RV College of Engineering®
Mysore Road, RV Vidyaniketan Post,
Bengaluru - 560059, Karnataka, India

*Go, change the world*

# RV COLLEGE OF ENGINEERING®

(Autonomous Institution Affiliated to Visvesvaraya Technological University, Belagavi)

# UNIT 3

## RESEARCH METHODOLOGY

**22IM21T**

**EL REPORT**

Submitted by


**Bindu Priya R (1RV23SCS02)**

**Sanjana Otihal(1RV23SCS10)**

**Trupti Hegade(1RV23SCS17)**

Under the guidance
of

**Dr. MANONMANI S**

Dept. of CSE

RV College of Engineering

Bengaluru - 560059


***Submitted in partial fulfillment of the requirements for the award of a degree of***

**MASTERS OF TECHNOLOGY IN
COMPUTER SCIENCE ENGINEERING**

**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING**

**2023-24**

# UNIT 3- RESEARCH DESIGN FOR DATA ACQUISITIONS

# CHAPTER 10: MEASUREMENT DESIGN

The introduction emphasizes the importance of data and measurement in research. It highlights that the quality of research is directly tied to the quality of data collected, especially for hypothesis testing. While measuring physical entities like length and weight is straightforward, measuring abstract concepts in management research (such as attitudes, morale, or job satisfaction) is complex. These concepts are often multidimensional, making the measurement process more intricate. Measurement involves the use of scales, which place measured entities along a continuum. Data can be obtained in three ways: using a pre-existing validated instrument, creating and validating a new instrument, or utilizing already recorded data. The chapter will cover key aspects of measurement, including definitions, data types, scale construction, measurement errors, and ensuring validity and reliability.

## 10.1 PRIMARY TYPES OF MEASUREMENT SCALES:

Measurement is the process of assigning numbers or symbols to represent quantities of properties or behaviours in research. It involves three key characteristics:

1. **Order** – Numbers are arranged sequentially.
2. **Distance** – Differences between numbers are ordered.
3. **Origin** – The series has a unique starting point, often represented by zero.

In social sciences, **scaling** refers to the process of quantifying abstract concepts like attitudes or perceptions. Measurement data in management research can be obtained by using a pre-developed instrument, creating a new instrument, or extracting data from existing records.

**Scaling** is a procedure for attempting to determine quantitative measures of subjective abstract concepts.

**Nominal scale**: This is a measurement procedure to classify

objects, events, and individuals into categories.

There are four primary types of measurement scales:

1. **Nominal Scales:** Nominal scales are the simplest form of measurement, used for categorizing objects, events, or individuals into distinct groups. The assigned numbers or labels do not imply any order, distance, or origin. These scales are non-quantitative, and statistical operations such as mean or standard deviation cannot be applied. Nominal scales are widely used in social sciences and business research where the focus is on classification.

- **Example 1**: Where do you live? (City, Town, Village)
- **Example 2**: Do you own a car? (Yes/No)

2. **Ordinal Scales:** Ordinal scales build on nominal scales by introducing a ranking or order to the categories. While the order is clear, the actual difference or distance between the ranks is not specified. Ordinal scales allow for comparisons like "greater than" or "less than," but they do not convey the extent of difference between ranks. Common statistical operations include medians and quartiles.

- **Example 1**: Ranking objectives by importance (e.g., Quality, Cost, Flexibility)

| Objectives | Rank |
|---|---|
| Quality | ______ |
| Cost | ______ |
| Flexibility | ______ |
| Dependability | ______ |

- **Example 2**: Ordering brands by preference or product costs without specifying the price gap.

| Objective | | Preference (Sample Answer) |
|---|---|---|
| 1 | 2 | |
| New product | New process | 1 |
| New product | Quality | 1 |
| New product | Cost | 1 |
| New process | Quality | 2 |
| New process | Cost | 1 |
| Quality | Cost | 1 |

| Derived Ranks | | |
|---|---|---|
| Objective | No. of times ranked first | Derived rank |
| New Product | 3 | 1 |
| Quality | 2 | 2 |
| New process | 1 | 3 |
| Cost | 0 | 4 |

3. **Interval Scales:** Interval scales are more advanced than nominal and ordinal scales because they not only order data but also have equal intervals between measurements. This allows for meaningful comparison of differences between values, but there is no true zero point, meaning ratios are not meaningful. For example, in temperature measurement, the difference between 20°C and 30°C is the same as between 30°C and 40°C, but 0°C does not indicate the absence of temperature.

- **Example**: Rating the strength of the link between departments (Very strong, Strong, Moderate, Weak, Very weak).

**Example** The link between the R&D and marketing departments in your organisation is:

| Very strong | Strong | Moderate | Weak | Very weak |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

4. **Ratio Scales:** A ratio scale is a type of measurement that includes all the properties of nominal, ordinal, and interval scales, with the added feature of a true zero point (natural origin). This allows for the measurement of absolute quantities, enabling comparisons in terms of equality, rank-order, intervals, and ratios. Ratio scales are widely used in the physical sciences for measuring variables like weight, length, and

time. In the social sciences, ratio scales are used for quantifying attributes such as income, age, and education, though behavioral traits are rarely measured this way. Ratio scales permit the use of all statistical analyses, including operations like multiplication and division.

- **Example**: What percentage of R&D expenditures is directed to new product development? (1–20%, 21–40%, etc.).

**Example** What percentage of R&D expenditures is directed to new product development?

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1–20 | 21–40 | 41–60 | 61–80 | 81–100 |

## 10.2 VALIDITY AND RELIABILITY IN MEASUREMENT:

Validity and reliability are essential criteria for evaluating the accuracy and consistency of measurement instruments in research. These aspects help ensure that measurements reflect the actual phenomenon being studied and that the results are dependable.

1. **Validity** refers to the degree to which a test or instrument measures what it is supposed to measure. It ensures the relevance and appropriateness of the instrument.
2. **Reliability** pertains to the accuracy and precision of a measurement. It is concerned with the consistency of results when the measurement is repeated under the same conditions.

**Sources of Errors in Measurement:** Measurement errors may arise from various sources:

- **Researcher**: Incorrect questions, inappropriate analysis, misinterpretation, and bias due to experimenter expectations.
- **Sample**: Use of the wrong target population, method, or individuals.
- **Interviewer**: Bias, incorrect interpretation, and carelessness.

