UNIT 6 : MEASUREMENT AND SCALING

UNIT STRUCTURE

6.1 Learning Objectives

6.2 Introduction

6.3 Concept of Measurement and Scaling

6.4 Primary Scale of Measurement

6.5 Comparative and Non Comparative Scaling Technique

6.6 Scale Evaluation: Measurement Error

6.7 Criteria for a Good Measurement.

6.8 Let Us Sum Up

6.9 Further Reading

6.10 Answers to Check your progress

6.11 Model Questions

6.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- discuss the concepts of measurement and scaling
- outline the primary scales of measurement and differentiate nominal, ordinal, interval, and ratio scales.
- classify scaling techniques
- describe the non comparative scaling techniques
- discuss the variations involved in constructing itemized rating scales.
- discuss the criteria used for scale evaluation and
- explain how to assess reliability and validity

6.2 INTRODUCTION

Once the type of research design has been determined and the exploratory and conclusive methods of research specified, the researcher can move on to the next phase of the research design: deciding on measurement and scaling procedures. This chapter describes the concepts of scaling and measurement and discusses four primary scales of measurement: nominal, ordinal, interval, and ratio. We next describe both comparative and non comparative scaling techniques and explain them in detail. The comparative techniques consist of paired comparison, rank order, and constant sum. The non comparative techniques are comprised of continuous and itemized rating scales. We discuss the popular itemized rating scales (the Likert, semantic differential, and Stapel scales), as well as the construction of multi-item rating scales. We show how scaling techniques should be evaluated in terms of reliability and validity, and consider how the researcher selects a particular scaling technique.

6.3 CONCEPT OF MEASUREMENT AND SCALING

Measurement means assigning numbers or other symbols to characteristics of objects according to certain set rules. Note that what we measure is not the object, but some characteristics of it. Thus, we do not measure consumers—only their perceptions, attitudes, preferences, or other relevant characteristics. In marketing research, numbers are usually assigned for two reasons. First, numbers permit statistical analysis of the resulting data. Second, numbers facilitate the communication of measurement rules and results. The most important aspect of measurement is that the researcher specifies rules for assigning numbers to the characteristics. The assignment process must be such that there is a one-to-one correspondence between the numbers and the characteristics being measured. For example, the same dollar figures are assigned to households with identical annual incomes. Only then can the numbers be associated with specific characteristics of the measured object, and vice versa. In addition, the rules for assigning numbers should be standardized and applied uniformly. They must not change over objects (e.g., stores) or time (e.g., quarter to quarter in quarterly surveys of customer satisfaction).

Scaling may be considered an extension of measurement. **Scaling** involves creating a continuum upon which measured objects are located. To illustrate, consider a scale for locating consumers according to the characteristic “attitude toward fast-food restaurants.” Each respondent is assigned a number indicating an unfavourable attitude (measured as 1), a neutral attitude (measured as 2), or a favourable attitude (measured as 3) toward a specific fast-food restaurant. Measurement is the actual assignment of 1, 2, or 3 to each respondent for each restaurant. Scaling is the process of placing the respondents on a continuum with respect to their attitude toward fast food restaurants. Suppose we measured attitudes toward five fast-food restaurants and then added the score of each respondent across the five restaurants to arrive at a total or overall attitude score. The total or overall score would vary from 5 (for a respondent who gave a rating of 1 to each of the five restaurants) to 15 (for a respondent who gave a rating of 3 to each of the five restaurants). In our example, scaling is the process by which respondents would be assigned an overall attitude score varying from 5 to 15.



CHECK YOUR PROGRESS

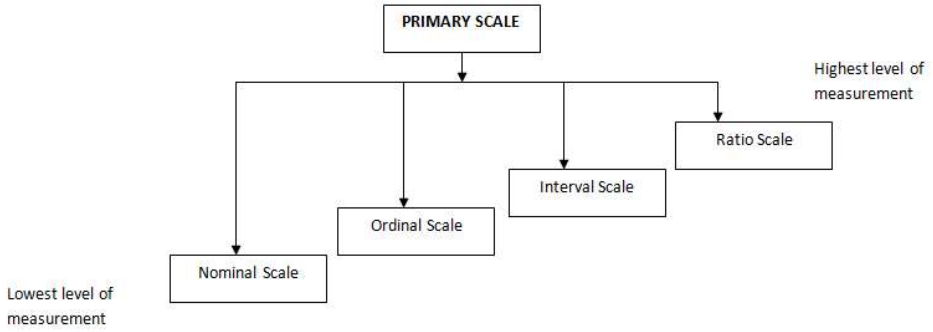
Q.1: Define Measurement and Scaling

.....

