

point and the third point indicates a higher degree as compared to the fourth and so on. Numbers for measuring the distinctions of degree in the attitudes/opinions are, thus, assigned to individuals corresponding to their scale-positions. All this is better understood when we talk about scaling technique(s). Hence the term 'scaling' is applied to the procedures for attempting to determine quantitative measures of subjective abstract concepts. Scaling has been defined as a "procedure for the assignment of numbers (or other symbols) to a property of objects in order to impart some of the characteristics of numbers to the properties in question."<sup>3</sup>

### Scale Classification Bases

The number assigning procedures or the scaling procedures may be broadly classified on one or more of the following bases: (a) subject orientation; (b) response form; (c) degree of subjectivity; (d) scale properties; (e) number of dimensions and (f) scale construction techniques. We take up each of these separately.

**(a) Subject orientation:** Under it a scale may be designed to measure characteristics of the respondent who completes it or to judge the stimulus object which is presented to the respondent. In respect of the former, we presume that the stimuli presented are sufficiently homogeneous so that the between-stimuli variation is small as compared to the variation among respondents. In the latter approach, we ask the respondent to judge some specific object in terms of one or more dimensions and we presume that the between-respondent variation will be small as compared to the variation among the different stimuli presented to respondents for judging.

**(b) Response form:** Under this we may classify the scales as categorical and comparative. Categorical scales are also known as rating scales. These scales are used when a respondent scores some object without direct reference to other objects. Under comparative scales, which are also known as ranking scales, the respondent is asked to compare two or more objects. In this sense the respondent may state that one object is superior to the other or that three models of pen rank in order 1, 2 and 3. The essence of ranking is, in fact, a relative comparison of a certain property of two or more objects.

**(c) Degree of subjectivity:** With this basis the scale data may be based on whether we measure subjective personal preferences or simply make non-preference judgements. In the former case, the respondent is asked to choose which person he favours or which solution he would like to see employed, whereas in the latter case he is simply asked to judge which person is more effective in some aspect or which solution will take fewer resources without reflecting any personal preference.

**(d) Scale properties:** Considering scale properties, one may classify the scales as nominal, ordinal, interval and ratio scales. Nominal scales merely classify without indicating order, distance or unique origin. Ordinal scales indicate magnitude relationships of 'more than' or 'less than', but indicate no distance or unique origin. Interval scales have both order and distance values, but no unique origin. Ratio scales possess all these features.

**(e) Number of dimensions:** In respect of this basis, scales can be classified as 'unidimensional' and 'multidimensional' scales. Under the former we measure only one attribute of the respondent or object, whereas multidimensional scaling recognizes that an object might be described better by using the concept of an attribute space of 'n' dimensions, rather than a single-dimension continuum.

<sup>3</sup> Bernard S. Phillips, *Social Research Strategy and Tactics*, 2nd ed., p. 205.

**(f) Scale construction techniques:** Following are the five main techniques by which scales can be developed.

- (i) *Arbitrary approach:* It is an approach where scale is developed on *ad hoc* basis. This is the most widely used approach. It is presumed that such scales measure the concepts for which they have been designed, although there is little evidence to support such an assumption.
- (ii) *Consensus approach:* Here a panel of judges evaluate the items chosen for inclusion in the instrument in terms of whether they are relevant to the topic area and unambiguous in implication.
- (iii) *Item analysis approach:* Under it a number of individual items are developed into a test which is given to a group of respondents. After administering the test, the total scores are calculated for every one. Individual items are then analysed to determine which items discriminate between persons or objects with high total scores and those with low scores.
- (iv) *Cumulative scales* are chosen on the basis of their conforming to some ranking of items with ascending and descending discriminating power. For instance, in such a scale the endorsement of an item representing an extreme position should also result in the endorsement of all items indicating a less extreme position.
- (v) *Factor scales* may be constructed on the basis of intercorrelations of items which indicate that a common factor accounts for the relationship between items. This relationship is typically measured through factor analysis method.

### Important Scaling Techniques

We now take up some of the important scaling techniques often used in the context of research specially in context of social or business research.

**Rating scales:** The rating scale involves qualitative description of a limited number of aspects of a thing or of traits of a person. When we use rating scales (or categorical scales), we judge an object in absolute terms against some specified criteria i.e., we judge properties of objects without reference to other similar objects. These ratings may be in such forms as “like-dislike”, “above average, average, below average”, or other classifications with more categories such as “like very much—like some what—neutral—dislike somewhat—dislike very much”; “excellent—good—average—below average—poor”, “always—often—occasionally—rarely—never”, and so on. There is no specific rule whether to use a two-points scale, three-points scale or scale with still more points. In practice, three to seven points scales are generally used for the simple reason that more points on a scale provide an opportunity for greater sensitivity of measurement.

Rating scale may be either a graphic rating scale or an itemized rating scale.

- (i) *The graphic rating scale* is quite simple and is commonly used in practice. Under it the various points are usually put along the line to form a continuum and the rater indicates his rating by simply making a mark (such as ✓) at the appropriate point on a line that runs from one extreme to the other. Scale-points with brief descriptions may be indicated along the line, their function being to assist the rater in performing his job. The following is an example of five-points graphic rating scale when we wish to ascertain people’s liking or disliking any product:

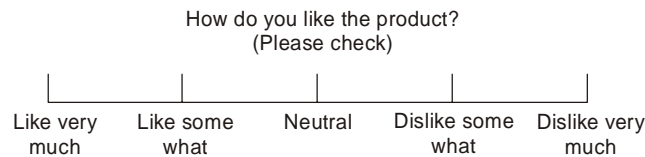


Fig. 5.1

This type of scale has several limitations. The respondents may check at almost any position along the line which fact may increase the difficulty of analysis. The meanings of the terms like “very much” and “some what” may depend upon respondent’s frame of reference so much so that the statement might be challenged in terms of its equivalency. Several other rating scale variants (e.g., boxes replacing line) may also be used.

- (ii) The *itemized rating scale* (also known as numerical scale) presents a series of statements from which a respondent selects one as best reflecting his evaluation. These statements are ordered progressively in terms of more or less of some property. An example of itemized scale can be given to illustrate it.

Suppose we wish to inquire as to how well does a worker get along with his fellow workers? In such a situation we may ask the respondent to select one, to express his opinion, from the following:

- He is almost always involved in some friction with a fellow worker.
- He is often at odds with one or more of his fellow workers.
- He sometimes gets involved in friction.
- He infrequently becomes involved in friction with others.
- He almost never gets involved in friction with fellow workers.

The chief merit of this type of scale is that it provides more information and meaning to the rater, and thereby increases reliability. This form is relatively difficult to develop and the statements may not say exactly what the respondent would like to express.

Rating scales have certain good points. The results obtained from their use compare favourably with alternative methods. They require less time, are interesting to use and have a wide range of applications. Besides, they may also be used with a large number of properties or variables. But their value for measurement purposes depends upon the assumption that the respondents can and do make good judgements. If the respondents are not very careful while rating, errors may occur. Three types of errors are common viz., the error of leniency, the error of central tendency and the error of halo effect. The error of leniency occurs when certain respondents are either easy raters or hard raters. When raters are reluctant to give extreme judgements, the result is the error of central tendency. The error of halo effect or the systematic bias occurs when the rater carries over a generalised impression of the subject from one rating to another. This sort of error takes place when we conclude for example, that a particular report is good because we like its form or that someone is intelligent because he agrees with us or has a pleasing personality. In other words, halo effect is likely to appear when the rater is asked to rate many factors, on a number of which he has no evidence for judgement.

*Ranking scales:* Under ranking scales (or comparative scales) we make relative judgements against other similar objects. The respondents under this method directly compare two or more objects and make choices among them. There are two generally used approaches of ranking scales viz.

**(a) Method of paired comparisons:** Under it the respondent can express his attitude by making a choice between two objects, say between a new flavour of soft drink and an established brand of drink. But when there are more than two stimuli to judge, the number of judgements required in a paired comparison is given by the formula:

$$N = \frac{n(n-1)}{2}$$

where  $N$  = number of judgements

$n$  = number of stimuli or objects to be judged.

For instance, if there are ten suggestions for bargaining proposals available to a workers union, there are 45 paired comparisons that can be made with them. When  $N$  happens to be a big figure, there is the risk of respondents giving ill considered answers or they may even refuse to answer. We can reduce the number of comparisons per respondent either by presenting to each one of them only a sample of stimuli or by choosing a few objects which cover the range of attractiveness at about equal intervals and then comparing all other stimuli to these few standard objects. Thus, paired-comparison data may be treated in several ways. If there is substantial consistency, we will find that if  $X$  is preferred to  $Y$ , and  $Y$  to  $Z$ , then  $X$  will consistently be preferred to  $Z$ . If this is true, we may take the total number of preferences among the comparisons as the score for that stimulus.

