
UNIT 7 RESEARCH TOOLS-I: QUESTIONNAIRE, RATING SCALE, ATTITUDE SCALE AND TESTS

Structure

- 7.1 Introduction
- 7.2 Scales of Data Measurement
- 7.3 Characteristics of a Good Research Tool
 - 7.3.1 Validity
 - 7.3.2 Reliability
 - 7.3.3 Usability
- 7.4 Types of Tools and their Uses
 - 7.4.1 Questionnaire and Schedules
 - 7.4.2 Rating Scale
 - 7.4.3 Attitude Scale
 - 7.4.4 Tests
- 7.5 Let Us Sum Up
- 7.6 Glossary
- 7.7 Answers to Check Your Progress Exercises

7.1 INTRODUCTION

Data collection is an important part of research. In order to collect the requisite data for any theme of research you have to devise appropriate tools and use suitable measuring techniques, and decide on the relevant attributes of the samples drawn. There are several research tools, varying in design, operation, complexity of features, and interpretation. In certain situations you may select from a list of available tools. In other situations you may find that existing research tools do not suit your purpose or objective of research and, therefore, you may like to modify them or develop your own. Each tool is appropriate for collecting a particular type of data or information which lends itself to a particular type of analysis and interpretation for drawing meaningful conclusions and generalizations. For this, you need to familiarize yourself with the nature, merits and limitations of various research tools. In this and the subsequent unit we shall focus on the characteristics, types, uses and limitations of some commonly used research tools – Questionnaires, interviews, observations, rating scales, attitude scales and tests.

Objectives

After studying this unit, you shall be able to :

- explain the characteristics of a good research tool,
- define a questionnaire and describe its various types,
- describe the characteristics, uses and limitations of a questionnaire,
- define a rating scale and describe its types, uses and limitations,
- define an attitude scale and describe its types, uses and limitations,
- define a test and describe the types, uses and limitations of tests, and
- choose appropriate techniques and use them efficiently in your research projects.

7.2 SCALES OF DATA MEASUREMENT

Research tools are the measuring devices. Every measuring device has some kind of graduation depending upon the system of measurement. For example, the FPS or

CGS systems measure length in foot or centimeter. Similarly weight is measured in pounds and grams. The foot rule that measures length is graduated in inches. There are two major attributes – (1) each inch is of equal length wherever it appears on the footrule, (2) two different measured as two inches, for example by same foot rule are of same length.

Just as FPS or CGS provides the basis for scaling for physical measurement, it is necessary to provide certain form of scale for nutritional/health measurements. Measurement of variables like demographic, and social, community infrastructure related or health related outcomes etc.

A scale is an instrument on which the characteristics are measured. It can be quantitatively calibrated in the usual sense, or can be qualitative also.

Generally three types of scales are used in epidemiological research. These are :

Nominal,

Ordinal, and

Metric scale

Let us review these scales.

Nominal scale : It is the most elementary form of the scale. As indicated by the name itself it is only nominal. This form of scale is largely used to classify people or object in certain categories like male-female, rural-urban, in other words, it labels object of measurement. For example, the observation such as gender (either male or female) is nominal, so are the criteria for defining aplastic anaemia from peripheral blood such as neutrophils, platelets and reticulocytes. The measurements on this scale do not have any specific order. To illustrate, gender is either male or female and none is higher or more than the other. The only way to associate numeric to such a scale is by way of assigning a code to each category. To illustrate assigning code 1 to male, code 2 to female.

Ordinal Scale : This is the second level of scale which is more sophisticated than the nominal scale, though remains in one of the cruder forms. In contrast to the nominal scale, this scale consists of ranks or ordering in the categories of a measurement. Disease severity is measured in ordered categories such none, mild, moderate, serious or critical. Wherever the sample of the research is arranged in ascending or descending order on the basis of data on a variable, we are using the ordinal scale, e.g.; when the students are ranked in a class on the basis of their achievement we are using ordinal scale e.g., the 10th rank in a class of 50 students is better than 11th rank but lesser than 9th. However, despite the ranking it does not indicate the difference between the 9th and 10th rank is equal to the difference between the 11th and 12th. In other words, the difference between the ranks are either unknown or unequal. The only information that is derived from this case is the relative position of a subject within the sampled population on a variable.

Metric Scale : The characteristics which can be exactly measured in terms of a quantity such as weight, height, haemoglobin level and heart rate are said to be measured on metric scale.

Exact measurements on metric scale are statistically preferable compared to the ordinal measurements. Yet, the irony is that circumstances lead to grouping of metric data into categories sometimes even after the exact data are obtained. Birth weight of a newborn may be categorized as – 2.4 kg, 2.5–2.9 kg and 3.0 kg and above. Such ordinal categories are sometimes easier to comprehend than exact metric measurements. In the process, however, valuable exact information is lost.

As mentioned above, there are three types of scales Nominal, Ordinal, and Metric. The choice of scaling technique depend upon the nature of the variable. Next, let us consider the characteristics of a good research tool.

7.3 CHARACTERISTICS OF A GOOD RESEARCH TOOL

There are mainly three characteristics of a good research tool. These include validity, reliability, and usability. In selecting tools for collecting data a researcher should evaluate them in terms of these characteristics. Let us discuss these one by one.

