Introduction to Psychological Testing

Sharanjit Kaur
Assistant Professor
Dept of Psychology
Hans Raj Mahila MahaVidyalaya
Jalandhar

Meaning of Psychological Tests

A **test** has been defined as a series of questions on the basis of which information is obtained.

*A **psychological test** is a standardized procedure to measure quantitatively or qualitatively one or more than one aspect of a trait by means of sample of behavior.*

There are **two purposes of psychological testing**:

- Comparing the same individual on two or more than two aspects of a trait.
- Two or more than two persons can be compared on the same trait.

According to **Bean (1953),** "a test is an organized succession of stimuli designed to measure quantitatively or to evaluate qualitatively some mental process, trait or characteristic."

Characteristics Of Psychological Tests

1. It is an **organized succession of stimuli** i.e. the items in the test are organized in a certain sequence and are based upon some principles of test construction.
2. Both **quantitative and qualitative measurements** are possible with the help of psychological tests.
3. A psychological test is based upon a **limited sample of behavior**. This means that any psychological test does not assess the totality of a person's behavior. Rather it focuses on the limited aspect of that behavior.
4. A psychological test usually provides scores or categories which are then interpreted with reference to a **standardization sample**. The standardization sample should be representative of the population for whom the test is meant for. Such psychological tests are known as **norm-referenced tests** which means that the results from such tests are interpreted with reference to the standardization sample. Some tests are **criterion-referenced** which are used to determine where an examinee stands in reference to a defined criteria or standard.

Classification of Tests

- **On the basis of the criterion of <u>administrative conditions</u>**

✓ **Individual tests:** These are those tests that are administered to one person at a time. Individual tests ore often used by school psychologists and counsellors to motivate children and to observe how they respond. Some individually administered tests are given orally, and they require the constant attention of the examiner. E.g. Kohs Block Design Test

Limitations:

Such tests are time-consuming

Such tests require the services of trained and experienced examiners.

✓ **Group tests** are tests which can be used among more than one person or in a group at a time. E.g. Bell Adjustment Inventory

- **On the basis of the criterion of <u>scoring</u>**

- ✓ **Objective tests:** These are those tests whose items are scored by competent examiners or observers in such a way that no scope for subjective judgment or opinion exists and thus, the scoring remains unambiguous. Tests having multiple-choice, true-fake and matching items are usually called objective tests. In such items the problem as well as its answer is given along with the distractor. The problem is known as the stern of the item. A distracter answer is one which is similar to the correct answer but is not actually the correct one. Such tests are also known as new-type tests or limited-answer tests.

- ✓ **Subjective tests:** These are the tests whose items are scored by the competent examiners or of in a way in which there exists some scope for subjective judgement and opinion. As a consequence, some elements of vagueness and ambiguity remain in their scoring. These are also called essay tests.

- **On the basis of the criterion of <u>time limit in producing the response</u>**

- ✓ **Power tests**: These are those tests in which there is a generous time limit so that most examines are able to complete every item of the test.in such tests, usually the items are arranged in an increasing order of difficulty. E.g. most of the intelligence tests. Such tests demonstrate how mush knowledge an examinee has.

- ✓ **Speed tests**: These are those tests which have a severe time limit but the items are comparatively easier and the difficulty level of all the items is more or less the same. Such tests reveal how rapidly the examinees can respond within a given time limit. E.g most of the clerical aptitude tests.

- **On the basis of the criterion of the <u>nature or contents of items</u>**

- ✓ **Verbal tests**: A verbal test is the one whose items emohasize reading, writing and oral expression as the primary mode of expression. The instructions ore printed or written. These are read by the examinees and the items are answered accordingly. E.g. Jalota Group General Intelligence Test.

- ✓ **Nonverbal tests**: These are those tests that emphasize but don’e altogether eliminate the use of language by using symbolic materials like pictures, figures, etc. language is used in giving instructions but the items don’t use language. Such tests are commonly used with young children. E.g. Raven progressive matrices.

- ✓ **Performance tests**: These are those tests which require the examinees to perform a task rather than answer some questions. The use of language in items is prohibited. Occasionally, oral language is used in giving instruction or the instructions may also be given in the form of gestures. Such tests are usually administered individually so that the examiner can count the errors and the time taken by the examinee in the performance of the test.

- ✓ **Nonlanguage tests**: These are those tests which don’t depend upon any form of written, spoken or reading communication. Such tests are completely independent of the ability to use language in any way. The instructions are usually given through gestures. The examinees respond by pointing at or manipulating objects such as pictures, blocks, puzzles, etc.

❑ **On the basis of the criterion of <u>purpose or objective</u>**

- ✓ **Intelligence tests**: Such tests intend to assess the intelligence of the examinees.
- ✓ **Aptitude tests**: Such tests assess potentials or aptitudes of the persons.
- ✓ **Personality tests**: Such tests assess traits, adjustments, interests, values, etc.
- ✓ **Achievement tests**: Such tests assess what the persons have acquired in the given area as a function of some training or learning.