.....

6.4 PRIMARY SCALES OF MEASUREMENT

There are four primary scales of measurement: nominal, ordinal, interval, and ratio (Figure below). Their properties are summarized in Table 5.1 and discussed in the following sections. We illustrate all the primary scales in measuring preference and patronage of fast-food restaurants as given in Table 6.1.The following diagram represents the various types of measurement scales:



1. Nominal Scale: A **nominal scale** is a figurative labeling scheme in which the numbers serve only as labels or tags for identifying and classifying objects. For example, the numbers assigned to respondents in a study constitute a nominal scale. When a nominal scale is used for the purpose of identification, there is a strict one-to-one correspondence between the numbers and the objects. Each number is assigned to only one object, and each object has only one number assigned to it. Common examples include PAN numbers and numbers assigned to Cricket Players. Thus a PAN number can be assigned to one and only one individual, and each individual can be assigned one and only one PAN number. In marketing research, nominal scales are used for identifying respondents, brands, attributes, restaurants, and other people and objects.

When used for classification purposes, the nominally scaled numbers serve as labels for classes or categories. For example, based on gender, you might classify the males as group 1 and the females as group 2. The classes are mutually exclusive and collectively exhaustive. Thus, a person will be classified as either a male or a female, but he or she cannot be classified into both categories because these categories are mutually exclusive. Also, every person will be classified as either a male or a female

because these classes are collectively exhaustive; we cannot have a person who does not fall into either of these categories. The objects in each class are viewed as equivalent with respect to the characteristic represented by the nominal number. Thus, all males are considered equivalent with respect to gender, as indeed are all females. All objects in the same class have the same number (i.e., all males are assigned the number 1 and all females are assigned the number 2) and no two classes have the same number (no male can be assigned a number 2 and no female can be assigned the number 1).

The numbers in a nominal scale do not reflect the amount of the characteristic possessed by the objects. For example, a high PAN number does not imply that the person is in some way superior to those with lower PAN numbers, or vice versa. The same applies to numbers assigned to classes. Thus, males (assigned a number 1) are in no way superior or inferior to females (assigned a number 2). The numbers 1 and 2 merely denote whether a person is a male or female, respectively. The nominal scale represents the lowest level of measurement and is illustrated in the following.

In a fast-food restaurant project, numbers 1 through 5 were assigned to the five restaurants considered in the study. Thus, restaurant number 4 referred to KFC. It did not imply that KFC was in any way superior or inferior to McDonald's, which was assigned the number 2. Any reassignment of the numbers, such as transposing the numbers assigned to KFC and McDonald's, would have no effect on the numbering system because the numerals do not reflect any characteristics of the restaurants.

Table 6.1 : Primary Scales of Measurement

Primary Scale	Basic Characteristics	Common Examples	Marketing Examples
Nominal	Numbers Identify and classify objects	Social Security numbers; numbering of football players	Brand numbers; store types; gender classification
Ordinal	Numbers indicate the relative positions of the objects but not the magnitude of differences between them	Quality rankings; rankings of teams in a tournament	Preference rankings; market position; social class
Interval	Differences between objects can be compared; zero point is arbitrary	Temperature (Fahrenheit; Celsius)	Attitudes; opinions; index Numbers
Ratio	Zero point is fixed; ratios of scale values can be computed	Length; weight	Age; income; costs; sales; market shares

2. Ordinal Scale: An ordinal scale is a ranking scale in which numbers are assigned to objects to indicate the relative extent to which the objects possess some characteristic. An ordinal scale allows you to determine whether an object has more or less of a characteristic than another object, but not how much more or less. Thus an ordinal scale indicates relative position, not the magnitude of the differences between the objects. The object ranked first has more of the characteristic compared to the object ranked second, but whether the object ranked second is a close second or a poor second is not known. Measurements of this type include "greater than" or "less than" judgments from the respondents. Thus, an ordinal scale conveys more information than a nominal scale that does not contain "greater than" or "less than" judgments. Common examples of ordinal scales include quality rankings, rankings of teams in a tournament, socioeconomic class, and occupational status. In marketing research, ordinal scales are used to measure relative attitudes, opinions, perceptions, and preferences. Another common form of ordinal scales is categorical scales (nominal scales) that show order, or more or less of a characteristic such as education levels, age categories, income categories, and so on. In an ordinal scale, equivalent objects receive the same rank. Ordinal scales can be transformed in any way as long as the basic ordering of the objects is maintained because the differences in numbers are void of any meaning other than order. Such transformations result in equivalent scales.