7.3.1 Validity

A tool used for collecting data must provide information that is not only relevant but free from systematic errors. In other words, it must produce only valid information and measure what it claims to measure. For example, an achievement test in nutrition must measure knowledge of students in nutrition alone. It should not turn out to be a language test. If a question on food frequency questionnaire is asked, and a certain student well versed in the English language writes a good 'essay' on it, the researcher should not end up measuring the language ability of the student. A tool, however, does not possess universal validity. It may be valid in one situation but not in another. The tool useful in deciding in a particular research situation may have no use at all for a different situation. So, instead of asking, "Is this research tool valid?" It is important to ask the more pertinent question, "How valid a particular tool is for collecting information which the researcher needs to gather?" Or, more generally, "For what decision is this tool valid?" There are three types of validity, (i) content validity; (ii) criterion-related validity, and (iii) construct validity.

Let us get to know them.

Content validity

Content validity relates to the relevance of the content of a research tool to the objective and nature of a research problem. For example, in the case of tests of achievement (for example nutrition knowledge test), content validity refers to the effectiveness of the test items in measuring the basic nutrition knowledge and estimated by evaluating the relevance of the test items to the instructional objectives, the actual subject studied, and the knowledge acquired individually and as a whole.

Content validity of a research tool is based on the judgment of several experts in the field concerned, careful analysis of objectives of the subject of research and the hypotheses, if any, to be tested. Content validity is also known as rational or logical validity or face validity.

Criterion-related validity

In decision making situations, selection or classification is based on an individual's expected performance as predicated by a research toll. For example, a psychological test or rating scale which predicts the kind of behaviour it was intended to predict, is said to possess 'predictive validity'. The prediction may be regarding eating behaviour or feeding practice. This validity, refers to the association between present result as indicated by a particular research tool and future behaviour. In order to determine the predictive validity of a tool, the results from it must be compared with the actual performance or outcome in the future. For example, if a test is designed to select students for a certain medical course, scores on the test must indicate a significant positive relationship with their ultimate success in the medical profession. A researcher studies predictive validity if his or her primary interest is in the outcome which he or she wants to improve by some professional decisions.

In some research situations, a researcher may wish to develop a new tool as a substitute for an already existing cumbersome tool (technique or method). If the existing tool is considered useful for decision making and we want to test the validity of the new one, the key question to ask is whether the new tool agrees with the information sought through the existing cumbersome technique. If they disagree, the new one cannot be substituted for the original tool. The agreement between the newly developed tool and the already existing cumbersome technique for which the tool has been developed, is estimated by an empirical comparison. Both, the newly developed tool and the original one are applied to the same sample groups, and the results are compared. This type of empirical check on agreement is called *concurrent validation*, as the information obtained through the two tools ought to give nearly the same results. The validity of the new tool thus established is called its 'concurrent validity'. Let us suppose that a researcher has developed an achievement test in nutrition. The scores on this test may be compared with scores given by the nutrition teacher to the sample students. If the two tests show nearly the same result the concurrent validity of the researchers newly developed tool can be established.

In case of predictive validity, the measure of the outcome is termed 'criterion'. While estimating concurrent validity the newly developed tool is proposed as a substitute for the existing technique or method, and the information obtained through the existing technique acts as the criterion. Since in both the cases the information sought through the newly developed tool is related to a criterion, the two types of validation are also termed 'criterion-related validity'.

Construct validity

A construct is a trait, attribute or quality that cannot be observed directly, but that can be inferred from testing.

Construct validity is concerned with the extent to which a test measures a specific trait, attribute or construct. This type of validity is essential for those tests which are used to assess individuals on certain psychological traits and abilities. Examples of common constructs are anxiety, intelligence, motivation, attitude, critical thinking etc. Construct validity is established by relating a presumed measure of a construct with some behaviour that it is hypothesized to underlie.

Next, let us review reliability another characteristics of a good research tool.

7.3.2 Reliability

A tool used for data collection must be reliable, that is, it must have the ability to consistently yield the same results when it is repeatedly administered to the same individuals under the same conditions. For example, if an individual records his/her responses on various items of a questionnaire and thus provides a certain type of information, he/she should provide approximately the same type of responses when the questionnaire is administered to him/her on the second occasion. If an achievement test is administered to learners and then readministered after a gap of fifteen days without any special coaching in that subject, within these fifteen days, the learners must show similar range of scores on readministration of the test.

Repeated measure of an attribute, characteristic or a trait by a tool may provide different results. They may be due either to a real change in the individual's behaviour/attributes/measures or to the unreliability or inconsistency of the tool itself. If the variation in the results is due to a real change in attribute being studied, the reliability of the tool is not to be doubted. However, if the variation is due to the tool itself, then the tool is to be discarded.

Tools for Research

There are various procedures to assess the reliability of a tool. These include (i) the test-retest method, (ii) the alternate or parallel-form method, (iii) the split half method, and (iv) the rational equivalence method.

i) *The test-retest method*

In this method the same tool is re-administered to the same sample of population shortly after its first administration. The relationship or agreement between the information or data sought through the two administrations provides the measure of reliability of the tool. To illustrate, in a research study to assess the haemoglobin levels using the cyanamethemoglobin method, the researcher, may collect two samples from the same subject(s) and assess the agreement between the two samples and analyzed. The chief disadvantage of this method is that if the time between two administrations of the tool is short, the immediate memory effects, practice and the confidence induced by familiarity with the tool may give a wrong measure of its reliability. This is true for a tool designed to assess nutrition knowledge. On the other hand, if the interval is too long, the real changes in behaviour in terms of growth may underestimate the reliability of the tool.