❑ On the basis of criterion of standardization

✓ Standardized tests

Standardized tests are those which have been subjected to the procedure of standardization which includes at least the following conditions:

- The first condition for standardization is that there must be a standard manner of giving instructions so that uniformity can be maintained in the evaluation of all those who take the test.
- The second condition for standardization is that there must be uniformity of scoring and an index of fairness of correct answer through the procedure of item analysis should be available.
- The third condition is that reliability and validity of the test must be established and the individuals for whom the-test is intended should be explicitly mentioned.
- The fourth condition, a controversial one, is that a standardized test should have norms. According to Cronbach (1970) a test even without norms may be called a standardized test. But the majority of psychologists favour the idea that a standardized test should have norms as well.

These are the tests which are administered and scored under standard and uniform testing conditions so that the results obtained from different samples may be compared. Items of standardized tests are fixed and cannot be modified.

✓ **Teacher-made tests**

- These tests are those that are constructed by teachers for use largely within their classrooms.
- The effectiveness of such tests depends upon the skill of the teacher and his knowledge of test construction.
- Items may come from any area of curriculum and they may be modified according to the will of the teacher.
- Rules for administration and scoring are determined by the teacher.
- Such tests are largely evaluated by the teachers themselves and no particular norms are provided; however, they may he developed by the teacher for his own class. Thus we find that the tests have been classified in terms of various criteria.
- These tests are used for a variety of purposes.

CHARACTERISTICS OF A GOOD TEST

1. **Objectivity**

- A test must have the trait of objectivity, i.e., it must be free from the subjective element so that there is complete interpersonal agreement among experts regarding the meaning of the items and scoring of the test.
- **Objectivity relates to two aspects of the test:**

✓ **Objectivity of items** - it means that the items should he phrased in such a manner that they are interpreted in exactly the same way by those who take the test. For ensuring objectivity of items, items must have uniformity of order of presentation i.e. either ascending or descending order.

✓ **Objectivity of scoring** - it means that the scoring of the test should be a standard one so that complete uniformity is obtained when the test is scored by the examiners at different time.

2. Reliability

- Reliability refers to self-correlation of the test.
- It shows the extent to which the results obtained are consistent when the test is administered once or more than once on the same sample with a reasonable time gap.
- Consistency in results obtained in a single administration is the index of internal consistency of the test and consistency in results obtained upon testing and retesting is an index of temporal consistency.
- Reliability, thus, includes both internal consistency as well as temporal consistency.
- For a test to be called sound it must be reliable because reliability indicates the extent to which the scores obtained in the test are free from such internal defects of standardization which are likely to produce errors of measurement

3. Validity

- It is another prerequisite for a test to be sound.
- Validity indicates the extent to which the test measures what it intends to measure, when compared with some outside independent criterion.
- In other words, it is the correlation of the test with some outside criterion.
- The criterion should be an independent one and should be regarded as the best index of trait or ability being measured by the test.

4. Norms

- A test must also be guided by certain norms.
- Norms refer to the average performance of representative sample on a given test.
- There are four common types of norms - age norms, grade norms, percentile norms and standard score norms.
- Depending upon the purpose and use, a test Constructor prepares any of these norms for his test. Norms help in interpretation of the scores.

5. Practicability

A test must also be practicable from the point of view of the time taken in its completion, length, scoring, etc.

Test Reliability

MEANING OF RELIABILITY

- The term reliability in psychological research refers to the consistency of a research study or measuring test.
- Reliability refers to the **repeatability of findings**. If the study were to be done a second time, would it yield the same results? If so, the data are reliable. If more than one person is observing behavior or some event, all observers should agree on what is being recorded in order to claim that the data are reliable.
- Reliability also applies to individual measures. When people take a vocabulary test two times, their scores on the two occasions should be very similar. If so, the test can then be described as **reliable**.