It should be remembered that paired comparison provides ordinal data, but the same may be converted into an interval scale by the method of the *Law of Comparative Judgement* developed by L.L. Thurstone. This technique involves the conversion of frequencies of preferences into a table of proportions which are then transformed into  $Z$  matrix by referring to the table of area under the normal curve. J.P. Guilford in his book “Psychometric Methods” has given a procedure which is relatively easier. The method is known as the *Composite Standard Method* and can be illustrated as under:

Suppose there are four proposals which some union bargaining committee is considering. The committee wants to know how the union membership ranks these proposals. For this purpose a sample of 100 members might express the views as shown in the following table:

**Table 5.1:** Response Patterns of 100 Members’ Paired Comparisons of 4 Suggestions for Union Bargaining Proposal Priorities

	Suggestion			
	A	B	C	D
A	–	65*	32	20
B	40	–	38	42
C	45	50	–	70
D	80	20	98	–
TOTAL:	165	135	168	132

\*Read as 65 members preferred suggestion B to suggestion A.

Contd.

Rank order	2	3	1	4
$M_p$	0.5375	0.4625	0.5450	0.4550
$Z_j$	0.09	(-).09	0.11	(-).11
$R_j$	0.20	0.02	0.22	0.00

Comparing the total number of preferences for each of the four proposals, we find that *C* is the most popular, followed by *A*, *B* and *D* respectively in popularity. The rank order shown in the above table explains all this.

By following the composite standard method, we can develop an interval scale from the paired-comparison ordinal data given in the above table for which purpose we have to adopt the following steps in order:

- (i) Using the data in the above table, we work out the column mean with the help of the formula given below:

$$M_p = \frac{C + .5(N)}{nN} = \frac{165 + .5(100)}{4(100)} = .5375$$

where

$M_p$  = the mean proportion of the columns

$C$  = the total number of choices for a given suggestion

$n$  = number of stimuli (proposals in the given problem)

$N$  = number of items in the sample.

The column means have been shown in the  $M_p$  row in the above table.

- (ii) The  $Z$  values for the  $M_p$  are secured from the table giving the area under the normal curve. When the  $M_p$  value is less than .5, the  $Z$  value is negative and for all  $M_p$  values higher than .5, the  $Z$  values are positive.\* These  $Z$  values are shown in  $Z_j$  row in the above table.
- (iii) As the  $Z_j$  values represent an interval scale, zero is an arbitrary value. Hence we can eliminate negative scale values by giving the value of zero to the lowest scale value (this being (-).11 in our example which we shall take equal to zero) and then adding the absolute value of this lowest scale value to all other scale items. This scale has been shown in  $R_j$  row in the above table.

Graphically we can show this interval scale that we have derived from the paired-comparison data using the composite standard method as follows:

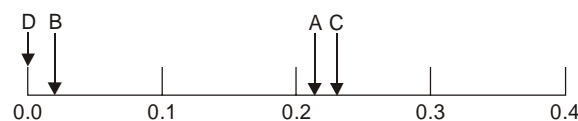


Fig. 5.2

\*To use Normal curve area table for this sort of transformation, we must subtract 0.5 from all  $M_p$  values which exceed .5 to secure the values with which to enter the normal curve area table for which  $Z$  values can be obtained. For all  $M_p$  values of less than .5 we must subtract all such values from 0.5 to secure the values with which to enter the normal curve area table for which  $Z$  values can be obtained but the  $Z$  values in this situation will be with negative sign.

**(b) Method of rank order:** Under this method of comparative scaling, the respondents are asked to rank their choices. This method is easier and faster than the method of paired comparisons stated above. For example, with 10 items it takes 45 pair comparisons to complete the task, whereas the method of rank order simply requires ranking of 10 items only. The problem of transitivity (such as *A* prefers to *B*, *B* to *C*, but *C* prefers to *A*) is also not there in case we adopt method of rank order. Moreover, a complete ranking at times is not needed in which case the respondents may be asked to rank only their first, say, four choices while the number of overall items involved may be more than four, say, it may be 15 or 20 or more. To secure a simple ranking of all items involved we simply total rank values received by each item. There are methods through which we can as well develop an interval scale of these data. But then there are limitations of this method. The first one is that data obtained through this method are ordinal data and hence rank ordering is an ordinal scale with all its limitations. Then there may be the problem of respondents becoming careless in assigning ranks particularly when there are many (usually more than 10) items.

### Scale Construction Techniques

In social science studies, while measuring attitudes of the people we generally follow the technique of preparing the opinionnaire\* (or attitude scale) in such a way that the score of the individual responses assigns him a place on a scale. Under this approach, the respondent expresses his agreement or disagreement with a number of statements relevant to the issue. While developing such statements, the researcher must note the following two points:

- (i) That the statements must elicit responses which are psychologically related to the attitude being measured;
- (ii) That the statements need be such that they discriminate not merely between extremes of attitude but also among individuals who differ slightly.

Researchers must as well be aware that inferring attitude from what has been recorded in opinionnaires has several limitations. People may conceal their attitudes and express socially acceptable opinions. They may not really know how they feel about a social issue. People may be unaware of their attitude about an abstract situation; until confronted with a real situation, they may be unable to predict their reaction. Even behaviour itself is at times not a true indication of attitude. For instance, when politicians kiss babies, their behaviour may not be a true expression of affection toward infants. Thus, there is no sure method of measuring attitude; we only try to measure the expressed opinion and then draw inferences from it about people's real feelings or attitudes.

With all these limitations in mind, psychologists and sociologists have developed several scale construction techniques for the purpose. The researcher should know these techniques so as to develop an appropriate scale for his own study. Some of the important approaches, along with the corresponding scales developed under each approach to measure attitude are as follows:

\* An information form that attempts to measure the attitude or belief of an individual is known as opinionnaire.

**Table 5.2:** Different Scales for Measuring Attitudes of People

<i>Name of the scale construction approach</i>	<i>Name of the scale developed</i>
1. Arbitrary approach	Arbitrary scales
2. Consensus scale approach	Differential scales (such as Thurstone Differential scale)
3. Item analysis approach	Summated scales (such as Likert Scale)
4. Cumulative scale approach	Cumulative scales (such as Guttman's Scalogram)
5. Factor analysis approach	Factor scales (such as Osgood's Semantic Differential, Multi-dimensional Scaling, etc.)

A brief description of each of the above listed scales will be helpful.

### **Arbitrary Scales**

Arbitrary scales are developed on *ad hoc* basis and are designed largely through the researcher's own subjective selection of items. The researcher first collects few statements or items which he believes are unambiguous and appropriate to a given topic. Some of these are selected for inclusion in the measuring instrument and then people are asked to check in a list the statements with which they agree.

The chief merit of such scales is that they can be developed very easily, quickly and with relatively less expense. They can also be designed to be highly specific and adequate. Because of these benefits, such scales are widely used in practice.

At the same time there are some limitations of these scales. The most important one is that we do not have objective evidence that such scales measure the concepts for which they have been developed. We have simply to rely on researcher's insight and competence.

### **Differential Scales (or Thurstone-type Scales)**

The name of L.L. Thurstone is associated with differential scales which have been developed using consensus scale approach. Under such an approach the selection of items is made by a panel of judges who evaluate the items in terms of whether they are relevant to the topic area and unambiguous in implication. The detailed procedure is as under:

- (a) The researcher gathers a large number of statements, usually twenty or more, that express various points of view toward a group, institution, idea, or practice (i.e., statements belonging to the topic area).
- (b) These statements are then submitted to a panel of judges, each of whom arranges them in eleven groups or piles ranging from one extreme to another in position. Each of the judges is requested to place generally in the first pile the statements which he thinks are most unfavourable to the issue, in the second pile to place those statements which he thinks are next most unfavourable and he goes on doing so in this manner till in the eleventh pile he puts the statements which he considers to be the most favourable.
- (c) This sorting by each judge yields a composite position for each of the items. In case of marked disagreement between the judges in assigning a position to an item, that item is discarded.

- (d) For items that are retained, each is given its median scale value between one and eleven as established by the panel. In other words, the scale value of any one statement is computed as the 'median' position to which it is assigned by the group of judges.
- (e) A final selection of statements is then made. For this purpose a sample of statements, whose median scores are spread evenly from one extreme to the other is taken. The statements so selected, constitute the final scale to be administered to respondents. The position of each statement on the scale is the same as determined by the judges.