- **Instrument**: Issues with the scale (e.g., rounding off) or the questionnaire (e.g., ambiguity, positional bias).
- **Respondent**: Inconsistencies, ego or humility, fatigue, lack of commitment, or random responses.

**Practicality** refers to factors like economy, convenience, and interpretability when using an instrument.

**Measurement Accuracy and Error:**

- **Measurement accuracy** can be calculated as the ratio of recorded value (r) to the true value (t):

  $$\text{Measurement accuracy} = r/t$$

- **Measurement error** can be defined as

  $$\text{Measurement error} = (1 - r/t)$$

Validity focuses on whether t (the true value) is being measured correctly, while reliability is concerned with the consistency of r (the recorded value). Reliable measurements show low variability in r, and valid measurements closely align r to t.

**10.2.1 Validity of Measurement**

After designing a measuring instrument, it is important to assess its usefulness through **validity**—the degree to which the instrument measures what it is intended to measure. Validity ensures that differences in measured values reflect true differences in the characteristic being studied. There are several types of validity:

1. **Internal Validity**: Refers to how accurately the instrument measures the characteristic it is supposed to. Three key forms of internal validity are:
   - **Content Validity**: Measures whether the instrument provides adequate coverage of the topic under study. Content validity can be evaluated by defining the relevant dimensions of the topic and ensuring the instrument covers them appropriately.

- **Criterion-Related Validity**: Reflects the ability of the instrument to predict an outcome (predictive validity) or categorize a current condition (concurrent validity). It is used for empirical estimating purposes and must meet qualities like relevance, freedom from bias, and reliability.
- **Construct Validity**: Tests how well the results from an instrument fit the underlying theory. It ensures that the instrument captures the intended abstract characteristic (such as attitude or personality). Construct validity can be assessed through **convergent validity** (where different instruments measuring the same concept show high correlation) and **discriminant validity** (where variables predicted to be uncorrelated remain so).

2. **External Validity**: Refers to the generalizability of research findings to broader populations, settings, and variables.

To fully establish **construct validity**, researchers may use methods like the **multi-trait multi-method matrix (MTMM)**, which helps verify convergent and discriminant validity.

The **Multi-Trait Multi-Method Matrix (MTMM)** is a tool used to assess both **convergent** and **discriminant validity**. Developed by Campbell and Fiske in 1959, it involves measuring multiple traits using multiple methods and analyzing the correlation among these measures.

Here's a breakdown of how the MTMM works using a correlation matrix:

**Basic Concept**

- **Convergent Validity**: The correlation between different measures of the same trait should be higher than the correlation between measures of different traits.
- **Discriminant Validity**: The correlation between different measures of different traits should be lower than the correlation between different measures of the same trait.

**Example Matrix (Table 10.2)**

| Construct (trait) | Measurement Method | Measure | Correlation Matrix |
|---|---|---|---|
| | M1 | M2 | M3 |
| **C1** | MM1 | 1.00 | 0.92 |
| **C1** | MM2 | 1.00 | 0.50 |
| **C2** | MM3 | 1.00 | 0.83 |
| **C2** | MM4 | 1.00 | |

**Interpretation:**

- **Convergent Validity**: For trait C1, correlations between different measurement methods (M1 with M2) are relatively high (0.92). This indicates that different methods are yielding similar results for the same trait.
- **Discriminant Validity**: The correlations between different traits measured by the same method (e.g., M1 with M3) are lower (0.30), which suggests that the instrument is distinguishing between different traits effectively.

### 10.2.2 Reliability of Measurement

**Reliability** refers to the consistency and stability of a measurement instrument. A reliable measure yields consistent results across different instances of use, ensuring that transient and situational factors minimally affect the outcome. Reliability is essential but not sufficient for validity; a valid measure must also be reliable.

**Aspects of Reliability**

1. **Stability**
   - **Definition**: The ability of a measure to produce consistent results over time, despite changes in conditions or the state of the respondents.
   - **Tests of Stability**:

- **Test-Retest Reliability**: This involves administering the same measure to the same group of respondents on two different occasions and assessing the correlation between the two sets of results. High correlation indicates good stability. Challenges include:
  - Short time intervals may lead to memory recall rather than true changes.
  - Transient factors might affect responses at the time of retesting.
  - Respondents may alter their views after the first measurement.

To address these issues:

  - Extend the time between measurements to reduce recall bias.
  - Randomly assign respondents to different groups and compare results.
  - Use statistical measures like standard deviation to assess variability.
- **Parallel-Form Reliability**: This involves using two different but equivalent forms of a measure that assess the same construct. High correlation between the two forms suggests minimal error due to wording or question order differences.

2. **Equivalence**
   - **Definition**: The reliability of measurements across different conditions or among different investigators. This aspect ensures that variations in measurement results are due to the constructs being measured, not due to differences in administration or sample items.
   - **Tests of Equivalence**:
     - **Inter-Rater Reliability**: Comparing measurements from different observers or interviewers to ensure consistency in observations.

- **Item Sample Equivalence**: Using alternative sets of questions that measure the same construct and comparing the results to check for consistency.

- **Split-Half Reliability**: This technique involves dividing a test into two halves and comparing the results of each half to assess consistency within the measure. It helps to determine if the instrument is internally consistent.

  The Spearman-Brown formula is applied:

  $R = n\,r/[1+(n-1)r]$

  Where,

  *R* = estimated reliability of the entire instrument

  *r* = correlation coefficient between measurements

  *n* = ratio of the number of items in the changed instrument to the number in the instrument

  When *n* = 2, split-half reliability, $R' = [2r/(1+r)]$

### 3. Inter-Item Consistency

- **Definition**: Measures the consistency of responses across multiple items that are intended to assess the same concept.
- **Popular Tests**:
  - **Cronbach's Alpha**: Assesses the internal consistency of a set of scaled items. It is calculated by comparing the average correlation between all pairs of items within the test. Higher values indicate better reliability. Cronbach's alpha is widely used for multipoint scaled items.
  - **Kuder-Richardson Formulas**: Used for dichotomous items (e.g., yes/no questions). They assess the internal consistency of tests with binary responses.
- **Calculation**:

- **Cronbach's Alpha**: Derived from a series of two random subsets of items, where the Spearman-Brown coefficient is calculated for each pair. The average of these coefficients gives the Cronbach alpha. SPSS or similar statistical software can calculate this efficiently.

