DESCRIPTIVE AND EXPERIMENTAL RESEARCH

DESCRIPTIVE RESEARCH

The descriptive research attempts to describe, explain and interpret conditions of the present i.e. —what is. The purpose of a descriptive research is to examine a phenomenon that is occurring at a specific place(s) and time. A descriptive research is concerned with 82 conditions, practices, structures, differences or relationships that exist, opinions held, processes that are going on or trends that are evident.

Types of Descriptive Research Methods

In the present unit, the following descriptive research methods are described in detail-

1. Survey Research
2. Co relational Research
3. Case Study
4. Ethnography
5. Document Analysis

SURVEY RESEARCH

Survey research involves collecting data to test hypotheses or to answer questions about people's opinions on some topic or issue. A survey is an instrument to collect data that describe one or more characteristics of a specific population. For example, researchers may ask teachers with one to three years of experience a series of questions to try to gather information about the aspects of their profession that new teachers find most challenging.

Survey research can be used to gather information about a group's beliefs, attitudes, behaviors, and demographic composition. Survey data are collected by asking members of a population a set of questions, which can be administered in a questionnaire that is mailed or emailed or in an interview over the phone or in person.

Surveys are either sample surveys or census surveys, usually the former. In a sample survey, as the name suggests, a researcher attempts to infer information about a population based on a representative sample drawn from that population. To be able to generalize sample survey data to an entire population, the sample responding to the survey should accurately represent all the subgroups within the population. In a census survey, researchers attempt to acquire information from every member of a population. Census surveys are usually conducted when a population is relatively small and readily accessible.

Survey researchers often seek information that is not already available, they usually need to develop an appropriate instrument (i.e., set of questions). If a valid and reliable instrument is available, researchers can certainly use it, but using an instrument just because it is readily available is not a good idea. If you want the appropriate answers, you have to ask the appropriate questions. Furthermore, survey researchers must be very careful to write or select questions that are clear and unambiguous. The researcher seldom has an opportunity to explain to participants who are filling out a questionnaire what a particular question or word really means. If researchers develop an instrument, they need to try it out and revise it as needed before collecting the research data.

Survey research design

Survey studies generally come in one of two designs—cross-sectional studies and longitudinal studies. The key difference between these two types is the number of times the survey is administered. In cross-sectional studies, a survey is administered to a population once. In longitudinal studies, surveys are administered to a population more than once with significant periods of time elapsing between each administration of the surveys.

Cross-Sectional Surveys

A cross-sectional survey is one in which data are collected from selected individuals at a single point in time. It is a single, stand-alone study. Cross-sectional designs are effective for providing a snapshot of the current behaviors, attitudes, and beliefs in a population. This design also has the advantage of providing data relatively quickly—you do not have to wait for years (as is often the case in longitudinal studies) before you have your data and can begin to analyze and draw conclusions. Cross-sectional studies are not effective if the researcher's goal is to understand trends or development over time. Furthermore, a single point in time often does not provide a broad enough perspective to inform decisions about changes in processes and systems reliably (e.g., to change the math curriculum in a school).

Longitudinal Surveys

In a longitudinal survey study, data are collected at two or more times. These surveys are extremely useful for studying the dynamics of a topic or issue over time. Longitudinal studies require an extended commitment by the researcher and the participants— some difficulties in conducting longitudinal studies include keeping track of sample members over time and maintaining sample members' willingness to participate in the study. Attrition (i.e., participants dropping out) is common.

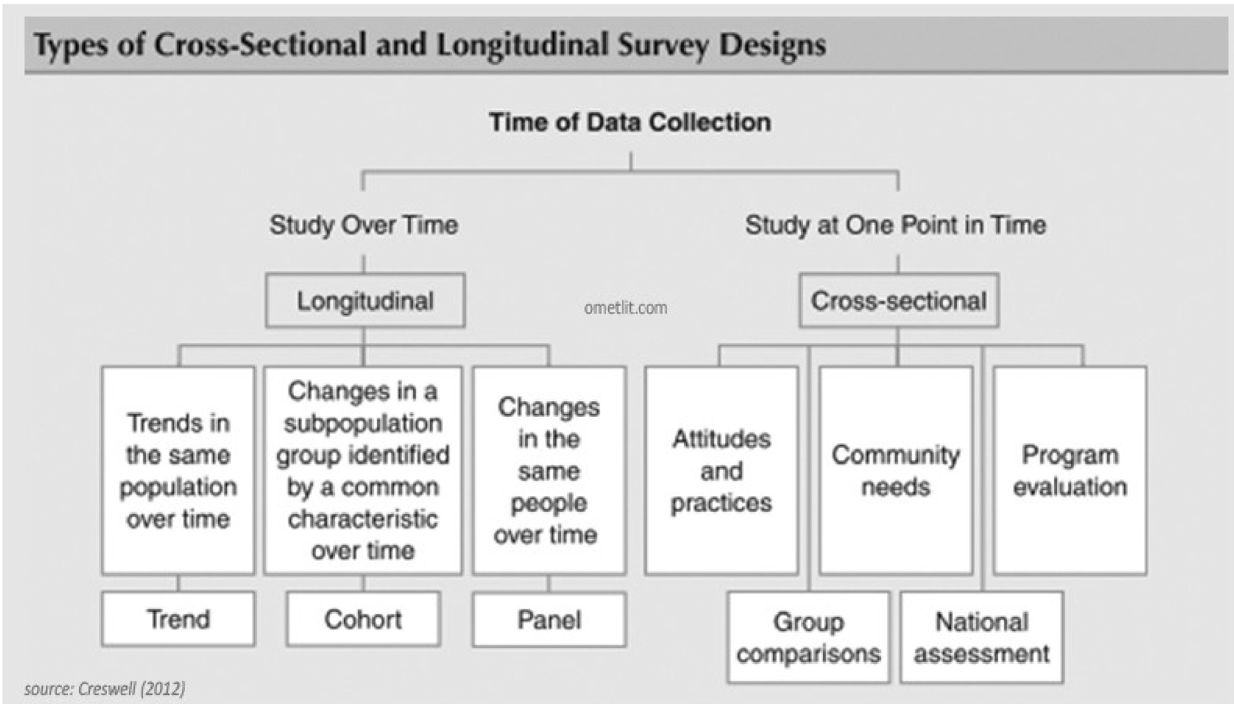
Longitudinal survey studies can be categorized into four basic types. All collect data multiple times; however, they differ in how the researcher samples the population and administers the survey.

A **trend survey** examines changes over time in a particular population defined by some particular trait or traits. Using a trend survey, the researcher is able to analyze changes in the attitudes, beliefs, or behaviors within that particular population over time. For example, assume a researcher wants to study trends in MA female students' attitudes toward gender equality. To provide information about the trend of the students' attitudes, the researcher would select a sample of the MA female students in the current year and then select another sample each successive year until the study is complete. In other words, the survey would be administered annually, and each annual sample would include MA female students pass out that year.

A **cohort survey** involves one population selected at a particular time period (e.g., MA female students of 2021—the first class to complete masters after having spent two years of University) but multiple samples taken and surveyed at different points in time. For example, the researcher could identify 1000 MA female students in 2021 and send surveys to 300 randomly selected participants. Then, in 2022, the researcher would return to the same population of 1000 students and again randomly select 300 participants to survey. Each sample could be composed of different students (although random sampling may result in some overlap), but all samples would be selected only from the population of MA female students from 2021.

A **panel survey** involves a sample in which the same individuals are studied over time. For example, in a 3-year panel study of MA female students of the class of 2000 who post graduated from Maharaja Sriram Chandra Bhanja Deo University, Baripada and Fakir Mohan University, Balasore the exact same individuals would be surveyed in each of the 2 years of the study. A frequent problem with panel studies (and cohort studies to a lesser degree) is loss of individuals from the study because of relocation, name change, lack of interest, or death. This attrition is especially problematic the longer a longitudinal study continues.

A **follow-up survey** addresses development or change in a previously studied population; some time after the original survey was given. For example, a researcher who wanted to study MA female students in Baripada a number of years after the original study was concluded would identify individuals who had participated in the original study and survey them again to examine changes in the attitudes, behaviors, or beliefs.



CONDUCTING SURVEY RESEARCH

Survey research requires the collection of standardized, quantifiable information from all members of a population or of a sample. To obtain comparable data from all participants, the researcher must ask them each the same questions. Surveys generally take one of two forms, questionnaires or interviews. A questionnaire is a written collection of survey questions to be answered by a selected group of research participants; an interview is an oral, inperson question-and-answer session between aresearcher and an individual respondent.

Questionnaires are usually mailed or emailed to potential participants. A questionnaire administered in this way is relatively inexpensive and usually permits collection of data from a much larger sample than an interview or a personally administered questionnaire. The disadvantages are that paper-and-pencil questionnaires mailed to participants do not allow any opportunity to establish rapport with respondents and the researcher cannot explain any unclear items.

Characteristics of Survey Research

Whether a survey design is longitudinal or cross-sectional, there are key characteristics of both that will help you design a survey or read and evaluate a published survey study. Survey researchers engage in the processes of:

a. Sampling from a population
b. Collecting data through questionnaires or interviews
c. Designing instruments for data collection
d. Obtaining a high response rate

Sampling from a population

Survey researchers typically select and study a sample from a population and generalize results from the sample to the population. We need to first define three terms: the population, the target population or sampling frame, and the sample. At the broadest level is the population, in which a group of individuals possesses one characteristic that distinguishes them from other groups. Researchers do not always study an entire population, either because they cannot identify the individuals or because they cannot obtain lists of names. (Lists are used when mailing out a questionnaire.) In practical, operational terms, researchers study a target population (sometimes called the sampling frame). This is the list or record of individuals in a population that a researcher can actually obtain. From the target population, researchers choose a sample. At the most specific level, researchers select a sample from the target population. These individuals are the people studied.

The most rigorous form of sampling is to use random sampling by employing a procedure such as using a random numbers table. In this process, the researcher selects a sample representative of the population so that claims or inferences can be drawn from the sample to the population.

In survey research, it is important to select as large a sample as possible so that the sample will exhibit similar characteristics to the target population. Also, in survey studies, it is sometimes difficult to obtain a good list of the target population.

It is also possible in survey research to study the entire population because it is small (e.g., members of literacy councils in a state) and can be easily identified. This type of survey study, sometimes called a census study, permits conclusions to be drawn about the entire population. Therefore, random sampling, hypothesis testing, and the use of inferential statistics are not necessary. For this type of study, survey researchers simply report descriptive statistics about the entire population.

Questionnaires and Interviews

Although many different forms of surveys exist, survey researchers typically collect data using two basic forms: questionnaires and interviews. Researchers need to consider the forms and weigh the advantages and disadvantages of each. A **questionnaire** is a form used in a survey design that participants in a study complete and return to the researcher. The participant chooses answers to questions and supplies basic personal or demographic information. An **interview survey**, however, is a form on which the researcher records answers supplied by the participant in the study. The researcher asks a question from an interview guide listens for answers or observes behavior, and records responses on the survey. In quantitative survey interviews, the investigator uses a structured or semi-structured interview consisting of mostly closed-ended questions, provides response options to interviewees, and records their responses. In qualitative survey interviews, an interviewer asks open-ended questions without response options and listens to and records the comments of the interviewee. Several different types of questionnaires and interviews are used in quantitative survey research. Here we will highlight the major types used in education:

- Mailed questionnaires
- Web-based questionnaires
- One-on-one interviews
- Focus group interviews
- Telephone interviews

Mailed Questionnaires

A **mailed questionnaire** is a form of data collection in survey research in which the investigator mails a questionnaire to members of the sample. Researchers might develop their own questionnaire, modify an existing one, or use one that they have located in the literature. The process consists of locating or developing a questionnaire, sending it out to the sample of the population, using repeated contacts with the sample to obtain a high response rate, checking for potential bias in responses, and analyzing the data.

A mailed questionnaire is a convenient way to reach a geographically dispersed sample of a population. The mail facilitates quick data collection, often in as little time as 6 weeks from the first mailing to the conclusion of data collection. A mailed questionnaire is economical because it involves only duplication and mailing expenses. The disadvantage of mailed questionnaires is that individuals may lack any personal investment in the study and decide not to return the instrument. Also, because the researcher does not have a means for explaining questions, participants may misinterpret items on the survey.

Web-Based Surveys or Questionnaires

With increased use of Web sites and the Internet, Web-based questionnaires are becoming popular. A **Web-based questionnaire** is a survey instrument for collecting data that is available on the computer. Several software programs are available for designing, gathering and analyzing survey data with sample questions and forms (e.g., see Qualtrix at **http://www.qualtrics.com/survey-software/** or Survey Monkey at **http://www .surveymonkey.com/**).

Educational researchers need to weigh the advantages and disadvantages of using a Web-based survey. On the positive side, such surveys can gather extensive data quickly, employ tested forms and sample questions rather than having to design them, and take advantage of the extensive use of the Web by individuals today, including its use as a site for social networking. Surveys are not based on random sampling so that drawing inferences to a general population is difficult. Web-based surveys may be biased toward certain demographic groups that tend to use computers. On the other hand, Web surveys may allow effective and economical surveying of the entire population and thereby skirt around the inference problem. Further, they saw a mixed system of Web-based and mailed surveys as promoting a high response rate.

One-on-One Interviews

One-on-one interviews are a form of survey data collection. In **one-on-one interviewing in survey research**, investigators conduct an interview with an individual in the sample and record responses to closed-ended questions. The process involves developing or locating an instrument and training the interviewer(s) in good interview procedures. This training consists of learning how to provide instructions during the interview, maintaining confidentiality about the interview, asking the exact question on the interview guide, completing the interview within the time allocated, being courteous, and not interjecting personal opinions into the interview. When multiple interviewers are used, researchers train all individuals to use the same procedure so that the mode of administration does not introduce bias into the study.

One-on-one interviews are useful for asking sensitive questions and enabling interviewees to ask questions or provide comments that go beyond the initial questions. Interviews lead to a high response rate because researchers schedule the interviews in advance and sample participants typically feel obligated to complete the interview. However, one-on-one interviews do not protect the anonymity of the participant as questionnaires do. Researchers may also prejudice participant answers, knowingly or unknowingly, through either comments or body language. Also, not all interviewees are comfortable disclosing information about themselves during the interview.

Focus Group Interviews

An alternative to a one-on-one interview is to administer a survey to a focus group. In quantitative **focus group interviews in survey research**, the researcher locates or develops a survey instrument, convenes a small group of people (typically a group of 4 to 6) who can answer the questions, and records their comments on the instrument. For example, this group might consist of parents who evaluate a new math or science curriculum in a school. Alternatively, international students provide views about cultural integration into an American university setting. During processes such as these, researchers ask the group questions on an instrument and record or take notes on the group conversation.

Focus groups provide for interaction among interviewees, collection of extensive data, and participation by all individuals in a group (Krueger, 1994). A disadvantage of focus group interviews is that they require the researcher to find consensus on questions so one score can be marked for all individuals in the group. In addition, some individuals may dominate the conversation, leading to responses that do not reflect the consensus of the group.

Telephone Interviews

In **telephone interview surveys**, the researcher records the participants' comments to questions on instruments over the telephone. The researcher develops or locates an instrument, obtains the telephone numbers of participants in the sample, conducts the telephone calls, and asks the participants to answer questions on the instrument. Telephone interviews allow the researcher easy access to interviewees who are geographically dispersed. However, the researcher cannot see any nonverbal communication on the part of the participant, and people often dislike telephone contacts because of their prior personal experiences with calls from survey firms asking for information.

Designing instruments for data collection

Designing good survey instruments is a challenging and complex process. Researcher should first consider whether a survey instrument is available to measure variables. S/he might also consider modifying an existing instrument. If neither of these approaches will work, design your own instrument. When survey researchers design an instrument for data collection, they typically perform the following steps:

a. Write different types of questions-These include personal, attitudinal, and behavioral questions; sensitive questions; and closed- and open-ended questions.
b. Use strategies for good question construction-This includes using clear language, making sure the answer options do not overlap, and posing questions that are applicable to all participants.
c. Perform a pilot test of the questions-This consists of administering the instrument to a small number of individuals and making changes based on their feedback.

Question Construction

Using good questions helps participants feel that they understand the question and can provide meaningful answers. Good questions are clear and unambiguous, and they do not confuse the participants. They also show respect for the participant by being sensitive to gender, class, and cultural needs of participants.

Followings are the potential question construction problems and some solutions for the development of survey questionnaire**-**

a. The question is unclear.

This usually occurs because words are vague or imprecise. Identify the unclear or vague words and replace them with words understood by participants in the study.

b. There are multiple questions.

Here, the question actually contains two or more questions, called a double- or triple-barreled question. Reduce the multiple questions to a single question.

c. The question is wordy.

When the question is too long, cut out unnecessary words to simplify and shorten the question. Look for excessive use of prepositions (e.g., more than three) or qualifying statements that lengthen the question.

d. The question is negatively worded or wordy.

If the question contains one or more negatives, such as "should not," the meaning becomes unclear. Also, reword the question if it leads the participants to one particular stance or another (e.g., using the word "pro-life"). Restate or reword the question to eliminate negative connotations or leading words.

e. The question includes jargon.

Jargon may not be familiar to all participants in a study. Eliminate the jargon and use words familiar to all participants.

f. There are overlapping responses.

This may lead to confusion when answering a question. Make sure that the response options do not overlap by creating distinct options.

g. There are unbalanced response options.

In this case, the responses may be unbalanced in terms of naturally occurring intervals. Response options may start with an "importance" word (e.g., "very important") and end with an "extent" word (e.g., "to a little extent"), rather than a matching adjective (e.g., "not important"). Decide on a single response option and use it consistently for all response categories for a question.

h. There is a mismatch between the question and the answers.

The responses may not match the "action" word used in the question. Identify the verb or adjective in the question that will be the basis for the response options and create options using this word. (E.g., if the question says "to what extent," the answer will say "a great extent.")

i. The question includes overly technical language.

When this occurs, the respondent may not have the level of understanding needed to respond to the question. Simplify the question so that all individuals will know the meaning of the words and can respond to the question.

j. Not all questions are applicable to all participants.

If some participants cannot answer the question, include "branching" or "contingency questions." These questions follow the original question and provide options to include all participants.

Pilot Testing the Questions

After good questions have been developed using principles of question construction, a researcher pilot test the questions. This helps determine that the individuals in the sample are capable of completing the survey and that they can understand the questions. A **pilot test** of a questionnaire or interview survey is a procedure in which a researcher makes changes in an instrument based on feedback from a small number of individuals who complete and evaluate the instrument. The participants in the pilot test provide written comments directly on the survey, and the researcher modifies or changes the survey to reflect those concerns. Because the pilot group provides feedback on the questionnaire, you exclude them from the final sample for the study.

For example, a survey of 100 middle school students' attitudes toward school might begin with a pilot test of an instrument with 50 questions. In this pilot test, the researcher selects 15 students to complete the instrument. The investigator then asks them to mark any problems on the survey, such as poorly worded questions, responses that do not make sense, or if it takes an excessive amount of time to complete the instrument. Based on student feedback, the researcher then revises the instrument before sending it out to the sample in the study.

Obtaining a high response rate

Survey researchers seek high response rates from participants in a study so that they can have confidence in generalizing the results to the population under study. When using interviews, the response rate is high because individuals interviewed typically consent to the interview in advance. However, when questionnaires are used, the number of responses returned (through mail or electronically) will vary. In either case, survey researchers place emphasis on obtaining a high response rate to their questionnaire or interview. On the instruments that are returned, the survey researcher is also concerned about whether the returned responses are biased. Even a small return rate may not be biased and be acceptable in survey research. Although response rate is important, bias is a larger concern than return rate because if the returned responses are biased, the database will be inadequate, regardless of the return rate.

Conducting a Questionnaire Study

The steps in conducting a questionnaire study are essentially the same as for other types of research, although data collection involves some unique considerations.

Stating the Problem

The problem or topic studied and the contents of the questionnaire must be of sufficient significance both to motivate potential respondents to respond and to justify the research effort in the first place. In defining the topic, the researcher should set specific objectives indicating the kind of information needed. Specific aspects of the topic, as well as the kind of questions to be formulated, should be described. For example, suppose a school superintendent wants to know how high school teachers perceive their schools. He wants to conduct a study to help identify areas in the high schools that can be improved. It is useful for the superintendent to begin by identifying important aspects of his general question; then he can select questions to address each aspect. He can perhaps focus on four subtopics: (a) respondent demographics (to compare the perceptions of males and females, experienced and new teachers, and teachers in different departments), (b) teacher perceptions of the quality of teaching, (c) teacher perceptions of available educational resources and (d) teacher perceptions of the school curriculum. Breaking the general topic into a few main areas helps to focus the survey and aid decision making in succeeding steps in the research sequence.