After developing the scale as stated above, the respondents are asked during the administration of the scale to check the statements with which they agree. The median value of the statements that they check is worked out and this establishes their score or quantifies their opinion. It may be noted that in the actual instrument the statements are arranged in random order of scale value. If the values are valid and if the opinionnaire deals with only one attitude dimension, the typical respondent will choose one or several contiguous items (in terms of scale values) to reflect his views. However, at times divergence may occur when a statement appears to tap a different attitude dimension.

The Thurstone method has been widely used for developing differential scales which are utilised to measure attitudes towards varied issues like war, religion, etc. Such scales are considered most appropriate and reliable when used for measuring a single attitude. But an important deterrent to their use is the cost and effort required to develop them. Another weakness of such scales is that the values assigned to various statements by the judges may reflect their own attitudes. The method is not completely objective; it involves ultimately subjective decision process. Critics of this method also opine that some other scale designs give more information about the respondent's attitude in comparison to differential scales.

### **Summated Scales (or Likert-type Scales)**

Summated scales (or Likert-type scales) are developed by utilizing the item analysis approach wherein a particular item is evaluated on the basis of how well it discriminates between those persons whose total score is high and those whose score is low. Those items or statements that best meet this sort of discrimination test are included in the final instrument.

Thus, summated scales consist of a number of statements which express either a favourable or unfavourable attitude towards the given object to which the respondent is asked to react. The respondent indicates his agreement or disagreement with each statement in the instrument. Each response is given a numerical score, indicating its favourableness or unfavourableness, and the scores are totalled to measure the respondent's attitude. In other words, the overall score represents the respondent's position on the continuum of favourable-unfavourableness towards an issue.

Most frequently used summated scales in the study of social attitudes follow the pattern devised by Likert. For this reason they are often referred to as Likert-type scales. In a Likert scale, the respondent is asked to respond to each of the statements in terms of several degrees, usually five degrees (but at times 3 or 7 may also be used) of agreement or disagreement. For example, when asked to express opinion whether one considers his job quite pleasant, the respondent may respond in any one of the following ways: (i) strongly agree, (ii) agree, (iii) undecided, (iv) disagree, (v) strongly disagree.



We find that these five points constitute the scale. At one extreme of the scale there is strong agreement with the given statement and at the other, strong disagreement, and between them lie intermediate points. We may illustrate this as under:

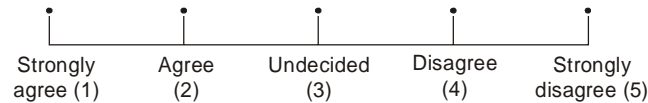


Fig. 5.3

Each point on the scale carries a score. Response indicating the least favourable degree of job satisfaction is given the least score (say 1) and the most favourable is given the highest score (say 5). These score—values are normally not printed on the instrument but are shown here just to indicate the scoring pattern. The Likert scaling technique, thus, assigns a scale value to each of the five responses. The same thing is done in respect of each and every statement in the instrument. This way the instrument yields a total score for each respondent, which would then measure the respondent's favourableness toward the given point of view. If the instrument consists of, say 30 statements, the following score values would be revealing.

$30 \times 5 = 150$  Most favourable response possible

$30 \times 3 = 90$  A neutral attitude

$30 \times 1 = 30$  Most unfavourable attitude.

The scores for any individual would fall between 30 and 150. If the score happens to be above 90, it shows favourable opinion to the given point of view, a score of below 90 would mean unfavourable opinion and a score of exactly 90 would be suggestive of a neutral attitude.

**Procedure:** The procedure for developing a Likert-type scale is as follows:

- (i) As a first step, the researcher collects a large number of statements which are relevant to the attitude being studied and each of the statements expresses definite favourableness or unfavourableness to a particular point of view or the attitude and that the number of favourable and unfavourable statements is approximately equal.
- (ii) After the statements have been gathered, a trial test should be administered to a number of subjects. In other words, a small group of people, from those who are going to be studied finally, are asked to indicate their response to each statement by checking one of the categories of agreement or disagreement using a five point scale as stated above.
- (iii) The response to various statements are scored in such a way that a response indicative of the most favourable attitude is given the highest score of 5 and that with the most unfavourable attitude is given the lowest score, say, of 1.
- (iv) Then the total score of each respondent is obtained by adding his scores that he received for separate statements.
- (v) The next step is to array these total scores and find out those statements which have a high discriminatory power. For this purpose, the researcher may select some part of the highest and the lowest total scores, say the top 25 per cent and the bottom 25 per cent. These two extreme groups are interpreted to represent the most favourable and the least favourable attitudes and are used as criterion groups by which to evaluate individual statements. This

way we determine which statements consistently correlate with low favourability and which with high favourability.

- (vi) Only those statements that correlate with the total test should be retained in the final instrument and all others must be discarded from it.

**Advantages:** The Likert-type scale has several advantages. Mention may be made of the important ones.

- (a) It is relatively easy to construct the Likert-type scale in comparison to Thurstone-type scale because Likert-type scale can be performed without a panel of judges.
- (b) Likert-type scale is considered more reliable because under it respondents answer each statement included in the instrument. As such it also provides more information and data than does the Thurstone-type scale.
- (c) Each statement, included in the Likert-type scale, is given an empirical test for discriminating ability and as such, unlike Thurstone-type scale, the Likert-type scale permits the use of statements that are not manifestly related (to have a direct relationship) to the attitude being studied.
- (d) Likert-type scale can easily be used in respondent-centred and stimulus-centred studies i.e., through it we can study how responses differ between people and how responses differ between stimuli.
- (e) Likert-type scale takes much less time to construct, it is frequently used by the students of opinion research. Moreover, it has been reported in various research studies\* that there is high degree of correlation between Likert-type scale and Thurstone-type scale.

**Limitations:** There are several limitations of the Likert-type scale as well. One important limitation is that, with this scale, we can simply examine whether respondents are more or less favourable to a topic, but we cannot tell how much more or less they are. There is no basis for belief that the five positions indicated on the scale are equally spaced. The interval between ‘strongly agree’ and ‘agree’, may not be equal to the interval between “agree” and “undecided”. This means that Likert scale does not rise to a stature more than that of an ordinal scale, whereas the designers of Thurstone scale claim the Thurstone scale to be an interval scale. One further disadvantage is that often the total score of an individual respondent has little clear meaning since a given total score can be secured by a variety of answer patterns. It is unlikely that the respondent can validly react to a short statement on a printed form in the absence of real-life qualifying situations. Moreover, there “remains a possibility that people may answer according to what they think they should feel rather than how they do feel.”<sup>4</sup> This particular weakness of the Likert-type scale is met by using a cumulative scale which we shall take up later in this chapter.

In spite of all the limitations, the Likert-type summated scales are regarded as the most useful in a situation wherein it is possible to compare the respondent’s score with a distribution of scores from some well defined group. They are equally useful when we are concerned with a programme of

\*A.L. Edwards and K.C. Kenney, “A comparison of the Thurstone and Likert techniques of attitude scale construction”, *Journal of Applied Psychology*, 30, 72–83, 1946.

<sup>4</sup>John W. Best and James V. Kahn, “Research in Education”, 5 ed., Prentice-Hall of India Pvt. Ltd., New Delhi, 1986, p. 183.

change or improvement in which case we can use the scales to measure attitudes before and after the programme of change or improvement in order to assess whether our efforts have had the desired effects. We can as well correlate scores on the scale to other measures without any concern for the absolute value of what is favourable and what is unfavourable. All this accounts for the popularity of Likert-type scales in social studies relating to measuring of attitudes.

**Cumulative scales:** Cumulative scales or Louis Guttman's scalogram analysis, like other scales, consist of series of statements to which a respondent expresses his agreement or disagreement. The special feature of this type of scale is that statements in it form a cumulative series. This, in other words, means that the statements are related to one another in such a way that an individual, who replies favourably to say item No. 3, also replies favourably to items No. 2 and 1, and one who replies favourably to item No. 4 also replies favourably to items No. 3, 2 and 1, and so on. This being so an individual whose attitude is at a certain point in a cumulative scale will answer favourably all the items on one side of this point, and answer unfavourably all the items on the other side of this point. The individual's score is worked out by counting the number of points concerning the number of statements he answers favourably. If one knows this total score, one can estimate as to how a respondent has answered individual statements constituting cumulative scales. The major scale of this type of cumulative scales is the Guttman's scalogram. We attempt a brief description of the same below.

