CHAPTER 3

RESEARCH PROCESS

The previous chapter emphasized the importance and need of a social research process to be in place to design, plan, implement and improve any developmental change process. The present chapter lists the various steps of the research process in detail starting with problem identification and conceptualization of the research plan to the preparation of the research report. Research as a process starts with the identification of the problem and moves ahead with the exploration of various probable solutions to that problem. It can be described as a researcher's quest to identify the problem and to solve it in the best possible way.

The first step in solving a research problem is the identification of the research problem and the various options/alternatives that are available to solve the problem in the best possible way (see Figure 3.1). The researcher should then identify/determine the information that is already available and the best possible research design needed to collect information, taking into account the time and cost factors. Finally, the information obtained must be assessed/analysed objectively to help in making an informed decision about the best possible way of solving the problem. This systematic

FIGURE 3.1
Social Research Process

approach to decision-making is referred to as the research process. The process can be broadly defined as a combination of following steps:

a) Problem definition.
b) Selection of research design.
c) Finalization of research instruments.
d) Data collection.
e) Data processing and analysis.
f) Report preparation.

PROBLEM DEFINITION

As mentioned earlier, research as a process starts with problem identification. It may sound simple, but often it is the most difficult part. In the socio-development scenario, it is even more challenging to lay down the problem area and research needs exactly because of non-controllable extraneous factors and externalities involved in the process. More so, as every socio-development process has a social and human angle attached to it, which makes the task even more difficult. For example, it is very challenging to establish linkages between environment and poverty, environment and health, or for that matter between natural resource degradation and population, and it is even more difficult to lay down the research objective precisely, which can be assessed using an objective approach/ tools.

The identification of the problem area then leads to the formulation of the research objective, which is the core of the research process. All subsequent steps in the research process, that is, selection of research design, research instruments and analysis take their cue from the objectives.

SELECTION OF RESEARCH DESIGN

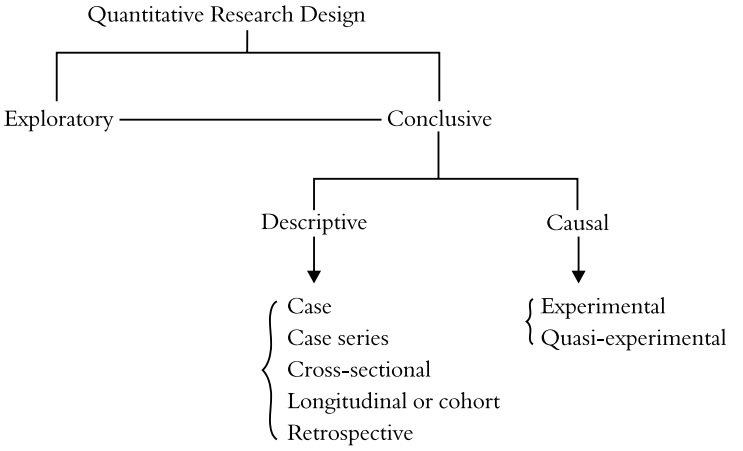
In quantitative research, the primary aim is to determine the relationship between an independent variable and another set of dependent or outcome variables in a population. Research design,[1] according to Kerlinger is the plan, structure and strategy of investigation conceived to obtain answers to research questions and to control variance (see Figure 3.2).

Quantitative research designs[2] can be broadly divided into two types, namely, exploratory research and conclusive research.

Exploratory Research

Exploratory research, as the name suggests, is often conducted to explore the research issue and is usually done when the alternative options have not been clearly defined or their scope is unclear. Exploratory research allows researchers to explore issues in detail in order to familiarize themselves

FIGURE 3.2
Quantitative Research Design Classification



with the problem or concept to be studied. Familiarization with the concept helps researchers in formulating research hypothesis.

Exploratory research is the initial research, which forms the basis of more conclusive research. It can even help in determining the research design, sampling methodology and data collection method. In some cases, exploratory research serves as the formative research to test concepts before they are put into practice.

Exploratory research, as mentioned earlier, explores the issue further, hence it relies more on secondary research, that is, the review of available literature and/or data, or qualitative research approaches such as informal discussions with primary and secondary stakeholders, project staff, donor agencies and more formal approaches, like in-depth interviews, focus groups or case studies. Exploratory research thus cannot provide a conclusive answer to research problems and usually are not considered useful for decision-making, but they can provide significant insights to a given situation. However, the results thus obtained cannot be generalized and should be interpreted with caution as they may or may not be representative of the population being studied.

Conclusive Research

Conclusive research can further be classified into descriptive research and causal research.

Descriptive Research

Descriptive research, as the name suggests, enumerates descriptive data about the population being studied and does not try to establish a causal relationship between events. This is also one of its major limitations as it cannot help determine what causes a specific behaviour or occurrence. It is

used to describe an event, a happening or to provide a factual and accurate description of the population being studied. It provides the number of times something occurs and helps in determining the descriptive statistics about a population, that is, the average number of occurrences or frequency of occurrences. In a descriptive study, things are measured as they are, whereas in an experimental study researchers take measurements, try some intervention and then take measurements again to see the impact of that intervention.

Descriptive research can be further classified into the following types:

a) Case study.
b) Case series study.
c) Cross-sectional study.
d) Longitudinal change.
e) Retrospective study.

Case is the simplest kind of descriptive study, which reports data on only one subject, individual or social process. For example, the study of an HIV patient or of a voluntary institution that is performing well. Case studies are now used worldwide as an accepted tool to document innovative approaches, success stories and failures. Case series is the descriptive study of a few cases. For example, studying success stories of resource-based self-help groups to identify their commonality would be a case series. Cross-sectional studies portray a snap shot of the prevalent situation as in these studies variables of interest in a sample are assessed only once to determine the relationships between them. The most commonly seen surveys use the cross-sectional design, which asks questions of people at one point in time. In the case of a longitudinal design, the same questions are asked at two or more points in time. Longitudinal design can be further classified into three subtypes: (i) trend study, (ii) cohort study and (iii) panel study.

a) *Trend study* can be defined as a repeated cross-sectional design where the same set of questions are asked to different sets of people/target population at different points in time. In trend studies, the variables of interest are assessed as a function of time.
b) *Cohort study* is a trend study that studies changes in cohorts, that is, the same set of people who experience the same kind of life or the same events over time. In prospective or cohort studies, some variables are assessed at the start of a study then after a period of time the outcomes are determined. For example, assessing the impact of a communication campaign on awareness levels of a target audience, would be a cohort study.
c) *Panel study* asks the same set of questions to the same people over time and is used to assess changes in the panel respondent's characteristics over times. In a nutshell, trend studies essentially look at how concepts/variables of interest change over time; cohort studies look at how the behaviour of the same set of people changes over time; and panel studies look at how people change over time.