4. **Combining Stability and Equivalence**

- **Method**: Using alternative parallel tests at different times and correlating the results provides a combined index of reliability. This approach accounts for both stability over time and equivalence, although it might yield a lower reliability coefficient due to additional sources of variation.

**Practicality**

- **Definition**: Concerns the feasibility of using a measurement tool in real-world settings, considering factors like economy, convenience, and interpretability.
- **Criteria**: A practical measurement tool should be cost-effective, easy to administer, and straightforward to interpret.

# CHAPTER 11: SAMPLE DESIGN

A sample is used when it is not possible or practical to make all possible observations of a phenomenon that is being studied. A sample is used when it's impractical to observe an entire population.

Often, in sampling design, the researcher uses a sampling frame (target, population), that is, a list of all objects in the population (for example, industrial directory for organisational selection, telephone directory for consumer surveys, and so forth). Generally, the sampling frame is smaller than the population.

- Sampling is the process of selecting a subset of individuals or items from a larger population to represent and make inferences about the entire population.
- In research, sampling is a critical step as it helps researchers gather data efficiently, economically, and with an acceptable level of accuracy.

- The **population** refers to the entire group of individuals, items, or events that the researcher is interested in studying and generalizing the findings to.
- A **sample** is a smaller, manageable subset of the population that is selected for data collection.
- The **sampling frame** is a list or collection of all the members of the population from which the sample will be drawn

Errors in sampling include random errors and bias, with precision measured by the standard error. Careful sample selection reduces bias. Sampling can be unrestricted (random) or restricted (controlled), and researchers often use a sampling frame (list of population members) for practical selection.

There are two aspects in sampling design. The first is the selection of elements from the population and the second is the basis on which representativeness of the sample is obtained.

**11.1 SAMPLING PROCESS:**

i. Defining the **population of concern.**
ii. Specifying a **sampling frame**, a set of items or events possible to measure.
iii. Specifying a **sampling method** for selecting items or events from the frame.
iv. Determining the **sample size**.
v. Implementing the sampling plan.
vi. Sampling and data collection

**Step 1- Definition of population:** A population is defined in terms of

(1) elements (2) sampling units (3) extent (4) time

Example: In a survey of manufacturing organisations, the population was defined as:

(1) Element—manufacturing, planning and control activity

(2) Sampling unit—batch manufacturing organisation in engineering industries

(3) Extent—have been manufacturing engineering goods

(4) Time—the past two years

**Step 2- Specification of sampling frame:**

For probability sampling one has to have a sampling frame. Errors may occur when the researcher fails to access the elements through telephone or from industrial directories

In non-probability sampling, convenience or referrals may suffice to specify the sample; the researcher always utilises his/her own sense of judgment

Example: Government publication on industries, listing of batch manufacturing engineering firms in the Bombay Stock Exchange Directory

**Step 3- Specification of the Sampling unit:**

It is the basic unit containing the elements of the population.

Example: Heads of manufacturing divisions like Directors, Vice Presidents, or General Managers were the sampling units directly approached in the selected organisations.

**Step 4- Selection of sampling method:**

Sampling method is the way in which the sample units are selected.

Example: Use of check sampling method in an exploratory study. The method includes the selection criteria or just availability, if the number in the population is not large

**Step 5- Determination of the sample size:**

The number to be sampled can be decided on statistical analysis when the sample size is large. It can be modified by considerations of availability, cost, and accessibility.

Example: Size of the sample is based on availability

## 11.2 SAMPLING METHODS

Sampling methods can be broadly classified into non-probability sampling and probability sampling, as given below:

| Non-probability sampling | Probability sampling |
|---|---|
| (a) Quota sampling | (a) Simple random sampling |
| (b) Convenience sampling | (b) Stratified random sampling (proportionate) |
| (c) Judgement sampling | (c) Stratified random sampling (disproportionate) |
| (d) Purposive sampling | (d) Cluster sampling (systematic) |
| | (e) Cluster sampling (area) |

## 11.3 NON-PROBABILITY SAMPLING:

It is a judgment-based procedure. It can be representative, but precision and confidence cannot be obtained. Samples obtained by them are not representative of the population. However, they may yield good parameter estimates; objective estimates of precision and confidence cannot be obtained.

### I. QUOTA SAMPLING:

- This is the most commonly used non-probability sampling method.
- Rough proportions of sub-classes (strata) in the population are estimated from an outside source (say, census). The number to be sampled in each sub-class is in exact proportion of sample size as the sub-class population is in the total population.
- The individual units in each strata/sub-class are chosen by the researcher based on judgement. Therefore, selection bias will be present in this type of sampling.
- Example The overall work force n of a firm may be classified into two or three skill categories or grades $n_g$. If a number of individuals from any grade $s_g$ are selected by judgement, then the sampling is quota sampling where S is the total number sampled

$$\frac{n_g}{n} = \frac{s_g}{S}$$

### II. JUDGMENT SAMPLING:

- In this sampling procedure, a sample is obtained on the basis of sound judgment or experience on the part of the sampler who adopts a particular data collection strategy.

- The intention is that typical or representative subjects should be chosen. However, is good for small samples only. Errors increase with the sample size.

**III. SNOWBALL SAMPLING:**

Snowball Sampling is a judgment-based sampling method used to study specific or rare characteristics within a population. Also known as multiplicative or mixed probability sampling, it starts by selecting initial subjects with the desired traits, either randomly or judgmentally. These initial participants then refer others with similar characteristics, expanding the sample through referrals. This method is particularly useful for identifying rare characteristics at a lower cost. For example, in a study of Bangalore-based software companies serving multinationals, initial experts are consulted, and through their recommendations, other relevant companies are identified.

**IV. PURPOSIVE SAMPLING:**

This is similar to judgement sampling, except that the sample is chosen so that a particular research purpose or objective is served and is adequate for it. The sample is typical rather than representative.

Example: In a study of informal technology transfer from large public sector undertakings to small scale units, the sample of small scale units is chosen mostly by judgements based on scanning the products produced and by some referrals to large companies

**V. CONVENIENCE SAMPLING:**

These sampling procedures are ad hoc procedures. They are also called accidental sampling. Whatever is easily accessible, subjects who are cooperative, or subjects who can articulate are chosen. There is no way of knowing whether the sample size is small or large. These are not representative and are not recommended for research.