However, this type of measurement is commonly used with questionnaires, observations, and interviews.

ii) *The equivalent or parallel-forms method*

This method requires that two equivalent or parallel forms of tool be prepared and administered to the same group of subjects. The items in these tests are parallel. Then, the results in terms of two sets of measures obtained by the use of the tool are correlated to measure the level of its reliability. An example of this method would be to assess the nutrient intake of a group of subjects using the 24-hour recall method and the food frequency questionnaire. The result obtained from these two sets of measures are correlated to measure the level of reliability of the tool under consideration.

In developing the parallel forms of a tool, care has to be taken to match the tool material with the content, the difficulty level and the form. The parallel-form method is widely used for determining reliability of a research tool. The reliability of psychological tests and attitude scales is usually estimated by this method.

iii) *The split-half method*

In this method, the tool is first divided into two equivalent 'halves'. If there are 50 items in a test, two equivalent halves are made of 25 items each. It may be done by having alternate items. The measure of the first half of the tool is correlated with the measure of the other half. The measures are correlated to find the reliability of tests and attitude scales. The main limitation of this method is that a tool can be divided into two halves in a number of ways and, thus, the estimate of the reliability may not have a unique value.

iv) *The rational equivalent method*

This method of measuring reliability is considered to be free from the limitations of the other methods discussed so far. Two forms of a tool are defined as equivalent when their corresponding contents are interchangeable.

Finally, let us review usability as a characteristic of a good research tool.

7.3.3 Usability

The usability of a tool depends on its objectivity, cost effectiveness, the time and effort required to administer it, and how easy it is to analyze and draw conclusions through its use.

A tool should yield objective information and results. In other words, the results should be independent of the personal judgment of the researcher. If it cannot yield objective

data, we say that it is not usable. If the tool can be administered in a short period of time, it is likely to gain cooperation of the subjects and save time of all those involved in its administration. The cost of construction and administration of the tool should be reasonable. The simplicity and ease of administration, the scores and interpretation are also important factors to be considered while selecting a tool, particularly, when the expert advice is not easily available. The tool should interest and fascinate the subjects so that it may gain their cooperation.

Check Your Progress Exercise 1

- 1) Explain the terms validity, reliability and usability in your own words.

.....

.....

.....

.....

.....

With the basic understanding regarding the characteristics of a good research tool, next we shall get to know about the various research tools and their uses.

7.4 TYPES OF TOOLS AND THEIR USES

Questionnaires, interviews, rating and attitude scales are the major data-gathering research tools. In the following sub-sections we shall discuss these research tools.

7.4.1 Questionnaire and Schedules

'Questionnaire' is a commonly used and frequently abused tool for gathering a variety of data. A questionnaire may include a series of questions for answer by an individual or group, with the aim of obtaining relevant data on the topic of research. This could be self-administered or could be administered by an interviewer as illustrated in Figure 7.1. The term 'schedule' is used when it contains a list of items on which information is to be collected. Let us learn about the types of questionnaires next.

Types of questionnaires

Questionnaires can be classified in various ways. Here we confine ourselves to *structured and unstructured* questionnaires.

Structured questionnaires are those which pose definite and concrete questions. They are prepared well in advance and not on the spot. Additional questions may be used only when there is a need to clarify vague or inadequate replies by respondents or when further details are needed. The form of questions may require responses which are either *closed* or *open*.

Close-ended questionnaires are used when categorized data are required. They include a set of questions to which a respondent can reply in a limited number of ways – 'yes', 'no', 'no-opinion', or an answer from a short list of possible responses. He/she is asked to put a tick (✓) mark in a space provided on the answer sheet or is requested to underline a response. Sometimes he/she is asked to insert brief answers of his/her own. The *open ended* responses, on the other hand, are free and spontaneous expressions by the respondent to the questions posed to him/her. The open-ended responses are used mainly for intensive study of a limited number of cases or preliminary exploration of new problems and situations. At times, the respondent is asked to write a descriptive essay and express his/her viewpoints or

report on details and events, without restrictions imposed as in the case of closed questions.

Unstructured questionnaires are frequently referred to as interview guides. They also aim at precision and contain definite issues that are covered while conducting an interview. Flexibility is the chief advantage of the unstructured questionnaire. It is designed to obtain viewpoints, opinions, attitudes and to show relationships between various types of information which might escape notice under more mechanical types of interrogation. No predetermined responses are provided; instead, free responses are solicited.

Next, we shall study the characteristics of a good questionnaire.