Case-control studies compare cases with a particular characteristic of control subjects, that is, subjects without the attribute in order to determine a causal effect, for example, cases of tuberculosis and the number of cigarettes smoked per day. Case-control studies are also known as retrospective studies because they focus on conditions, which might have resulted in subjects becoming cases.

Causal Research

Causal research is defined as a research design where the main emphasis is on determining a cause and effect relationship. It is undertaken to determine which variable might be causing a certain behaviour, that is, whether there is a cause and effect relationship between variables and if a relationship exists then what is the nature of the causal relationship. In order to determine causality, it is important to hold one variable constant to assess change in the other variable and then measure the changes in the other variable. Causal research by nature is not very easy as it is very difficult to ascertain the causal relationship between the observed variable and the variable of interest. In fact, the causal relationship could be due to other factors, especially when dealing with people's attitudes and perceptions. There are often much deeper psychological factors, which even the respondent may not be aware of while responding to a question.

There are two research methods/designs for exploring the cause and effect relationship between variables: (i) experimental studies and (ii) quasi-experimental studies.

Experimental Studies

Experimental studies are characterized by a control group and an experimental group and subjects are assigned randomly to either group. Researchers try to maintain control over all factors that may affect the result of an experiment as experimentation[3] is still believed to be and is used as one of the most important research designs for establishing causality between variables. It allows the researcher to manipulate a specific independent variable in order to determine what effect this manipulation would have on other dependent variables. Another important criterion, while following the experimental research design, is to decide on the setting of the experiment, that is, whether it should take place in a natural setting or in an artificial one.

Experimental studies/designs[4] are also known as longitudinal or repeated-measure studies. They are also referred to as interventions, because of the use of control and experimental groups.

Time series is the simplest form of experiment, where one or more measurements are taken on all subjects before and after a treatment and it could either be a single subject design or a multiple subject design. In the case of a single subject design, measurements are taken repeatedly before and after an intervention on one or a few subjects.

The very nature of a time series design can also pose some problems as any change that is observed could be due to something other than the treatment. The subjects might do better on the second test because of their experience/learning during the first test or there could be some other extraneous factors that may result in a difference between the results of the first and second test such as change in weather, change in aptitude or change in diet. Crossover design, where the subjects are normally given two treatments, one being the real treatment, the other a control or reference treatment, can solve this problem. In the case of a crossover design, as the name suggests, half the subjects first receive control treatment, whereas the other half receive experimental treatment and after a sufficient period of time, which should allow the treatment to wash out, the treatments are crossed over. Further, any effect of retesting or of change that happened during successive tests can then be subtracted out by an appropriate analysis and we can also use multiple crossover designs involving several treatments to sort out this problem.

In certain situations, the treatment effect is unlikely to wash out between measurements. It then becomes imperative to use a control group. In these designs, though all subjects are measured, only an experimental group receives the treatment and when subjects are measured again, then any change in the experimental group is compared with the change in the control group to assess the effect of the treatment.

In another case of experimentation, that is, in the case of a randomized controlled trial, subjects are assigned randomly to experimental and control groups. It minimizes the chance that either group is not representative of the population they belong to. Further, if the subjects are masked to the identity of the treatment, the design is called single blind controlled trial.

The term blind experiment means that the subjects do not know which group they belong to, that is, they do not know whether they belong to the experimental group or the control group. In a double blind experiment, even the researchers and facilitators do not know who is in which group. These precautions/measures are taken by the research team to avoid the Hawthorne effect[5] and the placebo effect. The Hawthorne effect is defined as the tendency of human beings to temporarily improve their performance when they are aware it is being studied, especially in a scenario where they think they have been singled out for some experimental treatment. The placebo effect refers to the tendency of some subjects to respond positively to a new treatment just because they expect it to work, although the treatment may be entirely ineffective.

In such a case, researchers first need to randomly select a control group, statistically equal to the treatment group. Though the subjects are assigned randomly to each group, it is important to ensure that both groups are from the same population. To do so, researchers should match population characteristics to ensure that the groups match in their distributional characteristics.

There is nothing sacrosanct about having only one control and treatment group and researchers may have more than one control or treatment group. Researchers can have full and partial treatment groups based on the nature of the treatment. The experiment procedure starts with a pre-test and ends with a post-test. It is important to point out that researchers can conduct multiple post-tests at any time during the experiment. Researchers need to analyse the findings based primarily on differences in the post-test scores of the experimental and control groups.

Quasi-experimental Studies

Quasi-experimental studies, as the name suggests, have some attributes of experimental research design as they involve some controls over extraneous variables when full experimental control is lacking. Thus, in some cases, where the situation demands partial control, these designs may be the only way to evaluate the effect of the independent variable of interest. Further, as quasi-experimental studies lack control, this research design is often used in the area of medicine where, for ethical reasons, it is not possible to create a truly controlled group. Quasi-experiment[6] is a type of quantitative research design conducted to explain relationships and/or clarify why certain events happen.

The objective of adopting a quasi-experimental design is to assess causality. It analyses the difference in treatment and control group to look for causality in situations when complete control is not possible. These designs were developed to examine causality in situations where it is not practical or possible to have complete control over the subjects.

Quasi-experiments are relatively strong in terms of internal validity and use matching instead of randomization. Thus, quasi-experimental designs lack at least one of the other two properties that characterize true experiments, namely, randomization and a control group.

FINALIZATION OF RESEARCH INSTRUMENTS

Desk Research

The first step to the finalization of research instruments is to do desk research. Desk research, as the name implies, is analysis/documentation of available information for preparing survey instruments, finalizing sampling and operation plans and developing a list of indicators for the study. It usually involves review of secondary literature, that is, related studies, schedules, etc.

Development of Research Instruments

Development of research tools/instruments forms the next step after the finalization of research design. Finalization of research instruments needs to be done in consultation with all research partners and the core study team members need to develop the research instruments in consonance with the research objectives and designated tasks.

Designing Research Tools/Instruments

Designing research instruments depends on various factors such as the research problem, type of survey design and nature of information that needs to be collected. In the case of a quantitative survey, structured questionnaires and schedules are preferred whereas in the case of qualitative research, semi-structured questionnaires or discussion guidelines are preferred. However, it is not as easy as it sounds. There are other factors that need to be considered.

Though survey is the most preferred option, it suffers from some problems too. Researchers can make inferences, but cannot be sure of the cause and effect relationship as they can be in the case of experimental or quasi-experimental research. Other weaknesses of the survey method include:

a) *Reactivity:* Respondents' bias arises because they may give morally desirable responses or feel good responses.
b) *Mismatched sampling frame:* In surveys it is difficult to ascertain the adequate number and type of people who are representative of the population.
c) *Non-response rate:* However hard researchers may try, there are always going to be people who will not participate in surveys. This leads to high a non-response rate.

d) *Measurement error:* Like respondent and interviewer bias, there is always going to be some bias because of the failure of the survey instrument or methodology in measuring the desired attribute.