Constructing the Questionnaire

Development of a valid questionnaire requires both skill and time. As a general guideline, a questionnaire should be attractive, brief, and easy to respond to. Respondents are turned off by sloppy, crowded, misspelled, and lengthy questionnaires, especially ones that require long written responses to each question. Turning people off is certainly not the way to get them to respond. No item should be included that does not directly relate to the topic of the study, and structured, selectiontype items should be used if possible. It is easier to respond by circling a letter or word than by writing a lengthy response.

Many types of items are commonly used in questionnaires, including scaled items (e.g., Likert and semantic differential), ranked items (e.g., "Rank the following activities in order of their importance"), checklist items (e.g., "Check all of the following that characterize your principal"), and free-response items (e.g., "Write in your own words the main reasons you became a teacher"). Most commonly, surveys consist of structured items (also called closed-ended items). A structured item requires a respondent to choose among the provided response options (e.g., by circling a letter, checking a list, or numbering preferences). Questionnaires rarely contain large numbers of free-response items, but they may include one or two to give respondents the opportunity to add information not tapped by the closed-ended items.

Administering the Questionnaire

Selecting Participants Survey participants should be selected using an appropriate sampling technique. Although simple random and stratified random samplings are most commonly used in survey research, cluster, systematic, and nonrandom samples are also used. In some rare cases, when the population is small, the entire group may make up the sample. The selected research participants must be able and willing to provide the desired information to the researcher.

The target population for the superintendent's study is likely to be all high school teachers in the state. Such a group is too large a group to survey reasonably, so the superintendent must select participants from the accessible population. In this case, the likely accessible population is high school teachers from the schools in the superintendent's district. A sample, perhaps stratified by gender and department, can be randomly selected and asked to complete the questionnaire.

Distributing the Questionnaire

An important decision faced by all survey researchers is, what method should I use to collect data? There are five approaches: mail, email, telephone, personal administration, and interview. Each approach has its advantages and disadvantages. The bulk of educational surveys rely on "snail" mailed or emailed questionnaires. Mailing questionnaires is relatively inexpensive, easily standardized, and confidential, but this method of administration is also subject to low response rates and suffers from the researcher's inability to ask probing or follow-up questions. Sending questionnaires by email has become a popular alternative. In addition to being speedy and efficient, this method shares both the advantages and disadvantages of mail questionnaires, with the additional disadvantage that not all potential respondents have email service. Telephone surveys tend to have high response rates and allow data to be collected fairly quickly, but they require lists of target phone numbers and administrator training as well as the willingness of a respondent to participate in a telephone survey—something that is becoming increasingly difficult in an era of outsourcing of telemarketing services. Personal administration of a prepared questionnaire is efficient if participants are closely situated, but it is time-consuming and also requires administrator training. Personal interviews allow rich, more complete responses, but they have the least standardization and take the longest to administer.

Tabulating Questionnaire Responses

The easiest way to tabulate questionnaire responses is to have participants mark responses to closedended questions on a scan able answer sheet. This option involves locating a scanner and possibly paying a fee to have questionnaires scanned. If scan able answer sheets are not an option, then each respondent's answers will have to be entered one by one into a computer spreadsheet (e.g., Excel or Lotus) or a statistical program (e.g., SPSS or SAS). If you design a questionnaire that will be hand tabulated, make sure that the format is easy to follow and allows respondents to mark answers clearly so that you can enter data quickly, without having to search for information.

If questionnaire contains open-ended questions, researcher will need to code answers according to patterns in the responses provided. With a qualitative software program, s/he can examine textual data, code it, and generate information regarding the frequency and nature of various codes. Many qualitative software programs also allow the researcher to export coded qualitative data into statistical programs, where advanced statistical analyses can be performed.

Analyzing Results

When presenting the results of a questionnaire study, you should include the total sample size and the overall percentage of returns along with the response rate for each item because not all respondents will answer all questions. The simplest way to present the results is to indicate the percentage of respondents who selected each alternative for each item (e.g., "On Item 4 dealing with possession of a master's degree, 50% said yes, 30% said no, and 20% said they were working on one").

Although item-by-item descriptions are a simple way to report the results of a survey, they can produce an overload of information that is difficult to absorb and condense. A better way to report is to group items into clusters that address the same issue and develop total scores across an item cluster. Possible explanations for certain attitudes and behaviors can be explored by identifying factors that seem to be related to certain responses. However, such comparisons can be made only if demographic information about the respondents is collected on the questionnaire.

Steps in conducting Survey Research

The steps in the process of conducting survey research follow the general process of research. Survey steps, however, address primarily the procedures for collecting data, analyzing data, and writing the final report.

1. **Decide if a Survey Is the Best Design to Use**

Researcher need to decide whether survey research is the best design to use in the study. Surveys help describe the trends in a population or describe the relationship among variables or compare groups. Instances where surveys are most suitable are to assess trends or characteristics of a population; learn about individual attitudes, opinions, beliefs, and practices; evaluate the success or effectiveness of a program; or identify the needs of a community.

There are several advantages to using surveys. S/he can administer them in a short time, they are economical as a means of data collection, and they can reach a geographically dispersed population. Further, canvass the participants anonymously, without biasing their responses. However, survey data is self-reported information, reporting only what people think rather than what they do. Sometimes the response rates are low and researchers cannot make claims about the representativeness of the results to the population. As mentioned earlier, surveys do not control for many variables that might explain the relationship between the independent and dependent variables, and they do not provide participants flexibility in responding to questions (unless open-ended questions are included).

2. **Identify the Research Questions or Hypotheses**

Researcher can address both research questions and hypotheses in a survey design. Surveys lend themselves to hypothesis testing because you will be studying a sample to draw inferences to a population. Forms of research questions or hypotheses are those that:

Describe the characteristics or trends of a population of people, such as the frequency of tobacco use among male high school students

Compare groups in terms of specific attributes, such as a comparison of teachers and administrators about attitudes toward "in-service" learning days

Relate two or more variables, such as a survey of teachers to relate "burnout" to number of years of teaching.

3. **Identify the Population, the Sampling Frame, and the Sample**

The process of survey research begins with identifying the population. This step requires defining the population, determining the number of people in it, and assessing whether you can obtain a list of names (i.e., the sampling frame) for the sample. Also, the population may need to be stratified before sampling, so select characteristics of the population (e.g., males and females) are represented in the sample.

Once Researcher identified the target population and compiled a list of its members, s/he can select the sample, preferably using random sampling procedures. S/he will need to identify an adequate sample size, using a sampling error formula.

4. **Determine the Survey Design and Data Collection Procedures**

The researcher must also determine if the survey study will be cross-sectional or longitudinal. The decision to use a longitudinal or cross-sectional design relates to the nature of the problem studied, access to participants, and the time available to the researchers for data collection. For example, learning about the longitudinal development of adolescent social skills in schools requires following adolescents over time and devoting extensive time to data collection. In contrast, examining parents' attitudes toward discipline in schools requires a cross-sectional study at one point in time to assess attitudes immediately and quickly.

Consider also whether data collection will be based on questionnaires (mailed or electronic) or interviews (individual, focus group, or telephone), and weigh the advantages and disadvantages of each form.

5. **Develop or Locate an Instrument**

You need an instrument to collect or measure the variables in a study. It is easier to locate an instrument than to develop one. Standards of reliability and construct validity need to be applied to scores from existing instruments before you select them for use. If a study addresses only a few variables, researchers can design their own instruments. A check for the reliability and validity of scores from this instrument during data analysis is most important.

6. **Administer the Instrument**

This step is perhaps the most time-consuming phase of survey research. It involves seeking and obtaining permission to conduct the survey and using procedures for data gathering, such as training interviewers or preparing questionnaires for mailing. It requires continually following up to obtain a high response rate, checking for response bias if questionnaires are used, and preparing the data for analysis by coding the information from the instruments into a computer file.

7. **Analyze the Data to Address the Research Questions or Hypotheses**

The data analysis procedures will reflect the types of research questions or hypotheses the researcher plans to address in the study. Analysis consists of noting response rates, checking for response bias, conducting descriptive analysis of all items, and then answering descriptive questions. It might also involve testing hypotheses or research questions using inferential statistics.

8. **Write the Report**

You should write the survey study using a standard quantitative structure that consists of an introduction, the review of the literature, the methods, the results, and the discussion. Specify in the "Methods" section of the study detailed information about the survey procedures. Include in the "Discussion" section comments about the generalizability of the results to the population

Criteria for Evaluating Survey Research

A good survey study includes the identification of the population and the sample, contains an adequate-sized sample systematically derived, employs a cross-sectional or longitudinal design, specifies the instruments (and includes sample questions), determines whether scores from them will likely be reliable and valid, uses appropriate data analysis procedures to answer the questions or hypotheses, and is written acknowledging ethical issues and using a standard structure.

Potential Ethical Issues in Survey Research

Ethical issues in survey research involve engaging in good practices. Often survey research is exempt by institutional review boards. During data collection, attention needs to be given to using appropriate incentives and delivering on benefits guaranteed. The survey data collection procedure cannot put data collectors at risk for their safety. Safety applies to the respondents or participants as well. Confidentiality of their responses needs to be protected, along with minimizing links between data respondents and participants. IDs linked to responses can be an effective means of protecting individual identity. Also, the researcher has an obligation to destroy survey instruments after the conclusion of the study.

CO-RELATIONAL RESEARCH METHOD

Co relational research describes what exists at the moment (conditions, practices, processes, structures etc.) and is therefore, classified as a type of descriptive method. Nevertheless, these conditions, practices, processes or structures described are markedly different from the way they are usually described in a survey or an observational study.

Correlational designs provide an opportunity for you to predict scores and explain the relationship among variables. In correlational research designs, investigators use the correlation statistical test to describe and measure the degree of association (or relationship) between two or more variables or sets of scores. In this design, the researchers do not attempt to control or manipulate the variables as in an experiment; instead, they relate, using the correlation statistic, two or more scores for each person (e.g., a student motivation and a student achievement score for each individual).

Co relational research comprises of collecting data to determine whether, and to what extent, a relationship exists between two or more quantifiable variables. Co relational research uses numerical data to explore relationships between two or more variables. The degree of relationship is expressed in terms of a coefficient of correlation. If the relationship exists between variables, it implies that scores on one variable are associated with or vary with the scores on another variable. The exploration of relationship of the relationship between variables provides insight into the nature of the variables themselves as well as an understanding of their relationships. If the relationships are substantial and consistent, they enable a researcher to make predictions about the variables.

Correlational research produces indexes that show both the direction and the strength of relationships among variables, taking into account the entire range of these variables. This index is called a correlation coefficient. The sign (+ or −) of the coefficient indicates the direction of the relationship. If the coefficient has a positive sign, this means that as one variable increases, the other also increases. For example, the correlation between height and weight is positive because tall people tend to be heavier and short people lighter. A negative coefficient indicates that as one variable increases, the other decreases. The correlation between outdoor air temperature during the winter months and heating bills is negative; as temperature decreases, heating bills rise.

The size of the correlation coefficient indicates the strength of the relationship between the variables. The coefficient can range in value from +1.00 (indicating a perfect positive relationship) through 0 (indicating no relationship) to −1.00 (indicating a perfect negative relationship). A perfect positive relationship means that for every z-score unit increase in one variable there is an identical z-score unit increase in the other. A perfect negative relationship indicates that for every unit increase in one variable there is an identical unit decrease in the other. Few variables ever show perfect correlation, especially in relating human characteristics.

Correlational research is aimed at determining the nature, degree and direction of relationships between variables or using these relationships to make predictions. Correlational studies typically investigate a number of variables expected to be related to a major, complex variable. Those variables which are not found to be related to this major, complex variable are omitted from further analysis. On the other hand, those variables which are found to be related to this major, complex variable are further analyzed in a causal-comparative or experimental study so as to determine the exact nature of the relationship between them.

In a correlational study, hypotheses or research questions are stated at the beginning of the study. The null hypotheses are often used in a correlational study.

Correlational study does not specify cause-and-effect relationships between variables under consideration. It merely specifies concomitant variations in the scores on the variables. For example, there is a strong relationship between students‘scores on academic achievement in Mathematics and their scores on academic achievement in Science. This does not suggest that one of these variables is the cause and the other is the effect. In fact, a third variable, viz., students’ intelligence could be the cause of students‘academic achievement in both, Mathematics and Science.

Uses of correlational research

Correlational research is useful in a wide variety of studies.

The most useful applications of correlation are

a. Assessing relationships
b. Assessing consistency and
c. Prediction.

a. Assessing relationships

Cor-relational research methods are used to assess relationships and patterns of relationship among variables in a single group of subjects. For instance, correlational research is used to answer questions such as the following: Is there a relationship between math aptitude and achievement in computer science? What is the direction and strength of this relationship, if any? You would most likely predict that a positive relationship would be found between scores on a math aptitude test and grades in computer science. A correlational study would determine the extent of any relationship between these variables.

b. Assessing Consistency

Correlation can be used to measure consistency (or lack thereof) in a wide variety of cases. For example, how consistent are the independently assigned merit ratings given by the principal and the assistant principal to teachers in a school? How much agreement is there among Olympic judges rating the performance of a group of gymnasts? When a researcher asks a group of teachers to rank the severity of disruption created by each item on a list of behavior disorders, to what extent do their rankings agree?

c. Prediction

If you find that two variables are correlated, then you can use one variable to predict the other. The higher the correlation, the more accurate the prediction. Prediction studies are frequently used in education. For example, correlational research has shown that high school grades and scholastic aptitude measures are related to college grade point average (GPA). If a student scores high on aptitude tests and has high grades in high school, he or she is more likely to make high grades in college than is a student who scores low on the two predictor variables. Researchers can predict with a certain degree of accuracy a student's probable freshman GPA based on high school grades and aptitude test scores. This prediction will not hold for every case because other factors, such as motivation, initiative, or study habits, are not considered. However, in general, the prediction is good enough to be useful to college admissions officers.

Designs of Correlational Research

The two primary correlation designs are explanation and prediction.

The Explanatory Design

Various authors refer to explanatory correlational research as "relational" research (Cohen & Manion, 1994, p. 123), "accounting-for-variance studies" (Punch, 1998, p. 78), or "explanatory" research (Fraenkel & Wallen, 2000, p. 360). Because one basic objective of this form of correlational research is to explain the association between or among variables, we will use the term explanatory research in this discussion. An explanatory research design is a correlational design in which the researcher is interested in the extent to which two variables (or more) co-vary, that is, where changes in one variable are reflected in changes in the other. Explanatory designs consist of a simple association between two variables (e.g., sense of humor and performance in drama) or more than two (e.g., pressure from friends or feelings of isolation that contribute to binge drinking).

The Prediction Design

Instead of simply relating variables—two variables at a time or a complex set such as in our last example—in a prediction design, researchers seek to anticipate outcomes by using certain variables as predictors. For example, superintendents and principals need to identify teachers who will be successful in their schools. To select teachers who have a good chance of success, the administrators can identify predictors of success using correlational research. Prediction studies, therefore, are useful because they help anticipate or forecast future behavior.

The purpose of a prediction research design is to identify variables that will predict an outcome or criterion. In this form of research, the investigator identifies one or more predictor variable and a criterion (or outcome) variable. A predictor variable is a variable used to make a forecast about an outcome in correlational research. In the case of predicting teacher success in a school, the predictor may be "mentoring" during teacher training or "years of experience teaching." In much prediction research, investigators often use more than one predictor variable.

The outcome being predicted in correlational research, however, is called the criterion variable. In our example, teacher success is the criterion variable. Although more than one outcome can be predicted, the typical educational study includes only one criterion variable.

Characteristics of Cor-relational Research

Correlation research includes specific characteristics:

- Displays of scores (scatterplots and matrices)
- Associations between scores (direction, form, and strength)
- Multiple variable analysis (partial correlations and multiple regression)

a. Displays of Scores

If you have two scores, in correlation research you can plot these scores on a graph (or scatterplot) or present them in a table (or correlation matrix).

Scatterplots

Researchers plot scores for two variables on a graph to provide a visual picture of the form of the scores. This allows researchers to identify the type of association among variables and locate extreme scores. Most importantly, this plot can provide useful information about the form of the association—whether the scores are linear (follow a straight line) or curvilinear (follow a U-shaped form). It also indicates the direction of the association (e.g., one score goes up and the other goes up as well) and the degree of the association (whether the relationship is perfect, with a correlation of 1.0, or less than perfect).

A plot helps to assess this association between two scores for participants. A **scatterplot** (or **scatter diagram**) is a pictorial image displayed on a graph of two sets of scores for participants. These scores are typically identified as *X* and *Y,* with *X* values represented on the horizontal axis, and *Y* values represented on the vertical axis. A single point indicates where the *X* and *Y* scores intersect for one individual.

A Correlation Matrix

Correlation researchers typically display correlation coefficients in a matrix. A **correlation matrix** presents a visual display of the correlation coefficients for all variables in a study.

b. Associations between Scores

After correlation researchers graph scores and produce a correlation matrix, they can then interpret the meaning of the association between scores. This calls for understanding the direction of the association, the form of the distribution, the degree of association, and its strength.

Direction of the Association

When examining a graph, it is important to identify if the points intersect, or move in the same or opposite directions. In a positive correlation (indicated by a "1" correlation coefficient) the points move in the same direction; that is, when X increases, so does Y or, alternatively, if X decreases, so does Y. In a negative correlation (indicated by a "–" correlation coefficient), the points move in the opposite direction; that is, when X increases, Y decreases, and when X decreases, Y increases. If scores on one variable do not relate in any pattern on the other variable, then no linear association exists.

Form of the Association

Linear Relationship

A positive linear relationship of scores is low (or high) scores on one variable relate to low (or high) scores on a second variable. For example, low scores on depression are associated with low scores on number of hours using the Internet per week.

A negative linear relationship result is low scores on one variable relate to high scores on the other variable. Low scores on depression, for example, might be associated with high scores on use of the Internet, suggesting a negative relationship.

Uncorrelated and Nonlinear Relationships

A particular score on one variable does not predict or tell us any information about the possible score on the other variable. In our example, a plot of the scores for depression and the scores for Internet use would be irregular, without any particular pattern.

A curvilinear distribution (or nonlinear relationship) shows a U-shaped relationship in scores. An increase, plateau, and decline in the Y-axis variable with increasing values of the X-axis variable. A decrease, plateau, and increase in the Y-axis variable, with increasing values of the X-axis variable. For example, it is possible that as Internet use increases, so does depression, up to a point at which the Internet actually becomes a coping mechanism for stress, and depression begins to decrease.

The correlation coefficient is useful for describing and measuring the association between two variables if the association is linear. If an r is used to estimate a curvilinear association, it would provide an underestimate of the correlation. Therefore, researchers use different statistics than the r to calculate the relationship between variables for a curvilinear distribution and for relating ranked data.