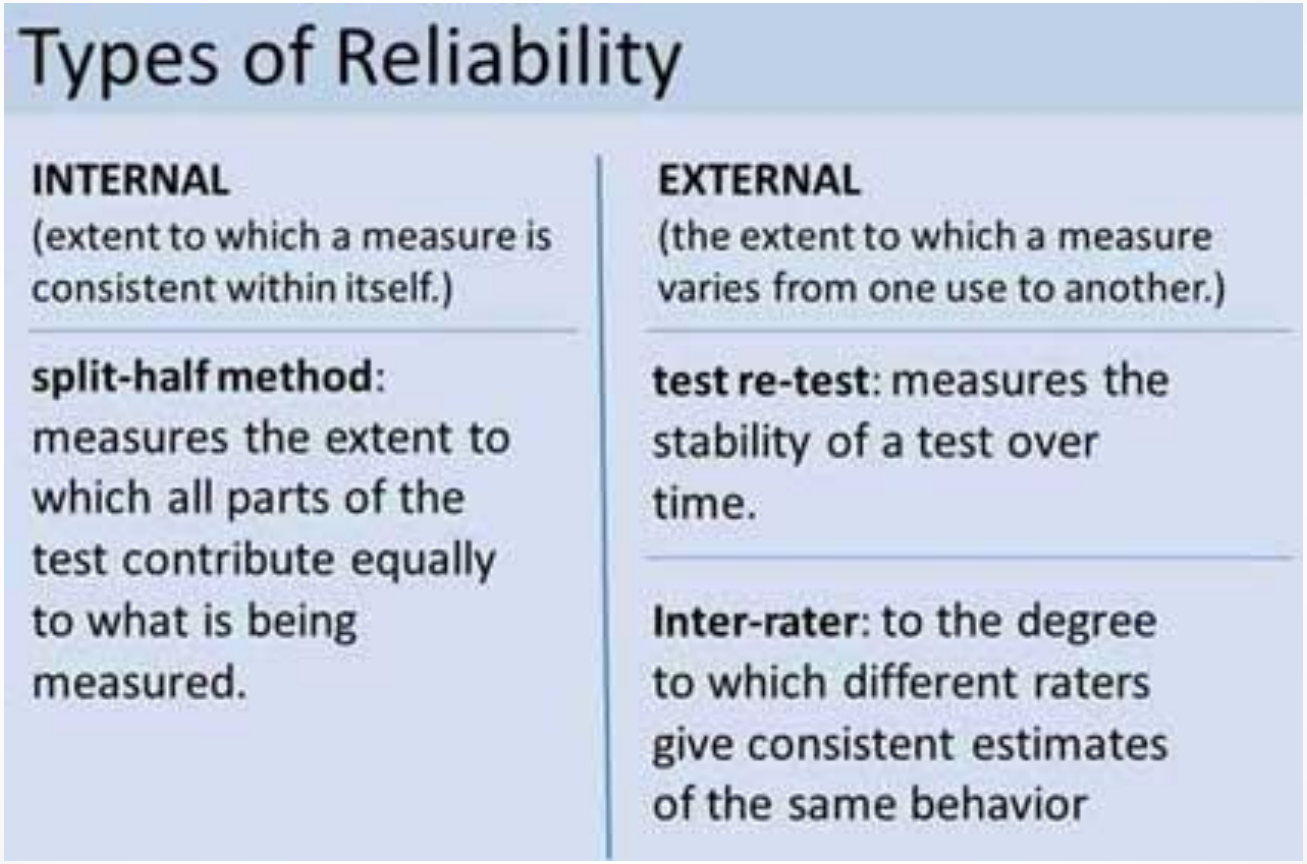
- A test is said to be consistent (when administered once) if the examinees who obtain high scores on one set of items also score high on an equivalent set of items and those who obtain a low score on one set of items also score low on an equivalent set of items.
- The consistency of scores obtained upon testing and retesting is referred to as the "temporal stability" of a test whereas consistency of scores obtained from two equivalent sets of items of a single test after a single administration is referred to as the "internal consistency'.
- According to **Anastasi & Uribina (1997)**, reliability refers to "the consistency of scores obtained by the same individuals when re-examined with test on different occasions, or with different sets of equivalent items, or under other variable examining conditions."

- The correlation coefficient indicating temporal stability is known as the coefficient of stability, and the correlation coefficient indicating internal consistency is known as the **coefficient of internal consistency** or the **alpha coefficient**.
- Any statistical measure of reliability must indicate both, the coefficient of stability as well as the alpha coefficient.

✓ For obtaining the ***coefficient of stability***, the two sets of measurements or scores, found upon testing and retesting are correlated with each other.