The next section looks at the important procedure of designing a survey instrument, which can contribute a lot in minimizing measurement error and interviewer error to some extent.

Survey Instrument

Survey instruments[7] can be broadly classified into two categories: (i) questionnaires and (ii) interviews. A questionnaire is almost always self-administered, allowing respondents to fill them out themselves. All the researcher has to do is to arrange for the delivery and collection of the questionnaires.

An interview is typically defined as a face-to-face discussion or communication via some technology like the telephone or computer between an interviewer and a respondent. There are three subtypes of interviews: (i) unstructured, which allows a free flow of communication in the course of the interview or questionnaire administration, (ii) structured, where the information that needs to be culled out from the respondents is already decided and (iii) semi-structured, which restricts certain kinds of communications but allows manoeuvring freedom on the discussion of certain topics.

Type of Question

Usually research questionnaires contain question of three basic types: (i) open-ended questions, (ii) dichotomous questions and (iii) multiple-response questions.

a) *Open-ended questions:* Open-ended questions are questions that do not have pre-coded options. These are used extensively in formative research or qualitative research when researchers want to capture the respondent's responses verbatim.
b) *Dichotomous questions:* Dichotomous questions have two possible answers like yes/no, true/false or agree/disagree responses. Surveys often use dichotomous questions when they are looking for a lead question.
c) *Multiple-response questions:* There are several questions that may have many probable answers, for example, knowledge regarding ways in which HIV/AIDS can spread. It is highly probable that most of the respondents would be aware of more than one probable way, thus it becomes imperative to frame questions as multiple-response questions.

Besides the type of question, there are various other attributes/norms that need to be adhered to while designing research instruments. These are discussed next.

Mutual Exclusivity

In the case of a multiple-response question, it is imperative to have response items that are mutually exclusive otherwise a bias could be introduced.

Non-exhaustive Response Set

Bias can also be introduced when the response alternatives available to the respondent leaves out valid choices they would otherwise make. The most common example is leaving out such responses as 'not applicable' or 'don't know' when, in fact, respondents may well be neutral or may actually not know, rather than be hiding their 'true' responses which the researcher is trying to force out by omitting these categories.

Order of Questions

The order of questions plays an important role in designing the survey instrument. The first paragraph should be clearly related to the announced purpose of the survey. Location details (state/district/village) could follow later together with background information questions. The survey should then proceed to attitude questions, sequencing from general and less sensitive items towards more specific and more sensitive ones.

Filter Items/Skip Questions

Filter items/skip questions are ones that allow for the elimination/filtering of unqualified respondents. For example, sometimes the researcher may have to ask the respondent one question in order to determine if they are qualified or experienced enough to answer a subsequent one. This can be done using a filter or skip question. For instance, the researcher may want to ask one question if the respondent has used a family-planning method and a different question if the respondent has not used any family-planning method. In this case, the researcher constructs a filter question to determine whether the respondent has ever used a family-planning method:

1) Have you ever used a family-planning method?
   a) Yes
   b) No
2) Please specify the family-planning method you have used?

Filter questions can get very complex. Sometimes, the researcher has to have multiple filter questions in order to direct the respondents to the correct subsequent questions. However, the researcher must always try to avoid having more than three levels for any question as too many filter questions and jump questions can confuse the interviewer and may even discourage the respondent from continuing with the survey. Researchers generally use graphics like arrows to indicate the question to which the skip question leads, or, alternatively, researchers can use instructions.

Cross-check Items

Cross-check items are check items which help researchers in tracking data consistency in research questionnaire. The researcher can ask the respondent's age at one point and date of birth at another

to cross-check both the survey items. Split-form interview is an extension of this, wherein the questionnaire is administered to different related respondents, for example, a husband and wife separately with a view towards cross-checking for consistency.

Caution Taken in Phrasing Questions

a) *Is the question really necessary?* First and foremost, researchers always need to ask whether the question is really necessary, that is, they must examine each question to see if the question needs to be asked at all and if it needs to be asked then what level of detail is required. For example, do you need the age of each child or just the number of children under 16?

b) *Double-barrelled questions/compounded items:* Researchers can often find a double-barrelled question by looking for the conjunction 'and' in the question. For example, the question 'What do you think of the proposed changes in the country's economic policy and foreign policy?' is a double-barrelled one.

Items with compound clauses may not be multidimensional, but may involve undue complexity. For example, the question, 'Have you ever faced complications after using the oral pill and whether you have consulted any doctor for the same?' is better broken into two items: 'Have you ever faced complications after using the oral pill?' and the follow-up question, 'If you have answered yes to the previous question, whether you have consulted any doctor for the complications?'

c) *Is the question specific or it is leading to ambiguity?* Sometimes we ask our questions too generally and the information we obtain is more difficult to interpret. Questions should be specific, avoiding generalities for if it is possible for respondents to interpret questions in dissimilar terms, they will.

d) *Is the question sufficiently general?* Sufficiently general question such as what is your opinion about India's policy should be avoided as it leaves scope of ambiguity. It is not clear in the question whether the researcher is asking about the country's economic policy, foreign policy, or domestic policy.

e) *Is the wording personal?* Personal wording in any scenario should be avoided, more so in the case of sensitive/controversial issues.

f) *Is the wording too direct?* Questions need to be specific but not direct as they may not get any response and may rattle the respondent.

g) *Loaded questions:* Sometimes the researcher's own biases can also creep in and may affect the wording of the questions. For example, the questions 'Do you still smoke?' or 'Do you still beat your wife' are loaded ones as the researcher has already loaded his bias into the questions to get a desired response.

h) *Recall items:* People's ability to recall the past is very limited. They may not be able to recall something, which happened more than six month ago. Thus, if recall is necessary, the time frame should be as recent as possible. In rural areas, festival like Holi, Diwali and Id or even Hindu calendar months like Sawan could be mentioned as a reference point.

i) *Unfamiliar terms and jargon:* Unfamiliar terms and jargon could cause a lot of problems for the respondents not only in understanding the question, but it may also confuse them. Take, for example, a question like 'Do you think India is moving away from a socialistic model of development?' Terms such as 'socialistic model of development' are not likely to be well understood by typical survey populations. When a term not in popular usage has to necessarily be used in an item, the interviewer must precede the item with a brief explanation. Wherever possible, familiar terms should be substituted for unfamiliar terms.

j) *Questions requiring inaccessible information:* Sometime, a question may use familiar terms, but require information most respondents would not know or would not like to share. Take, for example, a question such as 'What is your family income?'. Now, if the investigator asks this question to a family

member other than the head of household/chief wage earner in an agrarian economy setup, then he may not get an appropriate answer.

k) *Complexity and memory overload:* In complex research issues, sometimes the researcher tries to frame the questionnaire in a manner as complex as the research issue, without realizing that by doing so he is likely to overtax the respondent. The more complex the research objective or issue, the easier it is to overload memory. If there are over five alternatives, a show card should be used to allow the respondent to view the alternatives, not simply hear them orally in an interview situation.

l) *Problem of dangling alternatives:* For example, 'Would you say that you very strongly approve, strongly disapprove', presents 'dangling alternatives', which the respondent must memorize before even understanding the question. This can result in first-presented response, or in negativity. Besides grammar, it is important to ensure that the survey instrument takes into account the language and dialect people speak in the region where the survey is going to be conducted, either at the tool formulation stage or at the translation stage (see Box 3.1).

