B.Sc PSYCHOLOGY

3rd SEM CORE COURSE

UNIVERSITY OF CALICUT

PSY3B01-PSYCHOLOGICAL MEASUREMENT AND TESTING

2019 ADMISSION

Prepared by

Dhanya Nair

Assistant Professor

Department of Psychology

CPA College of Global Studies, Puthanathani

COURSE CODE	**PSY3B01**
TITLE OF THE COURSE	**PSYCHOLOGICAL MEASUREMENT AND TESTING**
SEMESTER IN WHICH THE COURSE TO BE TAUGHT	**3rd**
NO. OF CREDITS	**3**
NO. OF CONTACT HOURS	**48 (3hrs/week)**

Objectives of the course:

- **To offer foundation on psychological measurement and testing**
- **To provide the basis of test construction and to build up skills on developing psychometric test**
- **To familiarize the uses of psychological tests**
- **To make aware of ethical principals in testing**

Course Details

MODULE NO.	NAME OF MODULE	MODULE HOURS
1	Introduction to Measurement and Scaling Techniques	12
2	Nature and Use of Psychological Tests	10
3	Test Construction and Administration	12
4	Basics of Psychological research	14

MODULE 1: INTRODUCTION TO MEASUREMENT AND SCALING TECHNIQUES

- Measurement: psychological measurement similar to the equal-unit scales of physical measurement.
- Measurement is a relatively complex and demanding task, specially so when it concerns qualitative or abstract phenomena. The measurement mean the process of assigning numbers to objects or observations, the level of measurement being a function of the rules under which the numbers are assigned.
- It is easy to assign numbers in respect of properties of some objects, but it is relatively difficult in respect of others. For instance, measuring such things as social conformity, intelligence, or marital adjustment is much less obvious and requires much closer attention than measuring physical weight, biological age or a person's financial assets. In other words, properties like weight, height, etc., can be measured directly with some standard unit of measurement, but it is not that easy to measure properties like motivation to succeed, ability to stand stress and the like. We can expect high accuracy in measuring the length of pipe with a yard stick, but if the concept is abstract and the measurement tools are not standardized, we are less confident about the accuracy of the results of measurement.
- Technically speaking, measurement is a process of mapping aspects of a domain onto other aspects of a range according to some rule of correspondence.
- In measuring, we devise some form of scale in the range (in terms of set theory, range may refer to some set) and then transform or map the properties of objects from the domain (in terms of set theory, domain may refer to some other set) onto this scale. For example, in case we are to find the male to female attendance ratio while conducting a study of persons who attend some show, then we may tabulate those who come to the show according to sex. In terms of set theory, this process is one of mapping the

observed physical properties of those coming to the show (the domain) on to a sex classification (the range).

- Levels of measurement: nominal, ordinal, interval and ratio scales (measurement scales) or types of scales.
- Scales of measurement can be considered in terms of their mathematical properties. The most widely used classification of measurement scales are: (a) nominal scale; (b) ordinal scale; (c) interval scale; and (d) ratio scale.
- Nominal scale: Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of this is the assignment of numbers of basketball players in order to identify them.
- Nominal scales provide convenient ways of keeping track of people, objects and events. One cannot do much with numbers.
- Nominal scale is the least powerful level of measurement. It indicates no order or distance relationship and has no arithmetic origin. A nominal scale simply describes differences between things by assigning them to categories.
- Nominal data are, thus, counted data. The scale wastes any information that we may have about varying degrees of attitude, skills, understandings, etc. In spite of all this, nominal scales are still very useful and are widely used in surveys and other ex-post-facto research when data are being classified by major sub-groups of the population.
- Ordinal scale: The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule.
- Rank orders represent ordinal scales and are frequently used in research relating to qualitative phenomena. A student's rank in his graduation class involves the use of an ordinal scale.

- One has to be very careful in making statement about scores based on ordinal scales. For instance, if Ram's position in his class is 10 and Mohan's position is 40, it cannot be said that Ram's position is four times as good as that of Mohan. The statement would make no sense at all.
- Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values, and the real differences between adjacent ranks may not be equal. All that can be said is that one person is higher or lower on the scale than another, but more precise comparisons cannot be made. Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the median.
- A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to the non-parametric methods.
- Interval scale: In the case of interval scale, the intervals are adjusted in terms of some rule that has been established as a basis for making the units equal. The units are equal only in so far as one accepts the assumptions on which the rule is based.
- Interval scales can have an arbitrary zero, but it is not possible to determine for them what may be called an absolute zero or the unique origin. The primary limitation of the interval scale is the lack of a true zero; it does not have the capacity to measure the complete absence of a trait or characteristic.
- The Fahrenheit scale is an example of an interval scale and shows similarities in what one can and cannot do with it. One can say that an increase in temperature from 30° to

40° involves the same increase in temperature as an increase from 60° to 70°, but one cannot say that the temperature of 60° is twice as warm as the temperature of 30° because both numbers are dependent on the fact that the zero on the scale is set arbitrarily at the temperature of the freezing point of water.