The technique developed by Louis Guttman is known as scalogram analysis, or at times simply 'scale analysis'. Scalogram analysis refers to the procedure for determining whether a set of items forms a unidimensional scale. A scale is said to be unidimensional if the responses fall into a pattern in which endorsement of the item reflecting the extreme position results also in endorsing all items which are less extreme. Under this technique, the respondents are asked to indicate in respect of each item whether they agree or disagree with it, and if these items form a unidimensional scale, the response pattern will be as under:

**Table 5.3:** Response Pattern in Scalogram Analysis

	<i>Item Number</i>				<i>Respondent Score</i>
	4	3	2	1	
X	X	X	X	X	4
-	-	X	X	X	3
-	-	-	X	X	2
-	-	-	-	X	1
-	-	-	-	-	0

X = Agree

- = Disagree

A score of 4 means that the respondent is in agreement with all the statements which is indicative of the most favourable attitude. But a score of 3 would mean that the respondent is not agreeable to item 4, but he agrees with all others. In the same way one can interpret other values of the respondents' scores. This pattern reveals that the universe of content is scalable.

**Procedure:** The procedure for developing a scalogram can be outlined as under:

- (a) The universe of content must be defined first of all. In other words, we must lay down in clear terms the issue we want to deal within our study.
- (b) The next step is to develop a number of items relating the issue and to eliminate by inspection the items that are ambiguous, irrelevant or those that happen to be too extreme items.
- (c) The third step consists in pre-testing the items to determine whether the issue at hand is scalable (The pretest, as suggested by Guttman, should include 12 or more items, while the final scale may have only 4 to 6 items. Similarly, the number of respondents in a pretest may be small, say 20 or 25 but final scale should involve relatively more respondents, say 100 or more).

In a pretest the respondents are asked to record their opinions on all selected items using a Likert-type 5-point scale, ranging from 'strongly agree' to 'strongly disagree'. The strongest favourable response is scored as 5, whereas the strongest unfavourable response as 1. The total score can thus range, if there are 15 items in all, from 75 for most favourable to 15 for the least favourable.

Respondent opinionnaires are then arrayed according to total score for analysis and evaluation. If the responses of an item form a cumulative scale, its response category scores should decrease in an orderly fashion as indicated in the above table. Failure to show the said decreasing pattern means that there is overlapping which shows that the item concerned is not a good cumulative scale item i.e., the item has more than one meaning. Sometimes the overlapping in category responses can be reduced by combining categories. After analysing the pretest results, a few items, say 5 items, may be chosen.

- (d) The next step is again to total the scores for the various opinionnaires, and to rearrange them to reflect any shift in order, resulting from reducing the items, say, from 15 in pretest to, say, 5 for the final scale. The final pretest results may be tabulated in the form of a table given in Table 5.4.

**Table 5.4:** The Final Pretest Results in a Scalogram Analysis\*

Scale type	Item					Errors per case	Number of cases	Number of errors
	5	12	3	10	7			
5 (perfect)	X	X	X	X	X	0	7	0
4 (perfect)	–	X	X	X	X	0	3	0
(nonscale)	–	X	–	X	X	1	1	1
(nonscale)	–	X	X	–	X	1	2	2
3 (perfect)	–	–	X	X	X	0	5	0
2 (perfect)	–	–	–	X	X	0	2	0
1 (perfect)	–	–	–	–	X	0	1	0
(nonscale)	–	–	X	–	–	2	1	2
(nonscale)	–	–	X	–	–	2	1	2
0 (perfect)	–	–	–	–	–	0	2	0
			<i>n</i> = 5				<i>N</i> = 25	<i>e</i> = 7

\* (Figures in the table are arbitrary and have been used to explain the tabulation process only.)

The table shows that five items (numbering 5, 12, 3, 10 and 7) have been selected for the final scale. The number of respondents is 25 whose responses to various items have been tabulated along with the number of errors. Perfect scale types are those in which the respondent's answers fit the pattern that would be reproduced by using the person's total score as a guide. *Non-scale types* are those in which the category pattern differs from that expected from the respondent's total score i.e., non-scale cases have deviations from unidimensionality or errors. Whether the items (or series of statements) selected for final scale may be regarded a perfect cumulative (or a unidimensional scale), we have to examine on the basis of the coefficient of reproducibility. Guttman has set 0.9 as the level of minimum reproducibility in order to say that the scale meets the test of unidimensionality. He has given the following formula for measuring the level of reproducibility:

$$\text{Guttman's Coefficient of Reproducibility} = 1 - e/n(N)$$

where  $e$  = number of errors

$n$  = number of items

$N$  = number of cases

For the above table figures,

$$\text{Coefficient of Reproducibility} = 1 - 7/5(25) = .94$$

This shows that items number 5, 12, 3, 10 and 7 in this order constitute the cumulative or unidimensional scale, and with this we can reproduce the responses to each item, knowing only the total score of the respondent concerned.

Scalogram, analysis, like any other scaling technique, has several advantages as well as limitations. One advantage is that it assures that only a single dimension of attitude is being measured. Researcher's subjective judgement is not allowed to creep in the development of scale since the scale is determined by the replies of respondents. Then, we require only a small number of items that make such a scale easy to administer. Scalogram analysis can appropriately be used for personal, telephone or mail surveys. The main difficulty in using this scaling technique is that in practice perfect cumulative or unidimensional scales are very rarely found and we have only to use its approximation testing it through coefficient of reproducibility or examining it on the basis of some other criteria. This method is not a frequently used method for the simple reason that its development procedure is tedious and complex. Such scales hardly constitute a reliable basis for assessing attitudes of persons towards complex objects for predicting the behavioural responses of individuals towards such objects. Conceptually, this analysis is a bit more difficult in comparison to other scaling methods.

### Factor Scales\*

Factor scales are developed through factor analysis or on the basis of intercorrelations of items which indicate that a common factor accounts for the relationships between items. Factor scales are particularly "useful in uncovering latent attitude dimensions and approach scaling through the concept of multiple-dimension attribute space."<sup>5</sup> More specifically the two problems viz., how to deal

\* A detailed study of the factor scales and particularly the statistical procedures involved in developing factor scales is beyond the scope of this book. As such only an introductory idea of factor scales is presented here.

<sup>5</sup>C. William Emory, *Business Research Methods*, p. 264–65.

appropriately with the universe of content which is multi-dimensional and how to uncover underlying (latent) dimensions which have not been identified, are dealt with through factor scales. An important factor scale based on factor analysis is *Semantic Differential (S.D.)* and the other one is *Multidimensional Scaling*. We give below a brief account of these factor scales.

*Semantic differential scale*: Semantic differential scale or the S.D. scale developed by Charles E. Osgood, G.J. Suci and P.H. Tannenbaum (1957), is an attempt to measure the psychological meanings of an object to an individual. This scale is based on the presumption that an object can have different dimensions of connotative meanings which can be located in multidimensional property space, or what can be called the semantic space in the context of S.D. scale. This scaling consists of a set of bipolar rating scales, usually of 7 points, by which one or more respondents rate one or more concepts on each scale item. For instance, the S.D. scale items for analysing candidates for leadership position may be shown as under:

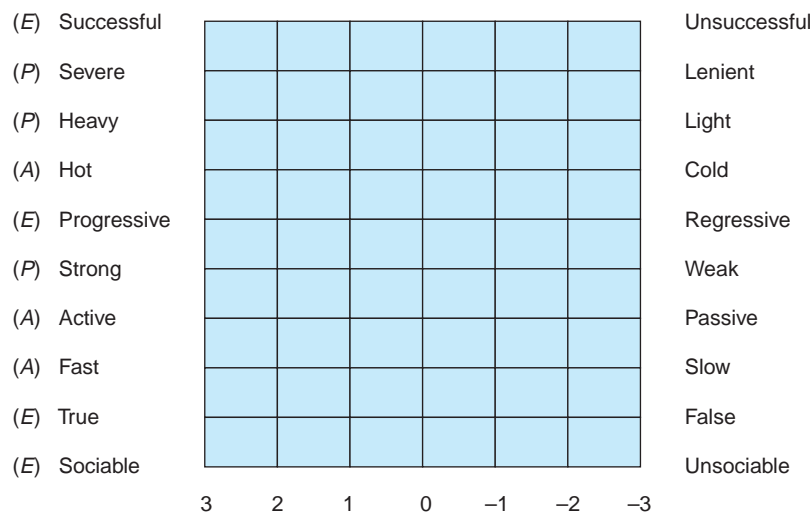


Fig. 5.4

Candidates for leadership position (along with the concept—the ‘ideal’ candidate) may be compared and we may score them from +3 to –3 on the basis of the above stated scales. (The letters, *E, P, A* showing the relevant factor viz., evaluation, potency and activity respectively, written along the left side are not written in actual scale. Similarly the numeric values shown are also not written in actual scale.)