Figure 7.1: Filling up a questionnaire

Characteristics of a good questionnaire

The characteristics of a good questionnaire can be analyzed by its:

1. *Purpose*

A good questionnaire must serve two purposes. First, it must translate the objectives of an investigation into specific questions, the answers to which will provide the data necessary to test the hypotheses and explore the area defined by the objectives. Each question should relate the corresponding objective so that the response obtained can be analyzed and interpreted accordingly. The research objectives and specifications of the data required must precede the construction of questionnaire. Secondly, the questionnaire must motivate the respondents to communicate the required information. It is essential to include a courteous and carefully constructed covering letter to explain the purpose and importance of the study. Easy-to-follow instructions to record response and explanatory note where needed are always helpful. The covering letter should assure the respondent that delicate information will be held in strict confidence.

2. *Language*

The language of a good questionnaire should be concise and directed towards producing uniformity of understanding among the respondents. The vocabulary should be simple and within the easy grasp of the least intelligent of the group under study. The syntax should be clear and straightforward. Vague phrases and expressions should be avoided. Technical expressions should be used only if the inquiry is directed to a select group which is well-versed in the technical language used. Proverbs and quotations should be avoided. Subjective words, such as 'bad', 'good', 'fair' and the like do not lend

themselves to quantitative measurements nor qualitative analysis unless they are used for comparisons on a rating scale.

3. *Frame of Reference*

The respondent's frame of reference influences his/her answers. Complex questions that require the respondent to go through several steps of reasoning before answering are undesirable and have often resulted in misleading information. For example in a question like 'should an anganwadi worker modernise their pre-school activities', the work 'modernise' may have different connotations for different people.

Questions on controversial issues should be broken down into components, so that the researcher can determine the respondent's feelings about various aspects of the problem, including those which he/she refuses to comment upon. A series of specific questions is needed so as to uncover degrees of intensity of feeling or conviction.

The required answers should be within the informational domain of the respondents. For example, a question 'Do you include "enhancers" in your diet may not bring the desired response as many respondents (say pregnant women) may not know about this terminology. The length of the questions and statements used should be governed by a reliable estimate of the respondent's comprehension level.

4. *Arrangement of Questions*

The arrangement or ordering of questions should receive special attention. It should appear logical to the respondents. The questions placed first in the questionnaire should be the easiest to answer. 'Interest-generating' questions should be asked at the beginning. A proper sequence of questions proceeds from the general to specific, from simple to complex ones, from those that will create favourable attitude to those that may be somewhat delicate or sensitive.

5. *Length of the Questionnaire*

A questionnaire should not be longer than necessary. The total number of questions must not be too large to tire or bore the respondents. If too many questions are asked and the respondent becomes tired, the questions at the end of the series may not be well answered. If it is necessary to include a large number of questions, it is advisable to have separate questionnaires.

6. *Form of Response*

The form in which the responses are recorded must be integrated with the form of the questions. There should be no hesitation in asking for responses in different forms in the same questionnaire, since it is frequently found that one form is better than another for questions about different aspects of the same subject. Questions requiring answers like "Yes" or "No" are subject to least bias. These responses are easy to tabulate. However, they do not always yield sufficient information on the subject under study. In such cases, the use of multiple choice responses is desirable. Questions that present multiple choices to the respondent are effective when the choices are few and easy to follow.

Besides the characteristics mentioned above, one another aspects require consideration when working with questionnaires. It is generally considered essential that all questionnaires, schedules etc. are tested for their efficacy before they are finally used for all the main study. This is called pre-testing. Let us get to know about this.

Questionnaire Pre-testing

As already mentioned above, once the questionnaire is prepared and before they are finally used for the main study, it is important to pre-test the questionnaire. Why? Many unforeseen problems or lacunae can be detected by such an exercise. The tool thus can be accordingly adjusted and improved.

Pre-testing can be done on small number of subjects, (similar to the study subjects), before the actual study based on a specified design. This kind of study undertaken on a small number of subjects is called a pilot study. This can reveal that the questions are sufficiently clear or not, whether the length of the interview is within limit or not, or whether the instructions are adequate or not.

Thus we have learnt that pre-testing the questionnaire is an important aspect in the use of questionnaires as a research tool. Next, let us review the issue of handling non-response.

With this we end our study of questionnaires and schedules. Let us assess our understanding of this topic by answering the questions included in the check your progress exercise 2.

Now that we are aware of the qualities of a good questionnaire let us next focus on the uses of questionnaire. Questionnaires are useful tool for data collection. The various uses of questionnaire are highlighted herewith.

Uses of questionnaire

1. A questionnaire is a popular means of collecting different kinds of data in research. It is widely used in nutritional research to obtain information about certain conditions and practices, and to inquire into opinions of an individual or a group.
2. A questionnaire is administered personally either individually or to a group of individuals or is mailed to them to save a great deal of time and money in travel. In the former situation, the person administering the tool has an opportunity to establish rapport with the respondents, to explain the purpose of the study to the respondents and to explain the meaning of questions which may not be clear to them. In the latter situation, mailed questionnaire is mostly used when the individuals cannot be contacted personally. The range of administration of a mailed questionnaire may be national or international.
3. Questionnaires are used both to initiate a formal inquiry and also to supplement and check data previously accumulated. They may pertain to studies of nutrition/health problems, measurement of opinion on nutrition/health issues or events, studies of nutrition policies and changes, child welfare, and numerous other issues.

Though questionnaire is useful tool in research it has few inherent limitations as highlighted next.