- The ratio of the two temperatures, 30° and 60°, means nothing because zero is an arbitrary point. Interval scales provide more powerful measurement than ordinal scales for interval scale also incorporates the concept of equality of interval.
- As such more powerful statistical measures can be used with interval scales. Mean is the appropriate measure of central tendency, while standard deviation is the most widely used measure of dispersion. Product moment correlation techniques are appropriate and the generally used tests for statistical significance are the 't' test and 'F' test.
- Ratio scale: Ratio scales have an absolute or true zero of measurement. The term 'absolute zero' is not as precise as it was once believed to be. We can conceive of an absolute zero of length and similarly we can conceive of an absolute zero of time.
- For example, the zero point on a centimeter scale indicates the complete absence of length or height. But an absolute zero of temperature is theoretically unobtainable and it remains a concept existing only in the scientist's mind.
- The number of minor traffic-rule violations and the number of incorrect letters in a page of type script represent scores on ratio scales. Both these scales have absolute zeros and as such all minor traffic violations and all typing errors can be assumed to be equal in significance. With ratio scales involved one can make statements like "Jyoti's" typing performance was twice as good as that of "Reetu."
- The ratio involved does have significance and facilitates a kind of comparison which is not possible in case of an interval scale. Ratio scale represents the actual amounts of variables.

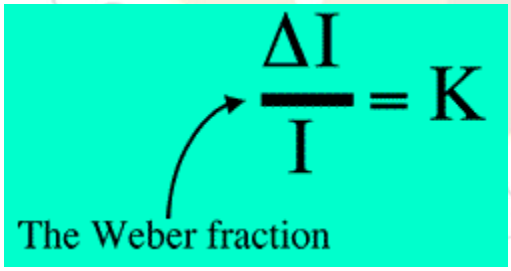
- Measures of physical dimensions such as weight, height, distance, etc. are examples. Generally, all statistical techniques are usable with ratio scales and all manipulations that one can carry out with real numbers can also be carried out with ratio scale values.
- Multiplication and division can be used with this scale but not with other scales mentioned above. Geometric and harmonic means can be used as measures of central tendency and coefficients of variation may also be calculated. Thus, proceeding from the nominal scale (the least precise type of scale) to ratio scale (the most precise), relevant information is obtained increasingly.
- If the nature of the variables permits, the researcher should use the scale that provides the most precise description. Researchers in physical sciences have the advantage to describe variables in ratio scale form but the behavioural sciences are generally limited to describe variables in interval scale form, a less precise type of measurement.
- Properties of scales of measurement: Magnitude, equal interval and absolute zero
- Magnitude. Magnitude is the property of "moreness." A scale has the property of magnitude if we can say that a particular instance of the attribute represents more, less, or equal amounts of the given quantity than does another instance. On a scale of height, for example, if we can say that John is taller than Fred, then the scale has the property of magnitude.
- A scale that does not have this property arises, for example, when a gym coach assigns identification numbers to teams in a league (team 1, team 2, and so forth). Because the numbers only label the teams, they do not have the property of magnitude. If the coach were to rank the teams by the number of games they have won, then the new numbering system (games won) would have the property of magnitude.
- Equal intervals. The concept of equal intervals is a little more complex than that of magnitude. A scale has the property of equal intervals if the difference between two

points at any place on the scale has the same meaning as the difference between two other points that differ by the same number of scale units. For example, the difference between inch 2 and inch 4 on a ruler represents the same quantity as the difference between inch 10 and inch 12: exactly 2 inches.

- As simple as this concept seems, a psychological test rarely has the property of equal intervals. For example, the difference between IQs of 45 and 50 does not mean the same thing as the difference between IQs of 105 and 110. Although each of these differences is 5 points, the 5 points at the first level do not mean the same thing as 5 points at the second. We know that IQ predicts classroom performance.
- However, the difference in classroom performance associated with differences between IQ scores of 45 and 50 is not the same as the differences in classroom performance associated with IQ score differences of 105 and 110. In later chapters we will discuss this problem in more detail.
- When a scale has the property of equal intervals, the relationship between the measured units and some outcome can be described by a straight line or a linear equation in the form Y _____ a _____ bX. This equation shows that an increase in equal units on a given scale reflects equal increases in the meaningful correlates of units.
- absolute 0. An absolute 0 is obtained when nothing of the property being measured exists. For example, if you are measuring heart rate and observe that your patient has a rate of 0 and has died, then you would conclude that there is no heart rate at all. For many psychological qualities, it is extremely difficult, if not impossible, to define an absolute 0 point. For example, if one measures shyness on a scale from 0 through 10, then it is hard to define what it means for a person to have absolutely no shyness
- Distinction between psychological measurement and physical measurement