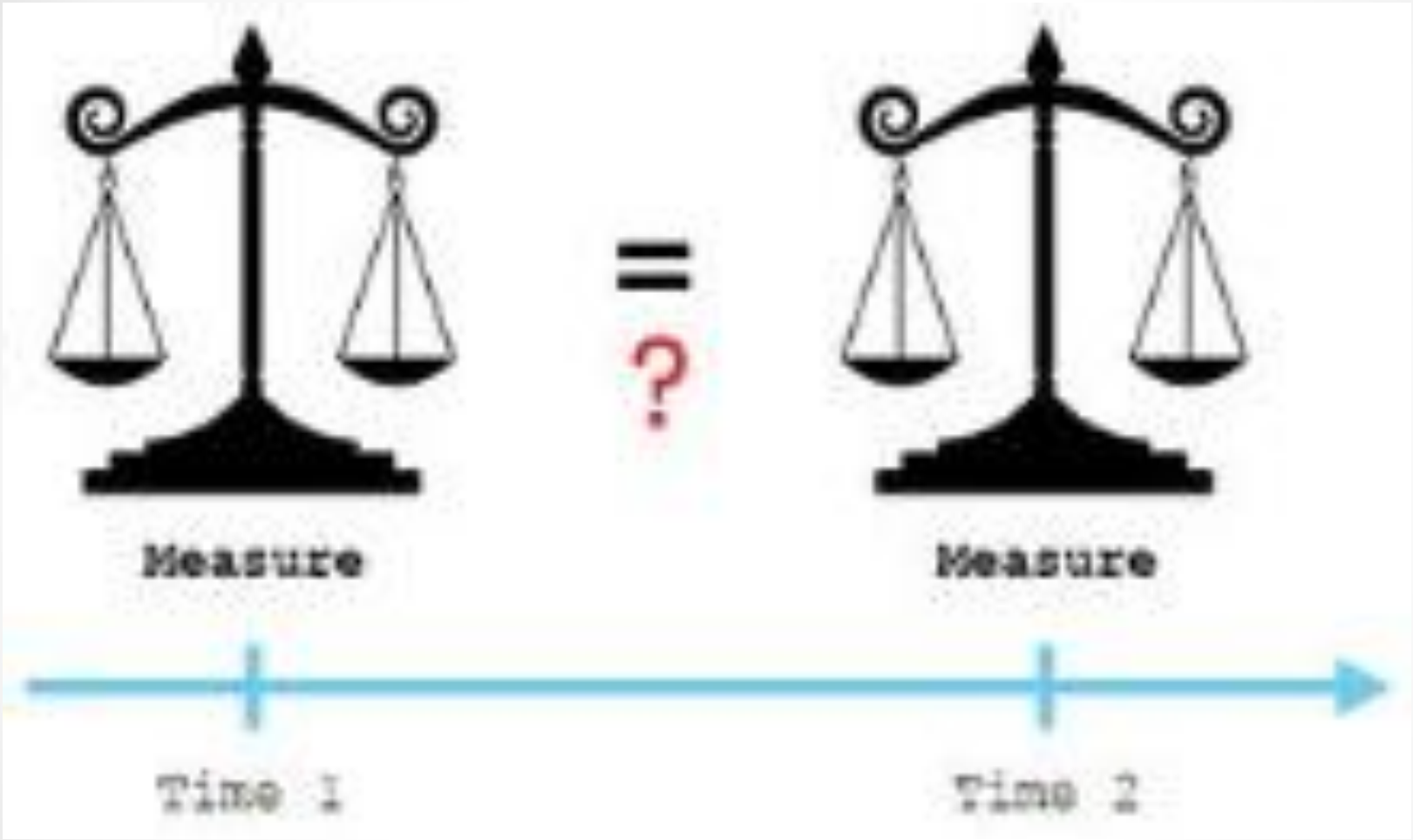
✓ For obtaining the ***alpha coefficient***, the two sets of measurements or scores by two equivalent sets of items of the same test after its single administration, are correlated with eachother.

Methods of assessing reliability or Types of reliability



Test-retest reliability

- The test-retest method assesses the external consistency of a test.
- Examples of appropriate tests include questionnaires and psychometric tests.
- It measures the stability of a test over time.
- A typical assessment would involve giving participants the same test on two separate occasions. If the same or similar results are obtained then external reliability is established.
- The disadvantages of the test-retest method are that it takes a long time for results to be obtained.
- The timing of the test is important; if the duration is too brief then participants may recall information from the first test which could bias the results. Alternatively, if the duration is too long it is feasible that the participants could have changed in some important way which could also bias the results.


=
?
Measure
Measure
Time 1
Time 2

Internal consistency reliability

- Internal consistency reliability indicates the homogeneity of the test. If all the items of the test measure the same trait, then the test is referred to as a homogeneous one and the internal consistency of such a test would be high.
- One of the most common methods of estimating the Internal consistency reliability of a test is the split-half method.
- It measures the extent to which all parts of the test contribute equally to what is being measured. This is done by comparing the results of one half of a test with the results from the other half. A test can be split in half in several ways, e.g. first half and second half, or by odd and even numbers. If the two halves of the test provide similar results this would suggest that the test has internal reliability.
- The split-half method is a quick and easy way to establish reliability. Guilford and Fruchter (1973) have described it as on-the spot reliability.

Alternate-forms reliability

- It is also known as parallel-forms, equivalent-forms and comparable-forms reliability.
- In parallel forms reliability you first have to create two parallel forms.
- One way to accomplish this is to create a large set of questions that address the same construct and then randomly divide the questions into two sets. You administer both instruments to the same sample of people. The correlation between the two parallel forms is the estimate of reliability.
- One major problem with this approach is that you have to be able to generate lots of items that reflect the same construct. This is often no easy feat.
- Furthermore, this approach makes the assumption that the randomly divided halves are parallel or equivalent. Even by chance this will sometimes not be the case.
- The parallel forms approach is very similar to the split-half reliability. The major difference is that parallel forms are constructed so that the two forms can be used independent of each other and considered equivalent measures.