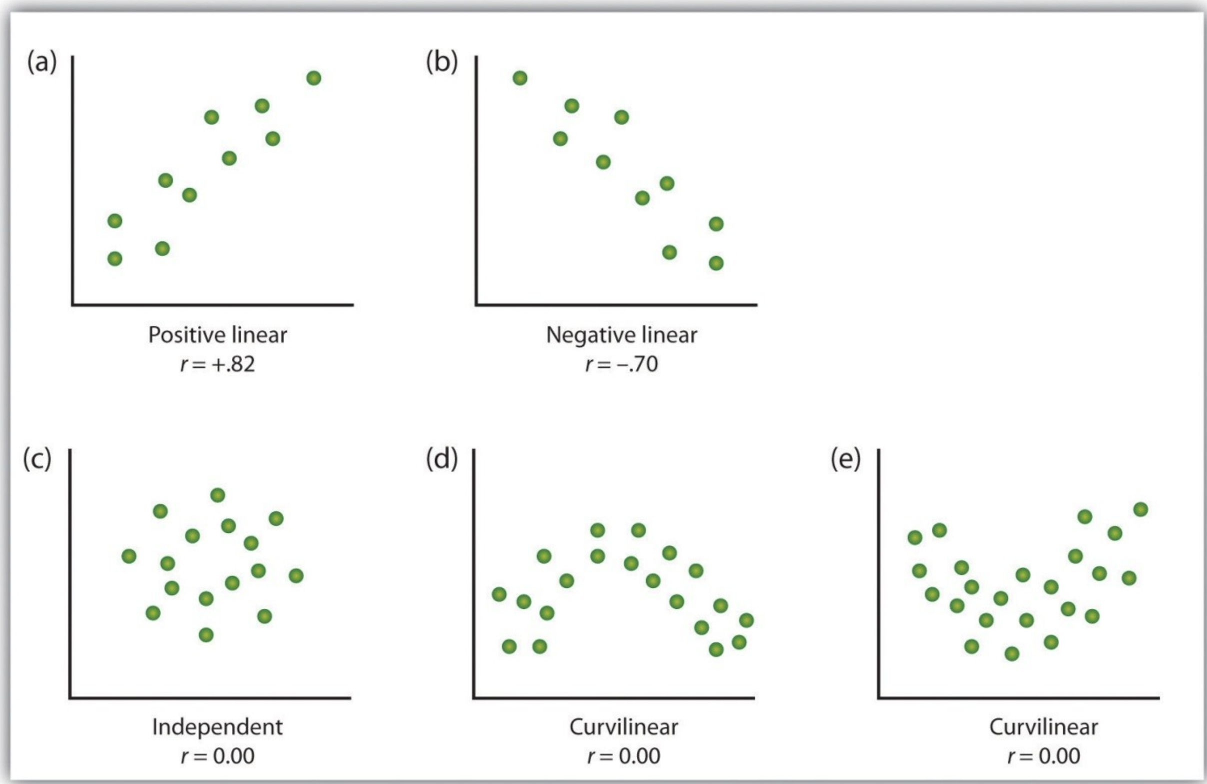
Instead of the r coefficient, researchers use the Spearman rho (r_s) correlation coefficient for nonlinear data and for other types of data measured on categorical (rankordered) scales. When you measure one variable on a continuous (interval or ratio) scale and the other is a categorical, dichotomous scale, the correlation statistic should not be the r but the point-biserial correlation. Assume that a researcher correlates continuous, interval scores on depression with males and females (a dichotomous variable). A point-biserial correlation statistic is used by converting the dichotomous variable (males, females) into numerical scores by assigning males = 1 and females = 2. Using these numbers and the formula for ordinal data, the researcher calculates the point-biserial correlation coefficient that measures the degree and direction of association between males and females on depression.

A variation of this theme of using different types of scales in assessing the association between two variables is the phi coefficient. The phi coefficient is used to determine the degree and direction of association when both variable measures are dichotomous. For example, males and females might be correlated with drug usage (no and yes). In this situation, the researcher also converts both dichotomous variables to numeric values (males = 1, females = 2; no to drugs = 1, yes to drugs = 2) and then uses the phi coefficient formula for converted scores.

Degree and Strength of Association

Degree of association means that the association between two variables or sets of scores is a correlation coefficient of –1.00 to +1.00, with 0.00 indicating no linear association at all. This association between two sets of scores reflects whether there is a consistent, predictable association between the scores.

Correlational researchers interpret the magnitude and direction of the correlations. With numbers indicating strength and valence signs indicating direction (+1.00 to –1.00), the statistic provides a measure of the magnitude of the relationship between two variables.



c. Multiple Variable Analysis

In many correlation studies, researchers predict outcomes based on more than one predictor variable. Thus, they need to account for the impact of each variable. Two multiple variable analysis approaches are partial correlations and multiple regression.

Partial Correlations

In many research situations, we study three, four, or five variables as predictors of outcomes. The type of variable called a mediating or intervening variable "stands between" the independent and dependent variables and influences both of them. This variable is different from a control variable that influences the outcome in an experiment. We use partial correlations to determine the amount of variance that an intervening variable explains in both the independent and dependent variables.

Multiple Regression

Correlation researchers use the correlation statistic to predict future scores. To see what impact multiple variables have on an outcome, researchers use regression analysis. Multiple regression (or multiple correlation) is a statistical procedure for examining the combined relationship of multiple independent variables with a single dependent variable. In regression, the variation in the dependent variable is explained by the variance of each independent variable (the relative importance of each predictor), as well as the combined effect of all independent variables (the proportion of criterion variance explained by all predictors), designated by R^2 (Kline, 1998).

Steps of a Correlational Research

1. **Selection of a Problem**

Correlational study is designed (a) to determine whether and how a set of variables are related, or (b) to test the hypothesis of expected relationship between among the set of two or more variables. The variables to be included in the study need to be selected on the basis of a sound theory or prior research or observation and experience. There has to be some logical connection between the variables so as to make interpretations of the findings of the study more meaningful, valid and scientific. A correlational study is not done just to find out what exists: it is done for the ultimate purpose of explanation and prediction of phenomena.

2. **Selection of the Sample and Tools**

Ideally, you should randomly select the individuals to generalize results to the population, and seek permissions to collect the data from responsible authorities and from the institutional review board. The group needs to be of adequate size for use of the correlational statistic, such as $N = 30$; larger sizes contribute to less error variance and better claims of representativeness. The sample is generally selected using one of the acceptable sampling methods. If the validity and the reliability of the variables to be studied are low, the measurement error is likely to be high and hence the sample size should be large. Thus it is necessary to ensure that valid and reliable tools are used for the purpose of collecting the data. Moreover, suppose you are studying the relationship between classroom environment and academic achievement of students. If your tool measuring classroom environment focuses only on the physical aspects of the classroom and not its psycho-social aspects, then your findings would indicate a relationship only between academic achievement of students and the physical aspects of the classroom environment and not the entire classroom environment since the physical aspects of the classroom environment is not the only comprehensive and reliable measure of classroom environment. Thus the measurement instruments should be valid and reliable.

3. **Design and Procedure**

The basic design of a correlational study is simple. It requires scores obtained on two or more variables from each unit of the sample and the correlation coefficient between the paired scores is computed which indicates the degree and direction of the relationship between variables.

4. **Analyze the Data and Represent the Results**

The objective in correlational research is to describe the degree of association between two or more variables. The investigator looks for a pattern of responses and uses statistical procedures to determine the strength of the relationship as well as its direction. A statistically significant relationship, if found, does not imply causation (cause and effect) but merely an association between the variables. More rigorous procedures, such as those used in experiments, can provide better control than those used in a correlational study.

The analysis begins with coding the data and transferring it from the instruments into a computer file. Then the researcher needs to determine the appropriate statistic to use. An initial question is whether the data are linearly or curvilinearly related. A scatterplot of the scores (if a bivariate study) can help determine this question. Also, consider whether:

- Only one independent variable is being studied (Pearson's correlation coefficient)
- A mediating variable explains both the independent and dependent variables and needs to be controlled (partial correlation coefficient)
- More than one independent variable needs to be studied to explain the variability in a dependent variable (multiple regression coefficient)

Based on the most appropriate statistical test, the researcher next calculates whether the statistic is significant based on the scores. For example, a *p* value is obtained in a bivariate study by:

- Setting the alpha level
- Using the critical values of an *r* table, available in many statistics books
- Using degrees of freedom of $N = 2$ with this table
- Calculating the observed *r* coefficient and comparing it with the *r*-critical value
- Rejecting or failing to reject the null hypothesis at a specific significance level, such as $p < 0.05$

In addition, it is useful to also report effect size (r^2). In correlational analysis, the effect size is the Pearson's correlation coefficient squared. In representing the results, the correlational researcher will present a correlation matrix of all variables as well as a statistical table (for a regression study) reporting the R and R^2 values and the beta weights for each variable.

5. **Interpret the Results**

The final step in conducting a correlational study is interpreting the meaning of the results. This requires discussing the magnitude and the direction of the results in a correlational study, considering the impact of intervening variables in a partial correlation study, interpreting the regression weights of variables in a regression analysis, and developing a predictive equation for use in a prediction study.

In all of these steps, an overall concern is whether your data support the theory, the hypotheses, or questions. Further, the researcher considers whether the results confirm or disconfirm findings from other studies. Also, a reflection is made about whether some of the threats discussed above may have contributed to erroneous coefficients and the steps that might be taken by future researchers to address these concerns.

In a study designed to explore or test hypothesized relationships, a correlation coefficient is interpreted in terms of its statistical significance. Co relational research is of the following two types:

(a) Relationship Studies:
These attempts to gain insight into variables that are related to complex variables such as academic performance, self-concept, stress, achievement motivation or creativity.

(b) Prediction Studies:
These are conducted to facilitate decisions about individuals or to aid in various types of selection. They are also conducted to determine predictive validity of measuring tools as well as to test variables hypothesized to be predictors of a criterion variable.

Criteria for Evaluating a Correlational Study

Evaluate a correlational study in terms of the strength of its data collection, analysis, and interpretations. These factors include adequate sample size, good presentations in graphs and matrices, clear procedures, and an interpretation about the relationship among variables.

Ethical Issues in Conducting Correlational Research

Ethical issues arise in many phases of the correlational research process. In data collection, ethics relate to adequate sample size, lack of control, and the inclusion of as many predictors as possible. In data analysis, researchers need a complete statement of findings to include effect size and the use of appropriate statistics. Analysis cannot include making up data. In recording and presenting studies, the write-up should include statements about relationships rather than causation, a willingness to share data, and publishing in scholarly outlets.

CASE STUDY

Case study research is descriptive research that involves describing and interpreting events, conditions, circumstances or situations that are occurring in the present. Case study seeks to engage with and report the complexities of social activity in order to represent the meanings that individual social actors bring to their social settings. It excels at bringing us to an understanding of a complex issue or object and can extend experience or add strength to what is already known through previous research. Case studies emphasize detailed contextual analysis of a limited number of events or conditions and their relationships. Darwin's theory of evolution was based, in essence, on case study research, not experimentation, for instance. In education, this is one of the most widely used qualitative approaches of research.

According to Odum "The case study method is a technique by which individual factor whether it be an institution or just an episode in the life of an individual or a group is analyzed in its relationship to any other in the group." Its distinguishing characteristic is that each respondent is (individual, family, classroom, institution, cultural group) is taken as a unit and the unitary nature of individual case is the focus of analysis. It seeks to engage with and report the complexity of social and/or educational activity in order to represent the meanings that individual actors in the situation bring to that setting. It assumes that social and/or educational reality is created through social interactions, situated in specific contexts and histories and seeks to identify and describe followed by analysing and theorising. It assumes that things may not be as they seem and involve in-depth analysis so as to understand a 'case' rather than generalizing to a larger population. It derives much of its philosophical underpinnings and methodology from ethnography, symbolic interactionism, ethnomethodology and phenomenology. It follows the 'social constructivism' perspective of social sciences.

Most case studies are usually qualitative in nature. Case study research excels at enabling us to understand a complex issue or object and can extend experience or add strength to what is already known through previous research. Case studies involve a detailed contextual analysis of a limited number of events or conditions and their relationships. Social scientists have made a wide use of this qualitative research method to examine contemporary real-life situations and provide the basis for the application of ideas and extension of methods. Yin defines the case study research method as an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used.

However, some case studies can also be quantitative in nature especially if they deal with cost-effectiveness, cost-benefit analysis or institutional effectiveness. Many case studies have been done by combining the qualitative as well as the quantitative approaches in which initially the qualitative approach has been used and data have been collected using interviews and observations followed by the quantitative approach. The approach of case studies ranges from general field studies to interview of a single individual or group. A case study can be precisely focused on a topic or can include a broad view of life and society. For example, a case study can focus on the life of a single gifted student, her/his actions, behaviors; abilities and so on in his school or it can focus on the social life of an individual including his entire background, experiences, motivations and aspirations that influence her/his behaviour society. Examples of case studies include a 'case' of curriculum development, of innovative training, of disruptive behavior, of an ineffective institution and so on.

Case studies can be conducted to develop a 'research-based' theory with which to analyze situations: a theory of, for and about practice. It is essential to note that since most case studies focus on a single unit or small number of units, the findings cannot be generalized to larger populations. However, its utility cannot be underestimated. A case study is conducted with a fundamental assumption that though human behaviour is situation-specific and individualized, there is a predictable uniformity in basic human nature.

A case study can be conducted to explore, to describe or to explain a phenomenon. It could be a synchronic study in which data are collected at one point of time or it could be longitudinal in nature. It could be conducted at a single site or it could be multi-site. In other words, it is inherently a very flexible methodology.

A case typically refers to a person a learner, a teacher, an administrator or an entity, such as a school, a university, a classroom or a programme. In some policy-related research, the case could be a country. Case studies may be included in larger quantitative or qualitative studies to provide a concrete illustration of findings, or they may be conducted independently, either longitudinally or in a more restricted temporal period. Unlike ethnographic research, case studies do not necessarily focus on cultural aspects of a group or its members. Case study research may focus on a single case or multiple cases.

Characteristics of a Case Study

Following are the characteristics of a case study:

1. It is concerned with an exhaustive study of particular instances. A case is a particular instance of a phenomenon. In education, examples of phenomena include educational programmes, curricula, roles, events, interactions, policies, process, and concept and so on. Its distinguishing feature is that each respondent (individual, class, institution or cultural group) is treated as a unit.
2. It emphasizes the study of interrelationship between different attributes of a unit.
3. According to Cooley, case study deepens our perception and gives us a clear insight into life… It gets at behaviour directly and not by an indirect or abstract approach.
4. Each case study needs to have a clear focus which may include those aspects of the case on which the data collection and analysis will concentrate. The focus of a study could be a specific topic, theme, proposition or a working hypothesis.
5. It focuses on the natural history of the unit under study and its interaction with the social world around it.
6. The progressive records of personal experience in a case study reveals the internal strivings, tensions and motivations that lead to specific behaviours or actions of individuals or the unit of analysis.
7. In order to ensure that the case study is intensive and in-depth, data are collected over a long period of time from a variety of sources including human and material and by using a variety of techniques such as interviews and observations and tools such as questionnaires, documents, artifacts, diaries and so on. 8. According to Smith, as cited by Merriam, (1998), these studies are different from other forms of qualitative of research in that they focus on a 'single unit' or a 'bounded system'. A system is said to be a bounded system if it includes a finite or limited number of cases to interviewed or observed within a definite amount of time.
8. It may be defined as an in-depth study of one or more instances of a phenomenon- an individual, a group, an institution, a classroom or an event- with the objective of discovering meaning, investigating processes, gaining an insight and an understanding of an individual, group or phenomena within the context in such a way that it reflects the real life context of the participants involved in the phenomena. These individuals, groups, institutions, classrooms or events may represent the unit of analysis in a case study. For example, in a case study, the unit of analysis may be a classroom and the researcher may decide to investigate the events in three such classrooms.

9. According to Yin, case studies typically involve investigation of a phenomenon for which the boundaries between the phenomenon and its context are not clearly evident. These boundaries should be clearly clarified as part of the case study. He further emphasizes the importance of conducting a case study in its real life context. In education, the classroom or the school is the real life context of a case study as the participants of such a case study are naturally found in these settings.
10. There are two major perspectives in a case study, namely, the etic perspective and the emic perspective. The etic perspective is that of the researcher (i.e. the outsider‘s perspective) whereas the emic perspective is that of the research participants including teachers, principals and students (i.e. the insider‘s perspective). This enables the researcher to study the local, immediate meanings of social actions of the participants and to study how they view the social situation of the setting and the phenomenon under study. A comprehensive case study includes both the perspectives.
11. A case study can be a single-site study or a multi-site study.
12. Cases are selected on the basis of dimensions of a theory (pattern-matching) or on diversity on a dependent phenomenon (explanation-building).
13. No generalization is made to a population beyond cases similar to those studied.
14. Conclusions are phrased in terms of model elimination, not model validation. Numerous alternative theories may be consistent with data gathered from a case study.
15. Case study approaches have difficulty in terms of evaluation of low-probability causal paths in a model as any given case selected for study may fail to display such a path, even when it exists in the larger population of potential cases.
16. Acknowledging multiple realities in qualitative case studies, as is now commonly done, involves discerning the various perspectives of the researcher, the case/participant, and others, which may or may not converge.

Components of a Case Study Design

According to Yin, following are the five component elements of a case study design:

a. Study questions
b. Study propositions (if any are being used) or theoretical framework
c. Identification of the units of analysis
d. The logical linking of the data to the propositions (or theory)
e. The criteria for interpreting the findings.

The purpose of a case study is a detailed examination of a specific activity, event, institution, or person/s. The hypotheses or the research questions are stated broadly at the beginning at the study. A study‘s questions are directed towards ‘how’ and ‘why’ considerations and enunciating and defining these are the first task of the researcher. The study‘s propositions could be derived from these ‘how’ and ‘why’ questions. These propositions could help in developing a theoretical focus. However, all case studies may not have propositions. For instance, an exploratory case study may give only a purpose statement or criteria that could guide the research process. The unit of analysis defines what the case study is focusing on, whether an individual, a group, n institution, a city, a society, a nation and so on. Linkages between the data and the propositions (or theory) and the criteria for interpreting the findings are usually the least developed aspects of case studies (Yin, 1994).

Types of Case Study Designs

Yin (1994) and Winston (1997) have identified several types of case study designs. These are as follows:

i. **Exploratory Case Study Design**

In this type of case study design, field work and data collection are carried out before determining the research questions. It examines a topic on which there is very little prior research. Such a study is a prelude to a large social scientific study. However, before conducting such an exploratory case study, its organizational framework is designed in advance so as to ensure its usefulness as a pilot study of a larger, more comprehensive research. The purpose of the exploratory study is to elaborate a concept, build up a model or advocate propositions.

ii. **Explanatory Case Study Design**

These are useful when providing explanation to phenomena under consideration. These explanations are patterns implying that one type of variation observed in a case study is systematically related to another variation. Such a pattern can be a relational pattern or a causal pattern depending on the conceptual framework of the study. In complex studies of organizations and communities, multivariate cases are included so as to examine a plurality of influences. Yin and Moore (1988) suggest the use of a pattern-matching technique in such a research wherein several pieces of information from the same case may be related to some theoretical proposition.

iii. **Descriptive Case Study Design**

A descriptive case study necessitates that the researcher present a descriptive theory which establishes the overall framework for the investigator to follow throughout the study. This type of case study requires formulation and identification of a practicable theoretical framework before articulating research questions. It is also essential to determine the unit of analysis before beginning the research study. In this type of case study, the researcher attempts to portray a phenomenon and conceptualize it, including statements that recreate a situation and context as much as possible.

iv. **Evaluative Case Study Design**

Often, in responsive evaluation, quasi-legal evaluation and expertise-based evaluation, a case study is conducted to make judgments. This may include a deep account of the phenomenon being evaluated and identification of most important and relevant constructs themes and patterns. Evaluative case studies can be conducted on educational programmes funded by the Government such as ―Sarva Shiksha Abhiyan or Orientation Programmes and Refresher Courses conducted by Academic Staff Colleges for college teachers or other such programmes organized by the State and Local Governments for secondary and primary school teachers.

Steps of Conducting a Case Study

Following are the steps of a case study-

1. Identifying a current topic which is of interest to the researcher
2. Identifying research questions and developing hypotheses (if any)
3. Determining the unit of sampling and the number of units
4. Identifying sources, tools and techniques of data collection
5. Evaluating and Analyzing Data
6. Report writing

1. **Identifying a current topic which is of interest to the researcher**

In order to identify a topic for case study research, the following questions need to be asked:

(i) What kind of topics can be addressed using the case study method?
(ii) How can a case study research be designed, shaped and scoped in order to answer the research question adequately?
(iii) How can the participation of individuals/institutions be obtained for the case study research?
(iv) How can case study data be obtained from case participants in an effective and efficient manner?
(v) How can rigor be established in the case study research report so that it is publishable in academic journals?

According to Maxwell, there are eight different factors that could influence the goals of a case study as follows:

a. To grasp the meanings that events, situations, experiences and actions have for participants in the study which is part of the reality that the researcher wants to understand.
b. To understand the particular context within which the participants are operating and its influence on their actions, in addition to the context in which one‘s research participants are embedded. Qualitative researchers also take into account the contextual factors that influence the research itself.
c. To identify unanticipated phenomena and influences that emerge in the setting and to generate new grounded theories about such aspects.
d. To grasp the process by which events and actions take place that lead to particular outcomes.
e. To develop causal explanations based on process theory (which involves tracing the process by which specific aspects affect other aspects), rather than variance theory (which involves showing a relationship between two variables as in quantitative research).

f. To generate results and theories that are understandable and experientially credible, both to the participants in the study and to others.
g. To conduct summative evaluations designed to improve practice rather than merely to assess the value of a final programme or product.
h. To engage in collaborative and action research practitioners and research participants.

