. Construct validity
2. Operational or Content Validity
3. Face validity
4. Translation validity
5. Criterion-related validity
6. Predictive validity
7. Concurrent validity
8. Convergent validity
9. Discriminant validity

#### 4.5.3.1 Construct validity:

It refers to the degree to which conclusions can really be made from the procedures of observation and measurement of the study to the theoretical constructs or ideas on which those procedures were based. Construct validity is related to external validity as both goes from specific to general. But the difference is that external validity generalizes from study context to other people, places or time, while construct validity involves generalization of methods or measures to the concept of the methods or measures.

In this method the procedures used for the measurement exactly reflects its construct. It assumes that the researcher has a clear explanation of the construct and he can also check the methods used against it. This relational approach of criterion validity to construct validity
assumes that the methods used should function in predictable ways in relation to other methods used based upon the theory of the construct.

In comprises of two broad categories such as theory and the observation. The theory is what goes on inside mind and the attempt to explain or articulate it to others. It is all of the ideas, theories, hunches and hypotheses. The theory lets the idea of the program or treatment as it should be and one finds the idea or construct of the outcomes or measures that is believed to affect. The observation consists of seeing the happenings around and the public manifestations. Here one finds the actual measures or observational procedures. Probably, the observation is based on the theories.

Construct validity is an assessment of how well one translates the ideas or theories into actual programs or measures. It can be viewed as a truth in labeling. It is representation of the concept. The idea of construct validity can be viewed as defintionalist perspective and the relationist perspective. According to the perspective of the person who defines construct validity assures that it should define the construct precisely. In his view, the researcher either operationalizes the construct correctly or doesn’t. It is an either-or type of thinking.

Devoid of relationalist perspective concepts are more or less related to each other. The meaning of terms or constructs differs relatively but not absolutely. The teaching style or type of program can be more or less like other teacher or other program. The particular measure might be capturing a lot of construct of self-confidence but it may not capture all of it. Relationalism suggests that meaning changes gradually. It rejects the idea that to rely on operational definitions is the basis for construct definition. In order to establish construct validity the following conditions are to be met:

1. To set the construct that is to be operationalized (e.g., self-esteem) with meaning. Hereby it is declared about what the construct is more or less similar to in meaning.
2. To enable the evidence of the control over the operationalization of the construct. Hereby it is tried that the operationalization looks like what they should theoretically look like. If self-esteem is to be measured, one needs to explain why the construct has been operationalized in the way it has been done.
3. To provide evidence that the data of the relations among the constructs supports the theoretical view. For eg: It can be said that self-esteem is closer in meaning to self-concept than it is to self-confidence. One should be able to reflect that measures of
self-esteem are more highly correlated with measures of self-concept than with self-confidence.

Construct validation means interpretation of a test in terms of numerous research findings. Construct validity of a test refers to the extent to which the test measures, a particular characteristic, of the individual. If a test is valid from the construct point of view, it can indicate the actual achievement of instructional objectives. Construct validity may be ascertained by the method of factorial analysis. This validity has special reference to the areas of abilities and personality. Experts working in these areas have demonstrated that each of the above areas may be reduced to statistical elements called factors.

4.5.3.2 Content Validity:

It is the process of matching the test items with the instructional objectives. It is non-statistical type of validity. Content validity may be defined as the extent to which a test measures a representative sample of the subject matter content and the behavioral changes under consideration. Lenon says, it is axiomatic that the content validity of a test must always be viewed in relation to a particular objective to be assessed. The measure is represented subjectively after a careful process of inspection comparing the content of the test with the objectives of the course of instruction.

4.5.3.3 Face Validity:

To improve the quality of face validity the researcher needs to do the assessment more systematically. For instance, if the researcher is trying to assess the face validity of a research ability measure, it would be more convincing if the researcher sends the test to a carefully selected sample of experts on research ability testing. They will undergo the review of the same and send their report back with the judgment that the measure appears to be a good measure of research ability. In case any rectification is to be done the researcher will follow the same.

