

PROBABILITY LEARNING (PL)

Probability Learning emerged from Probability-functionalism of Brunswick which states that behavior is best understood in terms of probable success in attaining goals. In his experiment he demonstrated that rats learn to tune their choice behavior to the probabilities of likely placement of food in either of same two directions. Using a T-maze / food was kept on 75% of trials say in right hand segment and 25% of trials on left hand one. Towards about the end of the experimental session rats were found to turn right or left in approximately the same ratio 75:25. It has been found that human performance in similar circumstances is not any different from this.

Historically this experiment of Brunswick can be said to be fore runner to PL researches. PL refers to a class of experiments which typically have a long series of discrete trials on each of which the experimental participant is supposed to indicate which of two cues will be presented. The cues are presented randomly with a fixed probability. In PL experiment 'problem matching' is said to occur if the probability with which the participant indicates that a cue will occur matches the probability that it does occur.

Humphreys planned an probabilistic set up in which human subjects engaged in predicting either of two experimenter determined alternatives.

In his study subjects were required to guess whether or not electric bulb would be lighted frequently either the bulb would or wouldn't be lit and the subjects' prediction would be automatically confirmed or rejected.

Each trial began with the signal (another light) the lighting of bulb was thoroughly arbitrary occurrence or non occurrence being in a chance random sequence. For instance occurrence and non occurrence of light would be in a proportion of 1:1. The result was that the output or the response probabilities matched the input or the event probabilities. The subjects soon learned to emit alternative responses in about the same proportion as the simultaneous events.

RELIABILITY

The term Reliability refers to the consistency of scores or measurement which is reflected in the reproducibility of the scores. A test is said to be consistent over a given period of time

when all the examinees retain their same relative ranks on two separate testings with the same test.

A test is also 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 & those who ~~score~~ obtain low scores on one set of items also score low on equivalent

set of items. The consistency of scores obtained upon testing & retesting is referred to as 'temporal stability' of a test whereas consistency of scores

obtained from two equivalent sets of items of a single test after single administration is referred as the 'internal consistency' of the test

scores. The correlation coefficient indicating temporal stability is known as coefficient of stability & the correlation coefficient indicating internal consistency is known as coefficient of internal consistency or the Alpha coefficient.

Ideally any statistical measure of reliability must indicate both stability coefficient & alpha coefficient.

Methods of Estimating Reliability /-

There are three common methods of estimating reliability coefficient of test scores. These methods are:

- (i) Test-Retest Method (or Temporal Stability Reliability)
- (ii) Split Half Method (or Internal consistency Reliability)
- (iii) Equivalent form Method (or Parallel form Reliability)
- Alternate " ")
- Comparable " ")

~~Test~~ (i) TEST RETEST METHOD

In this method the single form of a test is administered twice, on the sample with a reasonable time gap. This way, two time administration of the same test yield two independent sets of score. The two sets when correlated give the value of Reliability coefficient. The Reliability coefficient thus obtained is known as Temporal Stability Coefficient, and indicates to what extent the examinees retain their relative position as measured in the terms of test score over a given period of time. The reasonable time gap between the two tests is usually a fortnight. It is often seen unduly long time gaps reduce reliability coefficient & time gaps too short increases it.

The main shortcomings of Test-Retest

- ① It is cumbersome & time consuming procedure
- ② Maturation effects may operate in contributing to error variance.
- ③ Practice & Memory effect may operate
- ④ The examinees mental & physical state in both test situations may not be the same.

SPLIT-HALF METHOD

From a single administration of one form of a test it is possible to arrive at a measure of test reliability by various split half procedures. In this method, two scores are obtained for each individual by dividing the test into two comparable halves. Temporal stability of the scores does not enter into such a measure, since only one test session is involved.

The most important problem in this method is how to split the test in order to obtain the most nearly comparable halves. A procedure most commonly used is the Odd-Even Method. In this method scores on odd & even items of a test are found out. If the items were originally arranged in an approximate (i.e. equal) order of difficulty such a division yields nearly equivalent half scores.

The advantage of this method is that all

data necessary for computation of reliability coefficient are obtained in a single administration of a test. Thus the variability produced by the difference in two administrations of the same test is automatically eliminated. Not only this, the practice & memory effects are ruled out & also the method does not require subjects to frequently visit laboratories or other centres for testing & retesting.

