

Quadrant II – Transcript and Related Materials

Programme: Bachelor of Arts (Third Year)

Subject: Psychology

Paper Code: PSC 106

Paper Title: Psychological Testing

Unit: Test Development

Module Name: Test Construction and Scaling.

Name of the Presenter: Maria Muriel de Barros (Assistant Professor)

Rosary College of Commerce and Arts, Navelim-Goa

Notes

THE PROCESS OF TEST DEVELOPMENT

The process of Test Development involves five stage: Test conceptualization, test construction, test try-out, item analysis and test revision.

TEST CONSTRUCTION AND SCALING

Test construction is the second stage in the process of test development. During this stage the test items are drafted. This includes the processes of writing of test items, formatting items, item scaling, setting scoring rules, designing and building the test as a whole.

The process of test construction is carried out through the following three processes:

- a) Scaling: Setting rules for assigning numbers in measurement.
- b) Writing items: Formulating items/ item pool for the final test

c) Scoring: Assigning scores to test items/ responses.

A) SCALING

Scaling is a process by which measuring is designed and calibrated and by which numbers- scale values- are assigned to different amounts of the trait, attribute or characteristic being measured. In psychometrics: Scales are conceived as instruments to measure a trait, state or an ability.

Scales can be categorized along a continuum of levels of measurement as nominal, ordinal, interval and ratio. Also there are different types of scales based on the critical interest: Age-based scales, grade-based scales, stanine scales, unidimensional, multi-dimensional scales, comparative and categorical scales.

Test-taker have more or less of all characteristics that are measured by valid test. A higher or lower score indicates more or less of a certain characteristic that is being measured.

Rating scales are used to record the **test-takers judgements** of oneself, others, experiences, objectives. Individuals group words, statements or symbols and make judgements of the strength of a particular trait/ attitude/ opinion/ emotion. **For example:** The Morally Debatable Behaviour Scale- Revised (MDBS-R): The scale contains 30 items, each item contains a brief description of a moral issue or behaviour on which the test-taker expresses their opinion by means of a 10-point scale ranging from '*never justified*' to '*always justified*'

SCALING METHODS

1. **Summative scales:** The ratings of the test-taker on all the items is added together to obtain a final score. Five point scales are the most reliable

summative scales. For example: A test with 30 items to be rated on a 10-point scale: 30 would be the minimum score and 300 would be the maximum.

2. **Likert scale:** A type of rating scale, usually used to scale attitudes. For each item, the test-taker is presented with 5 alternative responses (sometimes 7) usually on an agree-disagree or approve-disapprove continuum. Likert scales usually results in ordinal-level data.
3. **Method of paired comparisons:** Test-takers are presented with pairs of stimuli which they are asked to compare. They must select one of the stimuli according to some rule. For example: which do they agree with or which is more appealing. For each pair of options, test-takers receive a **higher score** for selecting the **option deemed more justifiable** by the majority of a group of judges.

The **judges** would be asked to **rate the pairs of options** before the distribution of the test and a list of the options selected by the judges would be provided along with the **scoring instructions as an answer key**. The test score would reflect the number of times the choices of a test-taker agreed with those of the judges. The test-taker may receive a **point for selecting one options** and no points for selecting another option. This method forces test-takers to choose between items.

- **Sorting Tasks:** When sorting tasks are used, stimuli such as printed cards, photographs are presented to the test-taker for evaluation. This method results in ordinal information.
 1. ***Comparative scaling*** entails judgements of a stimulus in comparison with every other stimulus on the scale. Comparative scaling could also be accomplished by providing test-taker with a list of 30-items on a

sheet of paper and asking them to rank the justifiability of the items from 1 to 30. For example: 30 items printed on separate index cards and test-taker is asked to sort the cards from most justifiable to least justifiable.

2. **Categorical scaling:** Stimuli are placed in one of two or more alternative categories that differ quantitatively with respect to some continuum. For example: Test-taker is given 30 index cards with one item printed on each card. The test-taker has to sort the cards into three piles: Behaviours never justified, those sometimes justified and those always justified. For example:
3. **Guttman Scale:** Items on this scale range from weaker to stronger expressions of the attitude, belief or feeling being measured. A feature of the Guttman scale is that test-takers who agree with a stronger statement of an attitude also agree with milder statements. For example: The data gathered is analysed using a **scalogram analysis**: An item analysis procedure and approach to test development that involves a **graphic mapping of a test-taker's responses**. The objective is to obtain an arrangement of items wherein **endorsement of one item automatically connotes endorsement of less extreme positions**. Mainly used in consumer psychology: where purchasing one item influenced purchase of another item.