Face validity means that the given test appears or seems to measure what it is to measure. Fundamentally, the question of face validity is not one of validity in the usual sense but rather one of rapport and public relations. If the test content appears irrelevant or inappropriate, the result will be poor regardless of the actual validity of the test. To check the content validity of
the test a careful and critical examination of the test items in relation to the objectives should be made. It may be said that face validity is not to what the test measures, but what the test ‘appears to measure’. The contents of the test should not appear to be inappropriate, irrelevant. It is essentially based on the judgment of several experts of the subject and the test specialist.

In the present study the researcher gave the test to different experts of the subject. They rectified the items of the test and gave suggestions. The researcher made necessary changes as per the suggestions given by the experts.

The list of panel of experts is as under:

1. **Prof. Sonal Thareja**
   Principal (off g.) JG College of Education
   (Affiliated to Gujarat University)
   M.Com. M.Ed (Gold Medalist) NET (Education)

2. **Dr. Dipiti Bhatt**
   Prof. JG College of Education (PG) (affiliated to Gujarat University)
   Msc. M.Ed.Ph.D

3. **Dr. M Milind Mistry**
   Prof. Department of Education. (Gujarat University)
   M.A, M. Ed., Ph.D

4. **Dr. Ashwin B. Jansari**
   Head and Hon.Director(PGDCP), Psychology Department,
   School of Psychology Education and Philosophy,
   Gujarat University, Ahmedabad.

5. **Dr. Harikrishna A. Patel**
   Asst. Professor, Dept. of Education, KSV, Gandhinagar.

**4.5.3.4 Criterion-related Validity**
In criterion-related validity, the individual’s performance may be compared and judged against some criterion. In content validity, the criteria are the definitions of the construct itself and it is a direct comparison whereas in criterion-related validity, the researcher usually makes a guess about how the individual performs based on the theory of the construct. The different types of criterion related validity are based on different criteria the researcher uses as the standard for judgment. They are as follows.

**Predictive Validity**

In predictive validity, the ability of the measures to predict something that it should theoretically be able to predict is assessed. The predictive validity of a test is determined on the basis of an established criterion. A test has predictive validity, if scores on it predict future performance. The test is administered and scores are obtained with the criterion and the scores of the test are compared and correlated with the criterion.

**Convergent Validity**

The convergent validity tells us about the extent to which the methods of observation or measures are similar to the other methods or measures. It explains the observed facts or phenomenon. It can be explained by an example: one can correlate the scores of achievement with scores on other achievement test that purport to measure achievement. The high correlation would be evidence of convergent validity.

**Discriminant Validity**

The term discriminant itself explains that discrimination means differentiation. This validity examines the extent to which the methods of observation or measures differ from the other methods of observation or measures which explains that observed fact or phenomenon should not be similar to each other. For instance, to show the discriminant validity of a training program, one can collect that the proof which shows that the program is not similar to other developmental programs which are not the training programs or, to check the discriminant validity of a test explain about motor skills, one can correlate the scores of the test with scores on tests of cognitive skills. Low correlations give the evidence of discriminant validity.

**Concurrent Validity:**
Out of all these Criterion related validity the researcher has used Concurrent validity method to check the validity of the mental health inventory.

In concurrent validity, the methods of observation and measurement used to distinguish between groups assessed theoretically should be able to distinguish the groups. A test has concurrent validity when test gives an estimate of certain performance. The procedure is to administer the test and obtain scores. Then scores from some other performance are obtained and these two may be correlated. It means correlating the test scores with another set of criterion scores. This criterion is provided by a data-base of learner performance obtained on a test whose validity has been pre-established. The term concurrent may have the following meanings:

a. The two tests (the test held and its validity test) should cover the same content area at a given level and the same objectives.

b. The performance data on both the tests are obtained almost simultaneously.

c. The population for both the test remains the same and the two tests are administered in almost similar environment.