The obvious disadvantage is that since both the sets of scores are obtained on one occasion fluctuations due to changes in temporary conditions within the examinee as well as due to changes in external directions will operate in one direction i.e. either favorably or unfavorably, the result of which would be either an enhancement or the depression of real reliability coefficient.

Equivalent form Method/-

This method requires that the test be developed into two forms which should be comparable or equivalent. The two forms of the test are administered to the same sample with a time interval of usually a fortnight. Correlation between two sets of scores obtained from two equivalent forms represent the Reliability Coefficient.

is also called as coefficient of equivalence. The chief advantage of equivalent form method is that it automatically weeds out practice & memory effects. The biggest problem however is how to make both forms of the test equivalent in true sense. Freeman has listed following criteria to be adopted for ensuring equivalence of both forms of a test.

- (i) The no. of items in both forms should be the same.
- (ii) Items in both forms should have uniformity regarding the content & the range of difficulty.
- (iii) Distribution of the indexes of difficulty of items in both should be similar.
- (iv) Items in both forms should be equal in degree of homogeneity.
- (v) Means & standard deviations of both forms should be equal or nearly so.
- (vi) Mode of administration & scoring should be uniform.

The above criteria, in the opinion of Freeman must be met very closely if not perfectly for a test to be called as parallel test or equivalent test.

The limitations of this method are

- (i) If behavior functions under consideration are subject to large practice effects the use of parallel form will reduce & not eliminate such effect.

dis-equivalent forms are still unavailable for many tests because of the practical difficulties of constructing comparable forms.

Factors Influencing Reliability of Test Scores

All factors influencing the reliability of test scores can be categorised under two heads Extrinsic & Intrinsic factors. Extrinsic factors are those factors which lie outside of the test itself & tend to make the test reliable or unreliable. Intrinsic factors on the other hand refer to those factors which lie within the test itself & influence the reliability of the test.

Extrinsic factors/-

(1) ~~Group~~ Group Variability/- When the group of examinees ~~are~~ being tested are homogeneous in ability, the reliability of test scores is likely to be lowered but when they vary widely in the ~~abilities~~ ability, the reliability of test scores is likely to be high.

(2) Guessing/- Is an important source of unreliability. It makes reliability coefficient spuriously high.

Environmental Conditions/-

Variability in light, sound & other comforts tend to lower the reliability of the test. Likewise momentary fluctuations caused in the examinee due to changes in surrounding environment for eg breaking of a pencil while drawing a figure may also influence the test score, sometimes by inflating it & sometimes by depressing it.

Intrinsic Factors/-

- (i) Length of the test/- As the length of the test increases reliability is increased
- (ii) Homogeneity of Items/- Homogeneous items have high intra item correlation therefore high reliability
- (iii) Difficulty value of items/- In general, items having indexes of difficulty at 0.5 or close to it yield higher reliability than items of extreme indexes of difficulty.
- (iv) ~~Discrimination~~ ^{It is} Scorer's Reliability/- Also known as reader reliability & it ^{indicates} implies the extent of agreement among two or more scorers in scoring or rating of responses. If ~~scorer~~ scorers do not agree, the reliability is likely to be lowered.

on discriminative value. Items that have high discriminative value may favorably influence ~~total~~ ^{total test} ~~under item~~ correlation leading to high reliability.

VALIDITY

Validity refers to the degree to which test measures what it claims to measure. Validity is not the self correlation of a test. Rather it is the correlation of a test with some independent criteria, which are regarded by experts as the best measure of the trait or ability being measured by the test. For determining the validity of a test, the test must be compared with some ideal independent measures or criteria. The correlation coefficient computed between the test & the ideal measures or criteria is known as Validity coefficient.

"Independent criteria" refers to some measure of the trait or the group of traits that the test itself claims to measure. Validity has 3 important properties

- Validity is a relative term. A test is not generally valid. It is valid for a particular purpose as
- Validity is not a fixed property of the test & ~~the process~~ validation is not a fixed process. Rather it is an unending process.

- Validity like reliability is a matter of degree & is not all or none property.

Types of Validity /-

There are 3 types of Validity

- (i) Content or ~~Curricular~~ Curricular Validity
- (ii) Criterion related Validity
- (iii) Construct Validity.