2. Identifying research questions and developing hypotheses (if any)

The second step in case study research is to establish a research focal point by forming questions about the situation or problem to be studied and determining a purpose for the study. The research objective in a case study is often a programme, an entity, a person or a group of people. Each objective is likely to be connected to political, social, historical and personal issues providing extensive potential for questions and adding intricacy to the case study. The researcher attains the objective of the case study through an in-depth investigation using a variety of data gathering methods to generate substantiation that leads to understanding of the case and answers the research questions. Case study research is usually aimed at answering one or more questions which begin with "how" or "why." The questions are concerned with a limited number of events or conditions and their inter-relationships. In order to formulate research questions, literature review needs to be undertaken so as to establish what research has been previously conducted. This helps in refining the research questions and making them more insightful. The literature review, definition of the purpose of the case study and early determination of the significance of the study for potential audience for the final report direct how the study will be designed, conducted and publicly reported.

3. Determining the unit of sampling and the number of units

Sampling Strategies in a Case Study

In a case study design, purposeful sampling is done which has been defined by Patton as ‘selecting information-rich cases for study in-depth’. A case study research, purposeful sampling is preferred over probability sampling as they enhance the usefulness of the information acquired from small samples. Purposive samples are expected to be conversant and informative about the phenomenon under investigation. A case study requires a plan for choosing sites and participants in order to start data collection. The plan is known as an ‘emergent design’ in which research decisions depend on preceding information. This necessitates purposive sampling, data collection and partial, simultaneous analysis of data as well as interactive rather than distinct sequential steps.

During the phase of designing a case study research, the researcher determines whether to use single or multiple real-life cases to examine in-depth and which instruments and data collection techniques to use. When multiple cases are used, each case is treated as a single case. Each case/s conclusions can then be used as contributing information to the entire study, but each case remains a single case for collecting data and analysis. Exemplary case studies carefully select cases and carefully examine the choices available from among many research tools available so as to enhance the validity of the study. Careful selection helps in determining boundaries around the case. The researcher must determine whether to study 'unique cases', or 'typical cases'. S/He also needs to decide whether to select cases from different geographical areas. It is necessary at this stage to keep in mind the goals of the study so as to identify and select relevant cases and evidence that will fulfill the goals of the study and answer the research questions raised. Selecting multiple or single cases is a key element, but a case study can include more than one unit of embedded analysis. For example, a case study may involve study of a single type of school (For example, Municipal School) and a school belonging to this type. This type of case study involves two levels of analysis and increases the complexity and amount of data to be gathered and analyzed. Multiple cases are often preferable to single cases, particularly when the cases may not be representative of the population from which they are drawn and when a range of behaviors/profiles, experiences, outcomes, or situations is desirable. However, including multiple cases limits the depth with which each case may be analyzed and also has implications for the structure and length of the final report.

4. **Identifying sources, tools and techniques of data collection**

Sources of Data in a Case Study

A case study method involves using multiple sources and techniques in the data collection process. The researcher determines in advance what evidence to collect and which techniques of data analysis to use so as to answer the research questions. Data collected is normally principally qualitative and soft data, but it may also be quantitative also. Data are collected from primary documents such as school records and databases, students' records, transcripts and results, field notes, self-reports or thinkaloud protocols and memoranda. Techniques used to collect data can include surveys, interviews, questionnaires, documentation review, observation and physical artifacts. These multiple tools and techniques of data collection add texture, depth, and multiple insights to an analysis and can enhance the validity or credibility of the results.

Case studies may make use of field notes and databases to categorize and reference data so that it is readily available for subsequent re-interpretation. Field notes record feelings and intuitive hunches, pose questions, and document the work in progress. They record testimonies, stories and illustrations which can be used in reporting the study. They may inform of impending preconceptions because of the detailed exposure of the client to special attention or give an early signal that a pattern is emerging. They assist in determining whether or not the investigation needs to be reformulated or redefined based on what is being observed. Field notes should be kept separate from the data being collected and stored for analysis.

According to Cohen and Manion, the researcher must use the chosen data collection tools and techniques systematically and properly in collecting the evidence. Observations and data collection settings may range from natural to artificial, with relatively unstructured to highly structured elicitation tasks and category systems depending on the purpose of the study and the disciplinary traditions associated with it.

5. Evaluating and Analyzing Data

The case study research generates a huge quantity of data from multiple sources. Hence systematic organization of the data is essential in prevent the researcher from losing sight of the original research purpose and questions. Advance preparation assists in handling huge quantity of largely soft data in a documented and systematic manner. Researchers prepare databases for categorizing, sorting, storing and retrieving data for analysis. The researcher examines raw data so as to find linkages between the research object and the outcomes with reference to the original research questions. Throughout the evaluation and analysis process, the researcher remains open to new opportunities and insights. The case study method, with its use of multiple data collection methods and analysis techniques, provides researchers with opportunities to triangulate data in order to strengthen the research findings and conclusions. According to Creswell, analysis of data in case study research usually involves an iterative, spiraling or cyclical process that proceeds from more general to more specific observations.

The strategies used in analysis require researchers to move beyond initial impressions to improve the likelihood of precise and consistent findings. Data need to be consciously sorted in many different ways to expose or create new insights and will deliberately look for contradictory data to disconfirm the analysis. Researchers categorize, tabulate and recombine data to answer the initial research questions and conduct cross-checking of facts and incongruities in accounts. Focused, short, repeated interviews may be essential to collect supplementary data to authenticate key observations or check a fact.

6. Report writing

Case studies report the data in a way that transforms a multifarious issue into one that can be understood, permitting the reader to question and examine the study and reach an understanding independent of the researcher. The objective of the written report is to depict a multifaceted problem in a way that conveys an explicit experience to the reader. Case studies should present data in a way that leads the reader to apply the experience in his or her own reallife situation. Researchers need to pay exacting consideration to displaying adequate evidence to achieve the reader‘s confidence that all avenues have been explored, clearly communicating the confines of the case and giving special attention to conflicting propositions.

In general, a research report in a case study should include the following aspects:

- A statement of the study's purpose and the theoretical context.
- The problem or issue being addressed.
- Central research questions.
- A detailed description of the case(s) and explanation of decisions related to sampling and selection.
- Context of the study and case history, where relevant. The research report should provide sufficient contextual information about the case, including relevant biographical and social information (depending on the focus), such as teaching - learning history, students‘

and teachers' background, years of studying/working in the institution, data collection site(s) or other relevant descriptive information pertaining to the case and situation.

- Issues of access to the site/participants and the relationship between you and the research participant (case).
- The duration of the study.
- Evidence that you obtained informed consent that the participants' identities and privacy are protected, and, ideally, that participants benefited in some way from taking part in the study.
- Methods of data collection and analysis, either manual or computer-based data management and analysis (see weitzman & miles, 1995), or other equipment and procedures used.
- Findings, which may take the form of major emergent themes, developmental stages, or an in-depth discussion of each case in relation to the research questions; and illustrative quotations or excerpts and sufficient amounts of other data to establish the validity and credibility of the analysis and interpretations.
- A discussion of factors that might have influenced the interpretation of data in undesired, unanticipated, or conflicting ways.

A consideration of the connection between the case study and larger theoretical and practical issues in the field is essential to report. The report could include a separate chapter handling each case separately or treating the case as a chronological recounting. Some researchers report the case study as a story. During the report preparation process, the researcher critically scrutinizes the report trying to identify ways of making it comprehensive and complete. The researcher could use representative audience groups to review and comment on the draft report. Based on the comments, the researcher could rewrite and revise the report. Some case study researchers suggest that the report review audience should include the participants of the study.

Strengths of Case Study Method

a. It involves detailed, holistic investigation of all aspects of the unit under study.
b. Case studies data are strong in reality.
c. It can utilize a wide range of measurement tools and techniques.
d. Data can be collected over a period of time and is contextual.
e. It enables the researcher to assess and document not just the empirical data but also how the subject or institution under study interacts with the larger social system.
f. Case study reports are often written in non-technical language and are therefore easily understood by laypersons.
g. They help in interpreting similar other cases.

Weaknesses of Case Study Method

a. The small sample size prevents the researcher from generalizing to larger populations.
b. The case study method has been criticized for use of a small number of cases can offer no grounds for establishing reliability or generality of findings.
c. The intense exposure to study of the case biases the findings.

d. It has also been criticized as being useful only as an exploratory tool.
e. They are often not easy to cross-check.

Yet researchers continue to use the case study research method with success in carefully planned and crafted studies of real-life situations, issues, and problems.

ETHNOGRAPHY

Ethnographic studies are usually holistic, founded on the idea that human beings are best understood in the fullest possible context, including the place where they live, the improvements they have made to that place, how they make a living and gather food, housing, energy and water for themselves, what their marriage customs are, what language(s) they speak and so on.

Ethnography is a form of research focusing on the sociology of meaning through close field observation of socio-cultural phenomena. Typically, the ethnographer focuses on a community (not necessarily geographic, considering also work, leisure, classroom or school groups and other communities). Ethnography may be approached from the point of view of art and cultural preservation and as a descriptive rather than analytic endeavor.

It essentially is a branch of social and cultural anthropology. The emphasis in ethnography is on studying an entire culture. The method starts with selection of a culture, review of the literature pertaining to the culture, and identification of variables of interest - typically variables perceived as significant by members of the culture. Ethnography is an enormously wide area with an immense diversity of practitioners and methods. However, the most common ethnographic approach is participant observation and unstructured interviewing as a part of field research. The ethnographer becomes immersed in the culture as an active participant and records extensive field notes. In an ethnographic study, there is no preset limit of what will be observed and interviewed and no real end point in as is the case with grounded theory.

Hammersley and Atkinson define ethnography as, "We see the term as referring primarily to a particular method or sets of methods. In its most characteristic form it involves the ethnographer participating, overtly or covertly, in people's lives for an extended period of time, watching what happens, listening to what is said, asking questions—in fact, collecting whatever data are available to throw light on the issues that are the focus of the research. Johnson defines ethnography as "a descriptive account of social life and culture in a particular social system based on detailed observations of what people actually do."

Assumptions in an Ethnographic Research

According to Garson, these are as follows

a. Ethnography assumes that the principal research interest is primarily affected by community cultural understandings. The methodology virtually assures that common cultural understandings will be identified for the research interest at hand. Interpretation is apt to place great emphasis on the causal importance of such cultural understandings. There is a possibility that an ethnographic focus will overestimate the role of cultural perceptions and underestimate the causal role of objective forces.
b. Ethnography assumes an ability to identify the relevant community of interest. In some settings, this can be difficult. Community, formal organization, informal group and individual-level perceptions may all play a causal role in the subject under study and the importance of these may vary by time, place and issue. There is a possibility that an ethnographic focus may overestimate the role of community culture and underestimate the causal role of individual psychological or of sub-community (or for that matter, extracommunity) forces.
c. Ethnography assumes that the researcher is capable of understanding the cultural mores of the population under study, has mastered the language or technical jargon of the culture and has based findings on comprehensive knowledge of the culture. There is a danger that the researcher may introduce bias toward perspectives of his or her own culture.
d. While not inherent to the method, cross-cultural ethnographic research runs the risk of falsely assuming that given measures have the same meaning across cultures.

Characteristics of Ethnographic Research

According to Hammersley and Sanders, ethnography is characterized by the following features

a. People's behaviour is studied in everyday contexts.
b. It is conducted in a natural setting.
c. Its goal is more likely to be exploratory rather than evaluative.

d. It is aimed at discovering the local person's or —native's point of view, wherein, the native may be a consumer or an end-user.
e. Data are gathered from a wide range of sources, but observation and/or relatively informal conversations are usually the principal ones.
f. The approach to data collection is unstructured in that it does not involve following through a predetermined detailed plan set up at the beginning of the study nor does it determine the categories that will be used for analyzing and interpreting the soft data obtained. This does not mean that the research is unsystematic. It simply means that initially the data are collected as raw form and a wide amount as feasible.
g. The focus is usually a single setting or group of a relatively small size. In life history research, the focus may even be a single individual.
h. The analysis of the data involves interpretation of the meanings and functions of human actions and mainly takes the form of verbal descriptions and explanations, with quantification and statistical analysis playing a subordinate role at most.
i. It is cyclic in nature concerning data collection and analysis. It is open to change and refinement throughout the process as new learning shapes future observations. As one type of data provides new information, this information may stimulate the researcher to look at another type of data or to elicit confirmation of an interpretation from another person who is part of the culture being studied.

Guidelines for Conducting Ethnography

Following are some broad guidelines for conducting fieldwork

1. Be descriptive in taking field notes. Avoid evaluations.
2. Collect a diversity of information from different perspectives.
3. Cross-validate and triangulate by collecting different kinds of data obtained using observations, interviews, programme documentation, recordings and photographs.
4. Capture participants' views of their own experiences in their own words. Use quotations to represent programme participants in their own terms.
5. Select key informants carefully. Draw on the wisdom of their informed perspectives, but keep in mind that their perspectives are limited.
6. Be conscious of and perceptive to the different stages of fieldwork.
    a. Build trust and rapport at the entry stage. Remember that the researcher-observer is also being observed and evaluated.
    b. Stay attentive and disciplined during the more routine middle-phase of fieldwork.
    c. Focus on pulling together a useful synthesis as fieldwork draws to a close.
    d. Be well-organized and meticulous in taking detailed field notes at all stages of fieldwork.
    e. Maintain an analytical perspective grounded in the purpose of the fieldwork: to conduct research while at the same time remaining involved in experiencing the observed setting as fully as possible.
    f. Distinguish clearly between description, interpretation and judgment.

g. Provide formative feedback carefully as part of the verification process of fieldwork. Observe its effect.
h. Include in your field notes and observations reports of your own experiences, thoughts and feelings. These are also field data. Fieldwork is a highly personal experience.

Techniques Used in Conducting Ethnography

These include the following

a. Listening to conversations and interviewing. The researcher needs to make notes or audio-record these.
b. Observing behaviour and its traces, making notes and mapping patterns of behaviour, sketching of relationship between people, taking photographs, video-recordings of daily life and activities and using digital technology and web cameras.

Stages in Conducting Ethnography

According to Spradley, following are the stages in conducting an ethnographic study

1. Selecting an ethnographic project.
2. Asking ethnographic questions and collecting ethnographic data.
3. Making an ethnographic record.
4. Analyzing ethnographic data and conducting more research as required.
5. Outlining and writing an ethnography

Steps of Conducting Ethnography

According to Spradley, ethnography is a non-linear research process but is rather, a cyclical process. As the researcher develops questions and uncovers answers, more questions emerge and the researcher must move through the steps again.

According to Spradley, following are the steps of conducting an ethnographic study

1. **Locating a social situation**

The scope of the topic may vary from the "micro-ethnography" of a "single-socialsituation" to "macro-ethnography" of a complex society. According to Hymes, there are three levels of ethnography including (i) "comprehensive ethnography" which documents an entire culture, (ii) the "topic-oriented ethnography" which looks at aspects of a culture and (iii) "hypothesis-oriented ethnography" which beings with an idea about why something happens in a culture. Suppose you want to conduct research on classroom environment. This step requires that you select a category of classroom environment and identify social and academic situations in which it is used.

2. **Collecting data**

There are four types of data collection used in ethnographic research, namely

a. watching or being part of a social context using participant and non-participant observation and noted in the form of one observer notes, logs, diaries, and so on
b. asking open and closed questions that cover identified topics using semi-structured interviews
c. asking open questions that enable a free development of conversation using unstructured interviews and
d. Using collected material such as published and unpublished documents, photographs, papers, videos and assorted artifacts, letters, books or reports.

3. **Doing participant observation**

Formulate open questions about the social situations under study. Malinowski opines that ethnographic research should begin with "foreshadowed problems". These problems are questions that researchers bring to a study and to which they keep an open eye but to which they are not enslaved. Collect examples of the classroom environment. Select research tools/techniques. Spradley provides a matrix of questions about cultural space, objects, acts activities, events, time, actors, goals and feelings that researchers can use when just starting the study.

4. Making an ethnographic record. Write descriptions of classroom environment and the situations in which it is used.
5. Making descriptive observations. Select method for doing analysis.
6. Making domain analysis. Discover themes within the data and apply existing theories, if any, as applicable. Domain analysis requires the researcher to first choose one semantic relationship such as "causes" or "classes". Second, you select a portion of your data and begin reading it and while doing so, fill out a domain analysis worksheet where you list all the terms that fit the semantic relationship you chose. Now formulate structural questions for each domain. Structural questions occur less frequently as compared to descriptive questions in normal conversation. Hence they require more framing.
7. Making focused observations.
8. Making a taxonomic analysis. Taxonomy is a scientific process of classifying things and arranging them in groups or a set of categories (domains) organized on a single semantic relationships. The researcher needs to test his taxonomies against data given by informants. Make comparisons of two or three symbols such as word, event, constructs.
9. Making selected observations.
10. Making a componential analysis which is a systematic search for the attributes or features of cultural symbols that distinguish them from others and give them meaning. The basic idea in componential analysis is that all items in a domain can be decomposed into combinations of semantic features which combine to give the item meaning.
11. Discovering cultural themes.

A theme is a postulate or position, explicit or implicit, which is directly or indirectly approved and promoted in a society. Strategies of discovering cultural themes include

a. in-depth study of culture
b. making a cultural inventory
c. identifying and analyzing components of all domains
d. Searching for common elements across all domains such as gender, age, SES groups etc.
e. identifying domains that clearly show a strong pattern of behavior
f. Making schema of cultural scene and
g. identifying generic (etic) codes usually functional such as social conflict, inequality, cultural contradictions in the institutional social system, strategies of social control, managing interpersonal relations, acquiring status in the institution and outside, solving educational and administrative problems and so on.

12. Taking a cultural inventory.
13. Writing an ethnography

Guidelines for Interviewing

According to Patton, following are some useful guidelines that can be used for effective interviewing

1. Throughout all phases of interviewing, from planning through data collection to analysis, keep centered on the purpose of the research endeavor. Let that purpose guide the interviewing process.
2. The fundamental principle of qualitative interviewing is to provide a framework within which respondents can express their own understandings in their own terms.
3. Understand the strengths and weaknesses of different types of interviews: the informal conversational interview; the interview guide approach; and the standardized open-ended interview.
4. Select the type of interview (or combination of types) that is most appropriate to the purposes of the research effort.
5. Understand the different kinds of information one can collect through interviews: behavioural data; opinions; feelings; knowledge; sensory data; and background information.
6. Think about and plan how these different kinds of questions can be most appropriately sequenced for each interview topic, including past, present, and future questions.
7. Ask truly open-ended questions.
8. Ask clear questions, using understandable and appropriate language.
9. Ask one question at a time.
10. Use probes and follow-up questions to solicit depth and detail.
11. Communicate clearly what information is desired, why that information is important, and let the interviewee know how the interview is progressing.

12. Listen attentively and respond appropriately to let the person know he or she is being heard.
13. Avoid leading questions.
14. Understand the difference between a depth interview and an interrogation. Qualitative evaluators conduct depth interviews; police investigators and tax auditors conduct interrogations.
15. Establish personal rapport and a sense of mutual interest.
16. Maintain neutrality toward the specific content of responses. You are there to collect information not to make judgments about that person.
17. Observe while interviewing. Be aware of and sensitive to how the person is affected by and responds to different questions.
18. Maintain control of the interview.
19. Tape record whenever possible to capture full and exact quotations for analysis and reporting.
20. Take notes to capture and highlight major points as the interview progresses.
21. As soon as possible after the interview check the recording for malfunctions; review notes for clarity; elaborate where necessary; and record observations.
22. Take whatever steps are appropriate and necessary to gather valid and reliable information.
23. Treat the person being interviewed with respect. Keep in mind that it is a privilege and responsibility to peer into another person's experience.

Writing Ethnographic Research Report

The components of an ethnographic research report should include the following

1. Purpose / Goals / Questions.
2. Research Philosophy.
3. Conceptual/Theoretical Framework
4. Research Design / Model.
5. Setting/Circumstances.
6. Sampling Procedures.
7. Background and Experience of Researcher.
8. Role/s of Researcher.
9. Data Collection Method.
10. Data Analysis/Interpretation.
11. Applications/Recommendations.
12. Presentation Format and Sequence.

Advantages of Ethnography

These are as follows:

a. It provides the researcher with a much more comprehensive perspective than other forms of research.

b. It is also appropriate to behaviours that are best understood by observing them within their natural environment (dynamics).

Disadvantages of Ethnography

These are as follows:

a. It is highly dependent on the researcher's observations and interpretations
b. There is no way to check the validity of the researcher's conclusion, since numerical data is rarely provided.
c. Observer bias is almost impossible to eliminate.
d. Generalizations are almost non-existent since only a single situation is observed, leaving ambiguity in the study.
e. It is very time consuming.