Table 6.2 gives a particular respondent's preference rankings (see the first column under Preference Ranking). Respondents ranked five fast-food restaurants in order of preference by assigning a rank of 1 to the most preferred restaurant, a rank of 2 to the second most preferred restaurant, and so on. The lower number indicates that the restaurant is preferred to one with a higher number. Note that Pizza Hut (ranked 1), is preferred to McDonald's (ranked 2), but how much it is preferred we do not know. Also, it is not necessary that we assign numbers from 1 to 5 to obtain a preference ranking. The second ordinal scale, which assigns a number 10 to Pizza Hut, a number 25 to McDonald's, 30 to KFC, and so on, is an equivalent scale, as it was obtained by transforming the first scale while maintaining the order (see the second column under Preference Ranking). This transformation is based on the rule that the lower the numerical rank, the

higher the preference. The first scale, i.e., 1, 2, 3, 4, and 5 and the second scale, i.e., 10, 25, 30, 79, and 82, result in the same ordering of the restaurants according to preference and are therefore equivalent. This equivalence can be seen as follows. 1 is less than 2 and 10 is less than 25; 2 is less than 3 and 25 is less than 30; 3 is less than 4 and 30 is less than 79; 4 is less than 5 and 79 is less than 82. Both scales are based on the same rule, that is, the lower the numerical rank, the higher the preference and both result in the same rank order. As noted earlier, the differences in numbers are void of any meaning other than order. As a further illustration, *Fortune* magazine uses ordinal scaling to determine the world's most admired companies.

TABLE 6.2 : Illustration of Primary Scales of Measurement

Nominal Scale		Ordinal Scale		Interval Scale		Ratio Scale	
Store No	Store Name	Preference Ranking		Preference Ranking		Amt. Spent in Last Two Months	
				1-7	11 - 17	Dollars	Cents
1	Burger King	4	79	4	14	10	1000
2	McDonald's	2	25	7	17	50	5000
3	Subway	5	82	3	13	0	0
4	KFC	3	30	6	16	25	2500
5	Pizza Hut	1	10	7	17	75	7500

3. Interval Scale: In an **interval scale**, numerically equal distances on the scale represent equal values in the characteristic being measured. An interval scale contains all the information of an ordinal scale, but it also allows you to compare the differences between objects. The difference between any two adjacent scale values is identical to the difference between any other two adjacent values of an interval scale. There is a constant or equal interval between scale values. The difference between 1 and 2 is the same as the difference between 2 and 3, which is the same as the difference between 5 and 6. Thus, an interval scale conveys more information than an ordinal scale that does not contain information on distance. A common example in everyday life is a temperature scale. In marketing research, attitudinal data obtained from rating scales are often treated as interval data. For example, asking respondents to express their degree of agreement with the statement "I like Starbucks coffee" on a 5-point scale (1 = strongly disagree, 5 = strongly agree) will measure their attitude toward Starbucks coffee. Interval scales are widely used in marketing research to obtain data from respondents and are the most popular of all the primary scales. In a Football Tournament,

the points assigned to the countries with Spain receiving the highest points amounting to 1614 and Ecuador receiving 1066, the lowest, constituted an interval scale. In an interval scale, the location of the 0 point or the origin is not fixed. Both the origin and the units of measurement are arbitrary. Hence, it is permissible to shift the origin, for example, by adding a constant, and that will result in an equivalent scale. Because the 0 point is not fixed, it is not meaningful to take ratios of scale values. We illustrate this property of the interval scale by examining the information in Table 5.2, which shows a respondent's preferences for the five restaurants expressed on a 1 to 7 rating scale (see the first column under Preference Ratings). We can see that, although KFC received a preference rating of 6 and Subway a preference rating of 3, this does not mean that KFC is preferred two times as much as Subway. The ratings are then transformed to an equivalent 11 to 17 scale by shifting the origin from 1 to 11. This can be done by adding 10 to each scale value (see the second column under Preference Ratings). Thus, 1 becomes 11, 2 becomes 12, 3 becomes 13, and so on, and seven becomes 17. The transformed ratings for KFC and Subway become 16 and 13, and the ratio is no longer 2 to 1.