- Measurement When the quality and the quantity of anything is measured by using standard equipment or device, it is called Measurement. In the words of Chaplin (1975)- "Measurement means quantifications of variables." It is an important part of life, which can be divided into two types:
- Physical Measurement: It is a technique of measuring physical bodies using standard. For example, measurement of a the length of a desk or measurement of the volume of a glass of water, etc.
- Psychological Measurement: It is a technique of measuring attributes, attitudes, personality characteristics and abilities of living beings. It is of two types
    - Psychological tests.
    - Scales.
- Differences between Physical measurement and Psychological measurement
- In physical measurement, the measurement starts from true zero but in psychological measurement, zero is just as reference point. It is relative zero, not true zero. For example, psychological attributes such as reaction time can never be zero.
- The unit of physical measurement are equal and uniform throughout the measurement but there is a lack of equal and uniform unit in psychological measurement.
- Physical measurement is applicable on those things about which we get knowledge by the sense organs but in psychological measurement, we can't get knowledge of it's subject matter such as intelligence and personality characteristics by our sense organs. In other words, physical measurements is direct but psychological measurement is not direct.
- Physical measurement are to the point and it's measurement is exact but the psychological measurements are not precise
- Problems in psychological measurements.

- Concepts of Psycho physics: Psychophysics is a scientific approach to the measurement of mental processes. Psychophysics is the systematic study of sensory capacities by determining behavioral responses to physical changes in sensory stimuli.
- Threshold: the level or point at which you start to experience something, or at which something starts to happen.
- An absolute threshold is the smallest level of stimulus that can be detected, usually defined as at least half the time. The term is often used in neuroscience and experimental research and can be applied to any stimulus that can be detected by the human senses including sound, touch, taste, sight, and smell.
- The difference threshold was first described by a physiologist and experimental psychologist named Ernst Weber and later expanded upon by psychologist Gustav Fechner. Just Noticeable Difference (JND) in Psychology.
- Weber's law, also called Weber-Fechner law, historically important psychological law quantifying the perception of change in a given stimulus. The law states that the change in a stimulus that will be just noticeable is a constant ratio of the original stimulus.
- Weber's law also incorporates the just-noticeable difference (JND). This is the smallest change in stimuli that can be perceived.



- The fraction $\Delta I/I$ is known as the Weber fraction (aka Fechner fraction). If we rearrange the equation to $\Delta I = IK$, you can see that Weber's Law predicts a linear relationship

between the increment threshold and the background intensity. Below is a plot of some hypothetical data showing Weber's Law. The slope of the line is the Weber fraction.

- **Fechner's Law** is a generalization of Weber's theory. According to him, there is a broader relationship between sensory and physical intensity. **Fechner's law** states that the strength of the sensation becomes greater as the logarithm of stimulus intensity increases. In **Fechner**'s formulation JNDs provide the basic unit of perceived intensity.
- With these assumptions Weber's Law is valid and Fechner's Law can be stated as follows: S = K log1Q I, where
    - S is the subjective, perceived, psychological intensity
    - K is a constant associated with a specific sensory modality
- This formulation states that if physical intensity is increased in a logarithmic fashion the resultant psychological intensity will follow with equal steps.
- Fechner's Law emphasizes the compressed representation of the physical stimuli in the psychological domain. This compression implies that when varying signals or stimuli it may be necessary to make relatively large changes in the physical stimuli to render it psychologically perceptible.
- Psychophysical/ psychological scaling methods: - The methods traditionally used to determine the threshold were devised by G.T.Fechner(1801- 1887), a German physicist and philosopher, regarded as the founder of psychophysics.
- Fechner was concerned with the relation between physical stimulation and mental experience.
- Method of Average Error: The adjustment is repeated many times. This is also called the method of average error. In this method, the observers themselves control the

magnitude of the variable stimulus, beginning with a level that is distinctly greater or lesser than a standard one and vary it until they are satisfied by the subjective equality of the two.

- This method is also known as the Method of Adjustment.
- The intensity of the stimulus is under the observer's control i.e. the observer is required to adjust the intensity to a just detectable level.
- Once the observer adjusts the stimulus intensity until it is just detectable, the value of that intensity level defines the threshold.
- The difference between St and Co defines the error in each judgment.
- A large number of such judgments are obtained and the average of those judgments is calculated.
- Hence the name, „ method of average error "is given.
- Method of Minimal Changes: This method is also known as Method of Serial Exploration.
- For computing threshold by this method, two modes of presenting stimulus are adopted the increasing mode and the decreasing mode(the ascending series and the descending series).
- For computing RL(Reiz Limen,German equivalent to absolute threshold), standard stimulus is needed and the subject simply reports whether or not he has detected change in the stimulus presented in the ascending and descending series.
- RL maybe affected by two constant errors – the error of habituation and the error of anticipation. These two errors work in opposite direction. Alternate ascending and descending series control these errors to a great extent.