DOCUMENT ANALYSIS

Documentary Analysis is closely related to historical research since in such surveys we study the existing documents. But it is different from historical research in which our emphasis is on the study of the past; and in the descriptive research we emphasize on the study of the present. Descriptive research in the field of education may focus on describing the existing school practices, attendance rate of the students, health records, and so on.

The method of documentary analysis enables the researcher to include large amounts of textual information and systematically identify its properties. Documentary analysis today is a widely used research tool aimed at determining the presence of certain words or concepts within texts or sets of texts. Researchers quantify and analyze the presence, meanings and relationships of such words and concepts, then make inferences about the messages within the texts, the writer(s), the audience and even the culture and time of which these are a part.

Documentary analysis could be defined as a research technique for the objective, systematic, and quantitative description of manifest content of communications. It is a technique for making inferences by objectively and systematically identifying specified characteristics of messages. The technique of documentary analysis is not restricted to the domain of textual analysis, but may be applied to other areas such as coding student drawings or coding of actions observed in videotaped studies, analyzing past documents such as memos, minutes of the meetings, legal and policy statements and so on. In order to allow for replication, however, the technique can only be applied to data that are durable in nature. Texts in documentary analysis can be defined broadly as books, book chapters, essays, interviews, discussions, newspaper headlines and articles, historical documents, speeches, conversations, advertising, theater, informal conversation, or really any occurrence of communicative language. Texts in a single study may also represent a variety of different types of occurrences.

Documentary analysis enables researchers to sift through large amount of data with comparative ease in a systematic fashion. It can be a useful technique for allowing one to discover and describe the focus of individual, group, institutional or social attention. It also allows inferences to be made which can then be corroborated using other methods of data collection. Much documentary analysis research is motivated by the search for techniques to infer from symbolic data what would be too costly, no longer possible, or too obtrusive by the use of other techniques. These definitions illustrate that documentary analysis emphasizes an integrated view of speech/texts and their specific contexts. Document analysis is the systematic exploration of written documents or other artifacts such as films, videos and photographs. In pedagogic research, it is usually the contents of the artifacts, rather than say, the style or design, that are of interest.

Why analyze documents?

Documents are an essential element of day-to-day work in education. They include:

- Student essays
- Exam papers
- Minutes of meetings
- Module outlines
- Policy documents

In some pedagogic research, analysis of relevant documents will inform the investigation. If used to triangulate, or give another perspective on a research question, results of document analysis may complement or refute other data. For example, policy documents in an institution may be analyzed and interviews with staff or students and observation of classes may suggest whether or not new policies are being implemented. A set of data from documents, interviews and observations could contribute to a case study of a particular aspect of pedagogy.

How can documents be analyzed?

The content of documents can be explored in systematic ways which look at patterns and themes related to the research question(s). For example, in making a case study of deep and surface learning in a particular course, the question might be 'How has deep learning been encouraged in this course in the last three years?'

Minutes of course meetings could be examined to see whether or how much this issue has been discussed; Student handouts could be analyzed to see whether they are expressed in ways which might encourage deep learning. Together with other data-gathering activities such as student questionnaires or observation of classes, an action research study might then be based on an extended research question so that strategies are implemented to develop deep learning.

In the example of deep learning, perhaps the most obvious way to analyze the set of minutes would be to use a highlighting pen every time the term 'deep learning' was used. You might also choose to highlight 'surface learning' a term with an implied relationship to deep learning. You might also decide, either before starting the analysis, or after reading the documents, that there are other terms or inferences which imply an emphasis on deep learning. You might therefore go through the documents again, selecting additional references.

The levels of analysis will vary but a practitioner-researcher will need to be clear and explicit about the rationale for, and the approach to, selection of content.

Advantages and Disadvantages of document analysis

Robson (2002) points out the advantages and disadvantages of content analysis. An advantage is that documents are unobtrusive and can be used without imposing on participants; they can be checked and re-checked for reliability.

A major problem is that documents may not have been written for the same purposes as the research and therefore conclusions will not usually be possible from document analysis alone.

EXPERIMENTAL RESEARCH

An experiment is a scientific investigation in which the researcher manipulates one or more independent variables, controls any other relevant variables, and observes the effect of the manipulations on the dependent variable(s). An experimenter deliberately and systematically introduces change and then observes the consequences of that change. Only research problems that permit a researcher to manipulate conditions are appropriate for experimental research. The goal of experimental research is to determine whether a causal relationship exists between two or more variables. Because the experiment involves control and careful observation and measurement, this research method provides the most convincing evidence of the effect that one variable has on another. Experimental research is the description and analysis of what will be, or what will occur, under carefully controlled conditions.

Experimental research is the only type of research that can test hypotheses to establish cause–effect relations. It represents the strongest chain of reasoning about the links between variables. In experimental research the researcher manipulates at least one independent variable, controls other relevant variables, and observes the effect on one or more dependent variables. The researcher determines “who gets what“; that is, the researcher has control over the selection and assignment of groups to treatments.

The manipulation of an independent variable is the primary characteristic that differentiates experimental research from other types of research. The independent variable, also called the treatment, causal, or experimental variable, is that treatment or characteristic believed to make a difference. In educational research, independent variables that are frequently manipulated include method of instruction, type of reinforcement, arrangement of learning environment, type of learning materials, and length of treatment. This list is by no means exhaustive. The dependent variable, also called the criterion, effect, or posttest variable, is the outcome of the study, the change or difference in groups that occurs as a result of the independent variable. It gets its name because it is dependent on the independent variable. The dependent variable may be measured by a test or some other quantitative measure (e.g., attendance, number of suspensions, and time on task). The only restriction on the dependent variable is that it must represent a measurable outcome.

Experimental research is the most structured of all research types. When well conducted, experimental studies produce the soundest evidence concerning cause–effect relations. The results of experimental research permit prediction, but not the kind that is characteristic of cor-relational research. A cor-relational study predicts a particular score for a particular individual. Predictions based on experimental findings are more global and often take the form, “If you use Approach X, you will probably get different results than if you use Approach Y.” Of course, it is unusual for a single experimental study to produce broad generalization of results because any single study is limited in context and participants. However, replications of a study involving different contexts and participants often produce cause–effect results that can be generalized widely.

Characteristics of Experimental Method

There are four essential characteristics of experimental research

a. Control
b. Manipulation
c. Observation and
d. Replication

a. Control

Variables that are not of direct interest to the researcher, called extraneous variables, need to be controlled. Control refers to removing or minimizing the influence of such variables by several methods such as: randomization or random assignment of subjects to groups; matching subjects on extraneous variable(s) and then assigning subjects randomly to groups; making groups that are as homogenous as possible on extraneous variable(s); application of statistical technique of analysis of covariance (ANCOVA); balancing means and standard deviations of the groups.

b. Manipulation

Manipulation refers to a deliberate operation of the conditions by the researcher. In this process, a pre-determined set of conditions, called independent variable or experimental variable. It is also called treatment variable. Such variables are imposed on the subjects of experiment. In specific terms manipulation refers to deliberate operation of independent variable on the subjects of experimental group by the researcher to observe its effect. Sex, socio-economic status, intelligence, method of teaching, training or qualification of teacher, and classroom environment are the major independent variables in educational research. If the researcher, for example, wants to study the effect of 'X' method of teaching on the achievement of students in mathematics, the independent variable here is the method of teaching. The researcher in this experiment needs to manipulate 'X' i.e. the method of teaching. In other words, the researcher has to teach the experimental groups using 'X' method and see its effect on achievement.

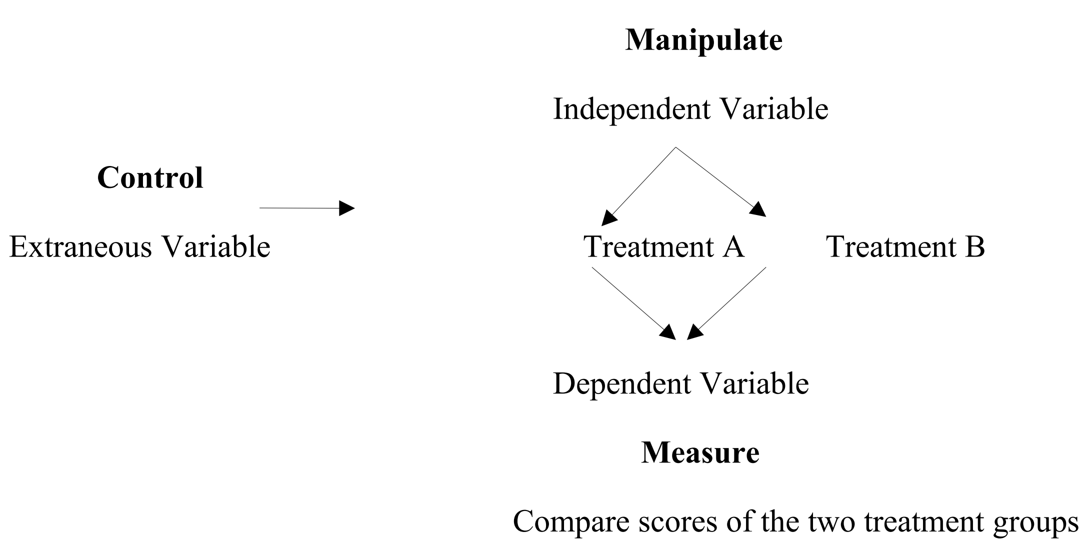
c. Observation

In experimental research, the experimenter observes the effect of the manipulation of the independent variable on dependent variable. The dependent variable, for example, may be performance or achievement in a task.

d. Replication

Replication is a matter of conducting a number of sub-experiments, instead of one experiment only, within the framework of the same experimental design. The researcher may make a multiple comparison of a number of cases of the control group and a number of cases of the experimental group. In some experimental situations, a number of control and experimental groups, each consisting of equivalent subjects, are combined within a single experiment.

Basic Elements of an Experiment



Experimental Comparison

An experiment begins with an experimental hypothesis, a prediction that the treatment will have a certain effect. The research hypothesis expresses expectations as to results from the changes introduced—that the treatment and no- treatment groups will differ because of the treatment's effects.

For the simplest experiment, you need two groups of subjects: the experimental group and the control group. The experimental group receives a specific treatment; the control group receives no treatment. Using a control group enables the researcher to discount many alternative explanations for the effect of treatment. For example, assume a researcher randomly assigns students to two groups. The experimental group receives a cash reward for successfully completing an object assembly task. If the experimental group does better on the task than the equivalent no- cash-reward group (control group), the researcher has evidence of the relationship between the cash reward and performance on the object assembly task.

More common than comparing a treatment group to a group receiving no treatment (true control group) is the situation in which researchers compare groups receiving different treatments. These are called comparison groups. The majority of educational experiments study the difference in the results of two or more treatments rather than the difference in the results of one treatment versus no treatment at all.

For example, it would be pointless to compare the spelling achievement of an experimental group taught by method A with a control group that had no spelling instruction at all. Instead, researchers compare groups receiving method A and method B treatments. Comparison of groups receiving different treatments provides the same control over alternative explanations, as does comparison of treated and untreated groups. To simplify subsequent discussions, we use the term control group to refer both to groups with no treatment and to groups with alternative treatments. Comparisons are essential in scientific investigation. Comparing a group receiving treatment with either an equivalent group receiving no treatment or an equivalent group or groups receiving alternative treatment makes it possible to draw well-founded conclusions from the results.

EXPERIMENTAL DESIGN

The term experimental design refers to the conceptual framework within which the experiment is conducted. The experimental design sets up the conditions required for demonstrating cause-and-effect relationships. These conditions are as follows: (1) Cause precedes effect in time, (2) the cause variable co varies (occurs together) with the effect, and (3) alternative explanations for the causal relationship can be ruled out.

Experimental designs differ in the number of independent variables that are manipulated. Some experimental designs have only one independent variable; other designs have two or more. Designs differ in the method of assigning subjects to different treatments. In randomized experiments, subjects are randomly assigned to the groups; in other cases, the design uses preexisting groups, or each subject may receive all the treatments. Designs also differ in how often dependent variable measures are made and whether all subjects receive all treatments or not.

An experimental design serves two functions:

a. It establishes the conditions for the comparisons required to test the hypotheses of the experiment
b. It enables the experimenter, through statistical analysis of the data, to make a meaningful interpretation of the results of the study.

The most important requirement is that the design must be appropriate for testing the previously stated hypotheses of the study. The mark of a sophisticated experiment is neither complexity nor simplicity but, rather, appropriateness. A design that will do the job it is supposed to do is the correct design. The hypothesis may state the expected effect of a single independent variable or the effect of two or more variables and the interaction among them. The experimenter's task is to select the design that best arranges the experimental conditions to test the stated hypotheses of the study.

A second requirement is that the design must provide adequate control so that the effects of the independent variable can be evaluated as unambiguously as possible. Unless the design controls extraneous variables, you can never be confident of the apparent relationship between the variables of the study. Randomization is the single best way to achieve the necessary control. Experimental studies utilizing randomization provide the best evidence for determining the effectiveness of educational practices and programs, and they are considered the gold standard for determining "what works" in educational research. Therefore, the best advice is to select a design that uses randomization in as many aspects as possible.

VALIDITY OF RESEARCH DESIGNS

Researchers must ask if the inferences drawn about the relationship between the variables of a study are valid or not. A very significant contribution to an understanding of the validity of experimental research designs was made by Campbell and Stanley (1963). They defined two general categories of validity of research designs: internal validity and external validity. Cook and Campbell (1979) elaborated this previous classification to four types of validity: internal validity, external validity, constructs validity, and statistical conclusion validity. For a more recent discussion of these four categories of experimental validity

According to Shadish, Cook, and Campbell (2002) following are the types of validity.

- Internal validity: The validity of the inferences about whether the effect of variable A (the treatment) on variable B (the outcome) reflects a causal relationship
- Statistical conclusion validity: The validity of the inferences about the co-variation between treatment and outcome

- Construct validity: The validity of the inferences about psychological constructs involved in the subjects, settings, treatments, and observations used in the experiment
- External validity: The validity of the inference about whether the cause–effect relationship holds up with other subjects, settings, and measurements

Any uncontrolled extraneous variables affecting performance on the dependent variable are threats to the validity of an experiment. An experiment is valid if results obtained are due only to the manipulated independent variable and if they are generalizable to individuals or contexts beyond the experimental setting. These two criteria are referred to, respectively, as the internal validity and external validity of an experiment.

Internal validity

Campbell and Stanley (1963) stated that internal validity is the basic requirement if one is to draw correct conclusions from an experiment. Internal validity refers to the inferences about whether the changes observed in a dependent variable are, in fact, caused by the independent variable(s) in a particular research study rather than by some extraneous factors. Internal validity is concerned with such questions as Did the experimental treatment cause the observed change in the dependent variable or was some spurious factor working? And Are the findings accurate? These questions of internal validity cannot be answered positively by the experimenter unless the design provides adequate control of extraneous variables.

If the design provides control of variables, you can eliminate alternative explanations of the observed outcome and interpret it as showing an intrinsic relationship between variables. Internal validity is essentially a problem of control. The design of appropriate controls is a matter of finding ways to eliminate extraneous variables that could lead to alternative interpretations and hence lower internal validity. Anything that contributes to the control of a design contributes to internal validity.

Campbell and Stanley (1963) identified eight extraneous variables that frequently represent threats to the internal validity of a research design. These variables are called threats because unless they are controlled, they may very well produce an effect that could be mistaken for the effect of the experimental treatment. If uncontrolled, these extraneous variables raise doubts about the accuracy of the experiment because they permit an alternative explanation of the experimental findings.

a. History

Specific events or conditions, other than the experimental treatment, may occur between the beginning of the treatment and the posttest measurement and may produce changes in the dependent variable. Such events are referred to as the history effect. In this case, history does not refer to past events but to extraneous events occurring at the same time that the experimental treatment is being applied and that could produce the observed outcome even without any treatment. These may be major political, economic, or cultural events or some rather minor disruptive factors that occur during the conduct of the experiment. The longer the period of time between the pre and post measurements on the subjects, the greater the history threat becomes.

Assume that a study was undertaken to investigate the use of films to change high school students' attitudes toward the use of selected drugs, including prescription drugs. The research plan called for a pretest to be administered to gauge initial attitudes, then a series of films to be shown over a 2-week period, and finally a posttest to determine what changes may have occurred. During those 2 weeks, the country was shocked by the tragic death of a talented, young movie actor whose death was believed to be due to an overdose of prescription drugs. The effect of the films on students' attitudes would be confounded with the effects of this tragic event so that the researcher would not know whether any change in attitudes occurred as a result of the films or because of the event and the subsequent extensive media coverage.

The effects of the unit and of the students' exposure to the media coverage are confounded, and it is impossible to know how much of the students' learning is caused by the unit and how much by events outside the experiment. In this case, history threatens the internal validity of the research.

In general, the use of a control group would eliminate the effects of history but only if both experimental and control groups are affected equally by the event. If one group was affected more than the other, internal validity would be threatened.

b. Maturation

The term maturation refers to changes (biological or psychological) that may occur within the subjects simply as a function of the passage of time. These changes threaten internal validity because they may produce effects that could mistakenly be attributed to the experimental treatment. Subjects may perform differently on the dependent variable measure simply because they are older, wiser, hungrier, more fatigued, or less motivated than they were at the time of the first measurements. Maturation is especially a threat in research on children because they are naturally changing so quickly. For example, it can be difficult to assess the effects of treatments for articulation problems among preschoolers because young children often naturally outgrow such problems. It has been difficult to assess the effects of compensatory programs such as Head Start on children's cognitive development because normal development ensures that children's cognitive skills will naturally improve over time.

c. Testing

Taking a test once may affect the subjects' performance when the test is taken again, regardless of any treatment. This is called the testing effect. In designs using a pretest, subjects may do better on the posttest because they have learned subject matter from a pretest, have become familiar with the format of the test and the testing environment, have developed a strategy for doing well on the test, or are less anxious about the test the second time. When an achievement test is used in the research, pretesting is a problem if the same form is used for both the pre- and posttest. We recommend using equivalent forms rather than the same test. Pretesting effects are less threatening in designs in which the interval between tests is large.

With attitude and personality inventories, taking a pretest may sensitize the subjects so that they think about the questions and issues raised and subsequently give different responses on the posttest (pretest sensitization). For example, assume a researcher administers an attitude scale toward an ethnic group, introduces a diversity awareness program, and then gives a posttest to determine whether attitudes changed. The attitude scale itself may stimulate subjects to think about their attitudes; this self-examination, rather than the program itself, may lead to improvements in attitudes.

d. Instrumentation

The instrumentation threat to internal validity is a result of a change in the instruments used during the study. The change in the way the dependent variable was measured from the first time to the second time, rather than the treatment, may bring about the observed outcome. Changes may involve the type of measuring instrument, the difficulty level, the scorers, the way the tests are administered, using different observers for pre and post measures, and so on. The best advice is to avoid any changes in the measuring instruments during a study. In classroom research, for example, a teacher should not use a multiple-choice pretest and an essay posttest, and the posttest should not be easier or more difficult than the pretest. Instrumentation is a problem in longitudinal research because the way measures are made may change over a period of time.

e. Statistical regression

The term statistical regression refers to the wellknown tendency for subjects who score extremely high or extremely low on a pretest to score closer to the mean (regression toward the mean) on a posttest. Statistical regression is a threat to internal validity when a subgroup is selected from a larger group on the basis of the subgroup's extreme scores (high or low) on a measure. When tested on subsequent measures, the subgroup will show a tendency to score less extremely on another measure, even a retest on the original measure. The subgroup will have a mean score closer to the mean of the original group.

You must always be aware of the effect of regression in designing your experiments. If dealing with extreme scores is an essential part of your research question, the best solution is to select a large group of extreme scorers and then randomly assign these individuals to two different treatments. Regression will occur equally for each group, so you can determine the effect of the treatment uncompounded with regression. You can reduce regression by using more reliable measures, which are less influenced by random error of measurement.

f. Selection bias

Selection is a threat when there are important differences between the experimental and control groups even before the experiment begins. A selection bias is a nonrandom factor that might influence the selection of subjects into the experimental or the control group. As a result, there is no assurance that the groups are equivalent. If they are not equivalent before the study, we cannot know whether any difference observed later is due to the treatment or to the pretreatment difference.

In a learning experiment, for example, if more capable students are in the experimental group than in the control group, the former would be expected to perform better on the dependent variable measure even without the experimental treatment. The best way to control selection bias is to use random assignment of subjects to groups. With random assignment, you cannot determine in advance who will be in each group; randomly assigned groups differ only by chance. We discuss random assignment later in this chapter.