To determine the concurrent validity of the mental health inventory, the correlation coefficient between mental health inventory prepared by the researcher and the mental health inventory prepared by Dr. Jagdish, Deptt. of Psychology, R. B. S. College, Agra and Dr. A. K. Srivastava, Deptt. of Psychology, Banaras Hindu University, Varanasi was found. Both these test were administered on 100 students each, keeping the gap of one day.

Correlation coefficient between the scores of both the mental health inventories was calculated. It was 0.9.

The table 4.7 shows the validity of both the test:

<table>
<thead>
<tr>
<th>Table 4.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity</td>
</tr>
<tr>
<td>Validity by concurrent method</td>
</tr>
</tbody>
</table>
From the above value of $r$ it can be interpreted that the test prepared by the researcher is valid.

### 4.6 Factors affecting the Validity

The validity of an evaluation tool can be influenced by a large number of factors. According to Gronlund (1981), the following factors influence the validity of the test:

1. **Factors in the test itself:** As we know that the test contains the items, we should check whether the test appears to measure the subject matter content and the mental function to be tested. The following factors lower the validity of the test items:
   
a. **Unclear Direction:** The students should be given the correct and clear instructions regarding the test items. If that is not done the validity of the test is affected.

b. **Reading Vocabulary and Sentences structure which are too difficult:** The test validity is lowered if too difficult sentences and too complicated words are put in the test.

c. **Inappropriate level of difficulty of the test items:** An inappropriate level of difficulty in the test items affects the validity of the tool.

d. **Poorly constructed test items:** The test items that provide the unintentional clues to the answer will also affect the validity.

e. **Improper arrangement of items:** If the test items are arranged in the order of difficult to easy pattern, the students will fail to reach the easy items as they will leave the test incomplete. Moreover, this arrangement will affect the validity of the test.

f. **Test too short:** In case the test is too short to become the representative of the course, then validity will be affected.
g. **Ambiguity**: Ambiguity leads to confusion and misinterpretation which will lower the validity of the test since a good student might get a wrong answer while the poorer can get the answer right.

2. **Functioning Content and Teaching Procedures**: The test is valid only if the test items function as intended.

3. **Factors in test administration and scoring**: These factors may affect teacher made test. For instance, factors like unfair help to individual students, insufficient time to complete the test, cheating and the unreliable scoring of easy type answers might lead to lower the validity of the test.

4. **Factors in pupil’s response**: The factors like the emotionally disturbed students, lack of students’ motivation, any child afraid of the test situation may not respond the test items normally affecting the validity of the test.

5. **Nature of the group and the criterion**: There are certain factors like age, sex, ability level; educational background and cultural background influence the test measures. Hence, the nature of the validation group should be mentioned in the test manuals. Moreover, the nature of the criterion used should be considered important while evaluating validity coefficient.

### 4.7 Relation between Validity and Reliability:

There is a close relationship between validity and reliability. In fact, they are the two dimensions of the same thing called test efficiency. While reliability is concerned with the stability of the test scores-self correlation of the test, validity is the correlation of the test with some outside independent criteria. Thus, a test which possesses poor reliability is not expected to yield high validity.

Hence, we can say that reliability is the prerequisite of validity. A high reliable test is always a valid measure of some function. This means reliability controls validity. Moreover, a test may be theoretically valid, but may be practically invalid as judged against its correlation with different independent criteria.

### 4.8 Usability of Mental Health Inventory

Usability means the degree to which the tool of evaluation can be successfully used by the educational practitioner. Apart from the three main criteria of a good test tool- Validity, reliability and objectivity- usability or practicability is another important characteristic. Thus
certain practical considerations such as ease of administration and scoring, ease of interpretation, availability of comparable forms and cost of testing is to be kept in mind while selecting a tool. Such practical considerations are referred to as the ‘usability’ of a tool of evaluation. It ensures successfully use of a tool by the school administrations and teachers.