Limitations of Questionnaires

The limitations of questionnaire include:

1. A questionnaire cannot be used with children and illiterates.
2. The return of the mailed questionnaire is often as low as 40 per cent to 50 per cent. As a result of this poor response, the data obtained are sometimes of limited validity. The respondents who return the questionnaires may not be representative of the entire group. It will make the sample a biased one and thus vitiate the findings.
3. Sometimes respondents may not like to respond in writing to questions of intimate and confidential nature or to questions involving controversial issues. For example, it has been experienced that people avoid questions related to marriage and sex and government servants avoid answering question about policy matters of the government.
4. It is sometimes difficult to formulate and phrase questions on certain complex and delicate problems.
5. There is no check on the respondent who misinterprets a question or gives incomplete or indefinite responses.

6. If the questionnaire is self-administered (by the respondent) the education and the attitudes of the respondents towards the survey can substantially influence the response.
7. Sometimes the respondent may modify his/her earlier/original responses to the questions when he/she finds that his/her responses to latter questions are contradicting the previous ones.

With this discussion we end our study of questionnaires. Let us assess our understanding of this topic by answering the questions included in the check your progress exercise 2.

Check Your Progress Exercise 2

- 1) Describe briefly the characteristics of a good questionnaire.

.....

.....

.....

.....

- 2) Differentiate between an open-ended and close-ended questionnaire.

.....

.....

.....

.....

- 3) List any two limitations and two uses of questionnaires.

.....

.....

.....

.....

7.4.2 Rating Scale

'Rating' is a term applied to an expression of opinion or judgment regarding some situation, object, character, or an attribute. In order terms, rating is the evaluation, assessment of something, in terms of quality, quantity or some contribution of both or 'Rating scale' refers to a 'scale' with a set of points which describe varying degrees of an attribute under investigation. Figure 7.2 illustrates the concept of rating.

Rating scales are broadly classified into five categories:

- i) numerical scales,
- ii) graphic scales,
- iii) standard scales,
- iv) rating by cumulative points, and
- v) forced choice ratings

We discuss them below in the same order.

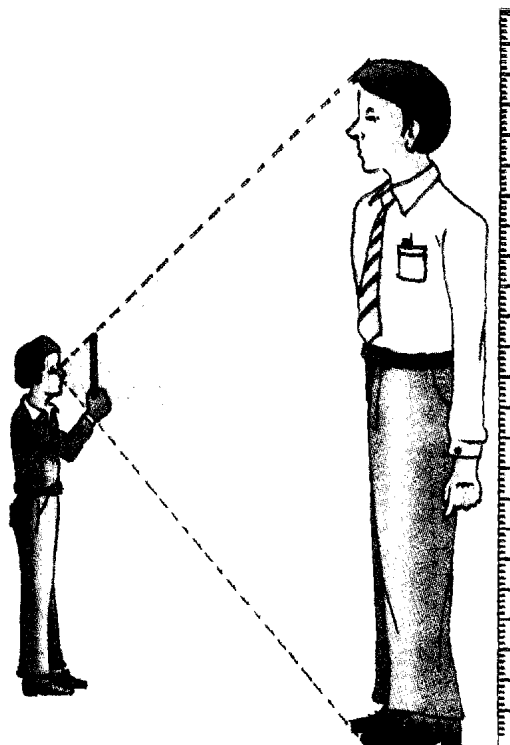


Figure 7.2: How tall are you

Numerical Scales

In a typical numerical scale, a sequence of defined numbers is supplied to the rater or the observer. He/she assigns to each stimulus to be rated, an appropriate number in line with these definitions or descriptions of the event or the stimulus. For example, the following scale may be used in obtaining ratings of the affective values of sensory testing of food products with individuals.

1. Most pleasant imaginable
2. Most pleasant
3. Extremely pleasant
4. Moderately pleasant
5. Mildly pleasant
6. Indifferent
7. Mildly unpleasant
8. Moderately unpleasant
9. Extremely unpleasant
10. Most unpleasant
11. Most unpleasant imaginable

The use of negative numbers is not favoured as those observers or raters who are not well versed in Algebra find it difficult to manage negative members.

Numerical rating scales are the easiest to be constructed. They are also the simplest in terms of handling the results. However, numerical scales have the limitations of biases.

Graphic Scales

The graphic scale is the most popular and the most widely used type of rating scale. In this scale a straight line is shown, vertically or horizontally, with various clues to

help the rater. The line is either segmented into units or continuous. If the line is segmented, the number of segments can be varied from case to case. Given below is an example of such a scale.

How effective was the nutrition education session?

Very Effective	Slightly Effective	Average	Slightly Ineffective	Very Ineffective
-------------------	-----------------------	---------	-------------------------	---------------------

There are many advantages in graphic scales. They are simple and easy to administer. Such scales are interesting to the rater and require little added motivation. However, scoring in the case of some formats of graphics scale is rather laborious.

Standard Scales

In standard scales a set of standards is presented to the rater. The standards are usually objects of the same kind to be rated with pre-established scale values. This type is like that of the scales for judging the quality of handwriting. The scales of handwriting provide several standard specimens that have previously been spread over on a common scale by some standardized procedure like equal-appearing intervals. With the help of these standard specimens, a new sample of handwriting can be equated to one of the standards, judged as being between two standards.