**11.4 PROBABILITY SAMPLING:**

In this method random sampling is used. To the selection of a sample from a population, when this selection is based on the **principle of randomization**, that is, random selection or chance.

**Principle of Randomization:** involves the random assignment of participants or subjects to different groups or conditions in a study. Randomization is used to minimize bias and ensure that the groups being compared are as similar as possible at the beginning of the study, except for the specific intervention or treatment being tested. This allows researchers to draw valid and reliable conclusions about cause-and-effect relationship

Probability sampling is a type of sampling method used in research, where each member of the population has a known and non-zero chance of being selected in the sample.

In probability sampling, every individual or item in the population has an equal or known probability of being included in the sample, ensuring that the sample is representative of the entire population.

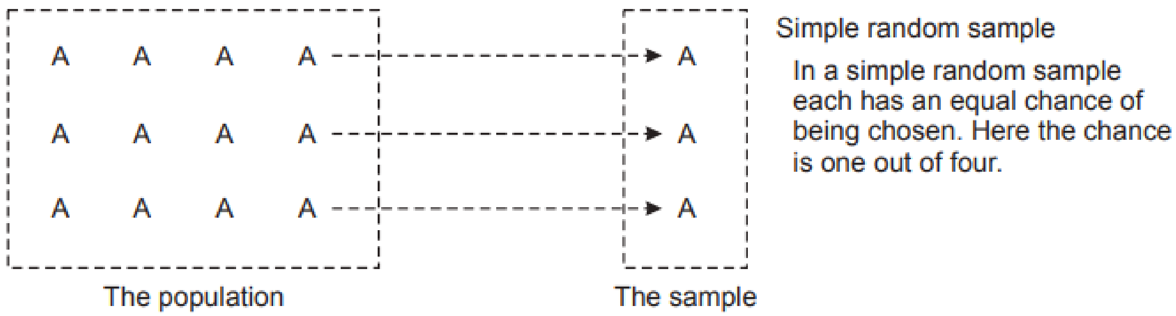
**I. SIMPLE RANDOM SAMPLING:**

In simple random sampling, every member of the population has an equal chance of being selected in the sample. This is typically achieved using random number generators or drawing lots to select the sample. Population is depicted in terms of parameters μ, $\sigma^2$, and p. The sample statistics are computed and used to estimate the population parameters with a stated confidence and precision.

The mean of all possible samples, the grand sample mean, is the mean of population

The variance of sampling mean, in the case of finite population, is related to the population mean as follows:

$\sigma_x^2 = [\sigma^2/\{n^*((N-n)/(N-1))\}]$ $\sigma_x$ = Sample standard deviation




## II. STRATIFIED RANDOM SAMPLING:

Stratified Random Sampling involves dividing the population into mutually exclusive subgroups based on certain characteristics (e.g., age, gender, region). Then, a random sample is drawn from each stratum proportionate to its size in the population. This method ensures representation of various subgroups in the sample.

**Step 1** Parent population is divided into mutually exclusive exhaustive subsets or groups/ classifications.

**Step 2** A simple random sample is independently drawn for each stratum. Any element belongs to one stratum only. Since the number of possible samples in each stratum is reduced, the possible number of samples that can be drawn in stratified sampling as against simple random sampling is considerably reduced. Two important statistics mean and variance are given below

1. Mean of a stratified sample = $\overline{X}_s = \sum_{g=1}^{L} N_g / N \times \overline{g}$

   Where,

   $N_g$ = number of elements in the population stratum 'g',

   $N$ = total size of population,

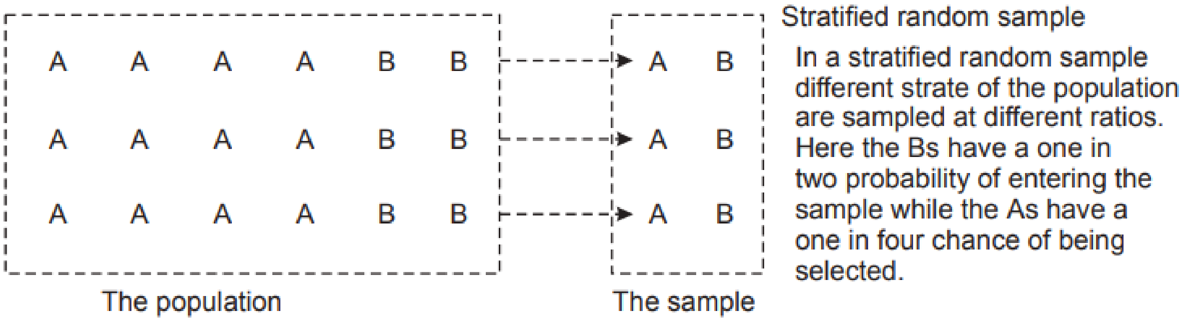
   $\bar{x}_g$ = sample mean of stratum $g$.

   $\bar{x}_s$ = sample mean of the total stratified sample, and

2. Variance of stratified sample is $S_{xs}{}^2$ is given by:

$$S_{xs}{}^2 = \sum_{g=1}^{n} (N_g / N)^2 * S_{xg}^2$$

   Where $S_{xg}{}^2$ is the variance of the stratum $g$.




### III. CLUSTER SAMPLING:

It involves dividing the population into clusters or groups (e.g., schools, households, villages). A random sample of clusters is selected, and then all members within the selected clusters are included in the sample.

The steps involved in cluster sampling are:

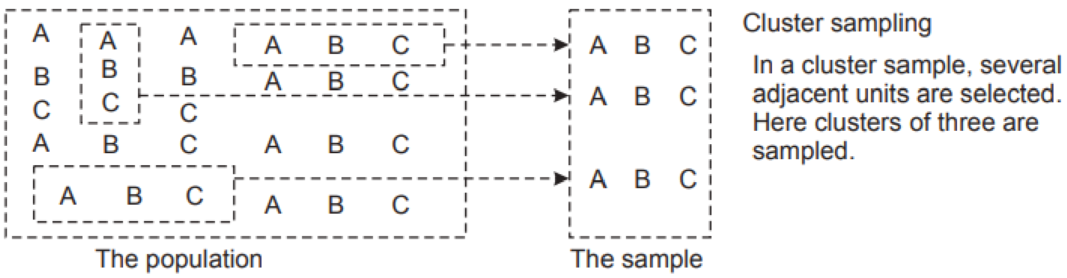
1. Parent population is divided into mutually exclusive/exhaustive sub-sets.