The present mental health inventory has good practical use. It can be used to know the mental health of IX standard students.

After standardization of the mental health inventory the researcher has prepared the manual of the final tool that gives the entire information of the inventory. The manual is as under:

4.9 Manual for Mental Health Inventory

Introduction

According to the present study mental health is not only the absence of diseases such as schizophrenia and other disorders, but the ability to cope with the life’s stressors, free from any kind of anxieties, mingle with others and have positive attitude towards their own life. The present mental health inventory has been constructed to check the positive mental health of an individual. That includes the healthy and positive thinking, good social interaction and free from mental illness.

Many psychologists have done research on mental health and different people have given different indicators to check the mental health level of an individual. In this inventory researcher has used the following indicators to know the mental health of the individuals. The researcher has considered five aspects of positive mental health. They are:

1. Introspection

2. Physical growth

3. Cope with stress

4. Autonomy
5. Environmental Mastery

**Different factors included for mental health study:**

The researcher has considered five aspects to check the positive mental health of an individual. Further the researcher has divided these five aspects into five sub topics under each main topic. They are as under:

(1) Introspection
   
   a. self concept
   
   b. how I feel about myself
   
   c. my attitude
   
   d. self analysis
   
   e. self regulation

(2) Physical growth
   
   a. Pace of physical growth
   
   b. Proportion
   
   c. energy quotient
   
   d. Physical fitness
   
   e. Physical appearance

(3) Cope with Stress
   
   a. stress causing factors
   
   b. coping ability
   
   c. my relation with stress and coping ability
   
   d. adjustment
   
   e. my reflection and stress
Mental health inventory has been constructed taking all the above aspects of mental health.

Hence in the present study mental health inventory is the instrument developed to check the five aspects of psychological health of an individual which would help the researcher to know whether the individual is contented in his life and how is his behavior with the people around and this would help the researcher to judge the mental health status of an individual.

Operational definition of five factors of mental health is as under:

1. **Introspection**: It includes how one can look into one’s own self and analyse the behavior by active observation of one’s own self.

2. **Physical Growth**: It includes how an individual grows at his own pace and accepts the undergoing changes during the process of his growth and development.

3. **Cope with Stress**: It includes how person carries himself under tremendous strain, tension and problematic situation.
4. **Autonomy:** It includes how one can take his own decisions, act independently and work freely.

5. **Environmental Mastery:** It includes the efficiency of an individual in managing a complex environment, adjusting with others and managing the responsibilities of daily life.

**Description of the inventory:**

MHI contains 80 items related to five dimensions of mental health as given above. The respondents are requested to mark tick to one option- ‘Always’, ‘Often’, ‘Sometimes’, ‘Rarely’ and ‘Never’. The respondents are asked to fill the details given in the beginning of the inventory. They are requested to read the instructions carefully before they respond to the questions in the inventory.

The time limit for this tool given was approximately 40 minutes.

**Administration of the inventory:**

MHI was administered on the students of different schools. They were made clear about the tool. They were explained about how their response to each item was important to the researcher. They had to tick in any one cell indicating the five alternatives, based on their mental health perception. Students should not discuss their responses with their friends. They should be given the assurance about the confidential nature of their responses.

**Development of the inventory:**

For the construction of the test items the contents of the statements were mainly taken from the available literature of mental health, particularly including Maslow and Mittel Mann (1951); Fromm (1955); Jahoda (1958) and Buch (1972). Some items were also framed on the basis of psychological state of an individual in different conditions.

In the beginning 125 items, each factors consisted of 25 items. These items were rectified by the experts of the subject. Pre- pilot test and pilot test was carried out and after item analysis, finally 80 items were selected for the final test. Out of these 80 items 55 were positive statements and 25 were negative statements. Finally data was collected using that prepared mental health inventory.

**Reliability of the inventory**
The reliability of the inventory was determined by Test Re-test method, split-half method, rulon method and Flanagan method.

The Table 1 below gives the reliability coefficients of different dimensions of mental health and over all.