- DETERMINATION OF RL OF SOUND INTENSITY BY THE METHOD OF LIMITS (Hypothetical data)

SOUND INTENSITY	A	D	A	D	A	D
12		+		+		+
11		+		+		+
10		+	+	+	+	+
9		+	+	+	-	+
8	+	+	-	-	-	+
7	+	+	-	-	-	-
6	-	-	-		-	-
5	-	-	-		-	
4	-	-	-		-	

DETERMINATION OF RL OF SOUND INTENSITY BY THE METHOD OF LIMITS

- Individual Thresholds: 6.5 6.5 8.5 8.5 9.5 7.5
- Mean Ascending Threshold: $\frac{6.5+8.5+9.5}{3} = 8.16$
- Mean Descending Threshold: $\frac{6.5+8.5+7.5}{3} = 7.5$
- Mean Absolute Threshold: $\frac{8.16+7.5}{2} = 7.83$

- DETERMINATION OF DL FOR WEIGHT LIFTING BY THE METHOD OF LIMITS (Hypothetical data)

Co St	A	D	A	D	A	D	A	D	A	D
109		+		+		+		+		+
108	+	+		+		+		+	+	+
107	=	+	+	+	+	+	+	+	=	+
106	=	=	=	+	=	=	=	+	=	+
105 (St)	-	=	=	=	=	=	=	=	-	=
104	-	-	-	=	-	-	-	=	-	-
103	-		-	-	-		-	-	-	
102	-		-		-		-		-	
101	-		-		-		-		-	

DETERMINATION OF DL FOR WEIGHT LIFTING BY THE METHOD OF LIMITS

Upper Threshold: 107.5 106.5 106.5 105.5 106.5 106.5 106.5 105.5 107.5 105.5

Lower Threshold: 105.5 104.5 104.5 103.5 104.5 104.5 104.5 103.5 105.5 104.5

Mean Upper Threshold: $\frac{1064}{10} = 106.4$

Mean Lower Threshold: $\frac{1045}{10} = 104.5$

- DETERMINATION OF DL FOR WEIGHT LIFTING BY THE METHOD OF LIMITS

- Interval of Uncertainty: Mean U Threshold – Mean L Threshold
  $106.4 - 104.5 = 1.9$
- Difference Limen: $\frac{\text{Interval of Uncertainty}}{2}$
  $\frac{1.9}{2} = 0.95$
- Point Of Subjective Equality: $\frac{\text{Mean U Threshold + Mean L Threshold}}{2}$
  $\frac{106.4 + 104.5}{2} = 105.45$
- Constant Error: PSE – St = 105.45 – 105 = 0.45

- Method of Constant Stimuli: This method requires a series of forced-choice trials.
- A fixed number of stimuli of different intensities, extending over a relatively wide range, are singly presented many times in random order.
- On each presentation the observer must make a detection response – either “Yes“(detection) or “No”(no detection).

- For each stimulus intensity, the percentage of trials in which the stimulus value is detected is computed.
- The intensity of the stimulus value detected on 50% of the trials is generally used as the measure of the Absolute Threshold.

DETERMINATION OF RL OF SOUND INTENSITY BY THE METHOD OF CONSTANT STIMULI(stimulus presented 100 times) (Hypothetical data)

Stimulus Value	f (detection)	%(detection)
50	98	98
48	80	80
46	70	70
44	65	65
42	43	43
40	30	30
38	20	20
36	15	15
34	12	12

DETERMINATION OF RL OF SOUND INTENSITY BY THE METHOD OF CONSTANT STIMULI

Calculation of RL by the method of Linear Interpolation:
50% point falls between the stimulus value of 42(43%) and 44(65%).
65 – 43 = 22
43+7=50%
By interpolation, we calculate(7/22)x2=.636 (as there are 2 step intervals between stimulus value of 44 and 42).
42+.636= 42.636 or 42.64
RL by interpolation method = 42.64

- Method Of Pair Comparison: The Paired Comparison Method describes values and compares them to each other. Learn how to create a Paired Comparison table.
- In pair comparison the stimuli are paired and the subject is required to make a comparative judgement by saying which member of each pair is preferred or possesses of the trait being scaled.
- Paired Comparison Method is a handy tool for decision making; it describes values and compares them to each other. It's often difficult to choose the best option when you have different ones that are far apart.
- Law of comparative judgement is defined as a series of assumptions which relate the proportion of times one stimulus is judged higher on a given attribute than the other stimulus to the scale values and discriminal dispersions of these two stimuli which are repeatedly being compared. The set of assumptions is derived from the following four postulates.
- Postulates 1: the stimulus presented to the subject produces a response, which has some value on the psychological continuum.

- Postulate 2: a given stimulus, if presented repeatedly, does not produce the same discriminal process
- Postulate 3: if any stimulus is presented to the subject a larger number of times, it will generate a frequency distribution, which is normal one.
- Postulate 4: the mean and standard deviation of the normal distribution of the discriminal process generated by a stimulus corresponds to its scale value and the discriminal dispersion respectively.
- Modal discriminal process: the process which is most frequently aroused by given stimulus.
- Method Of Rank Order: also called order of merit method, it's a simplest techniques for psychological scaling of objects, persons, events and other items.
- The method of rank order was developed by cattell (1903) and spearman (1904).
- The rank order is the computation of correlation. The ranks are converted as choice scores (C=n-r).
- The method of rank order has several advantages over its closely related rival called the rating scale.