CONTENT VALIDITY

It refers to the degree to which the test items are the representative samples of the 'knowledge' or 'performance' being measured. Content validity is needed in the test which are constructed to measure how well the examinee has mastered the specific skills or certain course of study. For ensuring full content validation it is essential that

- (i) The area of content should be specified explicitly.
- (ii) The relevance of contents or items should be established in the light of examinee's responses to those contents.

CRITERION RELATED VALIDITY

It is one that is obtained by comparing (or correlating) the test scores with the scores obtained on a criterion available at present or to be available in the future. The criterion is defined as an external & independent measure of essentially the same variable that the test claims to measure. There are two sub-types of criterion related validity.

- (a) Predictive Validity.
- (b) Concurrent Validity.

(a) Predictive Validity/- It is also designated as ^{empirical} or ^{statistical} validity. Predictive validity of a

is a extent to which it is efficient in forecasting or differentiating performance, in a specified area under actual working or living conditions. Predictive validity is need for tests which include long range forecast of academic achievement, forecast of vocational success & forecast of reaction to therapy.

Concurrent Validity/-

A kind of criterion related validity in which relation between test scores & the criterion scores is established at the same time. Although similar in spirit to predictive validity the procedure really uses a "post diction" than "prediction". The test is correlated with criterion which is available at the present time. A comparative study of predictive validity & concurrent validity reveals that for the same test predictive validity is usually lower than concurrent validity.

CONSTRUCT VALIDITY/-

It is also called as factorial validity & trait validity & Anastasi has defined it "as the extent to which the test maybe said to measure theoretical construct or trait". Thus construct validity is the validity of testing instrument based on the determination of the degree to which, the test items capture the hypothetical quality

or trait. A construct is a sort of a concept which is formally proposed with a definition & is related to empirical data (English & English). According to Nunnally, a construct indicates a hypothesis which tells us that "variety of behaviors will correlate with one another in the studies of individual differences &/or will be similarly affected by experimental treatment. Construct validity represents one of the most significant advances of modern measurement theory & practice. It is a significant advance because it unites psychometric notions with theoretical notions. Construct validation is more complex & difficult process than content or criterion related validation. Hence an investigator decides to compute construct validity, only when he is satisfied that ~~neither~~ any valid or reliable criterion is not available. In construct validation the expert tries to know what ~~test~~ ^{factors} psychological & other ~~properties~~ lie behind the test performance. According to Campbell & Fiske ~~satisfactory~~ ^{convergent} ~~and~~ ^{discriminant} validation are important for establishing satisfactory Construct validity. When the test correlates well with its expected referents, the process is known as Convergent Validation & when a test correlates poorly or not at all with measures with which it should not because it differs from those ~~of~~ ^{of} or measures, the procedure is called as Discriminant Val

Factors Influencing Validity

① Length of the test /- Homogenous lengthening of the test increases not only the reliability but also the validity. The longer test is more reliable & hence is more valid too.

② Sample Heterogeneity /- If the subjects have a wider range of ability, so that wider range of scores is obtained, the validity coefficients of the test would be enhanced & reverse will happen if the subjects have limited range of ability.

③ Ambiguous directions /- If the directions of the test are ambiguous, it is likely to be interpreted differently by different examinees. This would tend to encourage guessing & lower the validity.

④ Sociocultural differences /- Cultural differences among different societies are likely to affect the validity of a test. A particular test developed in one culture may not be valid for another culture because of differences in SES, sex roles, social norms etc.

Addition of Inappropriate items /- When inappropriate items, particularly vague ones whose difficulty values differ widely are added to the test, they are likely to lower both the Reliability & the Validity of the test.

(i) A test constructor should not always aim at having high reliability & high validity in the same test. If he does so, he is said to be working at cross purposes because sometimes the goals of reliability & validity are opposite to each other. High reliability requires items of equal difficulty & high intercorrelations between the items, whereas high validity requires items of different difficulty values & low intercorrelations among items (Guilford). Obviously attempting high validity as well as high reliability would imply working at cross purpose. Fortunately Tucker has provided a solution to this puzzle. According to him if inter-item correlations range from $+0.10$ to $+0.60$ one can expect to have a test with satisfactory reliability & validity.

(ii) Reliability increases with increased test length so does ~~not~~ the validity but at a slower rate.

(iii) For a good test reliability range should be around $.90 - .95$ & validity between $.50$ to $.60$.