4. Ratio Scale: A **ratio scale** possesses all the properties of the nominal, ordinal, and interval scales, plus an absolute 0 point. Thus, the ratio scale represents the highest level of measurement conveying the most information. In ratio scales, we can identify or classify objects, rank the objects, and compare intervals or differences. Unlike the nominal, ordinal, and interval scales, it is also meaningful to compute ratios of scale values. Not only is the difference between 2 and 5 the same as the difference between 14 and 17, but also 14 is seven times as large as 2 in an absolute sense. Common examples of ratio scales include height, weight, age, and income that are measured as exact numbers rather than as categories. In marketing, sales, costs, market share, and the number of customers are variables measured on a ratio scale.

Ratio scales allow only proportionate transformations that involve multiplication by a positive constant. One cannot add an arbitrary constant, unlike the case of an interval scale. An example of this transformation is provided by the conversion of yards to feet (multiplication by 3) or dollars to cents (multiplication by 100). The comparisons between the lengths of

objects are identical whether made in yards or feet. Ratio scales are not as popular as interval scales for obtaining information because respondents find it more taxing to provide ratio scale information than they do interval scale data. The ratio scale is further illustrated in the context of the fast-food restaurant example. In Table 6.2, a respondent is asked to indicate the dollar amounts spent in each of the five restaurants during the last two months. Note that, because this respondent spent $50 in McDonald's and only $10 in Burger King, this respondent spent five times as much in McDonald's as in Burger King. Also, the 0 point is fixed because 0 means that the respondent did not spend anything at that restaurant. Multiplying these numbers by 100 to convert dollars to cents results in an equivalent scale.



CHECK YOUR PROGRESS

Q.2: What are the Primary Scales of Measurement.

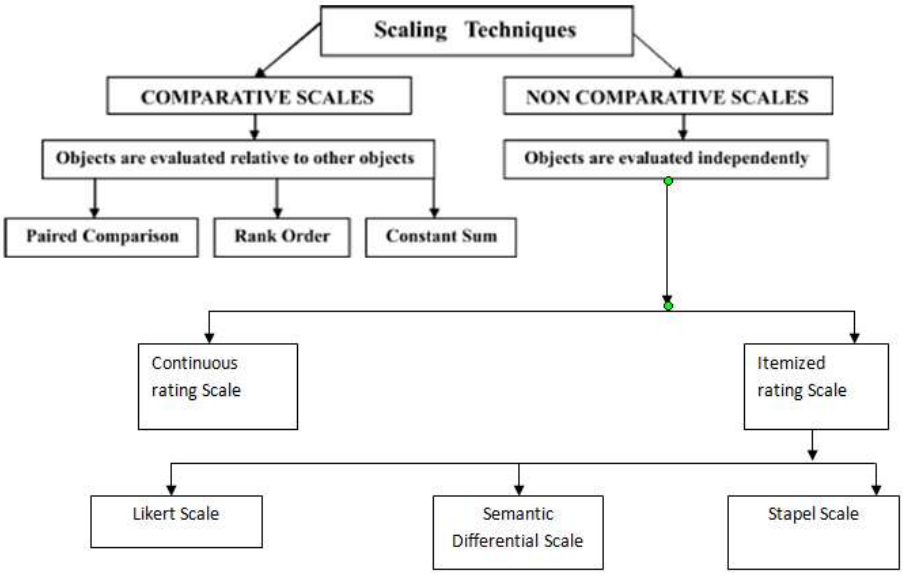
.....

.....

Q.3: Explain (i) Nominal Scale (ii) Ordinal Scale (iii) Interval Scale (iv) Ratio Scale

6.5 COMPARATIVE AND NON – COMPARATIVE SCALING TECHNIQUES

The scaling techniques commonly employed in marketing research can be classified into comparative and non comparative scales (see Figure below). **Comparative scales** involve the direct comparison of two or more objects; objects are evaluated in relation to other objects in the stimulus set. For example, respondents might be asked whether they prefer Coke or Pepsi. Comparative scale data must be interpreted in relative terms and have only ordinal or rank-order properties. For this reason, comparative scaling is also referred to as non metric scaling (non metric data have only ordinal properties). As shown in the figure below, comparative scales include paired



COMPARITIVE SCALING TECHNIQUES:

1. **Paired Comparison Scaling:** As its name implies, in **paired comparison scaling**, a respondent is presented with two objects and asked to select one according to some criterion. The data obtained are ordinal in nature. A respondent may state that she shops in JC Penney more than in Sears, likes Total cereal better than Kellogg's Product 19, or likes Crest more than Colgate. Paired comparison scales are frequently used when the stimulus objects are physical products. Paired comparison scaling is a widely used comparative scaling technique.