MODULE 2: NATURE AND USE OF PSYCHOLOGICAL TESTS

- Definition of psychological test- a psychological test is a standardized procedure to measure quantitatively or qualitatively one or more than one aspects of a trait by means of a sample of verbal or nonverbal behaviour.
- Historical perspective of psychological testing- early antecedents- major developments have occurred over the last century, many of them in the United States.
- Evidence suggests that the Chinese had a relatively sophisticated civil service testing program more than 4000 years ago (DuBois, 1970, 1972).
- Every third year in China, oral examinations were given to help determine work evaluations and promotion decisions.
- By the Han Dynasty (206 B.C.E. to 220 C.E.), the use of test batteries (two or more tests used in conjunction) was quite common.
- These early tests related to such diverse topics as civil law, military affairs, agriculture, revenue, and geography.
- Tests had become quite well developed by the Ming Dynasty (1368–1644 C.E.). During this period, a national multistage testing program involved local and regional testing centers equipped with special testing booths. Those who did well on the tests at the local level went on to provincial capitals for more extensive essay examinations.
- After this second testing, those with the highest test scores went on to the nation's capital for a final round. Only those who passed this third set of tests were eligible for public office.
- The Western world most likely learned about testing programs through the Chinese. Reports by British missionaries and diplomats encouraged the English East India Company in 1832 to copy the Chinese system as a method of selecting employees for

overseas duty. Because testing programs worked well for the company, the British government adopted a similar system of testing for its civil service in 1855.

- After the British endorsement of a civil service testing system, the French and German governments followed suit. In 1883, the U.S. government established the American Civil Service Commission, which developed and administered competitive examinations for certain government jobs. The impetus of the testing movement in the Western world grew rapidly at that time (Wiggins, 1973).
- **Three main features of psychological tests:**
    1. Standardized procedure: the test should be uniform to all examiners
    2. Behaviour sample: it must measure a behaviour or a trait
    3. Scores or categories: the attribute is assigned a number or label which assumes its intensity or quantity
- Historical perspective of psychological testing: psychological tests in their current form and nature came into being during the early part of 20th century. The major objective of psychological tests was related to recruitment of army or industry during World War I. during these time the most used tests are Binet-Simon intelligence scale(1905) and Rorschach inkblot test(1921).
- In Chinese Han Dynasty (206BC)people used test batteries to assess individuals abilities in arears like Military, Agriculture, Revenue etc. in 17th C, testing for humans abilities was a well-developed system in China. By copying Chinese system, British Govt devised a similar for its civil service in1855. Later German, French and US govt used similar methods of testing.
- Contributions of Galton, Wundt and Kraepelin: Galton studied about individual differences and coined the term ‘mental test’. Wilhem Wundt (1879), father of experimental psychology, in Leipzig- Germany, established a Laboratory dedicated

exclusively to experiments and observations in Psychology. Emil Kraepelin , an eminent psychiatrist from Germany helped in classify mental disorders according to their causes and characteristics. He also helped scientific psychology to attain growth in the direction of measuring & categorizing human mental attributes.

- Psychophysicists like Weber and Fechner interested in studying the relationship between physical sensation and corresponding psychological experience. Personalities like Edourad Seguin and Hermann Ebbinghaus have contributed to the development of the field of psychometrics.
- Uses of psychological test:
  - Psychological test as decision making tools
  - Psychological tests to help self-understanding
  - Psychological tests as tools for data collection
  - Clinical setting
  - Educational and guidance setting
  - Organizational setting
  - Counseling settings
- Characteristics of a good test:
  1. Objectivity: a good test should be free from all kinds of subjectivity with respect to testing procedures, scoring and interpretation
  2. Reliability: the temporal consistency of a test.
  3. Validity: a test measures what it intended to measure
  4. Norms: its standard references of a normative samples
  5. Practicability: a test which is reliable, valid and having norms without complex to administer and to take.

- Ethical issues in psychological testing
  - Ethics refers to rules or standards to be followed during conduction of psychological tests and helps in protect the participants from possible harms.
  - Informed consent: participants should give his consent for being the subject for testing
  - Confidentiality: the details of participants must be kept confident
  - Right to withdraw from testing: subject have all rights to withdraw from the test taking whenever they feel uncomfortable
  - Release of test result: participants have all rights to know their response and the result my comprehend in the language the participants can understand easily.
  - Labeling: investigators should not use stigmatized labels to describe participants instead can use of codes etc.
  - Competence of the examiner: test givers should receive proper training about the administration of test and its scoring method, reliability, validity etc.
  - Interpretation of test results: while interpretation of test result, investigators should keep in mind the influence of language, culture as well as the testing conditions
  - Deception: some times its better to not to say everything about the test towards test takers so as to not to form a conscious response every time.
  - Debriefing: it refers to summarizing various aspects of the test after the testing is completed.
- Factors influencing test administration
  1. Examiners
  2. Testing conditions
  3. Test taker
- Classification of psychological tests