NORMS OF The FINAL Test/-

The test constructor also prepares the norms test. Norms are defined as the average performance or score of a large sample, representative of a specified population. Norms are prepared to meaningfully interpret the scores obtained on the test as the Raw scores by themselves convey no meaning regarding the ability or trait being measured. Only when Raw scores are compared with the norm, a meaningful inference can be drawn. The preliminary considerations in developing the norms are -

- (i) Sample should represent a cross-section of the population.
- (ii) Sample size should be large.
- (iii) Sample should be randomly drawn.

The 3 commonly used norms are

- ① Age equivalent norm
- ② Grade equivalent norm
- ③ Percentile norm

① Age Equivalent Norm/-

It can be defined as the average performance of a representative sample of a certain age level on the measure of certain trait or ability. Age norms are most suited to these traits.

or abilities that increase systematically with age
eg Physical traits.

Disadvantages of age norms are

- ① They lack a standard uniform unit throughout the period of growth of physical & mental traits
- ② Growth rates of some traits are not comparable.

Grade Equivalent Norms/-

Grade Equivalent norms are defined as the average of a representative sample of a certain grade or class. The grade equivalent norms indicate the grade levels at which the performance of the representative groups is average. The disadvantages of these Grade Equivalent norms are

- ① Grade Equivalent norms of ~~different subjects~~ ^{the same} student in different subjects are not comparable.
- ② Grade norms assume that all students of a class or grade have more or less similar curriculum experiences.

Percentile Norms/-

Percentile norms are the most popular & common type of norms used in psychological and educational tests. Such norms can be prepared for either adults or children & for any type of tests. A percentile norm

indicates, for each raw score, percentage of the population sample that falls below that raw score. Percentile norms, thus represent provide a basis for interpreting an individual's score in terms of his own standing in a particular standardization sample. The main disadvantage ~~includes~~ includes the failure of the layman to distinguish between the percentile & percent scores & the inequality of units throughout the percentile scale & especially at the extremes of the distribution.

Preparation of Manual & Reproduction Test :

The last step in test construction is the preparation of a manual of the test. In the manual the constructor reports the psychometric information of the test, norms & references. It also includes information on the time limits, scoring & the procedures of the test. The test administrator finally the test constructor seeing the possible demand of the test orders the printing.

USES OF PSYCHOLOGICAL TESTS

Psychological tests have been devised & are used primarily for the determination & the ~~any~~ analysis of individual differences in general intelligence, specific aptitudes, educational achievement, vocational fitness & non intellectual personality traits. As the tests are extremely standardise & objective instruments they are less subject to bias & are able to bring out more cogently the individual differences. The following are some of the main areas of the use of psychological tests;

(i) Education / - Psychological tests especially those of general intelligence & of specific aptitudes are extensively used in educational classification, selection & planning right from primary education to universities.

In addition to pre-admission, selection & classification of students into gifted, normal, retarded, psychological tests are being increasingly used in programmes of remedial teaching. Vocational guidance is another important field which has been placed on more objective & scientific footing due to the development of standardised psychological tests.

(ii) INDUSTRY / - The ^{utility} use of psychological testing is ^{increasing} being felt in the industry as over the years the types & number of occupations have multiplied & specialization within the types has enormously increased. Of the various uses selection & placement is perhaps the most prominent one followed by the use of psychological test.

for the appraisal of the performance of the best employed in industrial organisation. Besides these functions psychological tests are also being employed for the purpose of

- counselling & guidance
- identifying special talents like leadership, etc. designing etc.
- evaluation of the success of training programmes especially Managerial Development programmes.
- measuring job attitudes & job satisfaction of the workers & the officers.
- checking the possibility of exaggerated claims by the job applicants.
- identifying the factors responsible for resistance to

CLINICAL PSYCHOLOGY - Today psychological tests are primarily being used for diagnosis of the factors associated with personal problem of learning, behavior, attitude &/or specific interpersonal relations.

SOCIAL SCIENCES RESEARCH - Tests have been used in a variety of psychological, educational, cultural, sociological & employment studies of group as well as individuals. Identification of psychological traits, measurements of group differences, investigation of biological & cultural factors associated with behavioral differences are some areas where tests have

Tests have also been useful in identifying changes within individuals due to age, education, psychotherapy, impact of propoganda, influence, mass media etc.

Thus we find tests & testing plays a significant role in wide variety of educational vocational & clinical situations. In most of these areas, they are being used along with other diagnostic & selection measures such as Case histories, Interviews, Performance Reports etc.