Scorer reliability or Inter rater reliability

- The test-retest method assesses the external consistency of a test.
- This refers to the degree to which different raters give consistent estimates of the same behavior. Inter-rater reliability can be used for interviews.
- It can also be called inter-observer reliability when referring to observational research. Here researcher when observe the same behavior independently (to avoid bias) and compare their data. If the data is similar then it is reliable.

Factors influencing reliability of test scores

The reliability of the scores of test can be influenced by a number of factors.

1. **Extrinsic factors**

- **Group variability** – when the group of examines being tested is homogeneous in ability, the reliability of the test scores is likely to be lowered. On the other hand, when the group of examinees being tested vary widely in their range of abilities, the reliability of the test scores is likely to be increased.
- **Guessing by the examinees** – guessing by examinees in the test is one of the sources of unreliability. There are two important effects of guessing upon the total scores:

✓ It tends to raise the total score and hence makes the reliability coefficient spuriously high.

✓ It contributes to the measurement error.

- **Environmental conditions** – the testing environment should be uniform. For example, arrangement of light, sound, etc should be made equal and uniform among all the examinees.

- **Momentary fluctuations in the examinee** – momentary fluctuations are the one which distract the examinee and may raise or lower the scores. For example, a broken pencil, noise, etc.

2. Intrinsic factors

- **Length of the test** – a longer test tends to yield a higher reliability in comparison to a shorter test.

- **Range of the total scores** – if there is lesser variability in the obtained total scores of the test, it is likely that the reliability of the test will be lowered. On the other hand, if the total scores of the test vary widely, the reliability of the test is increased.

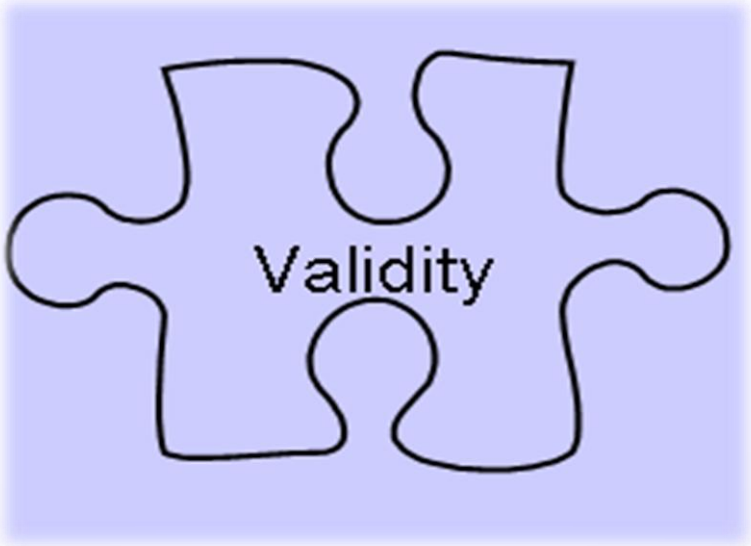
- **Homogeneity of items** – if all the items show good inter-item correlation, the reliability of the test is likely to be increased.

- **Difficulty value of items** – when the items of the tests are very difficult or very easy, the reliability of the test is likely to be lowered, whereas, when the item difficulty level is moderate, the reliability of the test is likely to be increased.

- **Discrimination value** – when the test is having discriminating items, the reliability of the test is likely to be increased.

- **Scorer reliability** - Scorer reliability refers to the extent to which the two or more scorers agree in scoring the same set of responses. If they do not agree, the reliability of the test is likely to be lowered.

Test Validity



Meaning of Validity

- **Validity** refers to the credibility or believability of the research.
- The word "valid" is derived from the Latin word ‘validus’ which means strong.
- The validity of a measurement tool is considered to be the degree to which the tool measures what it claims to measure; in this case, the validity is an equivalent to accuracy.
- It can also be defined as the degree to which evidence and theory support the interpretations of test scores.
- Validity is not the self-correlation of the test, rather it is the correlation with some outside independent criteria, which are regarded by experts as the best measure of the trait which is being measured by the test.
- The independent criteria is referred to as some measure of the trait that the test itself claims to measure.

- According to **Anastasi (1968),** “validity of a test is the accuracy with which it measures or as the degree to which it approaches infallibility in measuring what it purports to measure.

- According to **Kaplan and Saccuzzo (2001**), “validity is the agreement between a test score and the quantity it is believed to measure”.