Osgood and others did produce a list of some adjective pairs for attitude research purposes and concluded that semantic space is multidimensional rather than unidimensional. They made sincere efforts and ultimately found that three factors, viz., evaluation, potency and activity, contributed most to meaningful judgements by respondents. The evaluation dimension generally accounts for 1/2 and 3/4 of the extractable variance and the other two factors account for the balance.

**Procedure:** Various steps involved in developing S.D. scale are as follows:

- (a) First of all the concepts to be studied are selected. The concepts are usually chosen by personal judgement, keeping in view the nature of the problem.

- (b) The next step is to select the scales bearing in mind the criterion of factor composition and the criterion of scale's relevance to the concepts being judged (it is common practice to use at least three scales for each factor with the help of which an average factor score has to be worked out). One more criterion to be kept in view is that scales should be stable across subjects and concepts.
- (c) Then a panel of judges are used to rate the various stimuli (or objects) on the various selected scales and the responses of all judges would then be combined to determine the composite scaling.

To conclude, "the S.D. has a number of specific advantages. It is an efficient and easy way to secure attitudes from a large sample. These attitudes may be measured in both direction and intensity. The total set of responses provides a comprehensive picture of the meaning of an object, as well as a measure of the subject doing the rating. It is a standardised technique that is easily repeated, but escapes many of the problems of response distortion found with more direct methods."<sup>6</sup>

**Multidimensional scaling:** Multidimensional scaling (MDS) is relatively more complicated scaling device, but with this sort of scaling one can scale objects, individuals or both with a minimum of information. Multidimensional scaling (or MDS) can be characterized as a set of procedures for portraying perceptual or affective dimensions of substantive interest. It "provides useful methodology for portraying subjective judgements of diverse kinds."<sup>7</sup> MDS is used when all the variables (whether metric or non-metric) in a study are to be analyzed simultaneously and all such variables happen to be independent. The underlying assumption in MDS is that people (respondents) "perceive a set of objects as being more or less similar to one another on a number of dimensions (usually uncorrelated with one another) instead of only one."<sup>8</sup> Through MDS techniques one can represent geometrically the locations and interrelationships among a set of points. In fact, these techniques attempt to locate the points, given the information about a set of interpoint distances, in space of one or more dimensions such as to best summarise the information contained in the interpoint distances. The distances in the solution space then optimally reflect the distances contained in the input data. For instance, if objects, say  $X$  and  $Y$ , are thought of by the respondent as being most similar as compared to all other possible pairs of objects, MDS techniques will position objects  $X$  and  $Y$  in such a way that the distance between them in multidimensional space is shorter than that between any two other objects.

Two approaches, viz., the metric approach and the non-metric approach, are usually talked about in the context of MDS, while attempting to construct a space containing  $m$  points such that  $m(m-1)/2$  interpoint distances reflect the input data. The *metric approach to MDS* treats the input data as interval scale data and solves applying statistical methods for the additive constant\* which

<sup>6</sup> *Ibid.*, p. 260.

<sup>7</sup> Paul E. Green, "Analyzing Multivariate Data", p. 421.

<sup>8</sup> Jagdish N. Sheth, "The Multivariate Revolution in Marketing Research", quoted in "Marketing Research" by Danny N. Bellenger and Barnett A. Greenberg, p. 255.

\* Additive constant refers to that constant with which one can, either by subtracting or adding, convert interval scale to a ratio scale. For instance, suppose we know that distances, say  $a-b$ ,  $b-c$ ,  $c-d$  among stimuli on a ratio scale are 7, 6 and 3 respectively. If one were to subtract 3 from each of these distances, they would be 4, 3 and 0 respectively. The converted distances would be on an interval scale of measurement, but not on a ratio scale. Obviously, one can add 3 to all the converted distances and reach the ratio scale of distances. Thus 3 will be taken as the additive constant in this case. Well defined iterative approach is employed in practice for estimating appropriate additive constant.

minimises the dimensionality of the solution space. This approach utilises all the information in the data in obtaining a solution. The data (i.e., the metric similarities of the objects) are often obtained on a bipolar similarity scale on which pairs of objects are rated one at a time. If the data reflect exact distances between real objects in an  $r$ -dimensional space, their solution will reproduce the set of interpoint distances. But as the true and real data are rarely available, we require random and systematic procedures for obtaining a solution. Generally, the judged similarities among a set of objects are statistically transformed into distances by placing those objects in a multidimensional space of some dimensionality.

The *non-metric approach* first gathers the non-metric similarities by asking respondents to rank order all possible pairs that can be obtained from a set of objects. Such non-metric data is then transformed into some arbitrary metric space and then the solution is obtained by reducing the dimensionality. In other words, this non-metric approach seeks “a representation of points in a space of minimum dimensionality such that the rank order of the interpoint distances in the solution space maximally corresponds to that of the data. This is achieved by requiring only that the distances in the solution be monotone with the input data.”<sup>9</sup> The non-metric approach has come into prominence during the sixties with the coming into existence of high speed computers to generate metric solutions for ordinal input data.

The significance of MDS lies in the fact that it enables the researcher to study “the perceptual structure of a set of stimuli and the cognitive processes underlying the development of this structure. Psychologists, for example, employ multidimensional scaling techniques in an effort to scale psychophysical stimuli and to determine appropriate labels for the dimensions along which these stimuli vary.”<sup>10</sup> The MDS techniques, in fact, do away with the need in the data collection process to specify the attribute(s) along which the several brands, say of a particular product, may be compared as ultimately the MDS analysis itself reveals such attribute(s) that presumably underlie the expressed relative similarities among objects. Thus, MDS is an important tool in attitude measurement and the techniques falling under MDS promise “a great advance from a series of unidimensional measurements (e.g., a distribution of intensities of feeling towards single attribute such as colour, taste or a preference ranking with indeterminate intervals), to a perceptual mapping in multidimensional space of objects ... company images, advertisement brands, etc.”<sup>11</sup>

In spite of all the merits stated above, the MDS is not widely used because of the computation complications involved under it. Many of its methods are quite laborious in terms of both the collection of data and the subsequent analyses. However, some progress has been achieved (due to the pioneering efforts of Paul Green and his associates) during the last few years in the use of non-metric MDS in the context of market research problems. The techniques have been specifically applied in “finding out the perceptual dimensions, and the spacing of stimuli along these dimensions, that people, use in making judgements about the relative similarity of pairs of Stimuli.”<sup>12</sup> But, “in the long run, the worth of MDS will be determined by the extent to which it advances the behavioral sciences.”<sup>13</sup>

<sup>9</sup> Robert Ferber (ed.), *Handbook of Marketing Research*, p. 3–51.

<sup>10</sup> *Ibid.*, p. 3–52.

<sup>11</sup> G.B. Giles, *Marketing*, p. 43.

<sup>12</sup> Paul E. Green, *Analyzing Multivariate Data*, p. 421.

<sup>13</sup> Jum C. Nunnally, *Psychometric Theory*, p. 496.



### Questions

1. What is the meaning of measurement in research? What difference does it make whether we measure in terms of a nominal, ordinal, interval or ratio scale? Explain giving examples.
2. Are you in agreement with the following statements? If so, give reasons:
  - (1) Validity is more critical to measurement than reliability.
  - (2) Stability and equivalence aspects of reliability essentially mean the same thing.
  - (3) Content validity is the most difficult type of validity to determine.
  - (4) There is no difference between concept development and concept specification.
  - (5) Reliable measurement is necessarily a valid measurement.
3. Point out the possible sources of error in measurement. Describe the tests of sound measurement.
4. Are the following nominal, ordinal, interval or ratio data? Explain your answers.
  - (a) Temperatures measured on the Kelvin scale.
  - (b) Military ranks.
  - (c) Social security numbers.
  - (d) Number of passengers on buses from Delhi to Mumbai.
  - (e) Code numbers given to the religion of persons attempting suicide.
5. Discuss the relative merits and demerits of:
  - (a) Rating vs. Ranking scales.
  - (b) Summated vs. Cumulative scales.
  - (c) Scalogram analysis vs. Factor analysis.
6. The following table shows the results of a paired-comparison preference test of four cold drinks from a sample of 200 persons:

Name	Coca Cola	Limca	Goldspot	Thumps up
Coca Cola	–	60*	105	45
Limca	160	–	150	70
Goldspot	75	40	–	65
Thumps up	165	120	145	–

\* To be read as 60 persons preferred Limca over Coca Cola.

