Measurement Definition, Designing and Writing Items

Definition

Measurement is the process of systematically assigning numbers to objects and their properties, to facilitate the use of mathematics in studying and describing objects and their relationships.

❖ Some types of measurement are fairly concrete: for instance, measuring a person's weight in pounds or kilograms, or their height in feet and inches or in meters. Note that the particular system of measurement used is not as important as a consistent set of rules: we can easily convert measurement in kilograms to pounds, for instance.

❖ Measurement is not limited to physical qualities like height and weight. Tests to measure abstractions like intelligence and scholastic aptitude are commonly used in education and psychology, for instance: the field of psychometrics is largely concerned with the development and refinement of methods to test just such abstract qualities.

❖ The measurement process begins with formulation of research problem or hypothesis.

❖ Every research problem contains terms – concepts or variables.

❖ The ultimate goal of measurement is to specify clearly observable referents of the terms contained in one's hypothesis.

❖ Thus, the entire measurement process consists of moving from the abstract (concepts) to the concrete (measures of the concept.

Levels of Measurement

1. **Nominal Data:** With nominal data, as the name implies, the numbers function as a name or label and do not have numeric meaning. For instance, you might create a variable for gender, which takes the value 1 if the person is male and 0 if the person is female. The 0 and 1 have no numeric meaning but function simply as labels in the same way that you might record the values as "M" or "F."

2. **Ordinal Data:** Ordinal data refers to data that has some meaningful order, so that higher values represent more of some characteristic than lower values. For instance, in medical practice burns are commonly described by their degree, which describes the amount of tissue damage caused by the burn. A first-degree burn is characterized by redness of the skin, minor pain, and damage to the epidermis only, while a second-degree burn includes blistering and involves the dermis, and a third-degree burn is characterized by charring of the skin and possibly destroyed nerve endings. These categories may be ranked in a logical order.

3. **Interval Data:** Interval data has a meaningful order and also has the quality that equal intervals between measurements represent equal changes in the quantity of whatever is being measured. The most common example of interval data is the Fahrenheit temperature scale. If we describe temperature using the Fahrenheit scale, the difference between 10 degrees and 25 degrees (a difference of 15 degrees) represents the same amount of temperature change as the difference between 60 and 75 degrees.

4. **Ratio Data:** Ratio data has all the qualities of interval data (natural order, equal intervals) plus a natural zero point. Many physical measurements are ratio data: for instance, height, weight, and age all qualify.

Tests of Sound Measurement/ Standards to Evaluate Measurement

1. **Reliability:** is concerned with questions of stability and consistency.

• An ex. of a highly reliable measuring instrument is a steel tape measure. A cloth tape measure would be somewhat less reliable.

• If we have a technician measure the same part 10 times, using the same instrument, will the measurements be similar each time? In each case, if the answer is yes, we can say the test, scale, or instrument is reliable..

• Unreliable things cannot be valid.

• Considerations of reliability are not limited to educational testing: the same concepts apply to many other types of measurements including opinion polling, satisfaction surveys, and behavioural ratings

2. **Validity:** Validity refers to how well a test or rating scale measures what is it supposed to measure. Some researchers define validation as the process of gathering evidence to support the types of inferences intended to be drawn from the measurements in question. Categorization of validity that recognizes four types: content validity, construct validity, concurrent validity, and predictive validity.

Types of Errors in Measurement

Observed value= true value + systematic error+ random error The first source of variation is true differences in the concept the operation is intended to measure. Ex. IQ test ought to reflect only true differences in intelligence and nothing else.

1. **Systematic measurement error** results from the factors that systematically influence either the process of measurement or the concept being measured.

2. **Random measurement error** is unrelated to true differences in the concept being measured. It is the result of temporary, chance factors. Ex. A tired and bored respondent may give erroneous responses by not attending carefully to the questions asked. Similarly an ambiguously worded question will produce random errors by eliciting responses that vary according to respondent's interpretation of the question's meaning. • Such a error is random because its presence, extent and direction are unpredictable.

Sources of Error in Measurement

a) Respondent

b) Situation

c) Measurer

d) Instrument

Designing Items in Measurement

❖ Multi-Item Scale

Definition: A multi-item scale is designed to measure the respondent's attitude towards more than one attribute related to the stimulus object. It is a rating scale comprising of multiple items, where each item is a single question or a statement that needs to be evaluated.