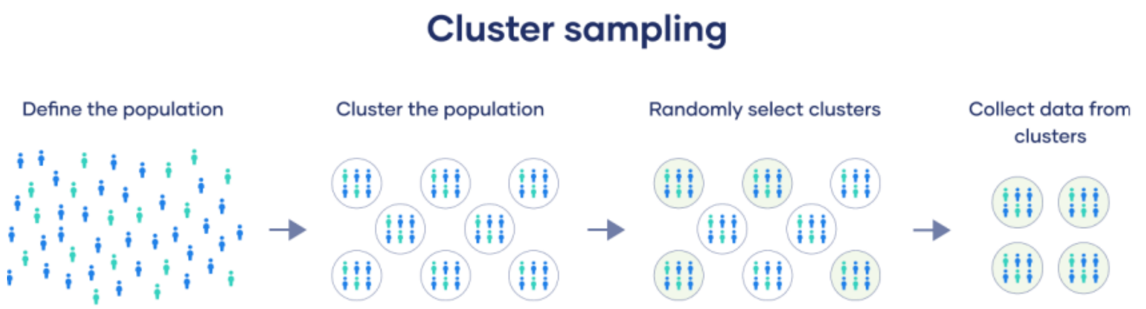
2. Random sampling of the sub-sets is made

   (i) One Stage Cluster Sampling Aggregation of individual samples is done before selecting the random sample.

   (ii) Two-Stage cluster sampling A random selection of the subsets is made and subjects from each sub-set is selected randomly

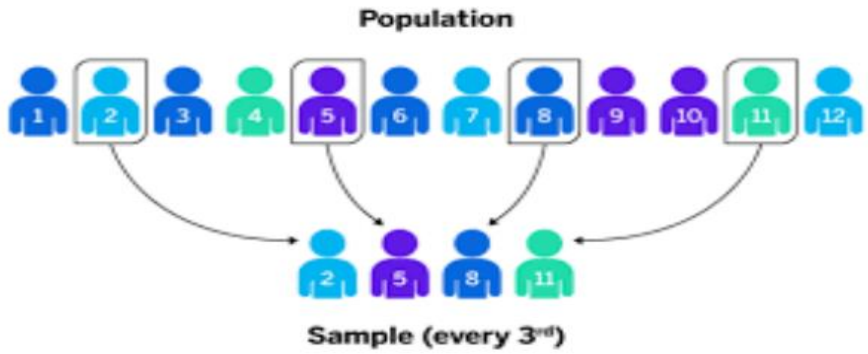
Cluster sampling is similar to stratified sampling, but the criteria used for stratified sampling are different. The most important distinction is that in stratified sampling a random sample is selected from each group/stratum. In cluster sampling, the sub-groups are sampled. The subgroups should be very nearly small-scale populations, that is, as heterogeneous as possible.






## IV. SYSTEMATIC RANDOM SAMPLING:

In systematic random sampling, a random beginning (first term choice) is made and every Lth item starting from the first item is selected. It is equivalent to single cluster sampling. It is an easy sampling procedure and often can be more representative, since it cuts across the population




## V. AREA SAMPLING / MULTI-STAGE SAMPLING

Multistage sampling combines two or more sampling methods. For example, a researcher might use stratified sampling to select clusters and then use simple random sampling within each cluster.

This pertains to the primary sampling of geographic areas, like districts, towns, regions, or factories. If only one level of sampling takes place then the elements or sample from it is a single stage area sample. If one or more successive samples are taken within a larger area, then it becomes a multi-stage area sample.

Example:

(i) Multi-stage sample

Stage 1—Random sample of location (nationwise)

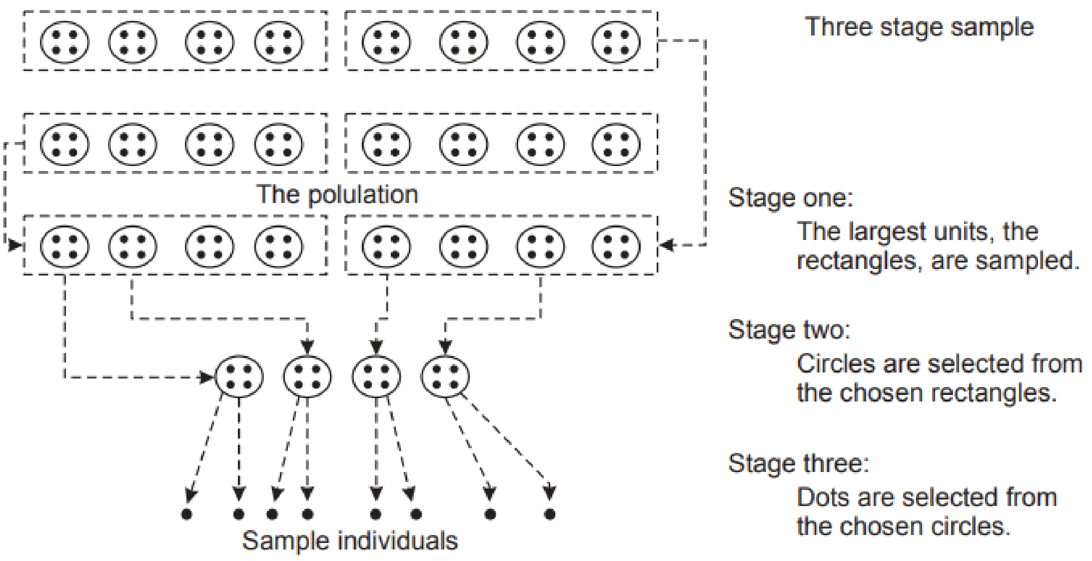
Stage 2—Random sample of blocks …(locationwise)

Stage 3—Random sample of households (blocks/segmentwise)

(ii) One-stage sample

(a) Random sample of blocks

(b) Random sample of households in each block




# Chapter-12

## DATA COLLECTION PROCEDURES

## INTRODUCTION

- Management research data can be classified into primary and secondary data. Primary data is collected by the researcher and agents through experiments, surveys, interviews, or observations.
- Secondary data is collected by other researchers for their own use and is less expensive and time-consuming. However, problems may arise when using secondary data, such as different units of measurement, aggregating data, classifications not matching the problem, and the availability period not fitting the overall research needs.
- Matching primary data with secondary sources can be difficult, and data accuracies may not be known.