- (a) How do these brands rank in overall preference in the given sample.
  - (b) Develop an interval scale for the four varieties of cold drinks.
7. (1) Narrate the procedure for developing a scalogram and illustrate the same by an example.  
 (2) Workout Guttman's coefficient of reproducibility from the following information:
  - Number of cases ( $N$ ) = 30
  - Number of items ( $n$ ) = 6
  - Number of errors ( $e$ ) = 10
 Interpret the meaning of coefficient you work out in this example.
8. Write short notes on:
  - (a) Semantic differential scale;
  - (b) Scalogram analysis;

- (c) Likert-type scale;
  - (d) Arbitrary scales;
  - (e) Multidimensional scaling (MDS).
9. Describe the different methods of scale construction, pointing out the merits and demerits of each.
10. “Scaling describes the procedures by which numbers are assigned to various degrees of opinion, attitude and other concepts.” Discuss. Also point out the bases for scale classification.

# 6

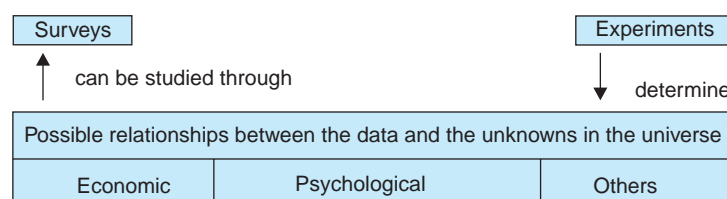
## Methods of Data Collection

The task of data collection begins after a research problem has been defined and research design/plan chalked out. While deciding about the method of data collection to be used for the study, the researcher should keep in mind two types of data viz., primary and secondary. The *primary data* are those which are collected afresh and for the first time, and thus happen to be original in character. The *secondary data*, on the other hand, are those which have already been collected by someone else and which have already been passed through the statistical process. The researcher would have to decide which sort of data he would be using (thus collecting) for his study and accordingly he will have to select one or the other method of data collection. The methods of collecting primary and secondary data differ since primary data are to be originally collected, while in case of secondary data the nature of data collection work is merely that of compilation. We describe the different methods of data collection, with the pros and cons of each method.

### COLLECTION OF PRIMARY DATA

We collect primary data during the course of doing experiments in an experimental research but in case we do research of the descriptive type and perform surveys, whether sample surveys or census surveys, then we can obtain primary data either through observation or through direct communication with respondents in one form or another or through personal interviews.\* This, in other words, means

\* An experiment refers to an investigation in which a factor or variable under test is isolated and its effect(s) measured. In an experiment the investigator measures the effects of an experiment which he conducts intentionally. Survey refers to the method of securing information concerning a phenomena under study from all or a selected number of respondents of the concerned universe. In a survey, the investigator examines those phenomena which exist in the universe independent of his action. The difference between an experiment and a survey can be depicted as under:



that there are several methods of collecting primary data, particularly in surveys and descriptive researches. Important ones are: (i) observation method, (ii) interview method, (iii) through questionnaires, (iv) through schedules, and (v) other methods which include (a) warranty cards; (b) distributor audits; (c) pantry audits; (d) consumer panels; (e) using mechanical devices; (f) through projective techniques; (g) depth interviews, and (h) content analysis. We briefly take up each method separately.

## Observation Method

The observation method is the most commonly used method specially in studies relating to behavioural sciences. In a way we all observe things around us, but this sort of observation is not scientific observation. Observation becomes a scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability. Under the observation method, the information is sought by way of investigator's own direct observation without asking from the respondent. For instance, in a study relating to consumer behaviour, the investigator instead of asking the brand of wrist watch used by the respondent, may himself look at the watch. The main advantage of this method is that subjective bias is eliminated, if observation is done accurately. Secondly, the information obtained under this method relates to what is currently happening; it is not complicated by either the past behaviour or future intentions or attitudes. Thirdly, this method is independent of respondents' willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or the questionnaire method. This method is particularly suitable in studies which deal with subjects (i.e., respondents) who are not capable of giving verbal reports of their feelings for one reason or the other

However, observation method has various limitations. Firstly, it is an expensive method. Secondly, the information provided by this method is very limited. Thirdly, sometimes unforeseen factors may interfere with the observational task. At times, the fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively.

While using this method, the researcher should keep in mind things like: What should be observed? How the observations should be recorded? Or how the accuracy of observation can be ensured? In case the observation is characterised by a careful definition of the units to be observed, the style of recording the observed information, standardised conditions of observation and the selection of pertinent data of observation, then the observation is called as *structured observation*. But when observation is to take place without these characteristics to be thought of in advance, the same is termed as *unstructured observation*. Structured observation is considered appropriate in descriptive studies, whereas in an exploratory study the observational procedure is most likely to be relatively unstructured.

We often talk about participant and non-participant types of observation in the context of studies, particularly of social sciences. This distinction depends upon the observer's sharing or not sharing the life of the group he is observing. If the observer observes by making himself, more or less, a member of the group he is observing so that he can experience what the members of the group experience, the observation is called as the *participant observation*. But when the observer observes as a detached emissary without any attempt on his part to experience through participation what others feel, the observation of this type is often termed as *non-participant observation*. (When the observer is observing in such a manner that his presence may be unknown to the people he is observing, such an observation is described as *disguised observation*.)

There are several merits of the participant type of observation: (i) The researcher is enabled to record the natural behaviour of the group. (ii) The researcher can even gather information which could not easily be obtained if he observes in a disinterested fashion. (iii) The researcher can even verify the truth of statements made by informants in the context of a questionnaire or a schedule. But there are also certain demerits of this type of observation viz., the observer may lose the objectivity to the extent he participates emotionally; the problem of observation-control is not solved; and it may narrow-down the researcher's range of experience.

Sometimes we talk of *controlled* and *uncontrolled observation*. If the observation takes place in the natural setting, it may be termed as uncontrolled observation, but when observation takes place according to definite pre-arranged plans, involving experimental procedure, the same is then termed controlled observation. In non-controlled observation, no attempt is made to use precision instruments. The major aim of this type of observation is to get a spontaneous picture of life and persons. It has a tendency to supply naturalness and completeness of behaviour, allowing sufficient time for observing it. But in controlled observation, we use mechanical (or precision) instruments as aids to accuracy and standardisation. Such observation has a tendency to supply formalised data upon which generalisations can be built with some degree of assurance. The main pitfall of non-controlled observation is that of subjective interpretation. There is also the danger of having the feeling that we know more about the observed phenomena than we actually do. Generally, controlled observation takes place in various experiments that are carried out in a laboratory or under controlled conditions, whereas uncontrolled observation is resorted to in case of exploratory researches.

## Interview Method

The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. This method can be used through personal interviews and, if possible, through telephone interviews.

(a) *Personal interviews*: Personal interview method requires a person known as the interviewer asking questions generally in a face-to-face contact to the other person or persons. (At times the interviewee may also ask certain questions and the interviewer responds to these, but usually the interviewer initiates the interview and collects the information.) This sort of interview may be in the form of direct personal investigation or it may be indirect oral investigation. In the case of direct personal investigation the interviewer has to collect the information personally from the sources concerned. He has to be on the spot and has to meet people from whom data have to be collected. This method is particularly suitable for intensive investigations. But in certain cases it may not be possible or worthwhile to contact directly the persons concerned or on account of the extensive scope of enquiry, the direct personal investigation technique may not be used. In such cases an indirect oral examination can be conducted under which the interviewer has to cross-examine other persons who are supposed to have knowledge about the problem under investigation and the information, obtained is recorded. Most of the commissions and committees appointed by government to carry on investigations make use of this method.

The method of collecting information through personal interviews is usually carried out in a structured way. As such we call the interviews as *structured interviews*. Such interviews involve the use of a set of predetermined questions and of highly standardised techniques of recording. Thus,

the interviewer in a structured interview follows a rigid procedure laid down, asking questions in a form and order prescribed. As against it, the *unstructured interviews* are characterised by a flexibility of approach to questioning. Unstructured interviews do not follow a system of pre-determined questions and standardised techniques of recording information. In a non-structured interview, the interviewer is allowed much greater freedom to ask, in case of need, supplementary questions or at times he may omit certain questions if the situation so requires. He may even change the sequence of questions. He has relatively greater freedom while recording the responses to include some aspects and exclude others. But this sort of flexibility results in lack of comparability of one interview with another and the analysis of unstructured responses becomes much more difficult and time-consuming than that of the structured responses obtained in case of structured interviews. Unstructured interviews also demand deep knowledge and greater skill on the part of the interviewer. Unstructured interview, however, happens to be the central technique of collecting information in case of exploratory or formulative research studies. But in case of descriptive studies, we quite often use the technique of structured interview because of its being more economical, providing a safe basis for generalisation and requiring relatively lesser skill on the part of the interviewer.