Rating by Cumulated Points

The unique and distinctive feature of rating by cumulative points is its immense use and ease of scoring. The rating score for an attribute, object or individual is the sum or average of the weighted or unweighted points. The 'check-list method' and the 'guess-who technique' belong to this category of rating. 'Check list methods' are applicable in the evaluation of the performance of an individual or an achievement scale in a job. The weights of 1 and -1 are assigned to every favourable and unfavourable trait, characteristics or attribute and the individual's score is the algebraic sum of the weights. In 'guess-who technique', some statements like "here is the one who is always doing the wrong things to make others sad", are constructed and each individual is asked to list all the members of his/her family who fitted such description, mentioning the same individual as many times as necessary. Each individual (in the family) scores a point for each favourable or unfavourable description applied to him/her, and the total score is the sum total of all such points.

Forced Choice Ratings

In 'forced-choice rating' methods, the rater is asked, not to say whether the rate has a certain trait or how much of it the rate has, but to essentially say whether he/she has some or one trait or another of a pair. For example instead of deciding whether an individuals' leadership qualities are superior or above average, it may be asked if the person

- exerts strong influence on his/her associates,
- is able to make others act, and
- asserts during functions.

With a brief review of the different rating scales let us get to know about the uses and the limitations of the rating scales.

Uses of Rating Scales

The uses of rating scales include:

- i) Rating methods consume much less time than other methods of scaling like 'pair comparison' and 'rank ordering'.

- ii) Rating methods are quite interesting to the raters, especially if graphic methods are used.
- iii) Best rating can be obtained by presenting one stimulus to a rater at a time.
- iv) Rating scales can be used with large numbers of stimuli to a rater at a time.
- v) Rating scales can be used with raters who have very little training for the purpose.
- vi) Rating methods can be used with large numbers of stimuli.
- vii) Rating scales have much wider range of application and can be used for tutor-ratings, personality ratings, school appraisal etc.

Limitations of Rating Scales

Rating scales have several limitations. Some of them are discussed as under:

- i) *Error of leniency.* There is a constant tendency among the raters to rate those whom they know well, higher than they should. Such raters are called 'easy raters'. Some raters become aware of their easy rating and consequently rate individuals lower than they should. Such raters are called 'hard raters'. The leniency error refers to a general and consistent tendency for a rater to rate too high or too low for whatever reasons.
- ii) *Effort of central tendency.* Most of the raters hesitate to rate the individuals on the extremes of the scales, instead they tend to rate the individuals on the middle of the scale. Obviously, the results get distorted.
- iii) *Halo-effect.* Halo-effect is an error which obscures the clusters of traits within an individual. The rater forms a general opinion about the person's merit and his/her ratings on specific traits are greatly influenced by this general impression. It results in a spurious positive correlation among the traits which are rated. If a learner likes a tutor, he/she will rate the tutor high on all traits without considering the meaning attached to a particular trait.
- iv) *The logical error.* The logical error is due to the fact that judges are likely to give similar ratings for traits which they feel are logically related to each other.
- v) *The contrast error.* The contrast error is due to a tendency of a rater to rate others in the opposite direction (contrasting) from himself/herself in a trait.
- vi) *The proximity error.* It has been seen that adjacent traits on a rating scale tend to inter-correlate higher than the remote ones, their degree of actual similarity being approximately equal. This error may be counteracted to some extent by placing similar traits farther apart and the dissimilar ones closer.

Next, let us focus on attitude scale.

7.4.3 Attitude Scale

'Attitude' is defined as the degree of positive or negative effect associated with a certain psychological entity. In other words it is the pre-disposition of an individual towards a entity – maybe an institution, practice, ideal, symbol, phrase, slogan, job or idea towards which people respond positively or negatively. The inquiry form that attempts to assess the attitude or belief of an individual is known as an opinionnaire or attitude scale depending on the way the questions are put and responses sought.

Types of Attitude Scales

Various scaling techniques have led to the development of different types of attitude scales which provide quick and convenient measure of attitudes. However, the method of 'equal-appearing intervals' (Thurstone Scales) and 'method of summated ratings' (Likert Scales) have been extensively used in attitude or opinion research. The attitude scales that are developed using these scaling techniques consists of a number of carefully edited and selected items called 'statements'.

To illustrate, to assess the attitudes of mothers regarding infant feeding practices, the statements can be constructed and attitudes rated as follows:

“It is essential that we feed colostrums (the first milk) to the babies”.

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

“Children should be fed frequently 4-5 times a day”

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

The method of ‘equal-appearing intervals’ was originally developed by Thurstone and Chave (1929). The attitude score of an individual obtained by this method has an absolute interpretation in terms of the psychological continuum of scale value of the statement making up the scale. If this score falls in the middle range of the psychological continuum, the attitude of the individual is described as “neutral”. If it falls towards the favourable end of the continuum, it is described as “favourable” and if it falls towards the unfavourable end, it is described as “unfavourable”.

In the ‘method of summated ratings’ developed by Likert, the item score is obtained by assigning arbitrary weights of 5, 4, 3, 2, and 1 for strongly agree (SA), agree (A), undecided (U), disagree (D), and strongly disagree (SD) respectively, for the statements favouring a point of view. On the other hand, the scoring weights of 1, 2, 3, 4 and 5 are given for the respective responses for statements opposing this point of view. An individual’s score on a particular attitude scale is the sum of his/her rating on all the items.