- The correlation coefficient computed between the test and the ideal independent criteria is called as the **validity coefficient.**

Properties of validity

- **It is a relative term** – a test is not generally valid. It is only valid for a specific purpose. For example, a test for mathematical ability is only valid for assessing mathematical ability, and it valid for assessing some other ability.
- **Validity is not a fixed property** – validation is not a fixed process rather it is an unending process. A test becomes meaningless with the discovery of meanings and concepts. Therefore, a test has to be validated again in the light of mew concepts and meanings.
- **It is a matter of degree** – it is not an all-or-none property rather a matter of degree. A test meant for measuring a particular trait cannot be said to be completely valid or not valid at all.

Aspects of validity

A test is valid for a particular purpose and since there are many uses of testing, therefore, there are different aspects of validity.

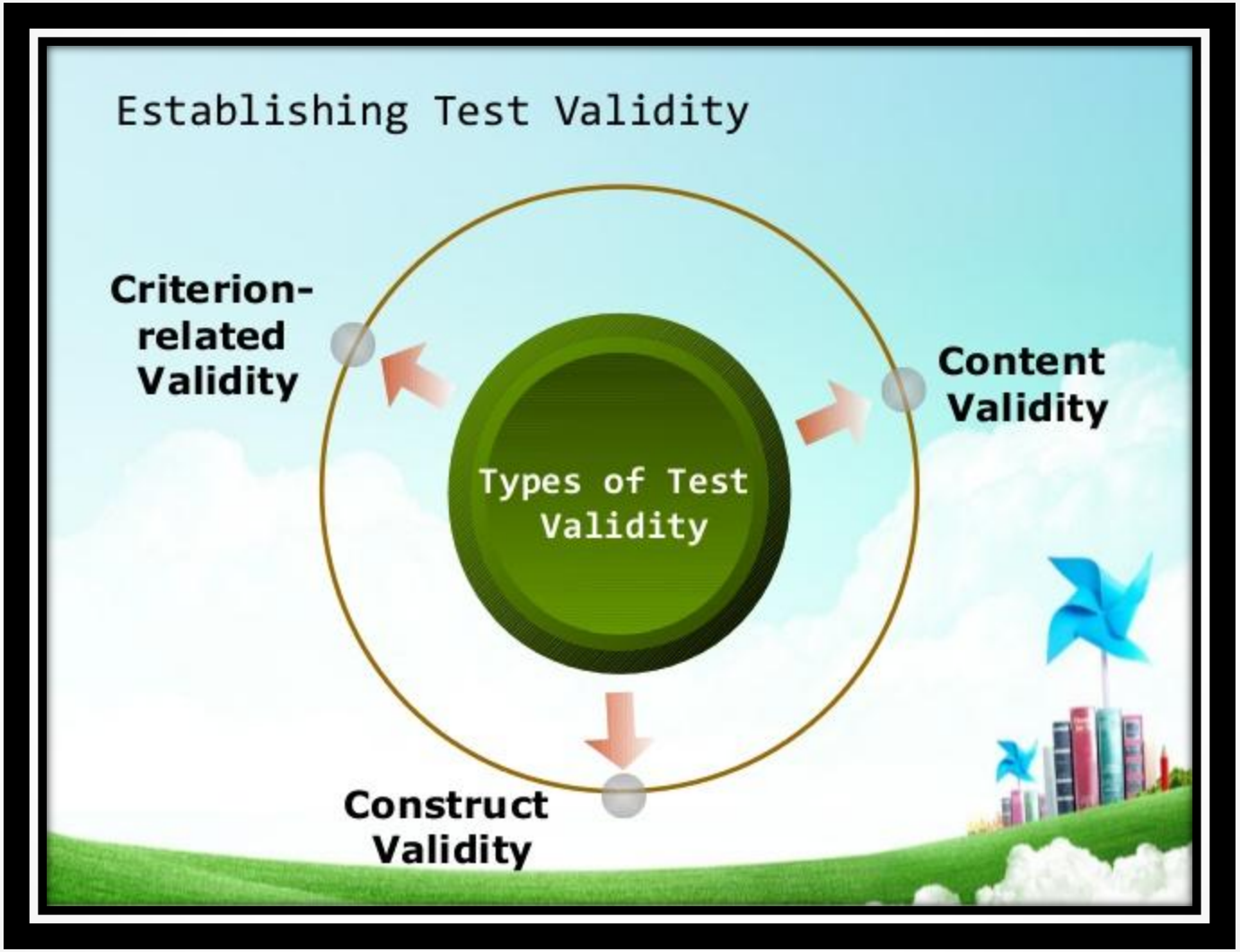
Following are the three main purposes of testing:

- **Representation of a certain specified area of content** – a tester may wish to determine the performance of an examinee in a sample of situations that a test claims to represent. For example, through an English spelling test, a tester may determine the present level of performance of the examinee among school children.
- **Establishment of a functional relationship with a variable available at present or in future** – with the help of testing, an examinee's present performance as well as a prediction about his future performance can be made. For example, with the help of a mechanical test, an individual's mechanical aptitude can be measured and his future performance in the job of a mechanic can be predicted.