We may as well talk about focussed interview, clinical interview and the non-directive interview. *Focussed interview* is meant to focus attention on the given experience of the respondent and its effects. Under it the interviewer has the freedom to decide the manner and sequence in which the questions would be asked and has also the freedom to explore reasons and motives. The main task of the interviewer in case of a focussed interview is to confine the respondent to a discussion of issues with which he seeks conversance. Such interviews are used generally in the development of hypotheses and constitute a major type of unstructured interviews. The *clinical interview* is concerned with broad underlying feelings or motivations or with the course of individual's life experience. The method of eliciting information under it is generally left to the interviewer's discretion. In case of *non-directive interview*, the interviewer's function is simply to encourage the respondent to talk about the given topic with a bare minimum of direct questioning. The interviewer often acts as a catalyst to a comprehensive expression of the respondents' feelings and beliefs and of the frame of reference within which such feelings and beliefs take on personal significance.

Despite the variations in interview-techniques, the major advantages and weaknesses of personal interviews can be enumerated in a general way. The chief merits of the interview method are as follows:

- (i) More information and that too in greater depth can be obtained.
- (ii) Interviewer by his own skill can overcome the resistance, if any, of the respondents; the interview method can be made to yield an almost perfect sample of the general population.
- (iii) There is greater flexibility under this method as the opportunity to restructure questions is always there, specially in case of unstructured interviews.
- (iv) Observation method can as well be applied to recording verbal answers to various questions.
- (v) Personal information can as well be obtained easily under this method.
- (vi) Samples can be controlled more effectively as there arises no difficulty of the missing returns; non-response generally remains very low.
- (vii) The interviewer can usually control which person(s) will answer the questions. This is not possible in mailed questionnaire approach. If so desired, group discussions may also be held.

- (viii) The interviewer may catch the informant off-guard and thus may secure the most spontaneous reactions than would be the case if mailed questionnaire is used.
- (ix) The language of the interview can be adopted to the ability or educational level of the person interviewed and as such misinterpretations concerning questions can be avoided.
- (x) The interviewer can collect supplementary information about the respondent's personal characteristics and environment which is often of great value in interpreting results.

But there are also certain weaknesses of the interview method. Among the important weaknesses, mention may be made of the following:

- (i) It is a very expensive method, specially when large and widely spread geographical sample is taken.
- (ii) There remains the possibility of the bias of interviewer as well as that of the respondent; there also remains the headache of supervision and control of interviewers.
- (iii) Certain types of respondents such as important officials or executives or people in high income groups may not be easily approachable under this method and to that extent the data may prove inadequate.
- (iv) This method is relatively more-time-consuming, specially when the sample is large and recalls upon the respondents are necessary.
- (v) The presence of the interviewer on the spot may over-stimulate the respondent, sometimes even to the extent that he may give imaginary information just to make the interview interesting.
- (vi) Under the interview method the organisation required for selecting, training and supervising the field-staff is more complex with formidable problems.
- (vii) Interviewing at times may also introduce systematic errors.
- (viii) Effective interview presupposes proper rapport with respondents that would facilitate free and frank responses. This is often a very difficult requirement.

*Pre-requisites and basic tenets of interviewing:* For successful implementation of the interview method, interviewers should be carefully selected, trained and briefed. They should be honest, sincere, hardworking, impartial and must possess the technical competence and necessary practical experience. Occasional field checks should be made to ensure that interviewers are neither cheating, nor deviating from instructions given to them for performing their job efficiently. In addition, some provision should also be made in advance so that appropriate action may be taken if some of the selected respondents refuse to cooperate or are not available when an interviewer calls upon them.

In fact, interviewing is an art governed by certain scientific principles. Every effort should be made to create friendly atmosphere of trust and confidence, so that respondents may feel at ease while talking to and discussing with the interviewer. The interviewer must ask questions properly and intelligently and must record the responses accurately and completely. At the same time, the interviewer must answer legitimate question(s), if any, asked by the respondent and must clear any doubt that the latter has. The interviewers approach must be friendly, courteous, conversational and unbiased. The interviewer should not show surprise or disapproval of a respondent's answer but he must keep the direction of interview in his own hand, discouraging irrelevant conversation and must make all possible effort to keep the respondent on the track.

(b) *Telephone interviews*: This method of collecting information consists in contacting respondents on telephone itself. It is not a very widely used method, but plays important part in industrial surveys, particularly in developed regions. The chief merits of such a system are:

1. It is more flexible in comparison to mailing method.
2. It is faster than other methods i.e., a quick way of obtaining information.
3. It is cheaper than personal interviewing method; here the cost per response is relatively low.
4. Recall is easy; callbacks are simple and economical.
5. There is a higher rate of response than what we have in mailing method; the non-response is generally very low.
6. Replies can be recorded without causing embarrassment to respondents.
7. Interviewer can explain requirements more easily.
8. At times, access can be gained to respondents who otherwise cannot be contacted for one reason or the other.
9. No field staff is required.
10. Representative and wider distribution of sample is possible.

But this system of collecting information is not free from demerits. Some of these may be highlighted.

1. Little time is given to respondents for considered answers; interview period is not likely to exceed five minutes in most cases.
2. Surveys are restricted to respondents who have telephone facilities.
3. Extensive geographical coverage may get restricted by cost considerations.
4. It is not suitable for intensive surveys where comprehensive answers are required to various questions.
5. Possibility of the bias of the interviewer is relatively more.
6. Questions have to be short and to the point; probes are difficult to handle.

## COLLECTION OF DATA THROUGH QUESTIONNAIRES

This method of data collection is quite popular, particularly in case of big enquiries. It is being adopted by private individuals, research workers, private and public organisations and even by governments. In this method a questionnaire is sent (usually by post) to the persons concerned with a request to answer the questions and return the questionnaire. A questionnaire consists of a number of questions printed or typed in a definite order on a form or set of forms. The questionnaire is mailed to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself. The respondents have to answer the questions on their own.

The method of collecting data by mailing the questionnaires to respondents is most extensively employed in various economic and business surveys. The merits claimed on behalf of this method are as follows:

1. There is low cost even when the universe is large and is widely spread geographically.



2. It is free from the bias of the interviewer; answers are in respondents' own words.
3. Respondents have adequate time to give well thought out answers.
4. Respondents, who are not easily approachable, can also be reached conveniently.
5. Large samples can be made use of and thus the results can be made more dependable and reliable.

The main demerits of this system can also be listed here:

1. Low rate of return of the duly filled in questionnaires; bias due to no-response is often indeterminate.
2. It can be used only when respondents are educated and cooperating.
3. The control over questionnaire may be lost once it is sent.
4. There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been despatched.
5. There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.
6. It is difficult to know whether willing respondents are truly representative.
7. This method is likely to be the slowest of all.

Before using this method, it is always advisable to conduct 'pilot study' (Pilot Survey) for testing the questionnaires. In a big enquiry the significance of pilot survey is felt very much. Pilot survey is in fact the replica and rehearsal of the main survey. Such a survey, being conducted by experts, brings to the light the weaknesses (if any) of the questionnaires and also of the survey techniques. From the experience gained in this way, improvement can be effected.

*Main aspects of a questionnaire:* Quite often questionnaire is considered as the heart of a survey operation. Hence it should be very carefully constructed. If it is not properly set up, then the survey is bound to fail. This fact requires us to study the main aspects of a questionnaire viz., the general form, question sequence and question formulation and wording. Researcher should note the following with regard to these three main aspects of a questionnaire:

1. *General form:* So far as the general form of a questionnaire is concerned, it can either be structured or unstructured questionnaire. Structured questionnaires are those questionnaires in which there are definite, concrete and pre-determined questions. The questions are presented with exactly the same wording and in the same order to all respondents. Resort is taken to this sort of standardisation to ensure that all respondents reply to the same set of questions. The form of the question may be either closed (i.e., of the type 'yes' or 'no') or open (i.e., inviting free response) but should be stated in advance and not constructed during questioning. Structured questionnaires may also have fixed alternative questions in which responses of the informants are limited to the stated alternatives. Thus a highly structured questionnaire is one in which all questions and answers are specified and comments in the respondent's own words are held to the minimum. When these characteristics are not present in a questionnaire, it can be termed as unstructured or non-structured questionnaire. More specifically, we can say that in an unstructured questionnaire, the interviewer is provided with a general guide on the type of information to be obtained, but the exact question formulation is largely his own responsibility and the replies are to be taken down in the respondent's own words to the extent possible; in some situations tape recorders may be used to achieve this goal.