- Based on different criteria like
- 1. Mode of Administration: Individual test- A test can be said individual test in the sense that they can be administered to only one person at a time and group test- group testing is the administration of psychological tests such as intelligence measures, achievement measures etc to groups of people rather than to individuals.
- 2. rate of Performance: speed tests- Speed tests are designed to assess how quickly a test taker is able to complete the items within a set time period and power test- those in which the difficulty level of items increases as the test progresses.
- 3. Behavioural Dimension Measured: personality test
    - Intelligence
    - Aptitude
    - Achievement
    - Creativity
    - Interest inventories
    - Value tests
    - Attitude tests
- 4. Medium used: paper-pencil test- here we use language and they usually of statements or questions to be answered; questionnaires and Performance /Situational – subject is required to perform something like block designing etc.
- 5. Nature of items: Verbal/Objective test- verbal means related to words. It might be multiple choice response and non-verbal/ Essay Type- materials like picture, figures or design blocks etc.
- 6. Interpretation: Norms referenced and criterion referenced
- 7. mode of scoring: self-scored tests and expert-scored test
- 8. scope or applicability: culture-specific tests and culture-free tests

Module 3: Test construction and administration

Introduction to steps of test construction

- Planning: planning about the nature and content of the test is very important. We have to select the domain to be assessed, form the objectives, collect reviews on the related studies.
- Writing: here we do write the items for the test construction of specific domain. Item refers to individual units of interactions in a test.
- Meaning and purpose of item analysis: its of two kinds assessment like item difficulty and item discriminability. These helps to identify the item difficulty between each item and the discriminability between individuals in possessing a trait or attribute.
- Administration: it can be considered as pilot study. Only through pilot administration a test could be assessed for its practicability, reliability and validity. After pilot study it is tried with very small sample for preliminary try-out and identify the major difficulties or uncertain of items in the test. Later the test provided for actual try-out for sample size of 400 participants and evaluation is done through a process known item analysis.
    - Item analysis: is the best method for selection of items into a test finally (Guilford, 1954). Two kinds of assessment
    - 1. Item discriminability/ discrimination index of an item: defines whether an item is able to discriminate between individuals who actually possess a trait/ attribute and those who do not. There are many methods to evaluate some are
    - (a). The extreme group method: the method of compares the performance of high scores of the test on each item with that of the low scores.
    - (b). The point Biserial Method: the point biserial correlation between an item and a total test score is $r_{pbis} = [\frac{(\bar{Y}_1 - \bar{Y})}{S_Y}] \sqrt{\frac{P_X}{1-P_X}}$

r_{pbis} = the point biserial correlation or index of dicriminability

$\bar{Y}_1$ = The mean score on the test for those who got item 1 correct

$\bar{Y}$ = the mean score on the test for all persons

S_Y = the standard deviation of the exam score for all persons

P_X = the proportion of persons getting the item correct

- Item characteristics curve: this is a pictorial representation of how an item is performing with respect to individual variations in abilities /knowledge.
- Item response theory: newer approach which have popularity among psychometricians. Here we can make an ability judgment without subjecting the tesr taker to all items of the test.
- 2. Item difficulty/ difficulty index of an item: denoted by the letter 'p'. the value of p ranges from 0 to 1. This helps in understand difficulty index as the 'easiness' of an item. Researcher suggest that the range of an item difficulty can best contribute to the discrimination capacity of the test.

- Standardization: this is a process by which we ensure consistency and objectivity of how test are administered and scored. This process includes following
  - Estimation of reliability: reliability refers to the overall consistency of a test. It is the capacity of a test to yield similar results when administered at different points of time.

    Reliability coefficient: the value of reliability coefficient ranges from 0 to 1. Higher the value of correlation coefficient greater would be the reliability of a test.

    Different methods to assess reliability

    1. Test -rest method: method of comparing the scores obtained by a group of individuals on two different occasions

2. Parallel form method: method requires 2 equivalent forms of a test
3. Split half method: technique used to evaluate the internal consistency of a test
4. Kuder-richardson formula: kuder and Richardson published KR_{20} formula. The formula is applicable for tests with dichotomous items-with score 1 and 0. $KR_{20} = r = \frac{N}{N-1}\left(\frac{S^2-\varepsilon pq}{S^2}\right)$
5. Cronbach alpha: Also called coefficient alpha. $r = \propto = \binom{N}{N-1}\left(\frac{S^2-\varepsilon S_i^2}{S^2}\right)$

- Estimation of validity: validity is about its accuracy. The different aspects of validity are face or surface validity, content validity, criterion related validity – concurrent validity and predictive validity, construct validity – convergent validity and discriminant validity.
- Face validity: subjective impression about the extent to which a test covers the concept which it is intended to measure.
- Content validity: how adequately the test represents the concepts based on which it is developed. To establish content validity the test items should be carefully evaluated by expert judges.
- Criterion related validity: evidences show how well proposed test scores correlates with an already established or standard criterion. This helps to predict various behaviour by evaluating its correspondence with other measures.
- Two types of criterion: concurrent validity-shows the degree to which the scores on a test are related to the scores on another pre-established test administered at the same time whereas predictive validity is the degree to which a test can actually predict the individual performance in future situations.