Selection bias is most likely to occur when the researcher cannot assign subjects randomly but must use intact groups (quasi-experiment). An intact group is a preexisting group such as a class or a group set up independently of the planned experiment.

g. Experimental mortality (attrition)

The experimental mortality (attrition) threat occurs when there is differential loss of participants from the comparison groups. This differential loss may result in differences on the outcome measure even in the absence of treatment. If, for example, several of the lowest scorers on a pretest gradually drop out of the experimental group, the remaining subjects will have a higher mean performance on the final measure because the lowest scoring subjects are underrepresented when the posttest is administered. Attrition is not usually a serious threat unless the study goes on for a long time or unless the treatment is so demanding that it results in low-performing participants dropping out.

h. Selection–maturation interaction

Some of these threats may interact to affect internal validity. For example, selection and maturation may interact in such a way that the combination results in an effect on the dependent variable that is mistakenly attributed to the effect of the experimental treatment. Such interaction may occur in a quasi-experimental design in which the experimental and control groups are not randomly selected but instead are preexisting intact groups, such as classrooms. Although a pretest may indicate that the groups are equivalent at the beginning of the experiment, the experimental group may have a higher rate of maturation than the control group, and the increased rate of maturation accounts for the observed effect. If more rapidly maturing students are “selected” into the experimental group, the selection–maturation interaction may be mistaken for the effect of the experimental variable.

Although Campbell and Stanley (1963) originally listed only eight threats to internal validity, Cook and Campbell (1979) suggested that at least three more should be considered.

i. Experimenter effect

Experimenter effect refers to unintentional effects that the researcher has on the study. Personal characteristics of the researcher, such as gender, race, age, and position, can affect the performance of subjects. Sometimes the actual implementation of the experiment inadvertently gives the experimental group an unplanned advantage over the control group. For example, in an experiment comparing the effectiveness of two teaching methods, the more capable teacher may be assigned to the experimental group. Internal validity is threatened if the experimenter has expectations or a personal bias in favor of one method over another. These preferences and expectancies on the part of the experimenter may be unconsciously transmitted to subjects in such a way that their behavior is affected.

j. Subject effects

Subjects' attitudes developed in response to the research situation called subject effects can be a threat to internal validity. For instance, in a classic study of the effects of various levels of lighting on worker productivity at the Hawthorne, Illinois, plant of the Western Electric Company, researchers observed that both increases and decreases in light intensity resulted in increased productivity. The researchers concluded that the attention given to the employees and the employees' knowledge that they were participating in an experiment—rather than any changes in lighting—were the major factors leading to the production gains. This tendency for subjects to change their behavior just because of the attention gained from participating in an experiment has subsequently been referred to as the Hawthorne effect.

This effect can be a problem in educational research that compares exciting new teaching methods with conventional methods. Sometimes subjects may react to what they perceive to be the special demands of an experimental situation. That is, subjects react not as they normally might but as they think the more "important" researcher wants them to act.

The opposite of the Hawthorne effect is the John Henry effect. This effect, also called compensatory rivalry, refers to the tendency of control group subjects who know they are in an experiment to exert extra effort and hence to perform above their typical or expected average. They may perceive that they are in competition with the experimental group and they want to do just as well or better. Thus, the difference (or lack of difference) between the groups may be caused by the control subjects' increased motivation rather than by the experimental treatment. This effect is likely to occur in classroom research in which a new teaching technique is being compared to a conventional method that may be replaced by the new method.

Another subject effect, called compensatory demoralization, occurs when subjects believe they are receiving less desirable treatment or are being neglected. Consequently, they may become resentful or demoralized and put forth less effort than the members of the other group.

k. Diffusion

Diffusion occurs when participants in one group (typically the experimental group) communicate information about the treatment to subjects in the control group in such a way as to influence the latter's behavior on the dependent variable. Also, teachers involved with the experimental group may share information about methods and materials with teachers of the control group. Assume the subjects in the experimental group being taught math by an innovative method get so excited about the project that they share information with their friends in the control group. Later, the groups may perform similarly on the dependent variable not because the new method was ineffective but because its effects were disseminated to the control group as well. Deemphasizing the fact that an experiment is going on can lessen the likelihood of diffusion problems.

In summary, the preceding threats to internal validity represent specific reasons why a researcher's conclusions about a causal relationship between variables may be completely wrong. Researchers must systematically examine how each of the threats may have influenced the results of a study. If the threats can be ruled out, researchers can have more confidence that the observed results were caused by the different treatments.

Dealing with threats to internal validity

An experiment should be designed to avoid or at least minimize the effect of threats to internal validity. The researcher's first efforts must be directed toward controlling for any relevant preexisting differences between subjects in the comparison groups. Only in this way can you be fairly confident that any posttest experimental differences can be attributed to the experimental treatment rather than to preexisting subject differences. Six basic procedures are commonly used to control inter subject differences and increase equivalence among the groups that are to be exposed to the various experimental situations: (a) random assignment, (b) randomized matching, (c) homogeneous selection, (d) building variables into the design, (e) statistical control (analysis of covariance) and (f) use of subjects as their own controls.

(a) Random Assignment

The experimenter with an available supply of subjects has the task of dividing them into two groups that will be treated differently and then compared on the dependent variable. In assigning subjects to groups for the study, the experimenter needs a system that operates independently of personal judgment and of the characteristics of the subjects. For example, the known high scorers must not all be assigned to group A and the low scorers to group B. A system that satisfies this requirement is random assignment. Random assignment is the assignment of subjects to groups in such a way that for any given placement, every member of the population has an equal probability of being assigned to any of the groups. Chance alone determines whether subjects are placed in the experimental or the control group, thus eliminating selection bias. The term randomization is often used as a synonym for random assignment. Randomization is the most powerful method of control because only chance would cause the groups to be unequal with respect to any potential extraneous variables.

When subjects have been randomly assigned to groups, the groups can be considered statistically equivalent. Statistical equivalence does not mean the groups are absolutely equal, but it does mean that any difference between the groups is a function of chance alone and not a function of experimenter bias, subjects' choices, or any other factor. A subject with high aptitude is as likely to be assigned to treatment A as to treatment B. The same is true for a subject with low aptitude. For the entire sample, the effects of aptitude on the dependent variable will tend to balance or randomize out. In the same manner, subjects' differences in political viewpoints, temperament, achievement motivation, socioeconomic level, and other characteristics will tend to be approximately equally distributed between the two groups. The more subjects in the original sample, the more likely that randomization will result in approximately equivalent groups.

When random assignment has been employed, any pretreatment differences between groups are nonsystematic—that is, a function of chance alone. Because these differences fall within the field of expected statistical variation, the researcher can use inferential statistics to determine how likely it is that post treatment differences are due to chance alone.

(b) Randomized Matching

When random assignment is not feasible, researchers sometimes select pairs of individuals with identical or almost identical characteristics and randomly assign one member of the matched pair to treatment A and the other to treatment B. This procedure is called randomized matching. Note that randomized matching requires that the subjects be matched on relevant variables first and then randomly assigned to treatments. The researcher first decides what variables to use for matching. These may be IQ, mental age, socioeconomic status, age, gender, reading, pretest score, or other variables known to be related to the dependent variable of the study. If the groups are adequately matched on the selected variable(s), the resulting groups are reasonably equivalent. The major limitation of matching is that it is almost impossible to find subjects who match on more than one variable. Subjects are lost to the experiment when no match can be found for them. This loss, of course, reduces the sample size and introduces sampling bias into the study. Subjects for whom matches cannot be found are usually those with high or low scores. Therefore, these subjects would be underrepresented.

(c) Homogeneous Selection

Another method that can make groups reasonably comparable on an extraneous variable is to select samples that are as homogeneous as possible on that variable. This is called homogeneous selection. If the experimenter suspects that age is a variable that might affect the dependent variable, he or she would select only children of a particular age. By selecting only 6-year-old children, the experimenter would control for the effects of age as an extraneous independent variable. Similarly, if intelligence is likely to be a variable affecting the dependent variable of the study, then subjects would be selected from children whose IQ scores are within a restricted range—for example, 100 to 110. This procedure has thus controlled the effects of IQ. From this resulting homogeneous population, the experimenter randomly assigns individuals to groups and can assume that they are comparable on IQ. Beginning with a group that is homogeneous on the relevant variable eliminates the difficulty of trying to match subjects on that variable.

Although homogeneous selection is an effective way of controlling extraneous variables, it has the disadvantage of decreasing the extent to which the findings n be generalized to other populations. If a researcher investigates the effectiveness of a particular method with such a homogeneous sample, such as children with average IQs, the results could not be generalized to children in other IQ ranges. The effectiveness of the method with children of low intelligence or very high intelligence would not be known. As with matching, a true experiment requires that the subjects be selected first and then assigned randomly to treatments.

(d) Building Variables Into The Design

Some variables associated with the subjects can be built into the experimental design and thus controlled. For example, if you want to control gender in an experiment and you choose not to use the homogeneous selection technique just discussed, you could add gender as another independent variable. You would include both males and females in the study and then use analysis of variance to determine the effects of both gender and the main independent variable on the dependent variable. This method not only controls the extraneous gender variable but also yields information about its effect on the dependent variable, as well as its possible interaction with the other independent variable(s).

(e) Statistical Control

Analysis of covariance (ANCOVA) is a statistical technique used to control for the effect of an extraneous variable known to be correlated with the dependent variable. For example, consider an experiment to study the effects of two methods of teaching reading on reading achievement, the dependent variable. Subjects' reading ability before the experiment would be a variable that would certainly be related to the dependent variable of the study. You would expect that those who are good readers to begin with would score well on the reading posttest, whereas those who are poor readers would tend to score more poorly. After randomly assigning half of the subjects to method A and half to method B, you would administer a reading pretest to both groups. At the end of the experiment, ANCOVA would statistically adjust the mean reading posttest scores for any initial differences between the groups on the pretest. The ANCOVA technique removes the portion of each subject's posttest score that is in common with his or her pretest score. The resulting F value can then be checked for statistical significance. The variable used in ANCOVA to adjust scores (in this case, the reading pretest) is called the covariate. Using this technique, you are not considering a subject's posttest score per se. Instead, you analyze the difference between posttest scores and what you would expect the posttest score to be, given the score on the pretest and the correlation between pretest and posttest.

(f) Using subjects as their own controls

Still another procedure involves using subjects as their own controls—assigning the same subjects to all experimental conditions and then obtaining measurements of the subjects, first under one experimental treatment and then under another. For example, the same subjects might be required to learn two different lists of nonsense syllables—one list with high association value and the other with low association value. The difference in learning time between the two lists is found for each subject, and then the average difference in learning time for all subjects can be tested for significance. This method of control is efficient when feasible, but in some circumstances it cannot be used. In some types of studies, exposure to one experimental condition would make it impossible to use the subjects for the other experimental condition. You cannot, for example, teach children how to divide fractions one way and then erase their memory and teach it another way.

A useful strategy for this experiment would be to randomly divide the subjects into two groups—one group learning the high- association syllables first, and the other learning the low-association syllables first. This would "balance out" the effects of learning to learn or fatigue.

However, if learning high-association syllables first helps subjects to learn low-association syllables later and the reverse is not true, this can confound the interpretation of the results.

External validity

External validity refers to the extent to which the findings of a study can be generalized to other subjects, settings, and treatments. Any single study is necessarily performed on a particular group of subjects, with selected measuring instruments and under conditions that are in some respects unique. Yet researchers want the results of a study to furnish information about a larger realm of subjects, conditions, and operations than were actually investigated. To make generalizations from the observed to the unobserved, researchers need to assess how well the sample of events actually studied represents the larger population to which results are to be generalized. To the extent that the inferences about a causal relationship hold over changes in subjects, settings, and treatments, the experiment has external validity.

Threats to External Validity

a. Selection–treatment interaction (non-representativeness)

A major threat to external validity of experiments is the possibility of interaction between subject characteristics and treatment so that the results found for certain kinds of subjects may not hold for different subjects. This interaction occurs when the subjects in a study are not representative of the larger population to which one may want to generalize. For example, a researcher may conduct a study on the effectiveness of microcomputer-assisted instruction on the math achievement of junior high students. Classes available to the researcher (i.e., the accessible population) may represent an overall ability level at the lower end of the ability spectrum for all junior high students (i.e., the target population). If so, positive effect shown by the participants in the sample may be valid only for lower ability students, rather than for the target population of all junior high students. Similarly, if microcomputer-assisted instruction appears ineffective for this sample, it may still be effective for the target population.

Selection–treatment interaction, like the problem of differential selection of participants associated with internal validity, mainly occurs when participants are not randomly selected for treatments, but this threat can occur in designs involving randomization as well, and the way a given population becomes available to a researcher may threaten generalizability, no matter how internally valid an experiment may be.

b. Setting–Treatment Interaction (Artificiality)

Artificiality in the setting may limit the generalizability of the results. The findings of a contrived lab study of motivation may not be the same as one would obtain in a study conducted in a public school setting.

c. Pretest–Treatment Interaction

Using a pretest may increase or decrease the experimental subjects' sensitivity or responsiveness to the experimental variable and thus make the results obtained for this pretested population unrepresentative of effects of the experimental variable on the un-pretested population from which the experimental subjects are selected. In this case, you could generalize to pretested groups but not to un-pretested ones. Assume that you give a group of seventh-graders a questionnaire concerning their dietary habits and randomly divide the group into experimental and control groups. You expose the experimental group to a series of film presentations concerning good eating habits, whereas the control group views a series of health films unrelated to eating habits (placebo). The dependent variable is derived by observing the children's food selections in an actual free-choice situation. If the experimental group shows a significantly greater preference for healthful foods, you would like to conclude that the films are effective. Before reaching a conclusion, you must consider the possibility that the pretest caused the students to think about their eating habits and "set them up" to respond to the films. The same effect might not have been observed in an un-pretested group.

d. Multiple-Treatment Interference

Sometimes the same research participants receive more than one treatment in succession. Multipletreatment interference occurs when carryover effects from an earlier treatment make it difficult to assess the effectiveness of a later treatment. For example, suppose you were interested in comparing two different approaches to improving classroom behavior, behavior modification and corporal punishment. For 2 months, behavior modification techniques were systematically applied to the participants, and at the end of this period you found behavior to be significantly better than before the study began. For the next 2 months, the same participants were physically punished (with hand slap pings, spankings, and the like) whenever they misbehaved, and at the end of the 2 months behavior was equally as good as after the 2 months of behavior modification. Could you then conclude that behavior modification and corporal punishment are equally effective methods of behavior control? Certainly not. In fact, the goal of behavior modification is to produce self-maintaining behavior—that is, behavior that continues after direct intervention is stopped. The good behavior exhibited by the participants at the end of the corporal punishment period could well be due to the effectiveness of previous exposure to behavior modification; this good behavior could exist in spite of, rather than because of, exposure to corporal punishment. If it is not possible to select a design in which each group receives only one treatment, the researcher should try to minimize potential multiple-treatment interference by allowing sufficient time to elapse between treatments and by investigating distinctly different types of independent variables.

e. Specificity of Variables

Any given study has specificity of variables; that is, the study is conducted with a specific kind of participant, using specific measuring instruments, at a specific time, and under a specific set of circumstances. We have discussed the need to describe research procedures in sufficient detail to permit another researcher to replicate the study. Such detailed descriptions also permit interested readers to assess how applicable findings are to their situations. When studies that supposedly manipulated the same independent variable get quite different results, it is often difficult to determine the reasons for the differences because researchers have not provided clear, operational descriptions of their independent variables. When operational descriptions are available, they often reveal that two independent variables with the same name were defined quite differently in the separate studies. Because such terms as discovery method, whole language, and computer-based instruction mean different things to different people, it is impossible to know what a researcher means by these terms unless they are clearly defined. Generalizability of results is also tied to the clear definition of the dependent variable, although in most cases the dependent variable is clearly operationalized as performance on a specific measure. When aresearcher has a choice of measures to select from, he or she should address the comparability of these instruments and the potential limits on generalizability arising from their use.

Generalizability of results may also be affected by short- or long-term events that occur while the study is taking place. This threat is referred to as the interaction of history and treatment effects and describes the situation in which events extraneous to the study alter the research results. Short-term, emotion-packed events, such as the firing of a superintendent, the release of district test scores, or the impeachment of a president may affect the behavior of participants. Usually, however, the researcher is aware of such happenings and can assess their possible impact on results, and accounts of such events should be included in the research report. The impact of long-term events, such as wars and economic depressions, however, is more subtle and tougher to evaluate.

To deal with the threats associated with specificity, the researcher must operationally define variables in a way that has meaning outside the experimental setting and must be careful in stating conclusions and generalization.

f. Treatment Diffusion

Treatment diffusion occurs when different treatment groups communicate with and learn from each other. When participants in one treatment group know about the treatment received by a different group, they often borrow aspects from that treatment; when such borrowing occurs; the study no longer has two distinctly different treatments but rather has two overlapping ones. The integrity of each treatment is diffused. Often, the more desirable treatment—the experimental treatment or the treatment with additional resources— is diffused into the less desirable treatment.

g. Subject Effects

Attitudes and feelings of the participants that develop during a study may influence the generalizability of the findings to other settings. This threat is also called the reactive effect because subjects are reacting to the experience of participating in an experiment. Subjects' knowledge that they have been selected for an experiment and are being treated in a special way may affect the way they respond to the treatment. Thus, the treatment could appear to be more effective than it might be in the long term. This effect weakens generalization to situations in which people do not regard themselves as special. Closely related is a novelty effect that may happen in research that compares groups using innovative new methods to untreated control groups. A new instructional method may appear to be successful because it leads to excitement and enthusiasm among subjects that may affect the application of results to other groups. Likewise, the John Henry effect may occur when subjects in the untreated control group are determined to do as well as or better than the subjects in the experimental group. The teachers in the control group may feel threatened and may exert extra effort so that they and their students will not look bad compared to the experimental group.

h. Experimenter Effects

Another threat to external validity is the experimenter effect, which occurs when the experimenter consciously or unconsciously provides cues to subjects that influence their performance. The results of the study could be specific to an experimenter with a certain personality or other characteristics. Sometimes the presence of observers during an experiment may so alter the normal responses of the participating subjects that the findings from one group may not be valid for another group or for the broader population, and it would be hazardous to generalize the findings.

DEALING WITH THREATS TO EXTERNAL VALIDITY

Controlling the threats to external validity is not as straightforward as with internal validity. With the latter, the research design is the significant factor. Before you can assume external validity, you need to examine carefully and logically the similarities and differences between the experimental setting and the target setting with respect to subjects and treatments. A review of the literature would reveal if other research on the same question had used different kinds of subjects, settings, or methodology.

The following suggestions can help control threats to external validity:-

- Randomly sample the target population to select subjects for the study and then randomly assign them to treatment groups. If this is not possible because of the population size, then select subjects randomly from the experimentally accessible population and show the similarity of the experimentally accessible population and the target population.
- Identify the relevant characteristics of subjects in the target population, and determine the impact of these characteristics by incorporating them into the research study. For example, if you want to generalize to ethnically diverse urban high schools, you could include different ethnic groups in the study and examine the performance of each group separately to determine if the experimental treatment worked equally well with all groups or if there were differences. This kind of information would help determine the groups of students to whom the results could be generalized. The same could be done with gender, age, educational levels, and other characteristics. Factorial designs enable researchers to assess the effectiveness of the treatment at different levels of other variables such as race and gender.
- By controlling the problems arising from a pretest–treatment interaction by choosing a design that does not use a pretest.

- Replicate the research study in a new setting. This is a good way to determine if similar results will be found. If you find the same results with other populations and in other settings, you can have reasonable confidence that generalizations are valid.

EXPERIMENTAL RESEARCH DESIGNS

An experimental design is the general plan for carrying out a study with an active independent variable. The design is important because it determines the study's internal validity, which is the ability to reach valid conclusions about the effect of the experimental treatment on the dependent variable. Designs differ in their efficiency and their demands in terms of time and resources, but the major difference is in how effectively they rule out threats to internal validity. Obviously, one first chooses the design that is appropriate for testing the hypothesis of the study. From the appropriate designs, one must choose the one that will (a) ensure that the subjects assigned to the treatment and control groups do not differ systematically on any variables except those under consideration and (b) ensure that the outcome is a consequence of the manipulation of the independent variable and not of extraneous variables.