Structured questionnaires are simple to administer and relatively inexpensive to analyse. The provision of alternative replies, at times, helps to understand the meaning of the question clearly. But such questionnaires have limitations too. For instance, wide range of data and that too in respondent's own words cannot be obtained with structured questionnaires. They are usually considered inappropriate in investigations where the aim happens to be to probe for attitudes and reasons for certain actions or feelings. They are equally not suitable when a problem is being first explored and working hypotheses sought. In such situations, unstructured questionnaires may be used effectively. Then on the basis of the results obtained in pretest (testing before final use) operations from the use of unstructured questionnaires, one can construct a structured questionnaire for use in the main study.

2. *Question sequence:* In order to make the questionnaire effective and to ensure quality to the replies received, a researcher should pay attention to the question-sequence in preparing the questionnaire. A proper sequence of questions reduces considerably the chances of individual questions being misunderstood. The question-sequence must be clear and smoothly-moving, meaning thereby that the relation of one question to another should be readily apparent to the respondent, with questions that are easiest to answer being put in the beginning. The first few questions are particularly important because they are likely to influence the attitude of the respondent and in seeking his desired cooperation. The opening questions should be such as to arouse human interest. The following type of questions should generally be avoided as opening questions in a questionnaire:

1. questions that put too great a strain on the memory or intellect of the respondent;
2. questions of a personal character;
3. questions related to personal wealth, etc.

Following the opening questions, we should have questions that are really vital to the research problem and a connecting thread should run through successive questions. Ideally, the question-sequence should conform to the respondent's way of thinking. Knowing what information is desired, the researcher can rearrange the order of the questions (this is possible in case of unstructured questionnaire) to fit the discussion in each particular case. But in a structured questionnaire the best that can be done is to determine the question-sequence with the help of a Pilot Survey which is likely to produce good rapport with most respondents. Relatively difficult questions must be relegated towards the end so that even if the respondent decides not to answer such questions, considerable information would have already been obtained. Thus, question-sequence should usually go from the general to the more specific and the researcher must always remember that the answer to a given question is a function not only of the question itself, but of all previous questions as well. For instance, if one question deals with the price usually paid for coffee and the next with reason for preferring that particular brand, the answer to this latter question may be couched largely in terms of price-differences.

3. *Question formulation and wording:* With regard to this aspect of questionnaire, the researcher should note that each question must be very clear for any sort of misunderstanding can do irreparable harm to a survey. Question should also be impartial in order not to give a biased picture of the true state of affairs. Questions should be constructed with a view to their forming a logical part of a well thought out tabulation plan. In general, all questions should meet the following standards—(a) should be easily understood; (b) should be simple i.e., should convey only one thought at a time; (c) should be concrete and should conform as much as possible to the respondent's way of thinking. (For

instance, instead of asking, "How many razor blades do you use annually?" The more realistic question would be to ask, "How many razor blades did you use last week?"

Concerning the form of questions, we can talk about two principal forms, viz., multiple choice question and the open-end question. In the former the respondent selects one of the alternative possible answers put to him, whereas in the latter he has to supply the answer in his own words. The question with only two possible answers (usually 'Yes' or 'No') can be taken as a special case of the multiple choice question, or can be named as a 'closed question.' There are some advantages and disadvantages of each possible form of question. Multiple choice or closed questions have the advantages of easy handling, simple to answer, quick and relatively inexpensive to analyse. They are most amenable to statistical analysis. Sometimes, the provision of alternative replies helps to make clear the meaning of the question. But the main drawback of fixed alternative questions is that of "putting answers in people's mouths" i.e., they may force a statement of opinion on an issue about which the respondent does not in fact have any opinion. They are not appropriate when the issue under consideration happens to be a complex one and also when the interest of the researcher is in the exploration of a process. In such situations, open-ended questions which are designed to permit a free response from the respondent rather than one limited to certain stated alternatives are considered appropriate. Such questions give the respondent considerable latitude in phrasing a reply. Getting the replies in respondent's own words is, thus, the major advantage of open-ended questions. But one should not forget that, from an analytical point of view, open-ended questions are more difficult to handle, raising problems of interpretation, comparability and interviewer bias.\*

In practice, one rarely comes across a case when one questionnaire relies on one form of questions alone. The various forms complement each other. As such questions of different forms are included in one single questionnaire. For instance, multiple-choice questions constitute the basis of a structured questionnaire, particularly in a mail survey. But even there, various open-ended questions are generally inserted to provide a more complete picture of the respondent's feelings and attitudes.

Researcher must pay proper attention to the wordings of questions since reliable and meaningful returns depend on it to a large extent. Since words are likely to affect responses, they should be properly chosen. Simple words, which are familiar to all respondents should be employed. Words with ambiguous meanings must be avoided. Similarly, danger words, catch-words or words with emotional connotations should be avoided. Caution must also be exercised in the use of phrases which reflect upon the prestige of the respondent. Question wording, in no case, should bias the answer. In fact, question wording and formulation is an art and can only be learnt by practice.

*Essentials of a good questionnaire:* To be successful, questionnaire should be comparatively short and simple i.e., the size of the questionnaire should be kept to the minimum. Questions should proceed in logical sequence moving from easy to more difficult questions. Personal and intimate questions should be left to the end. Technical terms and vague expressions capable of different interpretations should be avoided in a questionnaire. Questions may be dichotomous (yes or no answers), multiple choice (alternative answers listed) or open-ended. The latter type of questions are often difficult to analyse and hence should be avoided in a questionnaire to the extent possible. There should be some control questions in the questionnaire which indicate the reliability of the respondent. For instance, a question designed to determine the consumption of particular material may be asked

\* Interviewer bias refers to the extent to which an answer is altered in meaning by some action or attitude on the part of the interviewer.

first in terms of financial expenditure and later in terms of weight. The control questions, thus, introduce a cross-check to see whether the information collected is correct or not. Questions affecting the sentiments of respondents should be avoided. Adequate space for answers should be provided in the questionnaire to help editing and tabulation. There should always be provision for indications of uncertainty, e.g., “do not know,” “no preference” and so on. Brief directions with regard to filling up the questionnaire should invariably be given in the questionnaire itself. Finally, the physical appearance of the questionnaire affects the cooperation the researcher receives from the recipients and as such an attractive looking questionnaire, particularly in mail surveys, is a plus point for enlisting cooperation. The quality of the paper, along with its colour, must be good so that it may attract the attention of recipients.

### COLLECTION OF DATA THROUGH SCHEDULES

This method of data collection is very much like the collection of data through questionnaire, with little difference which lies in the fact that schedules (proforma containing a set of questions) are being filled in by the enumerators who are specially appointed for the purpose. These enumerators along with schedules, go to respondents, put to them the questions from the proforma in the order the questions are listed and record the replies in the space meant for the same in the proforma. In certain situations, schedules may be handed over to respondents and enumerators may help them in recording their answers to various questions in the said schedules. Enumerators explain the aims and objects of the investigation and also remove the difficulties which any respondent may feel in understanding the implications of a particular question or the definition or concept of difficult terms.

This method requires the selection of enumerators for filling up schedules or assisting respondents to fill up schedules and as such enumerators should be very carefully selected. The enumerators should be trained to perform their job well and the nature and scope of the investigation should be explained to them thoroughly so that they may well understand the implications of different questions put in the schedule. Enumerators should be intelligent and must possess the capacity of cross-examination in order to find out the truth. Above all, they should be honest, sincere, hardworking and should have patience and perseverance.

This method of data collection is very useful in extensive enquiries and can lead to fairly reliable results. It is, however, very expensive and is usually adopted in investigations conducted by governmental agencies or by some big organisations. Population census all over the world is conducted through this method.

### DIFFERENCE BETWEEN QUESTIONNAIRES AND SCHEDULES

Both questionnaire and schedule are popularly used methods of collecting data in research surveys. There is much resemblance in the nature of these two methods and this fact has made many people to remark that from a practical point of view, the two methods can be taken to be the same. But from the technical point of view there is difference between the two. The important points of difference are as under:

1. The questionnaire is generally sent through mail to informants to be answered as specified in a covering letter, but otherwise without further assistance from the sender. The schedule