- **Measurement of a hypothetical trait or construct –** a tester may wish to determine the extent to which an individual possesses some trait as measured by the test. For example, an examinee's scores on abstract constructs, such as extroversion, honesty, etc can be determined.

On the basis of the above mentioned three purposes of testing, there are three types of validity which are as follows:

1. **Content Validity**
2. **Criterion Related Validity**
3. **Construct Related Validity**


Establishing Test Validity
Criterion-related Validity
Content Validity
Types of Test Validity
Construct Validity

CONTENT VALIDITY

- Content validity is also known as curricular validity, intrinsic validity, circular validity or relevance.
- When all the content of the test measures what the test claims to measure, it is said that the test is having content validity. Therefore, content validity is concerned with the relevance of the content of the test.
- According to **Anastasi (1968)**, “content validity involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured."
- Content validity is the degree to which a test measures an intended content area.

- Content validity requires both item validity and sampling validity

- **Item validity** is basically concerned with whether the test items represent measurement in the intended content area.
- **Sampling validity** refers to the extent to which the test samples the whole content area.

- For **example**, a test designed to measure knowledge of biology might have good item validity because all the items indeed deal with good biology facts but might have poor sampling validity, that is, all its items may deal only with vertebrates.

- Content validity of a test is examined in two ways:
- By the expert’s judgment
- By statistical analysis.

- Content validity is most appropriately applied to the achievement test or the proficiency test. For the aptitude test, the intelligence test and the personality test, content validity is not essential. It sometimes may be a misleading index because the contents of these tests have less intrinsic semblance or similarity to the trait or behaviour they are attempting to sample than do the achievement tests.

- Face validity is often confused with content validity, but in the strict sense it is quite different.
- **Face validity** refers not to what the test actually claims to measure but to what it appears to measure superficially.
- Face validity is the mere appearance that the test has validity (Kaplan & Sacuzzo, 2001).
- When a test item looks valid to the group of examinees, the test is said to have face validity.
- The purpose of face validity is to establish rapport and secure co-operation because when test items do not appear to be valid to the examinees, they may not co-operate in responding

CRITERION-RELATED VALIDITY

- A criterion-related validity is one which is obtained by comparing or correlating the scores obtained on a criterion available at present or to be available in the future.
- A criterion is an external and independent measure of essentially the same variable that the test claims to measure.
- There are two subtypes of criterion-related validity which are as following:

✓ Predictive validity

✓ Concurrent validity

❖ **Predictive Validity**

- ✓ In predictive validity, a test is correlated against the criterion to be made available sometime in the future. I
- ✓ n this, the test scores are obtained and then a time months or years is allowed to elapse, after which the criterion scores are obtained. Then, the test scores and the criterion scores are correlated and the obtained correlation serves as the index of validity coefficient.
- ✓ According to **Marshall & Hales (1972**), "The predictive validity coefficient is a Pearson product-moment correlation between the scores on the test and the appropriate criterion, where the criterion measure is obtained after the desired lapse of time".

EXAMPLE

Suppose one wants to predict success in an graduate class in terms of grades – A,B,C,D and E, where A is the best grade and E is the worst grade. The Investigator may administer a test of intelligence at the time of their admission to the graduate class and thus obtain a set of scores. After three years on the basis of class performance, students are graded according to the above categories. Here grade points would constitute the criterion. If the correlation is high, it can be said that scores on intelligence are directly predicting the future performance of the students in the graduate class. The correlation would be the index of the validity coefficient.

❖ **Concurrent Validity**

- ✓ Concurrent validity is very similar to predictive validity except that there is no time gap in obtaining test scores and criterion scores.
- ✓ The test is correlated with a criterion which is available at the present time.
- ✓ Scores on a newly constructed intelligence test may be correlated with scores obtained on an already standardized test of intelligence. The resulting coefficient of correlation will be an indicator of Concurrent validity.
- ✓ If the correlation is too high, it will indicate that the new test is a needless duplication of the previous one.
- ✓ Likewise, an intelligence test may be validated or correlated against the marks obtained in the previous examination. This will also be an example of Concurrent validity.
- ✓ Concurrent validity is most suitable to tests meant for diagnosis of the present status rather than for prediction of future Outcomes.

- ✓ This validity can be determined by establishing relationship or discrimination. The method is simple and it involves determination of the relationship between scores and scores on some other established criterion which are concurrently available.
- ✓ In this method, the **steps** involved are as follows:

- The test is administered to a defined group of individuals.
- The criterion or previously established valid test is also administered to the same group of individuals.
- Subsequently, the two sets of scores are correlated.
- The resulting coefficient indicates the concurrent validity of the test. If the coefficient is high, the test has good concurrent validity.

Factors affecting validity

1. **Length of the test** – a longer test yields a higher test validity in comparison to a shorter one.