Paired comparison data can be analyzed by calculating the percentage of respondents who prefer one stimulus over another. Paired comparison scaling is useful when the number of brands is limited because it requires direct comparison and overt choice. With a large number of brands, however, the number of comparisons becomes unwieldy. Paired comparisons bear little resemblance to the market place situation that involves selection from multiple alternatives. Also respondents may prefer one object over certain others, but they may not like it in an absolute sense (that is, it is disliked less). The following Research in Practice provides further insights into paired comparison scaling

2. **Rank Order Scaling:** After paired comparisons, the most popular comparative scaling technique is rank-order scaling. Rank-order scaling is similar to the ordinal scale discussed earlier in the section on primary scales. In **rank-order scaling,** respondents are presented with several objects simultaneously and asked to order or rank them according to some criterion. For example, respondents may be asked to rank brands of toothpaste according to overall preference. These rankings are typically obtained by asking the respondents to assign a rank of 1 to the most preferred brand, 2 to the second most preferred, and so on, until a rank of *n* is assigned to the least preferred brand. Like paired comparison, this approach is also comparative in nature, and it is possible that the respondent may dislike, in an absolute sense, the brand ranked first. Rank-order scaling also results in ordinal data. Rank-order scaling is commonly used to measure preferences for brands as well as attributes. Compared to paired comparisons, this type of scaling process more closely resembles the shopping environment. It also takes less time because the respondents have to make fewer judgments compared to paired comparisons. Another advantage is that most respondents easily understand the instructions for ranking. The major disadvantage is that this technique produces only ordinal data.



CHECK YOUR PROGRESS

Q.4) Compare between Comparative and Non Comparative Scaling Techniques ?

.....

.....

Q.5) What are the different types of Comparative Scaling Techniques ?

.....

.....

NON-COMPARATIVE SCALING TECHNIQUES:

Respondents using a non comparative scale do not compare the object being rated to either another object or some specified standard, such as "your ideal brand." They evaluate only one object at a time, and for this reason non comparative scales are often referred to as monadic scales. In evaluating the objects, the respondents employ whatever rating standard seems appropriate to them.

A. Continuous Rating Scale: In a **continuous rating scale**, also referred to as a graphic rating scale, respondents rate the objects by placing a mark at the appropriate position on a line that runs from one extreme of the criterion variable to the other. Thus, the respondents are not restricted to selecting from marks previously set by the researcher. The form of the continuous scale may vary considerably. For example, the line may be vertical or horizontal; scale points, in the form of numbers or brief descriptions, may be provided. The scale points, if provided, may be few or many.

Once the respondent has provided the ratings, the researcher divides the line into as many categories as desired and assigns scores based on the categories into which the ratings fall. In the fast-food restaurant project example, the respondent exhibits an unfavorable attitude toward McDonald's. These scores are typically treated as interval data. The advantage of continuous scales is that they are easy to construct. However, scoring is cumbersome and unreliable. Continuous scales also provide little new information. Hence, their use in marketing research has been limited. Recently, however, with the increased popularity of computer-assisted personal interviewing, Internet surveys, and other technologies, their use is becoming more frequent.

B. Itemized Rating Scales: In an **itemized rating scale,** the respondents are provided with a scale that has a number or brief description associated with each category. The categories are ordered in terms of scale position, and the respondents are required to select the specified category that best describes the object being rated. Itemized rating scales are widely used in marketing research. We first describe the commonly used itemized

rating scales—the Likert, semantic differential, and Stapel scales and then examine variations in the use of itemized rating scales.

(i) Likert Scale : Named after its developer, Rensis Likert, the **Likert scale** is widely used. This rating scale requires the respondents to indicate a degree of agreement or disagreement with each of a series of statements about the stimulus objects. Typically, each scale item (statement) has five response categories, ranging from "strongly disagree" to "strongly agree." We illustrate with a Likert scale for evaluating attitudes toward McDonald's.

Instructions Listed below are different opinions about McDonald's. Please indicate how strongly you agree or disagree with each by using the following scale and circling the appropriate number:

1 = Strongly disagree

2 = Disagree

3 = Neither agree nor disagree

4 = Agree

5 = Strongly agree

Statement	Strongly Disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly Agree (5)
1. McDonald's sells high-quality food.		✓			
2. McDonald's has poor in-store service.		✓			
3. I like to eat in McDonald's.			✓		
4. McDonald's does not offer a good variety of menu items.				✓	
5. McDonald's is where America eats.	✓				
6. I do not like the advertising done by McDonald's.				✓	
7. McDonald's charges fair prices.		✓			