- Construct validity: which the measure tests a theoretical construct. Convergent validity- which assess a same construct correspond to each other. Discriminant-related to the distinctiveness of a test.
- Development of norms: norms are used in psychological tests in order to convert the obtained scores to a value that is comparable to a standard.
- Types of norms: standard score: standard score is a converted raw score which expresses its position relative to the mean, in terms of standard deviation. It is denoted by 'Z'.
- T score: it is a standard score with mean 50 and standard deviation 10.
- stanine scores: also, the 'standard nine' score which transforms all the scores in distribution into a single digit number from 1 to 9.
- age equivalent norms: are used to compare the performance of an individual with that of a normative age group. Highly used in assessing children's abilities.
- grade norms: these are derived by locating the performance of a test taker within the norms of students at each grade level in the standardization sample.
- percentiles and percentile ranks: this indicates the percentage of scores lying below it. It determines at which one hundredth part of the distribution of scores an individual is placed.

- Test administration: a tendency to think that a score obtained by an individual is the true reflection of his/her ability or traits.
- Different factors like- related to the examiner, related to test taker, related to procedural and situational factors.
- Related to examiner:

1. Training received by the examiner: lack of training received by the examiner negatively influence test result.

2. Reinforcing responses: this is one of the procedural variables that affect test scores. When an administrator gives feedbacks to subject for their responses, it can result in increased/decreased performance, particularly when test belongs to 'ability' category such as intelligence and aptitude.
3. Expectancy effects: expectation from the part of examiner can also affect test scores.

- Factors related to test taker:
    1. Language- the person who is not good at English and Hindi find difficult to interpret the meaning of an item.
    2. Subject variables: when a test taker is very anxious, it can affect the result.
    3. Deception and bias: if participants intentionally give wrong responses, it is called deception.
- Procedural and situational factors:
    1. Mode of administration: researchers have found that scores on a test differs depending upon whether it is self- administered or trainer-administered.
    2. Rapport between examiner and the test taker: a good rapport has to be established with the test taker before the actual testing.
    3. Testing conditions: physical environment like noise, illumination, temperature etc affect test result.

Module 4: Basics of Psychological research

- Meaning and characteristic

  According to Kerlinger (1973), research is a systematic, controlled, empirical and critical investigation of hypothetical propositions about the presumed relations among natural phenomena.

  Characteristics of scientific research:

  Objectivity- to be objective one has to guard against his own biases, beliefs, wishes, values and preferences

  Testability- researchers can observe, weigh or measure the same and check the observations for their accuracy.

  Replicability- if one uses similar methods and procedures of data collection and analysis as described in a study, she must be able to come up with similar results.

  Rigor- this will eliminate bias, facilitate proper data collection and analysis, which in turn would lead to sound and reliable research findings.

  Predictability- to explain and predict the phenomenon being studied

  Generalizability- every research should have the characteristic of generalization.

- Types: different types

  1. *Basic research and applied research:*

     Any branch of science including psychology has evolved largely from fundamental research. This kind of research is necessarily aimed at expanding human knowledge, then being a work of 'invention'. In fundamental research, knowledge is created for its own sake, freed from any form of objective-practical judgements. However, results derived through this kind of research are used later by applied sciences where they acquire commercial and social values. Applied research is said to be problem-driven. These researches are carried out in order to find solutions for practical

problems and thereby improve human condition. Applied research gains from fundamental research and applies these insights to formulate answers to different problems of real life.

2. *Qualitative and quantitative research:*

   These are two types of approaches for collecting data collection and analyzing data. Where qualitative research yields textual results or inferences and quantitative research is carried out on numerical data and it gives you numerical results.

   Eg: Is there any relationship between two variables

   Eg: how music is used as persuading technique in commercial advertisements

In these two questions the first will be applicable to quantitative type and the other will be applicable to qualitative type. There are different methods by which data are collected in qualitative is used they are unstructured interviews, Ethnography, naturalistic observations, focus group discussions, protocol analysis and archival analysis. To analyze qualitative data methods used are grounded theory approach, thematic analysis, discourse analysis, conversational analyze.

3. *Descriptive research and experimental research:* the research help to find out the cause-effect relationship between variables.

- Independent variable: in an experiment is a controlled observation in which the investigator manipulates the level of one variable
- Dependent variable: the variable that measure its effect on another variable
- An experiment often consists of two groups namely experimental and control group.
- An experimental group is the group which the original experimental manipulations are administered whereas the latter is a similar group. The key concept of experimental research is randomization.

- Randomization involves 3 sequential processes: 1. Randomly selecting a sample from the population, 2. Randomly assigning subjects to each group, 3. Randomly selecting one group as experimental group and the other as control group.