## SOURCES OF SECONDARY DATA

The sources of secondary data are sources internal to a firm or industry and sources that are

external.

(a) The internal sources are:

• company accounting records

• company reports

• in-house journals

• miscellaneous reports

• internal computer databases

(b) The external sources are:

• computerised data bases

• reports of associations

• reports and publications of government agencies

• other publications

**Internal Sources**

Company accounting data : Company accounting data includes financial information stored in various sources such as computers, ledgers, inventory, purchase logs, shop order files, worker files, payment ledgers, and cashbooks.

Company reports : Company reports :These consist of annual reports and regular reports submitted to the board of directors or statutory reports submitted to the government. A great deal of routine data can be obtained through these reports on operational and performance aspects.

In-house journal : Most corporations usually support in-house journals and useful general data may be obtained from them.

Miscellaneous reports : Miscellaneous reports include consultancy, research-supported reports, special reports to top management, troubleshooting, and large operational improvements.

Company computer databases : The use of internal data for decision-making has increased so much in recent times that intranets (internal networks linking outward to the internet but forbidding access from outside) have begun to operate successfully.

Enterprise data warehouses and data marts : Enterprise data warehouses and data marts have gained a competitive edge in the last 20 years by effectively using available information rather than optimizing approaches. These systems involve process managers and analysts making informal decisions from tables.

**External Sources**

Public computer databases : Public computer databases are rapidly growing, offering access to financial information, product sales, marketing channels, performances, manufacturing values, and employment. They also provide bibliographic services and software updates, with internet-based data treatment available.

Reports of associations : Industry associations produce annual reports detailing sales growth, operating characteristics, and special reports. Subscription-based newspapers and magazines collect this information.

Reports of government agencies : Government departments in India provide extensive data on financial, operational, and R&D activities through various publications, including the Reserve Bank of India bulletin, industrial classifications, census, industrial surveys, and economic surveys, covering demographics, housing, wages, and employment.

Industrial syndicates :These organisations also provide data on industrial services, plantwise / areawise information on manufactured products, inventories, sales and movement through marketing channels, financial ratios, and the like.

Other publications: Secondary data sources include academic publications, professional journals, project reports, dissertation abstracts, computer search systems, and computerised bibliographic services, with a sample list provided in Appendix B.

**Computer Search for Secondary Data**

Research data is abundant in worldwide computer databases, electronic libraries, and the internet. However, researchers need different skills to integrate these sources into their research efforts due to constant change and the need for electronic searching sources. To find necessary information, researchers should follow well-planned procedures and utilize electronic databases, e-libraries, and the internet.

The following steps may be useful (Mckie, 1997).

1. Specify your data needs.

2. Select a keyword/search query.

3. Select a suitable electronic library search source/database/search engine for the internet search.

4. Save useful information resulting from the search.

5. If necessary, repeat steps (2) and (3) with modified keywords

Step 1—Specify information/data needs The problem of research should be defined as clearly as possible in order to be specific about the data/information needed. Since electronic searching requires 'keyword' use, the more specific/clear the research objective, the more suitable a keyword would be for close search.

Step 2—Select keywords A keyword is used for querying the database or web. Single keywords give very general information, which is not very useful. To be more efficient in search, use an appropriate 'key phrase'.

- Use Boolean search ("AND", "OR", "NOT") For example, instead of using keyword 'scheduling semiconductor', you may use "scheduling AND semiconductor" for searching for information on scheduling particularly in semiconductor manufacturing.

Step 3—Select a library database/website (search engine)/business database In selecting a particular website for particular data, the user has to look for scope of the data/ information and the authority of the source of the data/information.

- Library searches can be conducted electronically through internet connections, using tools like socio file, current contents, and silver platter. Secondary sources and questionnaires are available for students and faculty members, and CD-ROM-based products are also available.
- Web searches utilize various search engines with varying display, keyword options, and updating speed. These engines connect numerous websites, with a special site like http://www.windweaver.com/index.htm and Carleton.ca/~cmckie/research.htm providing useful search engine types.

- The specific types of information searches are:

Known item searches 'Where' searches (databases or websites),'what' searches, for example, the Google item.

Where searches: Lycos system gives geographical locations.

What searches: Alta Vista, Yahoo are useful.

Who searches: These provide information on people (all sites provide this information).

Some of the search engines are listed at the end of this section

Step 4—Store information The useful information obtained in each cycle of search should be saved and stored suitably so that at the end of the search phase, the several bits can be organised systematically into a meaningful review of literature or a researchable data set. A number of cycles of this search procedure (steps 2 to steps 4) may be necessary to obtain the required information for research.

Web addresses of a few search engine.

www.google.com

www.lycos.com

www.yahoo.com

www.hotmail.com

**PRIMARY DATA COLLECTION METHODS**

Numerous methods and procedures have been developed to assist in data acquisition, employing various methods for describing and quantifying data, each suitable for specific applications.

The trend in social science and management research is to use multiple methods to minimize bias and ensure data reliability. Primary data is collected through standard research methods like experiments, surveys, field studies, case studies, and projective techniques.

The employment of any of these tools of data collection depends on the type of research that is being undertaken and the time and resources available to the researcher. Table 12.1 shows tools for data collection predominantly used in different types research

**Table 12.1** **Tools for Data Collection Verses Types of Research**

| | *Tools of Data Collection* | | | | |
|---|---|---|---|---|---|
| *Type of Research* | *Observation* | *Questionnaire* | *Interview* | *Projective Techniques* | *Records* |
| Experimental | X | X | X | X | – |
| Exploratory | X | X | X | – | X |
| Descriptive | X | X | X | – | X |
| Causal | – | X | X | – | – |
| Case Study | X | X | X | | X |
| Model building | X | – | X | – | X |

**Observation:**

Observation is the most direct form of data collection. Usually data is gathered by observation when it can be gathered accurately only by this method. For example, behaviour of children (who cannot talk). There is a considerable difference between casual observation and scientific observation. Scientific observation is well planned, recorded, and checked for validity and reliability.

Conditions under which data can be effectively collected are:

1. Data must be accessible to the observer (private activity, motivation, and attitudes). Some behaviours are communicated through facial expression and body language. These are observable when the activity is repetitive, frequent, and predictable.