Note that some statements are positive (e.g., statement 1) in that they reflect positively on McDonald's, while others are negative (e.g., statement 2) and reflect negatively on McDonald's. There is a reason for having both positive and negative statements. This controls the tendency of some respondents, particularly those with very positive or very negative attitudes, to mark the right- or left-hand sides of the scale without reading the statements. To conduct the analysis, each statement is assigned a numerical score, ranging

from 1 to 5. (Sometimes each statement may be assigned a score ranging from “2 to +2.) The analysis can be conducted on an item-by-item basis (profile analysis), or a total (summated) score can be calculated for each respondent by summing across items (statements). Suppose the Likert scale in the fast-food restaurant example was used to measure attitudes toward McDonald’s as well as Burger King. Profile analysis would involve comparing the two restaurants in terms of the average respondent ratings for each item, such as quality of food, in-store service, variety of menu items, and so on. The summated approach is most frequently used; as a result, the Likert scale is also referred to as a summated scale. When using this approach to determine the total score for each respondent on each restaurant, it is important to use a consistent scoring procedure so that a high (or low) score consistently reflects a favorable response. This requires that the categories assigned to the negative statements by the respondents be scored by reversing the scale. Note that for a positive statement, an agreement reflects a favorable response, whereas for a negative statement, a disagreement represents a favorable response. Accordingly, a “strongly agree” response to a positive statement and a “strongly disagree” response to a negative statement would both receive scores of 5. In the scale shown above, if a higher score is to denote a more favourable attitude, the scoring of items 2, 4, and 6 will be reversed. The respondent in the fast-food restaurant project example has an attitude score of 16 (2 + 4 + 3 + 2 + 1 + 2 + 2). For the second statement, because it is negative, although the respondent has circled 2, the score assigned is 4. Likewise, the circled numbers for statements 4 and 6 are also reversed. Each respondent’s total score for each restaurant is calculated. Note that across seven items, the total scores will be in a range from 7 to 35. A respondent with the highest total score will have the most favorable attitude toward McDonald’s, whereas a respondent with the lowest total score will have the least favorable attitude toward McDonald’s.

The Likert scale has several advantages. It is easy to construct and administer. Respondents readily understand how to use the scale, making it suitable for mail, telephone, electronic, or personal interviews. The major disadvantage of the Likert scale is that it takes longer to complete than other itemized rating scales because respondents have to read each

statement (or the interviewer has to read each statement to them). The following example shows another use of a Likert scale in marketing research.

(ii) Semantic Differential Scale : The semantic differential scale is a 7-point rating scale with end points associated with bipolar labels or descriptors. In a typical application, respondents rate objects on a number of itemized, 7-point rating scales bounded or anchored at each end by one of two bipolar adjectives, such as *cold* and *warm*. Individual items on a semantic differential scale may be scored on either a "3 to +3 or a 1 to 7 scale. The resulting data are commonly analyzed through profile analysis. In profile analysis, mean values on each rating scale are calculated and compared by plotting or by statistical analysis. This helps determine the overall differences and similarities among the objects. To assess differences across segments of respondents, the researcher can compare mean responses of different segments. On the other hand, in cases when the researcher requires an overall comparison of objects, such as determining restaurant preference, the individual item scores are added to arrive at a total score, as already illustrated for the Likert scale.

Its versatility makes the semantic differential a popular rating scale in marketing research. It has been widely used in comparing brand, product, and company images. It has also been used to develop advertising and promotion strategies and in new product development studies. Several modifications of the basic scale have been proposed.

(iii) Stapel Scale: The , named after its developer, Jan Stapel, is a unipolar rating scale with ten categories numbered from "5 to +5, without a neutral point (0). This scale is usually presented vertically. Respondents are asked to indicate how accurately or inaccurately each term describes the object by selecting an appropriate numerical response category. The higher the number, the more accurately the term describes the object, as shown in the McDonald's example. In this example, McDonald's is evaluated as not having high-quality food and having somewhat poor service. The data obtained by using a Stapel scale can be analyzed in the same way as Likert or semantic differential data. The Stapel scale produces results similar to the semantic differential scale. The Stapel scale's advantages are that it does not require a pretest of the adjectives or phrases to ensure true

bipolarity, and it can be administered over the telephone. However, some researchers believe the Stapel scale is confusing and difficult to apply. Of the three itemized rating scales considered, the Stapel scale is used least. However, this scale merits more attention than it has received.



CHECK YOUR PROGRESS

Q.6: What are the different types of Non Comparative Scaling Techniques ?

.....

.....