Likert, Semantic Differential, and Stapel scales are the examples of multi-item scales, comprising of multiple items for evaluating the attitude of the respondent towards the object. For the development of a multi-item scale, a technical expertise is required. A paradigm for its development is shown below:
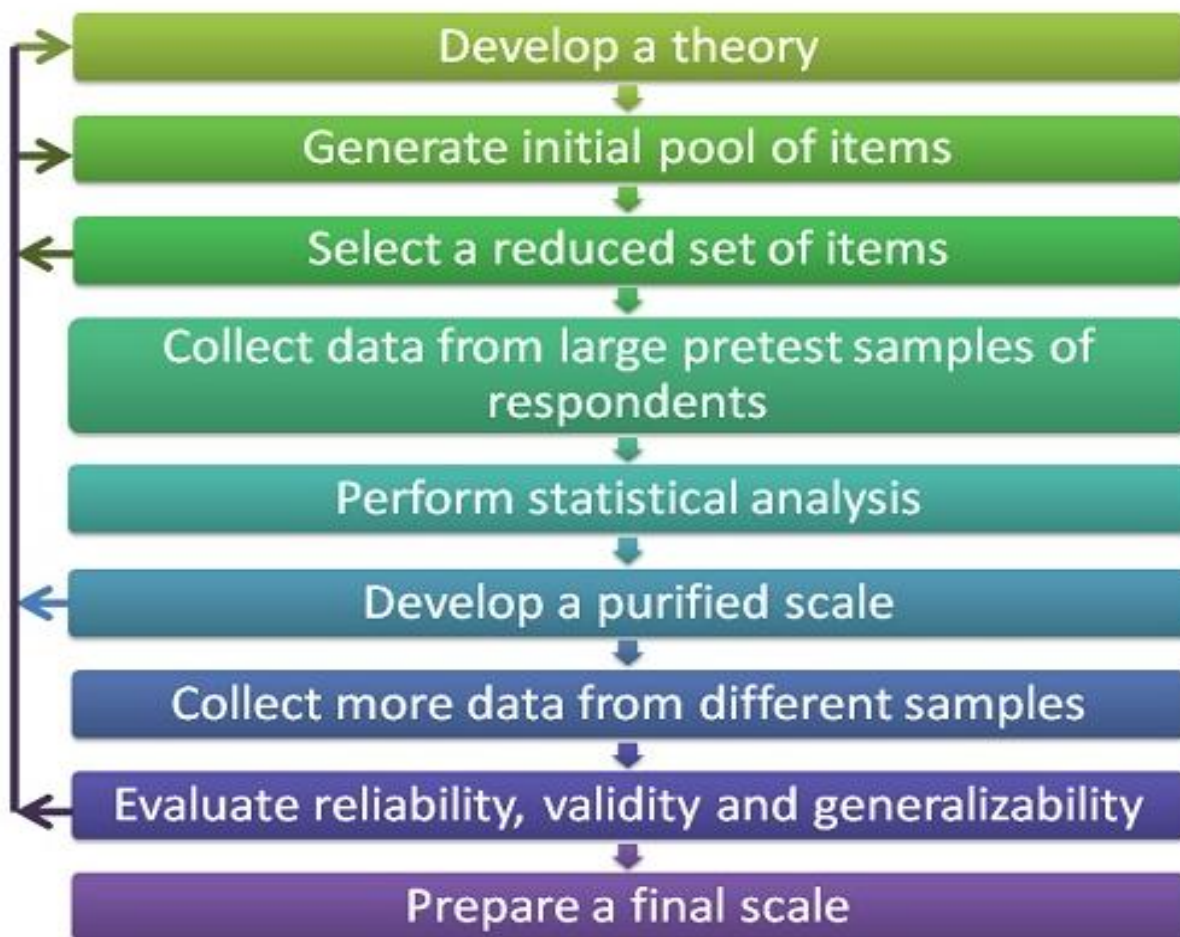
Steps are Discusses as:

1. The researcher begins with the development of construct of interest. The construct refers to the type of concept existing at the high level of abstraction than do everyday concepts, such as brand loyalty, attitude, product involvement, satisfaction, etc.

2. The construct must be operationalized in such a way that it complies with the theoretical definition. For example, in the context of tea, a consumer will be characterized as brand loyal if he shows a favourable attitude and must have purchased the same brand of tea for at least five out of last six purchase occasions.

3. In the next step, the initial pool of scale items is generated on the basis of theory, analysis of the secondary data and qualitative research. Out of this, a reduced set of prospective scale items is generated based on a qualitative judgment of a researcher and other expert individuals. The reduced scale items are further reduced in a quantitative manner since these are still too large to constitute a scale.

4. Once the reduced prospective set of scale items is obtained, the researcher collects the data on it from a large pretest sample of respondents. The data obtained is analyzed using the statistical techniques such as correlation, cluster analysis, confirmatory factor analysis, statistical tests, exploratory factor analysis and discriminated analysis.

5. As a result of the statistical analysis, further more items are eliminated, thereby resulting in a purified scale.

6. The purified scale is then evaluated for reliability, validity and generalizability by collecting more data from different samples.

7. As a result of such assessment, a final set of scale items is selected.