- Research process: scientific research is carried out systematically by following specific plans of action.

- the various steps are;

1. **identifying the research problem**: research problem is a question or matter involving doubt, uncertainty or difficulty that is proposed for solution or discussion.
2. **Reviewing related literature**: review of literature helps to redefine and refine the problem.
3. **Setting objectives and hypothesis**: research objectives precisely state what the researcher is trying to achieve. Objectives should be measurable, achievable, specific and realistic. A hypothesis can be defined as a tentative solution to the research question. Two types, 1. Null hypothesis H_0, 2. Alternate hypothesis H_1
4. **Formulating research design**: research design is the overall plan of the study, concerning the selection of samples and collection and analysis of data. There are different types of design like experimental design, case study design, descriptive design, action research design, etc.

- **Sampling**: the population selected for research is said to be sample. And based on the sample the researcher makes the assumptions about the entire sample. To make assumptions we need to select sample design like probability sampling or non-probability sampling.

- In Probability sampling each member of the population has a known probability of getting selected to the sample. It uses randomization to select the sample members. The

subdivisions of this design are **simple random sampling**- also known as chance sampling where each and every item in the population has an equal chance of being selected, **systematic sampling**- the selection process starts by picking some random point in the list and then every n^{th} element, **stratified sampling**- if the population from which a sample is to be drawn does not constitute a homogeneous group, then this technique is applied so as to obtain representative sample and **cluster sampling and area sampling**- this involves grouping the population and then selecting groups or clusters rather than individual elements for inclusion in the sample and in cluster sampling procedural potential for order bias and other sources of errors is usually higher.

5. **Collecting data:** in research data is defined as information collected for the purpose of analysis to produce original research findings. The different kinds of data are documents, audio/videotapes, test response, photographs, etc. the various dta collection methods in social science include observation, questionnaire, schedules, instruments and inventories, focus group discussions and archival records.
6. **Data analysis**: it is a kind of analysis in order to derive inferences or to obtain results from data collected. For qualitative data we use analysis technique which helps in analyze and interpret the data.
    1. **Content analysis:** procedure for classifying data into various categories for the purpose of summarizing, organizing and meaningfully interpreting it. The category might be words, phrases, sentences or themes. This method can be carried out with any kind of data like texts written, audio content, drawings etc.

2. **Discourse analysis**: also ‘conversation analysis’ this involves analyzing a naturally occurring conversation or any recorded interaction. It focuses on how language is used in everyday situations.
3. **Narrative analysis:** defined as ‘transcribed experience’. Characterized by ‘life story’. Researcher writes stories and later analyzes it.
4. **Grounded theory:** method of constructing theory from data by means of ‘induction’. Here an examination of a single case based on which a hypothesis or a general statement is formulated. This hypothesis is tested for its applicability across similar cases, successful completion of which will givw birth to a theory.
5. **Ethnographic analysis:** the method by which a distinct culture is studied in a holistic sense. This method relies on ‘field notes’ collected by researcher in the form of observation, interview summaries and reflections etc. which are later subjected to content analysis.

   Read on other methods of qualitative analysis such as framework, interpretive phenomenological, ethnomethodology etc.

- **Descriptive statistics:** it gives general idea about the trends in a data which may include- measures of central tendency (mean, median and mode), measures of dispersion (variance, standard deviation, range), frequencies and shape of distribution.
- **Inferential statistics**: helps to formulate generalization and predictions from the data by means of induction. This includes hypothesis testing (t-test, ANOVA, regression analysis), estimating population parameters (a parameter is a value that summarize data for a population like ‘census’).

- **Data analysis software packages**: there are many computers software packages available for the purpose of statistical data analysis for both qualitative and quantitative data. Data analysis is the penultimate stage in the research process.

7. **Report writing**: writing report is a important as carrying out a study because only through a report research could be communicated and transferred to both the scientific society and to the common people. While writing a report generally we follow APA (American Psychological Association) guidelines. The aim of research report is to convey why and how the researcher has conducted the research and the implication of the findings. A research report should have a clear beginning, middle and end.

APA style includes following sections;

1. Title page
2. Abstract
3. Introduction
4. Method
5. Results
6. Discussion
7. References
8. Tables or figures

References

- Singh, A.K.(2008). *Tests, Measurements and research Methods in Behavioural Sciences*(3rd ed.)
- Anastasi, A., & Urbina, S. (2005). *Psychological Testing* (7th ed.).New Delhi: Prentice –Hall Of India

- Kothari, C. R. (2009). *Research Methodology- Methods & Techniques*. (2nd ed.). India: Repro India Limited

Refer following for further reading on APA style.

- https://apastyle.apa.org/instructional-aids/reference-examples.pdf
- https://libguides.ecu.edu/c.php?g=982594&p=7463742
- APA Style
- APA Formatting and Citation (7th Ed.) | Generator, Template, Examples (scribbr.com)
- APA Format: Basic Rules You Must Follow (verywellmind.com)