Basic Non Comparitive Scales

Scale	Basic Characteristics	Examples	Advantages	Disadvantages
Continuous Rating Scale	Place a mark on a continuous line	Reaction to TV commercials	Easy to construct	Scoring can be cumbersome unless computerized
Itemized Rating Scales				
Likert Scale	Degree of agreement on a 1 (strongly disagree) to 5 (strongly agree) scale	Measurement of attitudes	Easy to construct, administer, and understand	More time consuming
Semantic Differential	Seven-point scale with bipolar labels	Brand, product, and company images	Versatile	Difficult to construct bipolar adjectives
Stapel Scale	Unipolar ten-point scale, −5 to +5, without a neutral point (zero)	Measurement of attitudes and images	Easy to construct and administer over telephone	Confusing and difficult to apply

Table 6.3: Non- Comparitive Scales

MULTI-ITEM SCALES : Itemized rating scales discussed earlier form the basic components of multi-item rating scales. As the name implies, multi-item scales consist of multiple items that need to be evaluated in order to obtain a given measurement. This was illustrated by the Likert scale measuring attitudes toward McDonald's that consisted of seven items. Likewise, the semantic differential and Stapel scales illustrated earlier in the context of McDonald's are also multi-item scales consisting of five and two items, respectively. The development of multi-item rating scales requires considerable technical expertise.

6.6 SCALE EVALUATION: MEASUREMENT ERROR

A multi-item scale should be evaluated for accuracy and applicability which would involves an assessment of reliability and validity of the scale. Both reliability and validity can be understood in terms of measurement, systematic, and random errors. **Measurement error** is the variation in the information sought by the researcher and the information generated by the measurement process employed. There are two sources of measurement error: systematic error and random error. **Systematic error** affects the measurement in a constant way. It represents stable factors that affect the observed score in the same way each time the measurement is made. Examples of systematic error are mechanical factors such as poor printing, overcrowding of items in the questionnaire, and poor design. **Random error**, on the other hand, is not constant. It represents transient factors that affect the observed score in different ways each time the measurement is made. Examples of random error are personal factors such as health, emotions, and fatigue. The distinction between systematic and random error is crucial to our understanding of reliability and validity.

6.7 CRITERIA FOR A GOOD MEASUREMENT

Reliability: Reliability refers to the extent to which a scale produces consistent results if repeated measurements are made. Systematic sources of error do not have an adverse impact on reliability because they affect measurement in a constant way and do not lead to inconsistency. In contrast, random error produces inconsistency, leading to lower reliability. Reliability can be defined as the extent to which a measure is free from random error. If there is no random error, the measure is perfectly reliable. Reliability is assessed by determining the association between scores obtained from different administrations of the scale. If the association is high, the scale yields consistent results and is therefore reliable.

Validity: The validity of a scale may be defined as the extent to which differences in observed scale scores reflect true differences among objects on the characteristic being measured, rather than systematic or random error. Perfect validity requires that there be no measurement error, that is, no systematic error and no random error. Validity can be assessed by

determining whether a scale performs as expected in relation to other selected variables. These variables may include demographic and psychographic characteristics, attitudinal and behavioral measures, or scores obtained from other scales. For example, a researcher may develop short forms of standard personality scales. The original scales and the short versions would be administered simultaneously to a group of respondents and the results compared to assess validity.

Relationship Between Reliability and Validity: The relationship between reliability and validity can be understood in terms of measurement error. If a measure is perfectly valid, it is also perfectly reliable. In this case, there is no measurement error. Thus, there is no systematic error and no random error. Thus, perfect validity implies perfect reliability. If a measure is unreliable, it cannot be perfectly valid because, at a minimum, random error is present. Systematic error may also be present. Thus, unreliability implies invalidity. If a measure is perfectly reliable, it may or may not be perfectly valid because systematic error may still be present, although random error is zero. While lack of reliability constitutes negative evidence for validity, reliability does not in itself imply validity. Reliability is a necessary but not sufficient condition for validity.



CHECK YOUR PROGRESS

Q.7: What is Measurement Error, Systematic Error and Random Error?

.....

.....

Q.8: What is Reliability and Validity?

.....

.....



6.8 LET US SUM UP

Measurement is the assignment of numbers or other symbols to characteristics of objects according to set rules. Scaling involves the generation of a continuum upon which measured objects are located. The four primary scales of measurement are nominal, ordinal, interval, and ratio. Of these, the nominal scale is the most basic in that the numbers are used only for identifying or classifying objects. In the ordinal scale, the next higher-level scale, the numbers indicate the relative position of the objects but not the magnitude of differences between them. The interval scale permits a comparison of the differences between the objects. Because it has an arbitrary 0 point, however, it is not meaningful to calculate ratios of scale values on an interval scale. The highest level of measurement is represented by the ratio scale in which the 0 point is fixed. The researcher can compute ratios of scale values using this scale. The ratio scale incorporates all the properties of the lower level scales.