BOX 3.1
Translation of Survey Items into Vernacular Languages

One of the key issues in a large nationwide study is translation of survey items into the regional/vernacular languages. Thus, while translating survey tools, researchers should take care of semantic equivalence, conceptual equivalence and normative equivalence of survey items (Behling and Law, 2000).

This is done through the translation/back translation method where independent translators translate from one language to another, and then back again, to see if the original and re-translated items remain the same.

Pre-testing and Finalization of Research Instruments

All the research instruments developed for the study should be thoroughly tested in order to ascertain their suitability in actual field conditions. Professionals with the support of field executives need to carry out the pre-testing exercise.

Pre-testing is considered an essential step in survey research. It is not only critical for identifying questionnaire problems but it also helps in removing ambiguities and other sources of bias and error. It can also highlight any problem interviewers may have regarding the language of the questions and the skip patterns.

Regarding pre-testing, Converse and Presser (1986) argue that a minimum of two pre-tests are necessary. They suggest that the first pre-test should have twice the number of items as the final pre-test, as one of the purposes of the pre-test is to identify weaker items and drop them from the survey. Items may also be dropped if the first pre-test shows that they have little variance to be accounted.

Pre-testing incorporates different methods or combinations of methods. These techniques have different strengths and weaknesses. Some of these techniques are highlighted in the next section.

Types of Pre-testing

Pre-testing techniques[8] can be further classified into two major categories based on the methodology and approach used for pre-testing: (i) pre-field techniques and (ii) field techniques. Pre-field

techniques are generally used at the initial stages of research instrument development through respondent focus groups and interviews.

In the field type of pre-testing, questionnaires are tested under field conditions and include techniques such as behaviour coding, interviewer and respondent debriefings and the analysis of non-response items and response distributions.

Pre-field Techniques

a) *Focus groups:* Focus group helps in identifying variations in questionnaire items, language, or interpretation of questions and pre-coded options. Self-administered questionnaires can be pre-tested in a focus group, to learn about the appearance and formatting of the questionnaires. Focus groups also produce information and insights that may be less accessible without the group.
b) *Cognitive laboratory interviews:* Cognitive laboratory interviews are also generally used early in the questionnaire.

Field Techniques

a) *Behaviour coding:* Behaviour coding, as the name suggests, depends on interactions between the respondent and the interviewer to decide about the relevancy of language, content and interpretation of questionnaire items.

   The focus of behaviour coding is on how the respondent answered the question and how the interviewer tried to ask the question. For example, if a respondent asks for clarification after hearing the question, it is likely that some aspect of the question may have caused confusion.
b) *Interviewer debriefing:* It tries to minimize interviewer bias by making questions simple and clear. It tries to assess whether the interviewer has understood the question correctly.

MEASUREMENT SCALES

There are four types of scales that are used in measurement: nominal, ordinal, interval, and ratio scales. In fact, they follow a hierarchy of measurement scales, nominal being at the lowest rung of the hierarchy and even application of statistical procedure are classified in relation to the scale used. They are categorized into two groups: categorical and continuous scale data, where nominal and ordinal scales are categorized together as categorical data while interval and ratio scales are grouped together as continuous data.

Nominal data having unordered scales are called nominal scales, for example, the gender categories male and female. Categorical data having ordered scales are called ordinal scale. In the case of continuous data, scales representing interval data are called interval scales and data having both equal intervals and an absolute zero point are called ratio scales.

a) *Nominal variables:* The values of the nominal variable data have no numeric meaning as no mathematical operation except counting can be done on the data. They are, in fact, used for classifying whether the individual items belong to some distinctively different categories. For example, we can say that individuals are different in terms of variables like gender, race, colour, caste, etc. However, apart from counting, no other mathematical operation can be carried out on these variables.

b) *Ordinal variables:* Ordinal variables, unlike nominal variables, allow us to rank the items we measure in terms of order and we can specify that higher order items definitely represent more of the quality extent represented by the variable, but we still cannot tell how much more than the other item. A typical example of an ordinal variable is the rating assigned to the impact of a programme, like excellent, average and poor. Now we can say that x per cent rated the programme as excellent, y per cent rated it as average and another z per cent rated it poor, but researchers cannot say for sure that the difference between excellent and average is same as that of average and poor. In the case of ordinal variables, only certain mathematical variables such as greater than or less than are feasible and only measures such as median and range can be calculated.

c) *Interval variables:* Interval variables provide more flexibility in terms of measurement as it not only allows us to rank the measured items but can also help in quantifying the size of the difference between them. For example, temperature, as measured in degrees Fahrenheit or Celsius, constitutes an interval scale. We can say that a temperature of 80 degrees is higher than a temperature of 40 degrees, but still we cannot say 80 degree is twice as hot as 40 degrees. Another example is the measure of time using the BC/AD system (here the initial point of reference is assumed to be zero). We have simply constructed a reference scale to measure time, which does not have a true or rational zero.

d) *Ratio variables:* Ratio variables measured by scale not only have equidistant points, but also have a rational zero. Thus, in addition to all the properties of interval variables, they feature an identifiable absolute zero point. A typical example of a ratio scale is the Kelvin temperature scale. In this case we can not only say that a temperature of 60 degrees is higher than one of 20 degrees, we can also specify that a temperature of 60 degrees is three times as hot as a temperature of 20 degrees. Most of the variables we use to measure in field situations conform to ratio scale properties, though most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales.

Attitudinal Scales

Attitudinal scales are composite scales, which try to bring objectivity into subjective concepts of aptitude and attitude. They measure underlying traits and behaviours such as trust, joy, patience, happiness or verbal ability. Thus, attitudinal scales are also defined as measures that try to quantify abstract and subjective behaviour and attitudes.

A scale is always unidimensional, which means it has construct and content validity. It is important to point out that the terms scale, index or benchmark should be used with caution. Index is a specialized scale, wherein highly correlated individual items are taken together to form a scale (see Box 3.2). The next section lists some of the most widely used attitude scales in social research.