2. They must be of short interval to reduce distortion due to recall.

3. It is desirable that the observational data is used to supplement the other methods of data collection.

Observational approaches can be classified as follows:

<u>Natural verses contrived</u> Observing how many cars pass a circle every hour is a natural observation, which is useful when phenomena occur frequently. A store worker checked by an observer disguised as a customer is a contrived observation

<u>Open verses disguised</u> An observation made on television or metering certain occurrences are examples of open observations. Observations in lab experiments are often disguised, that is, the participants are unaware of certain kinds of observations made by the experimenter.

<u>Structured verses unstructured</u> In structured observation what aspects are to be observed and recorded are exactly known. The others are ignored; an observation checklist is usually used (cf. questionnaire). When any aspect of a phenomenon, as in exploratory situations, is considered useful, the observation becomes unstructured

Direct observation observes current behavior, while indirect observation observes past behavior through physical traces. Direct observation is time-consuming and costly, not suitable for long-duration or rarely occurring phenomena, and data collected is not quantitatively quantified.

Mechanical verses human When devices like televisions, meters, video cameras, and photographic analysis are used for observation, it becomes mechanical. All other observations are human.

Non-behavioural and Behavioural Non-behavioural observations analyze historical or current records, including written, sound, and computer records. They can be used for content analysis, physical condition analysis, process and activity analysis, work, study, and manufacturing system analysis. Behaviour observation studies non-verbal, linguistic, extra-lingual, and spatial behavior.

**Questionnaires**

Questionnaires are preferred for surveys due to their cost-effectiveness, less administrative skills, and ability to handle large samples. They offer uniformity and anonymity, making them more objective. However, skilled interviewers may handle emotional responses better. Questionnaires also have low response rates, sometimes as low as 5-10%.

A questionnaire is a formalized set of questions designed to gather information about facts, knowledge, attitudes, needs, and motivations. It's considered an art rather than a science, with a simple approach and several steps to avoid ambiguous questions.

Steps in the design of a questionnaire:

Step 1 of questionnaire design involves specifying the information sought, defining the target population, outlining objectives, and selecting respondents to reduce surrogate information and respondent errors.

Step 2: Decide on questionnaire type, including one or more questions for each item of information. Analyze responses for useful questions. Include additional questions if necessary, and use an agree-disagree format for respondents' choice.

Example - Innovative tasks are mostly restricted to the R&D and Production Departments.

Strongly agree     Neutral     Strongly disagree

5     4     3     2     1

Two or more questions can be asked as one question. These are called ‘double-barreled’ questions.

Example -The extent to which you are satisfied regarding the remuneration package and recruitment policies of the organisation, to enable it to recruit managers, are:

Highly Satisfactory Moderately Satisfactory Highly Unsatisfactory

5 4 3 2 1

Questions

Demanding aggregations of several types of information should be avoided as far as possible. The main issue involved in the design of questions is to specifically consider the respondent and his difficulties in answering. Some of the difficulties may arise because of the following reasons.

• Uninformed respondent Many simple questions are preferred to a single complex or composite question.

• Forgetful respondents When memory-based answers are required, aided recalls may be used to make the answers less error prone (errors are omission, creation, and telescoping).

• Inarticulate respondents When ‘why’ questions are used, the respondent may be unable to answer, though willing. In such situations, it is desirable to use projective techniques.

• Personal and embarrassing questions Preferably, either a random response method or counter-biasing statements should be used. In the random response method, the question is mixed with other questions and a single answer is elucidated.

Example- A fairly unbiased percentage of rejection at the inspection stage can be obtained by asking the following questions.

(i) Acceptance rate at inspection for your product is per cent.

(ii) The acceptance rates at inspection for your company/s in general, in your sector is

per cent (100 - answer given in question (a)) = per cent rejection). Question(b) is inconsequential.

(iii) Some respondents may be unwilling to answer and may deliberately distort the information. This can be verified by alternate questioning methods

Example -Clearly laid down policies are necessary for promoting innovation in industry.

Questions

(i) Policies and procedures for R&D in your organisation are:

Very Clear Fairly Clear Moderate Ambiguous Very Obscure

5 4 3 2 1

(ii) If you have answered the above question (i) as Ambiguous or Very Obscure to what extent, in your opinion, are absence of clearly defined policies the cause for the poor rate of innovation in your organisation.

Very Large Extent Moderate Extent Very Small Extent

5 4 3 2 1

Step 3: Administration of questionnaire: Use interviews with questionnaire for past and present behavior, attitudes, and opinions, specifically for demographic properties like age, sex, income, and occupation.

Step 4: of the response format involves the freedom of the researcher and the respondent, typically using open-ended and multiple-choice questions. Open-ended questions allow the respondent to provide any answer, resulting in a wide range of responses. Multiple-choice questions reduce interviewer bias and allow for Gradations of answers. However, balanced designs are preferred. A split ballot approach can be used to determine the best method, offering multiple versions of the question to a small group of respondents.

Step 5:Wording and phrasing of questions In this regard, the basic principle is to make the questions simple and straightforward. Increasing the length of the questions for the same answer is desirable.

Example Capital productivity (where highly machine-oriented operations are involved), material productivity (when cost of material is a major portion of product cost) and labour productivity (when operations are mostly manual) are a few productivity indices used in manufacturing organisations.

Considering the most relevant one for your organisation, how would you rate your organisation's productivity?

| | Very High | | Moderate | | Very Low |
|---|---|---|---|---|---|
| Capital productivity | 5 | 4 | 3 | 2 | 1 |
| Material productivity | 5 | 4 | 3 | 2 | 1 |
| Labour productivity | 5 | 4 | 3 | 2 | 1 |

Step 6 : involves creating a sequence of questions, starting with simple, then moving to complex ones, ensuring they are free from bias, neutral, sensitive, and controversial, as this maximizes response efficiency.

Step 7:Layout of the questions The format and layout of the questionnaire should be physically designed so as to eliminate recording errors. Branching questions (like in computer programmes) should be avoided, particularly on mailed questionnaires. Sometimes, it is desirable to include codes with the questions.