Uses of Attitude Scales

The attitude scales have various uses as highlighted herewith:

- i) Attitude scales are used to measure the degree of positive or negative feeling associated with any slogan, person, institution, practice etc.
- ii) Attitude scales are used in public-opinion-surveys in order to make some important and crucial decisions. Nutrition and health educators seek knowledge of public beliefs and attitudes. Besides the uses attitude scales have certain limitations. Let us get to know them.

Limitations of Attitude Scales

The process of assessing attitude with the help of attitude scales has various limitations.

- i) An individual may conceal his/her real attitude, and express socially acceptable opinions only.
- ii) An individual may not really know how he/she feels about social issues and he/she may never have given the idea a serious consideration.
- iii) An individual may not be able to express his/her attitude towards an abstract situation unless he/she actually confronts with it in his/her real life.

Check Your Progress Exercise 3

- 1) List the types of rating scale.

.....

.....

2) Briefly comment on the uses and limitations of rating scales.

.....
.....
.....
.....

3) Describe briefly the uses and limitations of attitude scales.

.....
.....
.....

7.4.4 Tests

A test, in the narrowest sense, connotes the presentation of a standard set of questions to be answered. We obtain a measure (a numerical value) of a characteristics or attribute of a person pertaining to his/her potential knowledge of, say, nutrition, dietetics, health concepts from his/her answers to such a series of questions.

Types of Tests

Tests may be classified in different ways, some of which are described as follows:

A. *On the basis of administration:* We have three categories under this classifications:

- i) *Power vs. Speed Tests:* A 'power test' is a test in which every subject has a chance to attempt each item of the test. It has no time limit and the subject goes on attempting test items till he/she can no longer continue successfully. On the other hand, a 'speed test' is defined as one in which no subject has enough time to attempt all items. Most entrance tests held for admission into various university courses are speed tests. A large number of items are given to be attempted within a fixed time.
- ii) *Individual vs. Group Tests:* The tests which are administered on one individual at a time are known as 'individual tests'. These tests are useful in situations where a precise and detailed assessment of some characteristics of an individual is desired.

The test which is administered to many subjects at the same time is termed as a 'group test'. These tests are particularly useful when large numbers of subjects have to be tested at the same time.

- iii) *Paper-pencil and Performance Tests:* 'Paper-pencil tests' require the subject to respond to the item by writing his/her replies. They pose questions in the form of sentences, or designs, and require the subject to record his/her answer either by underlining, ticking or encircling one of the alternative answers or by writing a word, phrase or sentence in the blank space provided for this purpose.

On the other hand, in 'performance tests', problems are presented in a concrete form and the subject is required to respond not by writing but by manipulating toys, blocks or picture cards, etc., depending on the level of the testtaker.

B) *On the basis of standardization:* On the basis of standardization tests can be classified into two categories:

- i) Non-standardized teacher made tests and
- ii) Standardized tests.

Teachers use their own tests in classroom situations to assess the achievement of learners in different subjects/disciplines. Such tests are designed for specific use and their reliability or validity are not established by careful statistical controls. By contrast, in standardized tests as used in research, each item and total score are carefully analyzed. The content, administration and scoring in these tests are standardized.

C) *On the basis of traits and abilities to be measured:* Tests can also be classified in terms of their purpose, that is, the types of abilities and traits they describe and claim to measure. By this standard, we may distinguish five major classes of tests, each with many sub-classes. But in nutritional research, we may use the following two classes of tests.

- i) *Tests of general mental ability or intelligence:* Tests of general mental ability measure general ability which enters into performance of all activities and which differs in magnitude from individual to individual. The items in such tests assess the subjects' ability to perceive relationships, solve problems and apply knowledge in a variety of ways. Intelligence tests are classified as verbal and non-verbal tests, paper-pencil and performance tests, speed and power tests and individual and group tests. While studying the relationship between anaemia and cognitive functions, perhaps we may use the intelligence tests.
- ii) *Tests of attainment or achievement:* Tests which are conducted to measure present performance vis-à-vis the skill or knowledge that has been acquired as a result of training are called attainment or achievement tests. They are designed to measure an individual's level of learning in a particular discipline, subject or course at the end of instruction. If an achievement test is developed to assess whether a student possesses required skills in terms of a specified criterion at a particular time, the test is designated as 'criterion-referenced or mastery test'. If, on the other hand, the test is to assess the student's relative position in a group, the test is called 'discriminatory or norm-referenced test'.

Achievement tests may be classified as traditional or essay-type, and new-type or objective tests. New type or objective tests include multiple-choice, true-false, completion, matching and short answer items. Such tests of attainment are normally used to assess nutrition knowledge among subjects in nutrition intervention programmes.

Finally let us review the uses and limitation of tests.

Uses of Tests

Tests help in indentifying such types of activities:

- i) They help in (a) providing knowledge concerning the subjects behaviour and intelligence (b) evaluating the degree to which the objectives have been achieved and (c) determining, evaluating and refining the instructional techniques.
- ii) Tests serve various purposes in the process of counseling and guidance particularly in the context of counseling matters/other subjects. Sometimes test results are used to confirm a learners ideas about his/her skills, abilities or personality characteristics.
- iii) Tests play an important role in research specially in areas like nutrition dietetics. For example, in validation studies, tests are used to predict, and to measure a variable or construct.