Thurstone Scales

Thurstone scales, developed in 1929 by Thurstone and Chave, is one of the best-used techniques in attitude measurement for measuring a core attitude when there are multiple dimensions or concerns around that attitude. In Thurstone scaling, researchers usually ask a panel of judges to comment on relevant and conceivable questions (say 100 questions) to develop a scale.

The usual procedure of Thurstone scaling involves judges, who rank opinion statements into a set of order. Judges then sort out the statements as favourable or unfavourable vis-à-vis the variable of interest. When the judges are finished, for each judge the statement slips will be ordered into numbered piles. Each statement is allotted the number of its pile. Next, the slips are sorted out by

statement and the median pile value is determined for each statement. Statements are then sorted out into piles as per their median value. The researchers then select some statements from each pile to construct a scale, giving preference to those statements the judges agreed on while ranking. Further, researchers can administer the questionnaire to the panel to analyse inter-rater reliability. Researchers can also use the discrimination index to avoid non-homogenous items.[9]

BOX 3.2
Benchmark and Indexes

Benchmark: Benchmark, as the name suggests, is the standard or target value, accepted by professional associations or a group of organizations. It may be composed of one or more items. The observed values are compared against the benchmark value to ascertain the project's performance.
Indexes: Indexes are summative measures, constituting a set of items, which measure the latent underlying variable's characteristics. Further, all items in an index are highly correlated with each other.

Likert Scales or Summated Ratings Scale

The summated ratings scale/Likert scale was developed in 1932 by Rensis Likert as a five-point, bipolar response scale. It tries to assess people's agreement/disagreement, approval/disapproval on a five-point scale.

In constructing a Likert scale, a large number of statements are collected. In the next step, ambiguous, irrelevant statements are omitted. The remaining statements are then given to a few respondents who are asked to indicate their reaction to them using a five-point rating system: strongly approve, approve, undecided, disapprove and strongly disapprove. These categories are then assigned values of 5, 4, 3, 2 and 1 respectively. The correlation between statement scores and the total score is then ascertained. Those statements, which have a high correlation with the total score are then selected for the final scale. Researchers can also use index of discriminating power to select appropriate items for the scale (see Box 3.3).

BOX 3.3
Index of Discriminating Power

Index of discriminating power (DP): Index of discriminating power is used as a criterion for choosing more appropriate Likert items over other probable items. Scale items whose mean scores of the top 25 per cent of the respondent's score is different from the bottom 25 per cent of the respondent's scores have high DP coefficients.

Guttman Scaling

Guttman scaling, also known as scalogram analysis, was developed in the 1940s as a proposed method for scaling attitude items. It is based on the fact that attitudes can be arranged in an order that a respondent, who positively answers to a particular item, also responds positively to other items lower in rank.

It is based on the assumption that ordering of certain stimuli is possible. Thus, if an individual dominates a stimulus, he will also dominate other stimuli. These scales are also defined as ones in which the items constitute a one-dimensional series such that an answer to a given item predicts the answer to all previous items in the series. The scoring system is based on how closely they

follow a pattern of an ever-increasing hardened attitude towards some topic in the important questions.

Coefficient of scalability (which can be abbreviated as Cs)[10] is the standard method for assessing whether a set of items forms a Guttman scale. It usually follows that Cs should be .60 or higher for the items to be considered a Guttman scale.

$$Cs = 1 - E/X$$

Where E is the number of Guttman error
X is the number of errors expected by chance

Stouffer's H technique is a variant which gives greater stability to Guttman scale by basing each Guttman scale position on three or more items, rather than just one.

Mokken Scales

Mokken scales are similar to Guttman scales but they are probabilistic whereas Guttman scales are deterministic, that is, in a Mokken scale, a respondent answering all items positively will have a significantly greater probability than answering a null answer to a less difficult item. Whereas, in a perfect Guttman scale, answering an item positively indicates that the respondents will answer all less difficult items positively also.

Loevinger's H coefficient measures the conformity of a set of items to the Mokken scale. Loevinger's H is based on the ratio of observed Guttman errors to total errors expected under the null hypothesis.

Semantic Differential

The semantic differential scaling procedure was developed by Osgood in the 1950s to deal with attitudes such as emotions and feelings. It measures people's reactions to words and concepts in terms of ratings on polar scales and is based on the idea that people think dichotomously or in terms of polar opposites while forming opinion such as good-bad, or right-wrong.

In order to formulate a suitable semantic differential scale, several factors need to be considered the most important being the need to consider whether the scale is balanced or not. In fact, the semantic differential scale can be used with any adjective by collecting response patterns to analyse for scaling purposes. In order to quantify a semantic differential, a Likert-type scale is used and the endpoints are assumed to be extremes such as 'very bad' or 'very good' and another important consideration is to ensure that the scale is balanced, that is, either side of the indifferent cues have an equal number of cues.

Reliability and Validity

The terms reliability and validity are generally used as synonyms, though they have very different meanings when applied in statistics. Reliability and validity are two very important concepts that

deal with the psychological characteristics of measurement and its precision. As we all know, measurements are seldom perfect, especially in the case of questionnaire responses or processes, which are difficult to measure precisely and thus often result in measurement errors. Besides, reliability and validity, precision and accuracy of instruments/tests are two other terms that are often confused by people while reporting measured outcomes (see Box 3.4).

BOX 3.4
Precision and Accuracy

People often confuse the concepts of precision and accuracy, especially those who do not have a mathematical background. Precision signifies perfection in an instrument and assesses how finely an estimate is specified, whereas accuracy refers to how close an estimate is to the true value. Precision relates to the quality of a process through which a result is obtained, while accuracy relates to the quality of the result. It is important to note that estimates can be precise without being accurate, as in case of a computer output containing results specified to the fourth or sixth decimal place.

Reliability[11]

Reliability signifies the issue of consistency of measures, that is, the ability of a measurement instrument to measure the same thing each time it is used. There are three important factors involved in assessing reliability, the first being stability, which entails asking whether a measure is stable over time so that researchers can be confident that results relating to the measure for a sample of respondents will not fluctuate. The second issue is that of internal reliability, which seeks to assess whether the indicators that make up the scale or index are consistent. Inter-observer consistency is another key factor, which may arise due to the involvement of more than one observer in activities such as recording of observation or translation of data into categories.

Validity[12]

Validity tries to assess whether a measure of a concept really measures that concept, that is, the extent to which the concept measures the thing it was designed to measure. Thus, when people raise questions about the relation between a person's IQ test and his general level of intelligence, it signifies that they doubt the measurement validity of IQ tests in relation to the concept of intelligence. Thus, while IQ tests will have high reliability they might have low validity with respect to job performance. Thus, for a research study to be accurate, it is imperative that the findings are both reliable and valid.