2. **Range of ability** - when the group of examines being tested is homogeneous in ability, the validity of the test scores is likely to be lowered. On the other hand, when the group of examinees being tested vary widely in their range of abilities, the validity is likely to be increased.

3. **Ambiguous directions** – ambiguous directions would be interpreted differently by different examinees. Such directions encourage guessing on the part of examinees. Therefore, it tends to lower the validity of the tests.

4. Socio-cultural differences – Cultural differences among different societies also tend have an effect on test validity. A test which is developed in one culture may not be valid for other culture. On the other hand, when a test is cross-cultured, this factor does not affect the validity of the test.

5. Addition of inappropriate items in the test - A test having inappropriate vague items in it, would lower both the reliability as well as the validity of the test.

Test

Norms

Meaning of Norms

- In any psychological test, an individual's performance is recorded in terms of the raw scores which are expressed in terms of different units, such as- no. of errors, time taken, etc. These raw scores convey no meaning in themselves.
- There are two reference points which can be applied to interpret test scores:

1. **Norm-referencing** – when an examinee's test scores are compared with the score of a specific group of examinees on that test, the process is known as norm-referencing. **Norms are defined as the average performance on a particular test made by a standardization sample.** By a **standardization sample** is meant a sample which is a true representative of the population and takes the test for the purpose of providing data for comparison and interpretation of test scores.

2. Criterion-referencing – when an examinee's test scores are compared with an external standard or criterion, the process is known as criterion-referencing. In criterion-referenced tests, there is a fixed performance criteria. If an examinee qualifies that criteria, that examines is considered to be capable of performance demanded by the test.

Steps in developing norms

The following are the three important steps for developing norms:

1. **Defining the target population**

The first step in developing norms is to define the composition of the target group for which the test is meant. The composition of the target group will be determined by the use of the test. For example, if the test is meant for school children, then only school children would be taken in the target group.

2. **Selecting the sample**

After the target population has been defined by the test constructor, a representative sample from the population would be selected. Usually for constructing norms, a larger sample is preferred. For a larger sample, a completely random sample technique is considered to be the best one.

3. **Standardizing conditions for proper implementation of the test**

The conditions of test administration should be standardized for proper and valid comparisons of test scores with the norms.

Types of norms

1. **Age-equivalent norms**

- Age-norms are developed for any of the characteristics that change systematically with age, atleast upto a certain age level.
- Such norms are defined as the average performance of a representative sample of a certain age level on the measure of a certain trait or ability.
- For example, if the weight of a representative sample of 10 year old girl in Punjab is measured, then the average weight of the sample will become the age norms for the weights of 10 year olds.
- There are some **disadvantages of age norms:**

✓ Age norms lack a standard and uniform unit throughout the period of growth of physical and psychological development.

✓ Growth rate of some traits cannot be compared.

✓ Some traits cannot be expressed in terms of age-norms, for example, visual acuity.

2. **Grade-equivalent norms**

- Grade norm is the score or a range of scores that is representative of the typical level of achievement in the school population for any given grade.
- A student's score on an education achievement test are expressed in terms of grade equivalents.
- The Grade norms have certain limitation:

✓ Grade norms of the same student in different subjects cannot be compared.

✓ Grade norms assume that all students of a particular grade have more or less similar curriculum experiences.

✓ Grade norms cannot be prepared for such subjects in which the growth rate is rapid in elementary level and very slow in high school level.

3. Percentile norms

- Percentile norms can be used equally well with adults and children and are suitable for any type of test.
- Percentiles are merely a form of cumulative frequency distribution, but instead of being expressed in terms of accumulated scores from lowest to highest, the categorization is in terms of whole numbers of percentages of people.
- Percentile ranks tell what percent of the scores fall below a particular score.
- Percentiles are the specific scores or points within a distribution, which indicate the particular score below, which a defined percentage of score fall i.e. it indicate the individual's relative position in the standardization sample.
- For **example,** if a student's raw score of 40 corresponds to a percentile rank of 80, this means that 80 percent of the class obtained lower score than the student did, and that the 20 percent of the class obtained higher scores than the student.

4. Standard score norms

- A norm which is based upon a standard score is known as a standard score norm.
- Standard score norms are preferred in place of percentile norms as here units of the scale are equal so that they convey the same meaning throughout the scale.
- In order to understand a standard score norm, it becomes essential to know the meaning of standard score first.
- A standard score is a derived score having a specified mean and standard deviation.
- There are various types of standard scores, such as, z score, T score, stanine score, deviation IQ.

- **Standard scores are needed primarily for two reasons:**

i. To compare performance of the same person on different tests. This can be done by transforming the raw scores into standard scores.

ii. Standard scores have equal units of measurement and their size does not vary from distribution to distribution.

- **There are two ways in which the raw scores can be transformed into standard scores:**

i. Linear transformation

ii. Normalized transformation