Scaling techniques can be classified as comparative or noncomparative. Comparative scaling involves a direct comparison of stimulus objects. Comparative scales include paired comparisons, rank order, and constant sum. The data obtained by these procedures have only ordinal properties. In non comparative scaling, each object is scaled independently of the other objects in the stimulus set. The resulting data are generally assumed to be interval or ratio scaled. Non comparative rating scales can be either continuous or itemized. The itemized rating scales are further classified as Likert, semantic differential, or Stapel scales. Several variations of the original itemized scales are commonly used in marketing research.

Multi-item scales consist of a number of rating scale items. These scales should be evaluated in terms of reliability and validity. Reliability is the extent to which the measure is free from random error. Validity is the extent to which the measure is free from both random and systematic errors. The choice of particular scaling techniques in a given situation should be

based on theoretical and practical considerations. As a general rule, the scaling technique used should be the one that will yield the highest level of information feasible. Also, multiple measures should be obtained.

Respondents in many developed countries, because of higher education and consumer sophistication levels, are quite used to providing responses on interval and ratio scales. In developing countries, however, preferences can be measured best by using ordinal scales. Special attention should be devoted to determining equivalent verbal descriptors in different languages and cultures.



6.9 FURTHER READING

Malhotra, N.K. and Dash, S. (2011). *Marketing research: An Applied Orientation* (6th edition), New Delhi: Pearson Education.



6.10 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q1: Measurement means assigning numbers or other symbols to characteristics of objects according to certain set rules. Scaling may be considered an extension of measurement. Scaling involves creating a continuum upon which measured objects are located.

Ans to Q2. There are four primary scales of measurement: nominal, ordinal, interval, and ratio

Ans to Q3. (i) A Nominal scale is a figurative labeling scheme in which the numbers serve only as labels or tags for identifying and classifying objects.

(ii) In an interval scale, numerically equal distances on the scale represent equal values in the characteristic being measured.

(iii)An ordinal scale is a ranking scale in which numbers are assigned to objects to indicate the relative extent to which the objects possess some characteristic.

(iv) A ratio scale possesses all the properties of the nominal, ordinal, and interval scales, plus an absolute 0 point. Thus, the ratio scale represents the highest level of measurement conveying the most information.

Ans to Q4. **Comparative scales** involve the direct comparison of two or more objects; objects are evaluated in relation to other objects in the stimulus set. In **non comparative scales,** also referred to as monadic or metric scales, objects are evaluated independently of each other. The resulting data are generally assumed to be interval scale.

Ans to Q5. Types of Comparitive Scaling techniques:

a. Paired Comparison Scaling b. Rank Order Scaling

Ans to Q6. Types of Non Comparative Scaling Techniques:

a. Continuous Rating Scale b. Itemised Rating Scale

Ans to Q7. **Measurement error** is the variation in the information sought by the researcher and the information generated by the measurement process employed.

Systematic error affects the measurement in a constant way. It represents stable factors that affect the observed s score in the same way each time the measurement is made.

Random error, on the other hand, is not constant. It represents transient factors that affect the observed score in different ways each time the measurement is made.

Ans to Q8: **Reliability** refers to the extent to which a scale produces consistent results if repeated measurements are made. The **validity** of a scale may be defined as the extent to which differences in observed scale scores reflect true differences among objects on the characteristic being measured, rather than systematic or random error.



6.11 MODEL QUESTIONS

1) What is measurement?
2) What are the primary scales of measurement?
3) Describe the differences between a nominal and an ordinal scale.
4) What are the implications of having an arbitrary 0 point in an interval scale?
5) What are the advantages of a ratio scale over an interval scale? Are these advantages significant?
6) What is a comparative rating scale?
7) What is a paired comparison?
8) What are some advantages of rank-order scaling over paired comparison scaling?
9) What is a semantic differential scale? For what purposes is this scale used?
10) Describe the Stapel scale.
11) Briefly explain itemized rating scales and how they differ from the continuous rating scale.
12) Discuss the use of social media in implementing itemized rating scales.
13) Briefly explain the two types of measurement errors.
14) What is reliability?
15) What is validity?
16) What is the relationship between reliability and validity?
17) How would you select a particular scaling technique?

*****