It is important to point here that although reliability and validity are two different concepts, they are related in some way because validity presumes reliability, which means that if a measure is not reliable it cannot be valid, though the opposite is not true and a study can be reliable even if it is not valid. There are various threats to validity as well as reliability and some of these can be avoided if internal validity is ensured. This can be done if the researchers use the most appropriate research design for their study.

Methods of Measuring Reliability

There are four good methods of measuring reliability:

Test-retest Technique Test-retest technique is generally used to administer the same research instrument/test/survey or measure to the same group of people twice under the same conditions, but at different points in time. Reliability estimates are expressed in the form of a correlation coefficient, which is a measure of the correlation between two scores in the same group.

Multiple Forms Multiple forms are also known by other names such as parallel forms and disguised test-retest. It tests the reliability of the research instrument by mixing up the questions in the research instrument and giving it to the same respondents again to assess whether it results in any different responses.

Inter-rater Inter-rater reliability is used to assess the reliability of research tool instruments/tests when more than one rater/interviewer is involved in interviewing or content analysis. It is calculated by reporting the percentage of agreement on the same subject between different raters or interviewers.

Split-half Reliability In the case of the split-half reliability method, as the name suggests, half of the indicators, tests, instruments, or surveys, are analysed assuming it to be the whole thing. Then, the results of this analysis are compared with the overall analysis, to assess the reliability of the indicators, tests or instruments. Nowadays, researcher use Cronbach's alpha[13] to test internal reliability and it correlates performance on each item with an overall score. Kuder-Richardson coefficient[14] is another technique, which is used nowadays to measure internal reliability (see Box 3.5). These techniques can be easily calculated by using statistical packages such as Statistical Package for Social Sciences (SPSS), an example of which is discussed in Chapter 7.

BOX 3.5
Internal and Inter-rater Reliability

Cronbach's alpha is a commonly used test of internal reliability. It calculates the average of all possible split-half reliability coefficients and a computed alpha coefficient varies between 1, denoting perfect internal reliability, and 0, denoting no internal reliability. The figure of .75 or more usually is treated as a rule of thumb to denote an accepted level of reliability.

Tau-equivalent:[15] Different measures have identical true scores but need not have equal error variances. It is believed that for alpha to be a correct measure of reliability, the items constituting it need to be at least Tau-equivalent and if this assumption is not met, alpha is considered as a lower bound estimate of reliability.

Congeneric measures: Congeneric measures are based on the assumption that different measures have only perfect correlation among their true scores. Thus, it is not necessary that measures would have identical error variances or true score errors.

Inter-rater reliability tries to ascertain the reliability of the single rating. It is defined as the extent to which two or more individuals agree on a rating system. It addresses the consistency of the implementation of a rating system. There are various ways in which inter-rater reliability can be ascertained, one of which is to analyse the 'intra-class' correlation, which assumes that the raters have the same mean. For purposes such as planning power for a proposed

(*Box 3.5 continued*)

(*Box 3.5 continued*)

study, it does matter whether the raters to be used will be exactly the same individuals. Bland and Altman(1999: 135–60) proposed a very good methodology. They advise researchers to use two methods, whose difference in scores can be plotted against the mean for each subject.

Methods of Measuring Validity

Researchers should be concerned with both external and internal validity. External validity signifies the extent to which a research study can be generalized to other situations.

Internal validity refers to the true causes, which result in an outcome. In other words, it signifies the (i) the rigour with which the study was conducted and (ii) the extent to which the designers of a study have taken into account alternative explanations for any causal relationships they explore (Huitt, 1998). Internal validity is constituted of four broad sub-categories as discussed below:

Face Validity Face validity refers to validity that establishes the fact that the measure apparently reflects the content of the concept in question. Face validity is an intuitive process and is established by asking other people whether the measure seems to capture the concept that is the focus of attention. It is essentially an assertion on the part of the researchers that they have reasonably measured the concept they intended to measure.

Content Validity[16] Content validity, as the name suggests, tries to assess whether the content of the measurement technique is in consonance with the known literature on the topic. If the researcher has concentrated only on some dimensions of a construct or concept, then it is believed that other indicators were overlooked and thus the study lacks content validity. It can easily be estimated from a review of the literature on the concept/construct topic or through consultation with experts in the field of the concept. Thus, this process ensures that the researcher has covered all the conceptual space. Content validity is usually established by content experts. Thus, it is imperative to ensure that experts do not take their knowledge for granted and do not consider other people to have the same level of intelligence.

Criterion Validity Criterion validity is also known as instrumental validity. It draws an inference from test scores about performance and demonstrates the accuracy of a measure or procedure by comparing it with another standard valid procedure.

There are different forms of criterion validity: in concurrent validity, researchers seek to employ a criterion on which cases/subjects are known to differ and assess how well the criterion captures the actual behaviour; in the case of predictive validity, researchers use a future criterion measure to assess how well it estimates future events that have not happened yet.

Construct Validity In construct validity, researchers are encouraged to deduce the hypothesis from a theory that is relevant to the concept. Construct validity can be further segmented into two sub-categories: convergent validity and discriminate validity. In the case of convergent validity, validity is gauged by comparing it to measures of the same concept developed through other methods to assess how well the items are together (convergent validity) or distinguish different people on certain behaviours (discriminate validity).

DATA COLLECTION

Orientation of Professionals

Internal meetings-cum-workshops should be organized where all the professionals associated with the project are briefed on the objectives, methodology, research techniques, study instruments and guidelines for the training of field staff. This helps in creating a common understanding among all the professionals.

Recruitment of Field Staff

The project coordinator/principal research investigator in association with the core team members needs to look after the recruitment of the field staff. The recruitment needs to be done from the existing panel of field personnel and also from among fresh candidates applying for jobs at the local field office in response to advertisements in the local newspapers. Candidates having the desired qualifications and experience in conducting surveys should be recruited for the study. Recruitment should be 20 per cent more than the actual requirement to make up for attrition after training and the dismissal of candidates whose work is not found to be up to the mark.

Briefing to Field Staff

The professionals involved in the study should be involved in briefing the field staff. All the field persons engaged for the survey should be given extensive training. The training sessions should consist of instructions in interviewing technique, field procedures for the survey, time schedules, detailed instructions on schedules and manuals and each item in the questionnaire followed by mock exercises between the participants.

Field Work

Selection of Study Units and Respondents

Appropriate and required sample of respondent categories should be selected from survey areas using the sampling technique as commonly agreed. The core team members and the project associates need to be responsible for this exercise.

Operational Aspects of Fieldwork

The fieldwork for the study needs to be initiated immediately after the briefing/training of the field staff is complete. The entire team needs to work under the overall guidance of the project coordinator for the study.