### Table 1

**Reliability Coefficients of Mental health Inventory**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Method of Reliability</th>
<th>Sample (N)</th>
<th>r (_{tt})</th>
<th>S (_{Er})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test- Retest method</td>
<td>100</td>
<td>0.98</td>
<td>0.004</td>
</tr>
<tr>
<td>2</td>
<td>Split Half method</td>
<td>100</td>
<td>0.93</td>
<td>0.014</td>
</tr>
<tr>
<td>3</td>
<td>Rulon method</td>
<td>100</td>
<td>0.93</td>
<td>0.014</td>
</tr>
<tr>
<td>4</td>
<td>Flanagan method</td>
<td>100</td>
<td>0.94</td>
<td>0.012</td>
</tr>
</tbody>
</table>

From the above table it shows that the reliability of the mental health inventory is from 0.93 to 0.98, which shows that mental health inventory has high reliability.

**Validity of the inventory**

The face validity is obtained by giving the inventory to the experts of the subject.

To determine the concurrent validity of the mental health inventory, the correlation coefficient between mental health inventory prepared by the researcher and the mental health inventory prepared by Dr. Jagdish, Deptt. of Psychology, R. B. S. College, Agra and Dr. A. K. Srivastava, Deptt. of Psychology, Banaras Hindu University, Varanasi. Both these test were administered on 100 students each, keeping the gap of one day.

Correlation coefficient between the scores of both the mental health inventory was calculated. Correlation coefficient between the scores of mental health inventory prepared by
Scoring the responses to inventory items:

As discussed above 80 items were kept in the final tool out of which 55 were positive statements and 25 were negative statements. Positive statements were assigned the marking as: 5 marks for ‘Always’, 4 marks for ‘Often’, 3 marks for ‘Sometimes’, 2 marks for ‘Rarely’ and 1 mark for ‘Never’. While for negative statements marking was done in the reverse way, i.e. 1 marks for ‘Always’, 2 marks for ‘Often’, 3 marks for ‘Sometimes’, 4 marks for ‘Rarely’ and 5 mark for ‘Never’. Maximum score of the inventory is 400 and the minimum score of the inventory is 80. Interpretation of scores of mental health inventory is mentioned as follows:

<table>
<thead>
<tr>
<th>Scores</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>80- 160</td>
<td>Low mental health</td>
</tr>
<tr>
<td>161- 240</td>
<td>Medium mental health</td>
</tr>
<tr>
<td>241- 320</td>
<td>High mental health</td>
</tr>
<tr>
<td>321- 400</td>
<td>Very high mental health</td>
</tr>
</tbody>
</table>

Use of Mental Health Inventory

The inventory may be used for measuring mental health of individuals in various sphere of life. The present test is used to determine the mental health of the adolescent age group students. It is easy to administer, easy to score and easy to interpret. After knowing the poor mental health of an individual, one can be subjected to counseling and can be helped to come out of their troubles and lead a healthy life.

4.10 Final run of the inventory
Finally the inventory was ready with 80 statements. The researcher planned to go for final run of the inventory. For the administration of the inventory, the researcher went to the schools to take principals’ permission and requested them to allow researcher to collect data from the IX standard students of their school. After getting the permission, the researcher met the students and developed rapport with them and gave them complete instructions about how to respond to the statements of the inventory. The researcher took 10 minutes to give the instructions and the students were given 30 minutes to fill in their responses. Thus data was collected from 1600 students from 20 different schools.

4.11 Conclusion

In this chapter the researcher has calculated the reliability and the validity of the inventory to prove that the prepared mental health inventory is reliable and valid. Keeping in mind the variables of the study such as gender, area and type of Board, the norms of the inventory have been constructed using Percentile Rank and ‘T’ method. The table for the same is attached in the appendix IV. The researcher has also discussed about the usability of the inventory. In the next chapter the researcher would discuss about analysis and the interpretation of the data.
References