GROUP EXPERIMENTAL DESIGNS

The validity of an experiment is a direct function of the degree to which extraneous variables are controlled. If such variables are not controlled, it is difficult to interpret the results of a study and the groups to which results can be generalized. The term confounded is sometimes used to describe a situation in which the effects of the independent variable are so intertwined with those of extraneous variables that it becomes difficult to determine the unique effects of each. Experimental design strives to reduce this problem by controlling extraneous variables. Good designs control many sources that affect validity; poor designs control few. Two types of extraneous variables in need of control are participant variables and environmental variables. Participant variables include both organism variables and intervening variables. Organism variables are characteristics of the participants that cannot be altered but can be controlled for; the gender of a participant is an example. Intervening variables intrude between the independent and the dependent variable and cannot be directly observed but can be controlled for; anxiety and boredom are examples.

Control of Extraneous Variables

Randomization is the best way to control for many extraneous variables simultaneously; this procedure is effective in creating equivalent, representative groups that are essentially the same on all relevant variables. The underlying rationale for randomization is that if subjects are assigned at random (by chance) to groups, there is no reason to believe that the groups will be greatly different in any systematic way. In other words, they should be about the same on participant variables such as ability, gender, or prior experience, and on environmental variables as well. If the groups are the same at the start of the study and if the independent variable makes no difference, the groups should perform essentially the same on the dependent variable. On the other hand, if the groups are the same at the start of the study but perform differently after treatment, the difference can be attributed to the independent variable.

The use of randomly formed treatment groups is a unique characteristic of experimental research; this control factor is not possible with causal–comparative research. Thus, randomization is used whenever possible— participants are randomly selected from a population and randomly assigned to treatment groups.

Matching

Matching is a technique for equating groups on one or more variables, usually ones highly related to performance on the dependent variable. The most commonly used approach to matching involves random assignment of pairs, one participant to each group. In other words, the researcher attempts to find pairs of participants similar on the variable or variables to be controlled. If the researcher is matching on gender, obviously the matched pairs must be of the same gender. If the researcher is matching on variables such as pretest, ability scores, the pairing can be based on similarity of scores. Unless the number of participants is very large, it is unreasonable to try to make exact matches or matches based on more than one or two variables.

Once a matched pair is identified, one member of the pair is randomly assigned to one treatment group and the other member to the other treatment group. A participant who does not have a suitable match is excluded from the study. The resulting matched groups are identical or very similar with respect to the variable being controlled.

A major problem with such matching is that invariably some participants will not have a match and must be eliminated from the study. One way to combat loss of participants is to match less stringently. For example, the researcher may decide that if two ability test scores are within 20 points, they constitute an acceptable match. This approach may increase the number of subjects, but it can defeat the purpose of matching if the criteria for a match are too broad.

A related matching procedure is to rank all the participants from highest to lowest, based on their scores on the variable to be matched. The two highest ranking participants, regardless of raw score, are the first pair. One member of the first pair is randomly assigned to one group and the other member to the other group. The next two highest ranked participants (i.e., third and fourth ranked) are the second pair, and so on. The major advantage of this approach is that no participants are lost. The major disadvantage is that it is a lot less precise than pair-wise matching.

CLASSIFYING EXPERIMENTAL DESIGNS

Experimental designs may be classified according to the number of independent variables:

a. Single-variable designs and factorial designs
    - A single-variable design has one manipulated independent variable
    - Factorial designs have two or more independent variables, at least one of which is manipulated.

b. Experimental designs may also be classified according to how well they provide control of the threats to internal validity: pre-experimental, true experimental, and quasiexperimental designs.
   - Pre-experimental designs do not have random assignment of subjects to groups or other strategies to control extraneous variables.
   - True experimental designs (also called randomized designs) use randomization and provide maximum control of extraneous variables.
   - Quasi-experimental designs lack randomization but employ other strategies to provide some control over extraneous variables. They are used, for instance, when intact classrooms are used as the experimental and control groups.

Thus, true experimental designs have the greatest internal validity, quasi-experimental designs have somewhat less internal validity, and the pre-experimental designs have the least internal validity.

Followings are the terms and symbols that will use in different experimental designs

E – Experimental group

C – Control group

X – Independent variable

Y – Dependent variable

R – Random assignment of subjects to groups

Y_1 – Dependent variable measures taken before experiment / treatment (pre-test)

Y_2 – Dependent variable measures taken after experiment/ treatment (Post-test)

M_r – Matching subjects and then random assignment to groups.

In the paradigms for the various designs, the Xs and Ys across a given row are applied to the same people.

The left-to-right dimension indicates the temporal order, and the Xs and Ys vertical to one another are given simultaneously. A dash (—) indicates that the control group does not receive the X treatment or receives an alternative treatment.

PRE-EXPERIMENTAL DESIGNS

Following are two designs that are classified as pre-experimental because they provide little or no control of extraneous variables.

Design 1: One-Group Pretest–Posttest Design

The one-group pretest–posttest design usually involves three steps: (a) administering a pretest measuring the dependent variable; (b) applying the experimental treatment X to the subjects; and

(c) administering a posttest, again measuring the dependent variable. Differences attributed to application of the experimental treatment are then evaluated by comparing the pretest and posttest scores.

Design 1: One-Group Pretest–Posttest Design

Pretest	Independent	Posttest
Y_1	X	Y_2

To illustrate the use of this design, assume that an elementary teacher wants to evaluate the effectiveness of a new technique for teaching fourth-grade math. At the beginning of the school year, the students are given a standardized test (pretest) that appears to be a good measure of the achievement of the objectives of fourth-grade math. The teacher then introduces the new teaching technique and at the end of the semester administers the same standardized test (posttest), comparing students' scores from the pretest and posttest in order to determine if exposure to the new teaching technique made any difference. The limitation of this design is that because no control group is used, the experimenter cannot assume that any improvement in scores is due to the new technique.

Two obvious extraneous variables not controlled in this design are history and maturation. Things happen between pretest and posttest, other than the experimental treatment, that could affect learning. In the math example, widespread media interest in math education, increased emphasis on math in the school, or the introduction of a particularly effective teacher could increase student achievement in this area. Or an epidemic causing increased absences could depress achievement. Between pretest and posttest, children are growing mentally and physically, and they may have learning experiences that could affect their achievement. History and maturation become more threatening to internal validity as the time between pre- and posttest increases. Instrumentation and regression also present uncontrolled threats to internal validity of this design.

Another weakness is that Design 1 affords no way to assess the effect of the pretest. We know there is a practice effect when subjects take a test a second time or even take an alternate form of the test—or they may learn something just from taking the test and will do better the second time. To deal with this problem, some researchers have used Design 1 without the pretest. However, eliminating the pretest would only make a poor design worse.

The best advice is to avoid using Design 1. Without a control group to make a comparison possible, the results obtained in a one-group design are basically uninterruptable.

Design 2: Static Group Comparison

The static group comparison uses two or more preexisting or intact (static) groups, only one of which is exposed to the experimental treatment. Although this design uses two groups for comparison, it is flawed because the subjects are not randomly assigned to the groups and no pretest is used. The researcher makes the assumption that the groups are equivalent in all relevant aspects before the study begins and that they differ only in their exposure to X. To attempt to assess the effects of the X treatment, the researcher compares the groups on the dependent variable measure.

Design 2: Static Group Comparison

Group	Independent Variable	Posttest
E	X	Y_2
C	—	Y_2

Although this design has sometimes been used in educational research, it is basically worthless. Because neither randomization nor even matching on a pretest is used, we cannot assume that the groups are equivalent prior to the experimental treatment. Because of the possibility of initial differences between the groups, one could not conclude that the outcome is a result of the experimental treatment. In addition to selection bias, maturation and mortality are threats to the internal validity of this design.

TRUE EXPERIMENTAL DESIGNS

The designs in this category are called true experiments because subjects are randomly assigned to groups. Because of the control they provide, they are the most highly recommended designs for experimentation in education.

Design 3: Randomized Subjects, Posttest-Only Control Group Design

Randomized subjects, posttest-only control group design is one of the simplest yet one of the most powerful of all experimental designs. It has the two essential elements necessary for maximum control of the threats to internal validity: randomization and a control group. No pretest is used; the randomization controls for all possible extraneous variables and ensures that any initial differences between the groups are attributable only to chance and therefore will follow the laws of probability. After the subjects are randomly assigned to groups, only the experimental group is exposed to the treatment. In all other respects, the two groups are treated alike. Members of both groups are then measured on the dependent variable Y_2, and the scores are compared to determine the effect of X. If the obtained means of the two groups differ significantly (i.e., more than would be expected on the basis of chance alone), the experimenter can be reasonably confident that the experimental treatment is responsible for the observed result.

Design 3: Randomized Subjects, Posttest-Only Control Group Design

	Group	Independent Variable	Posttest
(R)	E	X	Y_2
(R)	C	—	Y_2

The main advantage of Design 3 is randomization, which ensures statistical equivalence of the groups before introduction of the independent variable. Recall that as the number of subjects increases, the likelihood that randomization will produce equivalent groups' increases. We recommend at least 30 subjects in each group. Design 3 controls for the main effects of history, maturation, regression, and pretesting; because no pretest is used, there can be no interaction effect of pretest and X. Thus, this design is especially recommended for research on changing attitudes. It is also useful in studies in which a pretest is either not available or not appropriate, such as in studies with kindergarten or primary grades, where it is impossible to administer a pretest because the learning is not yet manifest. Another advantage of this design is that it can be extended to include more than two groups if necessary. Possible threats to internal validity are subject effects and experimenter effects.

Design 3 does not permit the investigator to assess change. If such an assessment is desired, then a design that uses both a pretest and a posttest should be chosen. Because of the lack of a pretest, mortality could be a threat. Without having pretest information, preferably on the same dependent variable used as the posttest, the researcher has no way of knowing if those who dropped out of the study were different from those who continued (Shadish, Cook, & Campbell, 2002).

Design 4: Randomized Matched Subjects, Posttest-Only Control Group Design

Randomized matched subjects, posttest-only control group design is similar to Design 3, except that it uses a matching technique to form equivalent groups. Subjects are matched on one or more variables that can be measured conveniently, such as IQ or reading score. Of course, the matching variables used are those that presumably have a significant correlation with the dependent variable. Although a pretest is not included in Design 4, if pretest scores on the dependent variable are available, they could be used very effectively for the matching procedure. The measures are paired so that opposite members' scores are as close together as possible. The flip of a coin can be used to assign one member of each pair to the treatment group and the other to the control group.

Design 4: Randomized Matched Subjects, Posttest-Only Control Group Design

	Group	Independent Variable	Posttest
(M_r)	E	X	Y_2
	C	—	Y_2

Matching is most useful in studies in which small samples are to be used and where Design 3 is not appropriate. Design 3 depends completely on random assignment to obtain equivalent groups. With small samples the influence of chance alone may result in a situation in which random groups are initially very different from each other. Design 3 provides no assurance that small groups are really comparable before the treatments are applied. The matched-subjects design, however, serves to reduce the extent to which experimental differences can be accounted for by initial differences between the groups; that is, it controls preexisting inter-subject differences on variables highly related to the dependent variable that the experiment is designed to affect. The random procedure used to assign the matched pairs to groups adds to the strength of this design.

Design 4 is subject to the difficulties mentioned previously in connection with matching as a means of control. The matching of all potential subjects must be complete, and the members of each pair must be assigned randomly to the groups. If one or more subjects were excluded because an appropriate match could not be found, this would bias the sample. When using Design 4, it is essential to match every subject, even if only approximately, before random assignment. Design 4 can be used with more than two groups by creating matched sets and randomly assigning one member of each set to each group.

Design 5: Randomized Subjects, Pretest–Posttest Control Group Design

Design 5 is one of the most widely used true (randomized) experiments. In the randomized subjects, pretest–posttest control group design, one randomly assigns subjects to the experimental and control groups and administers a pretest on the dependent variable Y. The treatment is introduced only to the experimental subjects (unless two different treatments are being compared), after which the two groups are measured on the dependent variable. The researcher then compares the two groups' scores on the posttest. If there are no differences between the groups on the posttest, the researcher can then look at the average change between pretest and posttest (Y2−Y1) scores for each group to determine if the treatment produced a greater change (gain) than the control situation. The significance of the difference in the average pretest–posttest change for the two groups could be determined by a t test or F test. For reasons beyond the scope of this discussion, measurement experts have pointed out that technical problems arise when comparing gain scores.

Design 5: Randomized Subjects, Pretest–Posttest Control Group Design

	Group	Pretest	Independent Variable	Posttest
(R)	E	Y_1	X	Y_2
(R)	C	Y_1	—	Y_2

The recommended statistical procedure to use with Design 5 is an analysis of covariance (ANCOVA) with posttest scores as the dependent variable and pretest scores as the covariate to control for initial differences on the pretest. ANCOVA is a more powerful test and gives more interpretable results than does the comparison of gain scores for the two groups.

The main strength of this design is the initial randomization, which ensures statistical equivalence between the groups prior to experimentation; also, the fact that the experimenter has control of the pretest can provide an additional check on the equality of the two groups on the pretest, Y_1. Design 5 thus controls most of the extraneous variables that pose a threat to internal validity. Differential selection of subjects and statistical regression are also controlled through the randomization procedure.

The main concern in using Design 5 is external validity. Ironically, the problem stems from the use of the pretest, an essential feature of the design. As mentioned previously, there may be an interaction between the pretest and the treatment so that the results are generalizable only to other pretested groups. The responses to the posttest may not be representative of how individuals would respond if they had not been given a pretest.

Despite this shortcoming, Design 5 is widely used because the interaction between pretest and treatment is not a serious problem in most educational research. The pretests used are often achievement tests of some type and therefore do not significantly sensitize subjects who are accustomed to such testing. However, if the testing procedures are somewhat novel or motivating in their effect, then it is recommended that the experimenter choose a design not involving a pretest. Alternatively, whenever you suspect that the effect of the pretest might be reactive, it is possible to add a new group or groups to the study—a group that is not pretested. Solomon (1949) suggested two designs that overcome the weakness of Design 5 by adding an un-pretested group or groups.

Following are the brief presentation of these two designs

Design 6: Solomon Three-Group Design

The first of the Solomon designs uses three groups, with random assignment of subjects to groups. Note that the first two lines of this design are identical to Design 5. However, the Solomon three-group design has the advantage of employing a second control group labeled C_2 that is not pretested but is exposed to the treatment X. This group, despite receiving the experimental treatment, is functioning as a control and is thus labeled control group. This design overcomes the difficulty inherent in Design 5—namely, the interactive effect of pretesting and the experimental treatment. The posttest scores for the three groups are compared to assess the interaction effect.

Design 6: Solomon Three-Group Design

	Group	Pretest	Independent	Posttest
(R)	E	Y_1	X	Y_2
(R)	C_1	Y_1	—	Y_2
(R)	C_2	—	X	Y_2

If the experimental group has a significantly higher mean on the posttest Y_2 than does the first control group (C_1), the researcher cannot be confident that this difference is caused by X. It might have occurred because of the subjects' increased sensitization after the pretest and the interaction of their sensitization and X. However, if the post-test mean (Y_2) of the second control group (C_2) is also significantly higher than that of the first control group, then one can conclude that the experimental treatment, rather than the pretest–X interaction effect, has produced the difference because the second control group is not pretested.

Design 7: Solomon Four-Group Design

The Solomon four-group design provides still more rigorous control by extending Design 6 to include one more control group that receives neither pretest nor treatment.

Design 7: Solomon Four-Group Design

	Group	Pretest	Independent Variable	Posttest
(R)	E	Y_1	X	Y_2
(R)	C_1	Y_1	—	Y_2
(R)	C_2	—	X	Y_2
(R)	C_3	—	—	Y_2

Design 7, with its four groups, has strength because it incorporates the advantages of several other designs. It provides good control of the threats to internal validity. Design 7 has two pretested groups and two without a pretest; one of the pretested groups and one of the non-pretested groups receive the experimental treatment, and then all four groups take the posttest. The first two lines (as in Design 5) control extraneous factors such as history and maturation, and the third line (as in Design 6) controls the pretest–X interaction effect. When the fourth line is added to make Design 7, researcher has control over any possible contemporary effects that may occur between Y_1 and Y_2.

In Design 7, you can make several comparisons to determine the effect of the experimental X treatment. If the posttest mean of the E group is significantly greater than the mean of the first control group, C_1, and if the C2 post-test mean is significantly greater than that of C_3, you have evidence for the effectiveness of the experimental treatment. You can determine the influence of the experimental conditions on a pretested group by comparing the posttests of E and C_1 or the pre–post changes of E and C_1. You can find the effect of the experiment on an un-pretested group by comparing C_2 and C_3. If the average differences between posttest scores, $E - C_1$ and $C_2 - C_3$, are approximately the same, then the experiment must have had a comparable effect on pretested and un-pretested groups.

Design 7 actually involves conducting two experiments, one with pretests and one without pretests. If the results of these two experiments agree, as indicated previously, the investigator can have much greater confidence in the findings. The main disadvantage of this design is the difficulty involved in carrying it out in a practical situation. More time and effort are required to conduct two experiments simultaneously, and there is the problem of locating the increased number of subjects of the same kind that would be needed for the four groups.

FACTORIAL DESIGNS

The designs presented thus far have been the classical single-variable designs in which the experimenter manipulates one independent variable X to determine its effect on a dependent variable Y. However, in complex social phenomena several variables often interact simultaneously, and restricting a study to one independent variable may impose an artificial simplicity on a complex situation. The X variable alone may not produce the same effect as it might in interaction with another, so the findings from one-variable designs may be misleading. For instance, we might ask about the effectiveness of a particular method of teaching on students' learning. The answer may well be that the effectiveness depends on a number of variables, such as the age and ability level of the students, the personality of the teacher, the subject matter, and so on. Computer-assisted instruction, for example, may be more effective with below-average students than with bright ones. A classical one-variable design would not reveal this interactive effect of method and intelligence level. The information yield of an experiment can be markedly increased by using a factorial design. A factorial design is one in which the researcher manipulates two or more variables simultaneously in order to study the independent effect of each variable on the dependent variable, as well as the effects caused by interactions among the several variables. Some have said that the real breakthrough in educational research came with Fisher's (1925) development of factorial designs.

The independent variables in factorial designs are referred to as factors. Factors might be categorical variables such as gender, ethnicity, social class, and type of school, or they might be continuous variables such as aptitude or achievement. The researcher identifies the levels of each of these factors to be investigated. For example, aptitude might have two levels (high and low) or three levels (high, average, and low). Gender would have two levels (male and female), as would method of instruction (lecture and discussion).

Design 8: Simple Factorial Design

Factorial designs have been developed at varying levels of complexity. The simplest factorial design is the 2 × 2, which is read as "2 by 2." This design has two factors, and each factor has two levels.

Design 8: Simple Factorial Design

Variable 2 (X_2)	Variable 1 (X_1)	
	Treatment A	Treatment B
Level 1	Cell 1	Cell 3
Level 2	Cell 2	Cell 4

To illustrate, let us assume that an experimenter is interested in comparing the effectiveness of two types of teaching methods—methods A and B—on the achievement of ninth-grade science students, believing there may be a differential effect of these methods based on the students' level of science aptitude. Following table shows the 2 × 2 factorial design. The aptitude factor has two levels—high and low; the other factor (instructional method) also has two levels (A and B). The researcher randomly selects 60 Ss from the high-aptitude group and assigns 30 Ss to method A and 30 Ss to method B. This process is repeated for the low-aptitude group. Teachers are also randomly assigned to the groups. A 2 × 2 design requires four groups of subjects; each group represents a combination of a level of one factor and a level of the other factor.

Example of a Factorial Design

	Instructional Method (X_1)		
Aptitude (X_2)	Method A	Method B	Mean
High	75.0	73.0	74
Low	60.0	64.0	62
Mean	67.5	68.5	

The scores in the four cells represent the mean scores of the four groups on the dependent variable, the science achievement test. In addition to the four cell scores representing the various combinations of treatments and levels, there are four marginal mean scores: two for the columns and two for the rows. The marginal column means are for the two methods, or treatments, and the marginal row means are for the two levels of aptitude.

From the data given, you can first determine the main effects for the two independent variables. The main effect for treatments refers to the treatment mean scores without regard to aptitude level. If you compare the mean score of the two method A groups, 67.5, with that of the two method B groups, 68.5, you find that the difference between these means is only 1 point. Therefore, you might be tempted to conclude that the method used has little effect on the achievement scores, the dependent variable.