Collection of Secondary Data

Some studies also involve the collection of secondary data along with the main survey data. This exercise can be a simultaneous activity along with the main survey. Prior appointment needs to be taken from the designated officials.

The team should collect all necessary secondary information from the records, registers and documents available at the respective offices and individuals. The secondary information needs to be collected using a standard information sheet prepared in line with the objectives of the study and finalized in consultation with the client. The core team members should also visit the field to oversee the progress of secondary data collection and the quality and completeness of the information collected.

Quality Control of Field Data and Monitoring

For proper monitoring of fieldwork and ensuring the quality of data collected it is imperative that emphasis be given to the following aspects of field-work.

a) Observation of some of the interviews/discussions carried out by the field staff.
b) Spot checks to verify the accuracy of the information collected.
c) Back checks.
d) Maintenance of log sheets by field executives indicating team performance.
e) Visits by the concerned research professionals for monitoring fieldwork and providing technical guidance to the field staff.

DATA PROCESSING AND ANALYSIS

The system analyst needs to look after data processing and analysis. The project coordinator and the core team members should provide inputs at various stages of data processing and analysis.

Once the research data has been collected, the process of preparing it for analysis begins. Quantitative data will need to be sorted and coded and even qualitative data will need to be indexed or categorized, in preparation for analysis.

Coding

Coding is defined as the process of conceptualizing research data and classifying them into meaningful and relevant categories for the purpose of data analysis and interpretation. A number is assigned to each category, in the form of a code, for example, in the case of the gender variable, code 1 is assigned to males and code 2 is assigned to females. Coding formats may be included on the questionnaire, or can also be developed after the data have been collected in cases where respondents' replies do not fall into pre-coded response categories, that is, in the case of open-ended questions and for pre-coded questions which have an 'other' code.

The basic rules for the development of the coding scheme known as the coding frame for quantitative data are that the codes must be mutually exclusive, coding formats for each item must be comprehensive and the codes must be applied consistently whereas coding rules for qualitative data permit the allocation of responses to more than one category in order to facilitate conceptual development.

Interview data can be hand coded by the interviewer during or after the interview, that is, field coding can be done directly on to the paper questionnaire. However, it often requires coding in the office by a coder, or a team of coders.

a) *Coding boxes:* While designing questionnaires, coding boxes should be allocated for each question. It is important to point out that each coding box must contain only one number and for answers that have been allocated a two-digit code, two coding boxes need to be provided—one for each number.
b) *Coding transfer sheets:* In a majority of the cases, pre-coded questionnaires are used for data entry though in some cases coding transfer sheets for each questionnaire, containing the transferred codes from each question, are also used. Coding transfer sheets are used in cases where the investigator does not wish to clutter the questionnaire with numerical codes and coding boxes, but it doubles the administrative effort and entry costs.

   Irrespective of the method used, they should specify exactly where in the individual's computer records each item of data is to be placed. This is usually done by allocating variable names to each question, which are stored in the computer's data entry programme in a predefined sequence as well as on the coding frame.
c) *Numerical values for codes:* In the case of quantitative analysis, it is essential that the collected information is coded either quantitatively in the form of a measurement such as weight in kilograms or 'qualitatively' in the form of a category so that the numbers in each group can be counted. Thus, for gender the groups are male and female; for marital status the groups are married, single, widowed, divorced and separated. Further, each of the categorized groups to be analysed require a numeric value before they can be entered on to the computer, counted and analysed. For example, dichotomous responses such as male and female choices could be scored 1 and 2 respectively.
d) *Coding open questions:* Open-ended questions form an integral part of questionnaire as it allows respondents to use their own words and form their own response categories. Open-ended questions responses are then listed by the investigator after the data has been collected, which can be grouped by theme for the development of an appropriate coding framework. Even in the case of a structured questionnaire, pre-coded response options have the provision for the 'others' category thus making it imperative that a list is prepared to develop a coding frame for the various 'other' response choices that were offered to respondents and whose replies did not fit the codes given.

e) *Coding closed questions:* Closed-ended questions require that any groupings should be defined before the data are collected. The response is then allocated to the pre-defined category, with a number assigned. The response is then itself an item of data ready for transfer to coding boxes, data entry and analysis.

Data Entry on to the Computer

As with the coding, the process of verification of office data entry involves two data entry persons independently entering the data. This is also known as double data entry. Double data entry should be supported by the use of a computer programme, which can check for any differences in the two data sets, which then have to be resolved and corrected by a member of the research team.

Human coding and entry or direct electronic data entry are usually preferred. With the latter, the computer displays each question on the screen and prompts the interviewer to input the response directly, whereupon it is programmed to store it under the correct code. Coded data is stored in the form of a data table in a computer file. Statistical software packages contain facilities for entering the data, which can be read directly by that package, although many packages can translate data typed into other programmes. The semi-structured schedules and the in-depth interviews would be entered with the help of latest data entry software packages.[17] Integrated System for Survey Analysis (ISSA 6.0) is a very good software having inbuilt checks.

To ensure data quality, researchers usually key in all the data twice from raw schedule and the second data entry is done by a different key entry operator whose job also includes verifying mismatches between the original and second entries. It is observed that this kind of 'double entry' provides a high 99.8 per cent accuracy rate for all data entered.

Creation of the System File

The computer package chosen for analysis will have the facility for the creation of a system file before the data can be entered. For example, the Statistical Package for the Social Sciences (SPSS) system file will require the labelling of all the variables and their response choices, the number of columns to be assigned to each and determination of which codes are to be assigned as missing in the analyses. This should be done before the coding has been completed so that it is ready before the data entry is due to start.

Cleaning the Data

Once the data has been stored in computer readable form, the next task is to eliminate the more obvious errors that will have occurred during the data collection, coding and input stages. An edit

programme will need to be specified. This should look at missing values, skips, range checks and checks for inconsistency.

An edit programme will require a set of instructions for the computer package used that will automatically examine, and draw attention to, any record that appears to have an error in it.

a) *Range checks:* For data fields containing information about continuous variables like height and weight, observations should fall within a specified range. Thus, if the height of an adult male falls outside the normal range it should be checked.
b) *Consistency checks:* Often certain combinations of within-range values of different variables are either logically impossible or very unlikely. Data entry programme shall have some checks to ensure data consistency, for example, a person who has undergone sterilization should not be using the spacing method of birth control. These checks will not eliminate all the errors introduced during the data collection, coding and data input phases, but will certainly minimize the errors. There is no substitute for careful recording of data, coding, data entry and verification.
c) *Missing values and data checks:* There are two types of missing values: first, where a question is deliberately blank because it did not apply to the individual respondent or where a reply was expected but was not given, which is known as an inadequate response. Such missing cases can occur because the respondent refused or did not answer the question, or because the interviewer forgot to ask it.