Step 8:Iteration Like in any design function, it may be necessary to go through steps one to seven in one or two iterations to eliminate undesirable features of questions

Step 9:Pre-testing and revision of questions The pre-test is generally done on a small

sample of respondents who are similar to the respondents of the main study (undergraduates in campus in place of senior executives is to be avoided!). The main questions involved in pre-test are the following:

• Which are the items to be pre-tested? The guiding factor is to choose questions, in which the interviewer’s errors tends to be large.

• How is the pre-test conducted? It is essential to interview the respondents after answering the questionnaire and ascertain why the questions were answered the way they were. Alternatively, protocol analysing (loud thinking of the respondent during answering), and debriefing may also be used.

• Is the number of respondents sufficient? Adequate sample size should be determined for the pre-test (cf. Sample size determination)

**Interviews**

Interviews are versatile and representative, allowing for more representative sample sizes and a more customer-oriented approach than questionnaires. They allow for explanations of unclear questions, which is not possible with questionnaires. A skilled interviewer can overcome respondents' unwillingness to answer emotionally complex questions.

<u>Depth interviews</u> are commonly used to gather information on people's behavior, attitudes, needs, and characteristics, often used when direct questions may not be effective due to the subject's unwillingness or embarrassment.

<u>Individual depth interviews,</u> lasting between 1/2 and 45 minutes, allow the interviewer to create questions without specific questions, while the respondent has freedom in content and manner, without influencing them. These interviews generally yield high-quality ideas.

Individual depth interview is used:

• When detailed probing is necessary to elicit answers on behaviour, attitude, character, and needs.

• When the subject matter is confidential.

• When emotions or embarrassment may be evoked while answering the questions.

• When obtaining answers to questions which would be constrained in groups because of the subject's conformity to socially acceptable norms.

• When information on complex behaviour patterns is desired while interviewing professionals.

• When it may be desirable to allow the respondent to sketch anecdotes or tell stories

A focused group interview is a structured interview where 8 to 12 individuals are interviewed to discuss a specific behavior or characteristic. The moderator leads the group, establishes rapport, and sets rules for interaction. The interviewer's main function is to focus on a given experience and its effects, with topics and aspects known in advance. The interviewer controls the direction of the interview by confirming or guiding the respondents. The interviewer analyzes the situation and develops a guide to the interview, setting out major areas of inquiry and data to be obtained. Focused groups offer advantages such as sharpening opinions, more accurate information, and a sense of belonging. However, they also have disadvantages such as difficulty in obtaining a random sample, potential participant games, strong individual influences, moderator biases, and risky generalization from group to population.

**Projective Techniques**

The underlying theory behind projective techniques is that the description of vague and fuzzy objects requires interpretations based on one's experience, attitudes, and values. The more vague the object is, the greater the revelation of the respondent. There are four types of projective techniques:

(i) Association tests,
(ii) Completion tests,
(iii) Construction techniques, and
(iv) Expression techniques. All have originated from clinical psychology.

**Association tests** In these tests either the first thought or word that comes to mind when the researcher presents a word or phrase (free word association) or successive words or thoughts that come to mind (successive word organisation), is given by the respondent. The responses are analysed for measuring attitudes to certain category of stimuli.

**Completion tests** In sentence completion tests, the respondent completes an incomplete sentence with the phrase or word, that comes first to his mind (cf. open-ended question). In strong completion techniques, which are an extended version of the S.C.T., an incomplete sentence is completed based on the respondents' experiences and attitudes.

**Construction techniques** The respondent is asked to construct (that is, develop) a stray diagram or a description by just presenting a visual depiction of a context

**Expression techniques** This is mainly role-playing (by the respondents) of the behaviour of another person. All these techniques require a highly trained interviewer. They tend to be expensive and since the samples are small and are non-probable in nature, the respondents are generally non-representative.

**NON-SAMPLING ERRORS**

Sampling errors are statistical inferences from sample to population, while non-sampling errors are common in both research sample surveys and complete enumerative surveys. Non-sampling errors arise from data collection, processing, and responding elements, and can be estimated.

The sources of errors are (Murthy, 1967):

- Inadequate data specification.
- Inadequate or faulty methods of interview, observation, or measurement.
- Lack of trained investigators.
- Inadequate scrutiny of basic data.
- Errors in coding data entry and verification.
- Errors in presentation and printing.
- Non-coverage errors.
- Response errors.

**Non-Observation Errors**

Non-coverage errors occur in sampling surveys when certain population elements are not studied due to inadequate sampling frames or wrong procedures. Over coverage errors may result in duplicated items. Non-response errors occur when subjects do not respond to the survey, leading to refusal, delay in response testing, or researcher persistence beyond acceptable time limits.

**Observation errors,** caused by incorrect or careless responses to questions, can render questionnaires questionable, making interviews more effective than questionnaires.

**VALIDITY AND RELIABILITY OF DATA COLLECTION PROCEDURES**

Validity and reliability in data collection are crucial for accurate measurements and their relevance to the intended purpose. Factorial validity is established through confirmatory factor analysis, which confirms the existence of theorised dimensions. Reliability is tested for consistency and stability, with Cronbach's alpha being a reliable coefficient that indicates how well items in a set are positively correlated. Methods for consistency, reliability, and stability can be applied to data collection procedures. The split half reliability coefficient is used in specific situations to test for consistency when multiple scales, dimensions, or factors are assessed. Stability can be assessed through parallel form reliability and test-retest reliability, where high correlations between similar forms are obtained.

**Validity and Reliability of Interviews :**Establishing rapport is crucial for effective interviews, with women and younger individuals often successful. Experience enhances interviewing skills, and carefully designed structures are more valid. Experts' critical judgement aids in question determination, and altering questions or repeating the interview ensures reliability.

**Validity and Reliability of Observation :**Observation validity and reliability require identifying significant incidents of behavior. Researchers should consult experts and establish theories based on incident qualities. Independent observers trained in observation and recording are better. Initial try-outs and joint observation build reliability. Randomly selected time samples enhance validity and reliability by providing unbiased and representative samples.

**Validity and Reliability of Questionnaires:**

A possible procedure for validation and ensuring reliability of questionnaires is suggested below:

1. Ask the right questions.

2. Phrase them in the least ambiguous manner.

3. Clearly define the important terms.

4. Rate the content validity.

5. Obtain predictive validation by correlating questionnaire scores with observed behaviour or peer/superior ratings.

6. Administer the questionnaire to the same participants twice and compare the two responses.

7. Get the first four listed above checked by an expert.