Now examine the mean scores for the levels to determine the main effect of X_2, aptitude level, on achievement scores. The main effect for levels does not take into account any differential effect caused by treatments. The mean score for the two high-aptitude groups is 74, and the mean score for the two low- aptitude groups is 62; this difference, 12 points, is the effect attributable to aptitude level. The high-aptitude group has a markedly higher mean score; thus, regardless of treatment, the high-aptitude groups perform better than the low-aptitude groups. Note that the term main effects does not mean the most important effect but, rather, the effect of one independent variable (factor) ignoring the other factor. In the example, main effect for teaching method refers to the difference between method A and method B (column means) for all students regardless of aptitude. The main effect for aptitude is the difference between all high- and low-aptitude students (row means) regardless of teaching method.

A factorial design also permits the investigator to assess the interaction between the two independent variables—that is, the different effects of one of them at different levels of the other. If there is an interaction, the effect that the treatment has on learning will differ for the two aptitude levels. If there is no interaction, the effect of the treatment will be the same for both levels of aptitude. In the above table the method A mean is higher than the method B mean for the high-aptitude group, and the method B mean is higher for the low-aptitude group. Thus, some particular combinations of treatment and level of aptitude interact to produce greater gains than do some other combinations.

OTHER RANDOMIZED EXPERIMENTAL DESIGNS

The experimental designs we have discussed so far use at least two groups of subjects, one of which is exposed to the treatment (independent variable) and the other that does not receive the treatment or is exposed to another level of the treatment. The researcher then compares the dependent variable scores for the different treatment groups. The essential feature of these designs is that they compare separate groups of subjects in order to determine the effect of the treatment. When the independent variable is manipulated in this way, we have what is called a between-subjects design. For example, a researcher who compares reading achievement scores for students taught by one method with scores for an equivalent group of students taught by a different method is using a betweensubjects design.

However, the manipulation of an independent variable does not have to involve different groups of subjects. It is possible to use experimental designs in which the same participants are exposed to different levels of the independent variable at different times. For example, a researcher might measure the learning of nonsense syllables by one group of students under different levels of anxiety or the math performance scores of a group of students when music is played in the classroom versus no music. This type of design in which a researcher observes each individual in all of the different treatments is called a within- subjects design. It is also called a repeated-measures design because the research repeats measurements of the same individuals under different treatment conditions. The main advantage of a within-subjects design is that it eliminates the problem of differences in the groups that can confound the findings in betweensubjects research. Remember that one is not comparing one group of subjects to another; one is comparing each individual's score under one treatment with the same individual's score under another treatment. Each subject serves as his or her own control. Another advantage of within-subjects designs is that they can be conducted with fewer subjects. The disadvantage of these designs is the carryover effect that may occur from one treatment to another. To deal with this problem, researchers typically arrange for the participants to experience the different treatments in random or counterbalanced order.

QUASI-EXPERIMENTAL DESIGNS

In many situations in educational research, however, it is not possible to randomly assign subjects to treatment groups. Neither the school system nor the parents would want a researcher to decide to which classrooms students were assigned. In this case, researchers turn to quasi- an experiment in which random assignment to treatment groups is not used. **Quasi-experimental designs** are similar to randomized experimental designs in that they involve manipulation of an independent variable but differ in that subjects are not randomly assigned to treatment groups. Because the quasi-experimental design does not provide full control, it is extremely important that researchers be aware of the threats to both internal and external validity and considers these factors in their interpretation. Although true experiments are preferred, quasi- experimental designs are considered worthwhile because they permit researchers to reach reasonable conclusions even though full control is not possible.

Design 9: Nonrandomized Control Group, Pretest–Posttest Design

In a typical school situation, schedules cannot be disrupted nor classes reorganized to accommodate a research study. In such a case, one uses groups already organized into classes or other preexisting intact groups.

The **nonrandomized control group, pretest–posttest design** is one of the most widely used quasi-experimental designs in educational research. We can see that it is similar to Design 5 but with one important difference: Design 9 does *not* permit random assignment of subjects to the experimental and control groups.

Design 9: Nonrandomized Control Group, Pretest–Posttest Design

Group	Pretest	Independent Variable	Posttest
E	Y_1	X	Y_2
C	Y_1	—	Y_2

A researcher might be allowed to use two sections of freshman English at a high school for a study on vocabulary development. The researcher should select two sections that at least appear to be similar; for example, one should not choose a remedial class and an advanced class. Although subjects cannot be randomly assigned, one can flip a coin to determine which of the two intact groups will be the experimental group and which will be the control group. The researcher would give a vocabulary pretest to both classes, administer a program designed to improve vocabulary to the experimental group only, and then give a vocabulary posttest to both groups. If the experimental group shows significantly greater achievement on the posttest, can the researcher conclude that the new program was effective?

Without random assignment of subjects, Researcher does not know if the groups were equivalent before the study began. Perhaps the class designated the experimental group would have done better on the posttest without the experimental treatment. Thus, there is an initial *selection bias* that can seriously threaten the internal validity of this design. The pretest, the design's most important feature, provides a way to deal with this threat. The pretest enables you to check on the equivalence of the groups on the dependent variable before the experiment begins. If there are no significant differences on the pretest, Researcher can discount selection bias as a serious threat to internal validity and proceed with the study.

If there are some differences, the investigator can use ANCOVA to statistically adjust the posttest scores for the pretest differences.

Because both experimental and control groups take the same pretest and posttest, and the study occupies the same period of time, other threats to internal validity, such as maturation, instrumentation, pretesting, history, and regression (if groups are not selected on the basis of extreme scores), should not be serious threats to internal validity. Having the same person teach both English classes would be recommended. There are some possible internal validity threats, however, that this design does not control, namely threats resulting from an interaction of selection and some of the other common threats.

The nonrandomized control group, pretest–posttest design is a good second choice when random assignment of subjects to groups is not possible. The more similar the experimental and the control groups are at the beginning of the experiment, and the more this similarity is confirmed by similar group means on the pretest, the more credible the results of the nonrandomized control group pretest–posttest study become. If the pretest scores are similar and selection– maturation and selection–regression interactions can be shown to be unlikely explanations of posttest differences, the results of this quasi-experimental design are quite credible.

Design 10: Counterbalanced Design

A **counterbalanced design,** another design that can be used with intact class groups, rotates the groups at intervals during the experimentation. For example, groups 1 and 2 might use methods A and B, respectively, for the first half of the experiment and then exchange methods for the second half. The distinctive feature of Design 10 is that all groups receive all experimental treatments but in a different order. In effect, this design involves a series of replications; in each replication the groups are shifted so that at the end of the experiment each group has been exposed to each *X*. The order of exposure to the experimental situation differs for each group. The following shows a counterbalanced design used to compare the effects of two treatments on a dependent variable:

Design 10: A Counterbalanced Design with Two Treatments

	Experimental Treatments	
Replication	***X*1**	***X*2**
1	Group 1	Group 2
2	Group 2	Group 1
	Column mean	Column mean

A classroom teacher could use a counterbalanced study to compare the effectiveness of two methods of instruction on learning in science. The teacher could choose two classes and two units of science subject matter comparable in the nature of the concepts, difficulty of concepts, and length. It is essential that the units be equivalent in the complexity and difficulty of the concepts involved. During the first replication of the design, class (group) 1 is taught unit 1 by method $X1$ and class (group) 2 is taught unit 1 by method $X2$. An achievement test over unit 1 is administered to both groups. Then class 1 is taught unit 2 by method $X2$ and class 2 is taught unit 2 by method $X1$; both are then tested over unit 2. After the study, the column means are computed to indicate the mean achievement for both groups (classes) when taught by method $X1$ or method $X2$. A comparison of these column mean scores through an analysis of variance indicates the effectiveness of the methods on achievement in science.

A counterbalanced design may be used when several treatments are to be investigated. Following is a counterbalanced design with four treatments.

Each row in following table represents one replication. For each replication, the groups are shifted so that group A first experiences $X1$, then $X2$, $X3$, and finally $X4$. Each cell in the design would contain the mean scores on the dependent variable for the group, treatment, and replication indicated. The mean score for each column would indicate the performance of all four groups on the dependent variable under the treatment represented by the column.

Experimental Treatments

Replication	X_1	X_2	X_3	X_4
1	Group A	B	C	D
2	Group C	A	D	B
3	Group B	D	A	C
4	Group D	C	B	A
	Column mean	Column mean	Column mean	Column mean

Design 10 overcomes some of the weaknesses of Design 9; that is, when intact classes must be used, counterbalancing provides an opportunity to rotate out any differences that might exist between the groups. Because all treatments are administered to all groups, the results obtained for each X cannot be attributed to preexisting differences in the subjects. If one group should have more aptitude on the average than the other, each X treatment would benefit from this greater aptitude.

The main shortcoming of Design 10 is that there may be a carryover effect from one X to the next. Therefore, this design should be used only when the experimental treatments are such that exposure to one treatment will have no effect on subsequent treatments. This requirement may be difficult to satisfy in much educational research. Furthermore, one must establish the equivalence of learning material used in various replications. It may not always be possible to locate equivalent units of material. Another weakness of the counterbalanced design is the possibility of boring students with the repeated testing this method requires.

TIME-SERIES DESIGNS

Design 11: One-Group Time-Series Design

The **one-group time-series design** involves periodic measurement on one group and the introduction of an experimental treatment into this time series of measurements. As the design indicates, a number of measurements on a dependent variable are taken, *X* is introduced, and additional measurements of *Y* are made. By comparing the measurements before and after, rsearcher can assess the effect of *X* on the performance of the group on *Y*. A time-series design might be used in a school setting to study the effects of a major change in administrative policy on disciplinary incidents. Or a study might involve repeated measurements of students' attitudes and the effect produced by introducing a documentary film designed to change attitudes.

Design 11: One-Group Time-Series Design

$$Y_1 \quad Y_2 \quad Y_3 \quad Y_4 \quad X \quad Y_5 \quad Y_6 \quad Y_7 \quad Y_8$$

The major weakness of Design 11 is its failure to control history; that is, you cannot rule out the possibility that it is not *X* but, rather, some simultaneous event that produces the observed change. Perhaps such factors as seasonal or weather changes or such school events as examinations could account for the change. In a study designed to assess the effect of a lecture–film treatment on student attitudes toward minorities, to what extent would the attitude measurements be affected by a nationally publicized minority riot in a distant city? The extent to which history (uncontrolled contemporary events) is a plausible explanatory factor must be taken into account by the experimenters as they attempt to interpret their findings. Statistical interpretation can be a particular problem with time data. The usual tests of significance are not appropriate with a time design because they assume that observations are independent of one another; but time-series data are typically correlated with one another.

Design 12: Control Group Time-Series Design

The **control group time-series design** is an extension of Design 11 to include a control group. The control group, again representing an intact class, would be measured at the same time as the experimental group but would not experience the *X* treatment. This design overcomes the weakness of Design 11—that is, failure to control history as a source of extraneous variance. The control group permits the necessary comparison. If the E group shows a gain from *Y*4 to *Y*5 but the C group does not show a gain, then the effect must be caused by *X* rather than by any contemporaneous events, which would have affected both groups.

Design 12: Control Group Time-Series Design

Group									
E	Y_1	Y_2	Y_3	Y_4	X	Y_5	Y_6	Y_7	Y_8
C	Y_1	Y_2	Y_3	Y_4	—	Y_5	Y_6	Y_7	Y_8

Other variations of the time-series design include adding more control groups, more observations, or more experimental treatments.

COMPARISON OF SINGLE-SUBJECT AND GROUP DESIGNS

In both single-subject and group experiments, the goal of the experimenter is to establish as unequivocally as possible the connection between the manipulation of the independent variable (treatment) and its effect on the dependent variable (behavior). In group designs, random assignment of subjects to experimental or control groups eliminates many rival explanations of differences observed after treatment. Treatment effects (between groups) can be assessed relative to inter subject variability effects (within group) by using appropriate statistical tests. These tests determine whether chance alone is a credible explanation for the results. The single-subject design uses other methods to establish credibility.
The experimenter controls the amount of time in which baseline and treatment phases are in effect, and the length of the baseline period can be extended until the behavior stabilizes. For unambiguous interpretation, the baseline should be relatively fl at or the trend should be in the opposite direction from that expected after treatment. One drawback to experimenter control of the length of the treatment phase is the tendency to continue treatment until "something happens." If behavior change does not closely follow the beginning of treatment, it is possible that another, non experimental variable is the cause of the observed change.

Single-subject experimental designs do bypass one source of error of group designs—namely, inter subject variability. Each individual serves as his or her own control, so comparability is not a problem. The major means of control is replication, a feature seldom incorporated into group designs. The ABAB design involves a single replication using the same subject, whereas the multiple-baseline design replicates more than one treatment. Replication of the multiple-baseline design makes it less likely that effects attributed to treatment were in fact caused by extraneous event or subject variables.

Well-designed single-subject research can meet the criteria for internal validity. However, the question of external validity—the generalizability of experimental findings—is not as easily answered by designs that use only one or a few subjects. You can demonstrate that allowing a behaviorally disordered teenager to listen to rock music contingent on completing assignments increases the amount of schoolwork done by that particular teenager, but how can you determine whether this treatment will be successful with other teenagers or all behaviorally disordered teenagers? Although any one particular single-subject study will be low in external validity, a number of similar studies that carefully describe subjects, settings, and treatments will build the case for wide application of particular treatment effects.

Steps in Conducting Experimental Research

1. **Decide if an Experiment Addresses Research Problem**

The type of issue studied by experimenters is the need to know whether a new practice influences an outcome. Of all designs in education, it is the best design to use to study cause-and-effect relationships. However, to study these issues, you must be able to control the setting of the experiment as well as manipulate one level of the independent variable. An experiment is not the best choice when the problem calls for generalizing results to a population or when you cannot manipulate the conditions of the experiment.

2. **Form Hypotheses to Test Cause-and-Effect Relationships**

A hypothesis advances a prediction about outcomes. The experimenter establishes this prediction (in the form of a null or alternative hypothesis) and then collects data to test the hypothesis. Hypotheses are typically used in experimental research more than are research questions, but both can be used.

Independent variables should contain at least one variable with multiple levels, and the researcher needs to manipulate one of the levels. Dependent variables are outcomes, and experimenters often study multiple outcomes (e.g., student learning and attitudes).

Variables are measured on an instrument or recorded as observations. They need to produce valid and reliable scores. You need to give special attention to choosing measures that will result in scores with high construct validity.

3. **Select an Experimental Unit and Identify Study Participants**

One of the first steps in conducting an experiment is to decide on your experimental unit. An experimental unit of analysis is the smallest unit treated by the researcher during an experiment. When we use the term treated*,* we are referring to the experimental treatment. Researcher may collect data from individuals, but the experimental unit actually treated differs from one experiment to another. The experimental unit receiving a treatment may be a single individual, several individuals, a group, several groups, or an entire organization.

Who will participate in your experiment? Participants in an experimental study are those individuals tested by the researcher to determine if the intervention made a difference in one or more outcomes. Investigators may choose participants because they volunteered or they agreed to be involved. Alternatively, the researcher may select participants who are available in well-defined, intact groups that are easily studied.

How many people will you study? In an ideal experiment, the researcher forms at least one control and one experimental group (Bausell, 1994). In many experiments, the size of the overall number of participants (and participants per group) is dictated by practical issues of the number of volunteers who enroll for the study or the individuals available to the researcher. The researcher also uses statistics to analyze the data, and these statistics call for minimum numbers of participants.

How should the participants be chosen? Researcher should randomly select individuals for the experiment from the study population so that inferences can be made from the results to the population. This selection is accomplished through numbering the individuals in the population and randomly selecting participants using a random numbers table.

How should the individuals be assigned to groups? An optimal situation is to randomly assign the individuals to groups, but this procedure may not always be feasible. Also, to provide added control over extraneous factors, matching, blocking, selecting of homogeneous groups, and the use of covariates are recommended.

4. **Select an Experimental Treatment and Introduce It**

The key to any experimental design is to set levels of treatment and apply one level to each group, such as one level to an experimental group and another level to a control group. Then the groups are compared on one or more outcomes. Interventions may consist of programs or activities organized by the researcher. In deciding what intervention to use, you might consider several factors:

The experimental researcher should select an intervention of adequate "dosage" (Lipsey, 1998). This means that the intervention must last long enough and be strong enough to actually have an impact on the outcome.

A good intervention is one that has been used by other researchers and it should predict a change in the outcome. The review of the literature and an assessment of past theories as predictions for relationships help researchers locate an intervention that should predict change.

Experimental researchers should choose an intervention that can be implemented with as little intrusion in the setting and on the participants as possible. This means that the researcher needs to respect the school or nonschool setting being studied and gain the cooperation of sponsors at the site and of the participants in the study.

Choose an intervention based on a small pilot test. Select a group of participants in the population and provide the intervention to them. This approach may be a pre-experimental design with a single group (to facilitate ease of implementation) or an intervention of a short duration. It may involve as few as five or six subjects (Bausell, 1994). From this pilot, you can draw conclusions about the potential impact of the intervention for the final experiment.

5. **Choose a Type of Experimental Design**

One aspect of preparing for the experiment is choosing the design and providing a visual diagram of it. You need to make several decisions based on your experience with experiments, the availability of participants for the study, and your ability to practically control for extraneous influences in the project before choosing a design.

6. **Conduct the Experiment**

Conducting the experiment involves procedural steps consistent with the design selected. It may involve:

- Administering a pretest, if you plan to use one
- Introducing the experimental treatment to the experimental group or relevant groups
- Monitoring the process closely so that the threats to internal validity are minimized
- Gathering posttest measures (the outcome or dependent variable measures)
- Using ethical practices by debriefing the participants by informing them of the purpose and reasons for the experiment, such as asking them what they thought was occurring.

7. Organize and Analyze the Data

Three major activities are required at the conclusion of the experiment: coding the data, analyzing the data, and writing the experimental report. Coding the data means that the researcher needs to take the information from the measures and set up a computer file for data analysis. This procedure begins with cleaning the data to make sure that those who complete the instruments do not enter unusual data in the computer file through keystroke errors or errant mistakes. You can explore the database for these errors by running a descriptive analysis of it using a statistical analysis program and noting variables for which unusual data exist. This descriptive analysis can provide the first review of the outcomes of the study, and scanning the results can provide an understanding of the responses of all participants to the outcome measures. This step becomes the first phase of the data analysis.

After a descriptive analysis of all participants, the researcher begins the analysis of comparing groups in terms of the outcomes. This is the heart of an experimental analysis, and it provides useful information to answer the hypotheses or research questions in the study. The statistic of choice is a group comparison statistic, such as the *t* test or the family of parametric analysis of variance statistics (e.g., ANOVA, analysis of covariance [ANCOVA]).

8. Develop an Experimental Research Report

The experimental report follows a standard format. In the "Methods" or "Procedures" section of an experiment, the researcher typically includes information about:

- Participants and their assignment
- The experimental design
- The intervention and materials
- Control over extraneous variables
- Dependent measures or observations

As in a quantitative study, you write this report using standard terms for research (e.g., intervention, control, experimental group, pre- and posttest) and an objective, impartial point of view.

Evaluating an Experiment

A good experiment has a powerful intervention, groups few in number, derived in some systematic way, and where individuals will gain from the experiment. The scores on the measures are both valid and reliable because the researcher has attended to potential threats of validity.

Ethical Issues in Experimental Research

Ethical issues in conducting experiments relate to withholding the experimental treatment from some individuals who might benefit from receiving it, the disadvantages that might accrue from randomly assigning individuals to groups. This assignment overlooks the potential need of some individuals for beneficial treatment. Ethical issues also arise as to when to conclude an experiment, whether the experiment will provide the best answers to a problem, and considerations about the stakes involved in conducting the experiment.

Suggested Reading

Ary, D., Jacobs, L. C., Sorensen, C., and Razavieh, A. (2010), Introduction to Research in Education, Wadsworth Cengage Learning, Canada.

Best, J. W. and Kahn, J. V. (1995), Research in Education, Prentice Hall, New Delhi.

Cohen, L., Manison, L., and Morrison, K. (2018), Research Methods in Education, Routledge, London and New York.

Creswell, J. W. (2012), Educational Research Planning, Conducting, and Evaluating Quantitative and Qualitative Research, Pearson, New York.

Gay, L. R., Mills, G.E., and Airasian, P. W. (2012), Educational Research Competencies for Analysis and Application, Pearson, New York.

Kerlinger, Fred. N. (1978), Foundations of Behavioral Research, New York University.

Kothari, R.C. (2004), Research Methodology, New Delhi, New Age International (P) Limited, Publishers.

Langenbach, M., Vaughn, C., and Aagaard, L. (1993), An Introduction to Educational Research, Allyn and Bacon, Boston.