It is also customary to use 9 or 99 (as close ended codes for options) for questions, which do not apply to the respondent (NAs); for example, skips in the questionnaire will be employed so that men will not be asked about specific questions. The inadequate and do not apply response codes are then set as missing on the computer, so they are not routinely included in the analyses.

Checking for Bias in the Analyses

Response Bias

Response bias is one of the most common phenomenons, which is observed during data collection. As much information as possible should be collected about non-responders to research in order that the differences between responders and non-responders to a research study can be analysed, and the extent of any resulting bias assessed.

In order to check for age bias, for example, the investigator should compare the age structure of the respondents with that of the non-responders, or that of the study population as a whole.

Interviewer Bias

In practical situations, interviewer bias is more commonly observed than response bias. Where more than one enumerator, interviewer or observer has been used, comparisons should be made between the data collected by each one. As there is bound to be some variation in the way an interviewer asks a particular question and the only way to remove interviewer bias is to provide rigorous training followed by field exercise.

ANALYSIS OF DATA

The core team members and the system analyst under the guidance of the project coordinator shall prepare the analysis/tabulation plan. The tabulation plan will be finalized in consultation with the client. The required tables can then be generated using the latest version of analysis software[18] like Stata, SAS, or SPSS. Though in the case of qualitative analysis, researchers shall first focus on transcription of qualitative data, which can later be used during content analysis (see Box 3.6).

BOX 3.6
Transcription and Content Analysis

In qualitative research it is believed that data analysis is as good as transcription of raw data. Transcription is an important stage in qualitative data analysis and almost all qualitative research studies involve some degree of transcription. Transcription is not about jotting up or summing up what a researcher, interviewer or transcriber feels. It is all about what the respondent feels.

The recorded cassettes of the focus group discussion/in-depth interviews are transcripted in the desired language by the transcriptors/project associates with the guidance of the core team members. The content analysis of the focus group discussions and the final analysis of the qualitative schedule needs to be done by the core team members with the help of the in-house qualitative researchers.

PREPARATION OF REPORT

The next step after analysis of data is to prepare the report. Though it is not necessary that in a research project the report shall be submitted only after data analysis. As per the terms agreed upon or requirements of the study, the following reports could be prepared at different stages of the study:

a) Pre-testing report.
b) An inception report prior to initiation of fieldwork.
c) A mid-term evaluation report.
d) Draft report.
e) Final report.

A report generally consists of the following sections:

a) Executive summary.
b) Introduction/background.
c) Findings.
d) Summary, conclusion and recommendation.

The executive summary is the portion of the report that most people read. Though not as a rule but ideally this section should be around three to five pages long, with bullets providing as much information as possible. It should contain all relevant information starting from project background,

research methodology to findings and recommendations so that the reader has an overall understanding of the basics of the project.

The introduction should include all relevant information necessary to understand the context and implementation of the programme from its inception through the current reporting period. It clearly describes the goal and objectives that the project expects to achieve. Next, it should detail the research objectives of the study and research design and methodology adopted to assess the research objectives. It should also clearly specify the study area and sample size of the study, besides detailing out the timeline of the project.

The findings are the soul of the evaluation report. It presents the results of the various instruments described in the methodology section. Findings present in chapters shall align itself to research objectives of the study. Further findings may or may not be summed up in one chapter; it may be presented in two or three chapters depending on the objectives and complexities of the project.

The last section of the report shall comprise of conclusions, discussions and recommendations. It provides a final interpretation of success or failure of the project and how the programme can be improved. In presenting this information, some of the key points discussed are (i) whether the project achieved the desired result, (ii) certainty that the programme caused the results and (iii) recommendation to improve the programme.

Notes

1. A design is a plan that dictates when and from whom measurements will be gathered during the course of an evaluation. The first and obvious reason for using a design is to ensure a well-organized evaluation study: all the right people will take part in the evaluation at the right times (Fitz-Gibbon and Morris, 1978: 10).
2. Evaluation designs generally fall into one of four types: (i) experimental, (ii) quasi-experimental, (iii) survey, or (iv) naturalistic.
3. Experimentation represents the process of data collection and so refers to the information necessary to describe the interrelationships within a set of data. It involves considerations such as the number of cases, sampling methods, identification of variables and their scale types, identification of repeated measures and replications.
4. Cook and Campbell (1979) mention 10 types of experimental design, all using randomization of subjects into treatment and control groups.
5. The Hawthorne effect is so named because the effect was first observed while Elton Mayo was carrying out pioneering research in industrial psychology at the Hawthorne plant of Northern Electric.
6. Some authors, on the contrary, argue that case studies are a prototype of quasi-experimental design if pursued systematically.
7. Survey research is the method of gathering data from respondents thought to be representative of some population, using an instrument composed of closed structure or open-ended items and survey instrument is the schedule of questions or response items to be posed to respondents.
8. Pre-testing is used to assess reliability of the research instrument and if there is any remaining doubt about the reliability of one or more items, the researcher should consider split sample comparisons, where two versions of the same item appear on two different survey forms administered randomly.
9. Q-dispersion is usually used to have a measure of item ranking measurement, which is quite similar in nature to standard deviation.

10. The coefficient of reproducibility (Cr) is an alternative measure for assessing whether a set of items form a Guttman scale and is defined as:

    $Cr = 1–E/N$

    Where E refers to number of choices and N denotes the number of subjects.

11. For a brief but in-depth discussion of reliability, including statistical formulae for calculating reliability, see Thorndike et al. (1991).
12. The Joint Committee of the American Educational Research Association, American Psychological Association, adds that 'validity ... refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from data'.
13. Cronbach's alpha was popularized in 1951 by an article by Cronbach on the work in the 1940s by Guttman. The widely accepted cutoff is that alpha should be .70 or higher for a set of items to be considered a scale.
14. Kuder-Richardson is a special case of Cronbach's alpha for ordinal categories.
15. Psychometric literature classifies indicator sets for variables into three categories, that is, congeneric sets, which are presumed to have passed some test of convergent validity such as Cronbach's alpha. Tau equivalents also have equal variance for the indicators. Parallel indicators are Tau-equivalent having equal reliability coefficients.
16. Content validity is based on the extent to which a measurement reflects the specific intended domain of content (Carmines and Zeller, 1991: 20).
17. SPSS and SAS also have software for data entry such as SPSS Data Entry II and SAS FSPFSEDIT respectively.
18. Statistical packages are available to assist in the quantitative analysis of data. Most data base packages will be able to provide descriptive statistics like simple frequencies, average scores and so on, but specific package such as SPSS, Stata for quantitative analysis are frequently used (details are provided in Chapter 7).