Limitation of Tests

- i) Tests of intelligence or special aptitude should not be considered as the absolute measures of pure intelligence or creative thinking because the performance in such tests is partly determined by one's background and schooling.

- ii) Tests measuring cognitive processes can hardly measure higher mental processes such as ability to discover scientific laws and principles.
- iv) The strength and depth of understanding of an individual and appreciative reactions are hardly measured by tests.

7.5 LET US SUM UP

In this unit, we have described the characteristics of a good research tool, various types of research tools and their uses and limitations.

- A good research tool should be valid, reliable and usable.
- Validity pertains to the truthfulness of a research tool; reliability is the consistency of the results when a research tool is repeatedly administered on the same individual under similar conditions; and usability relates to the objectivity, economy of time and cost, ease of administration and analysis of results, etc.
- Content validity, criterion-related validity and construct validity are the three main types of validity.
- Generally, in assessing reliability of a research tool, we usually make use of test-retest, parallel-form, split-half or rational equivalence method.
- Questionnaires, interviews, rating scales, attitude scales, and tests, are the main data-gathering research tools or techniques.
- Questionnaires consist of a series of questions dealing with psychological, social, educational and other related issues. Questionnaires are either structured or unstructured. A good questionnaire is specific in purpose, simple in language, logical in arrangement of questions and moderate in length. It is administered personally or mailed to individuals.
- Rating scales are used to obtain judgement on a set of points which describe varying degrees of an attribute under observation. Numerical scales, graphic scales, rating by cumulative points and forced choice ratings are the commonly used rating techniques. Leniency halo-effect, logical error, contrast error and proximity error are the major error types which are faced in using rating scales.
- Attitude scales are used to assess the attitude of an individual towards another individual, slogan, religion, institution, a different mode of teaching etc. They are mostly used in opinion surveys. Thurstone's method of 'equal appearing intervals' and Likert's method of 'summated ratings' are extensively used in the construction of attitude scales.

7.6 GLOSSARY

- Tools** : an instrument used for gathering data or information pertaining to certain attributes or characteristics of an individual or a group; issues relating to society, systems etc.
- Validity** : the extent to which a tool does the job for which it used. This term has different connotations for various types of tools and, thus, a different type of validity evidence is appropriate for each.
- i) **Content Validity:** it relates to the relevance of a content of a research tool pertaining to the objectives and nature of a research problem.

- ii) Criterion-related Validity: the extent to which measures or scores on the tool are in agreement with (concurrent validity) or predict (predictive validity) some given criterion measure.
- iii) Construct Validity: the extent to which a test measures relatively abstract psychological traits or constructs.

Reliability	:	the extent to which a tool is consistent in measuring whatever it may measure.
Usability	:	the extent to which tool is objective, easy to administer and cost effective. Printed form containing a set of questions of open and closed types with spaces for filling in responses by the respondent.
Structured Questionnaire	:	a questionnaire in which questions are framed with a view to limiting the variety of responses made by the respondent.
Unstructured Questionnaire	:	a questionnaire in which the questions are left open with a view to providing freedom to respondents for giving responses.
Rating	:	a term applied to an expression of opinion or judgement regarding some situation, object or character.
Rating Scale	:	a scale with a set of points which describe varying degrees of the dimension of an attribute under observation.
Attitude	:	the degree of positive or negative affect associated with some psychological object. By psychological object we mean any institution, ideal, symbol, phrase, slogan, job or idea towards which people can differ in their opinion.
Attitude Scale	:	a device which provides a measure of attitudes. It consists of a series of short but carefully formulated statements or propositions dealing with several selected aspects or many appropriate aspects of issues, institutions or groups of people under study.

7.7 ANSWERS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress Exercise 1

- 1) *Validity* in the case of a research tool is its relevance to the objectives and nature of a research problem.

Reliability is the consistency of a research tool in measuring whatever it measures.

Usability is the extent to which a tool is objective, easy to administer and cost effective.

Check Your Progress Exercise 2

- 1) A good questionnaire must

- indicate its objectives through the questions,
 - use a direct, concise and simple language,
 - include a series of 'why', 'what', 'when' and 'how' questions,
 - place questions in a logical order, and include a relatively small number of questions and demand several forms of responses.
- 2) When a list of possible or expected responses can be given against the question or items, the questionnaire is said to be close ended. On the other hand, when the response is to be recorded verbatim then the question or item is called open ended.
 - 3) Refer to sub-section 7.4.1 under the heading uses and limitations and answer the question in your own words.

Check Your Progress Exercise 3

- 1) Rating scales are broadly classified into five types
 - i) numerical scales,
 - ii) graphic scales,
 - iv) standard scales,
 - v) rating by cumulative points, and
 - vi) forced choice ratings
- 2) Refer to sub-section 7.4.2 and answer the question in your own words.
- 3) Attitude scales:
Uses
 - The degree of positive or negative feelings associated with a slogan, person, institution, etc. be measured easily using such scales.
 - They help in taking informal decisions related to industrial, political and/or educational matters effecting public opinion.Limitations
 - An individual may conceal the real attitudes.