



---

---

## Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects

---

---

---

### Authors

**Hamed Taherdoost**  
University Canada West

*hamed.taherdoost@gmail.com*  
Vancouver, Canada

---

### Abstract

*One of the main stages in a research study is data collection that enables the researcher to find answers to research questions. Data collection is the process of collecting data aiming to gain insights regarding the research topic. There are different types of data and different data collection methods accordingly. However, it may be challenging for researchers to select the most appropriate type of data collection based on the type of data that is used in the research. This article aims to provide a comprehensive source for data collection methods including defining the data collection process and discussing the main types of data. The possible methodologies for gathering data are then explained based on these categories and the advantages and disadvantages of utilizing these methods are defined. Finally, the main challenges of data collection are listed and in the last section, ethical considerations in the data collection processes are reviewed.*

---

### Key Words

*Data Collection, Research Methodology, Data Collection Methods, Academic Research Paper, Data Collection Techniques.*

---

## I. INTRODUCTION

Different methods for gathering information regarding specific variables of the study aiming to employ them in the data analysis phase to achieve the results of the study, gain the answer of the research

questions or test the hypotheses are referred to as data collection.

Data collection as a main stage in research can overshadow the quality of achieving results by decreasing the possible errors which may occur during a research project. Therefore, alongside a good design for the study, plenty of quality time should be spent in the collection of data to gain appropriate results since insufficient and inaccurate data prevents assuring the accuracy of findings (Kabir, 2016). On the other hand, although a suitable data collection method helps to plan good research, it cannot necessarily guarantee the overall success of the research project (Olsen, 2012).

## II. TYPES OF DATA

Before selecting a data collection method, the type of data that is required for the study should be determined (Kabir, 2016). This section aims to provide a summary of possible data types to go through the different data collection methods and sources of data based on these categories. However, we need to understand what data is exactly? The embodied information in terms of figures or facts used to analyze for different calculations and finally gain a result to address the study question or hypothesis testing is known as data (Hurrell, 2005). Data can be categorized using different ways including quantitative and qualitative.

### A. Qualitative Data

Both nominal and descriptive non-numerical data which cannot be shown as numbers are known as qualitative data in words or sentences format. This type of data answers to "how and why" questions in a research study and mostly covers data regarding feelings, perceptions, and emotions using unstructured approaches such as interviews for data collection. Researchers use different methods such as using audiotapes, sketches, notes, and photographs to gather these data.

Although qualitative data can be suitable to achieve further information to explore and determine new effects and consequences of programs on the research, and finally enhance the quality of quantitative results, its implementation is dependent on spending a considerable amount of cost and time and the results may not be generalizable. It means the findings of case studies can be used just for the same issues as the general patterns for different studies. Qualitative methods encompass three main categories including observations, document reviews, and in-depth interviews in spite of the fact that there are less common ways to gather qualitative data. In the next section, all of the methods of data collection are discussed.

### B. Quantitative Data

Numerical data which is mathematically generated and computed is recognized as quantitative data. There are different scales for measuring quantitative data including nominal, ordinal, interval, and ratio

scales (Kabir, 2016). Scales can be categorized into two general types as “Rating Scales and Attitude Scales” as well. Rating scales assign a numerical value to the points or categories to evaluate them. On the other hand, more complex methods are attitude scales that determine the predisposition of people toward any individual, phenomenon, or object (Taherdoost, 2016b). A qualitative method addresses the “what” question type in a study. These approaches employ structured data collection methods and are based on random sampling. In comparison to the qualitative methods, these methods are regarded to be cheaper, and the findings can be standardized to achieve other results based on some criteria such as size. The findings can be easily generalized and summarized as well. A simple comparison between the results is also possible. Nevertheless, these methods also can face unexpected differences and some difficulties as the implementation and investigation capacity are limited in these methods. These approaches use different methods such as experiments and structured interviews for data collection which are discussed in the data collection methods section.

### III. DATA COLLECTION METHODS

Generally, data collection methods are divided to two main categories of Primary Data Collection Methods and Secondary Data Collection Methods. Figure I shows some of data collection methods for primary and secondary data.

Data that is not published yet and is the first-hand information which is not changed by any individual is known as primary data. In other words, researchers use different approaches to gather and collect primary data for a specific purpose. Thus, the validity, reliability, objectivity, and authenticity of data are more in primary data in comparison with the secondary data types. These qualities are important in some types of research methods such as statistical surveys as the use of the information is specific to a problem and cannot be provided from published references. Thus, although the research can be conducted based on secondary data, it is not possible to achieve a reliable result without using primary data as well. As secondary data is manipulated and changed by others. Using primary sources, helps to gain high-quality data which can improve results, and you also have the opportunity to add further data when required during the research procedures. Primary data collection; however, can face difficulties in defining different terms in collecting data for example, the reasons behind data collection, what to collect, when to collect data, and the type of data collection method. It is also an expensive approach, obtains the majority of research budget, and needs to provide funding resources from different agencies. You need to ensure the standard of collected data by accurately collecting them, eliminating unnecessary data and also not using fake and cooked-up ones. To achieve primary data, different sources can be used such as experiments, surveys, interviews, and questionnaires (Kabir, 2016; Taherdoost, 2021).

Secondary data is the data gathered from published sources meaning that the data is already gathered by someone else for another reason and can be used for other purposes in a research as well. In all papers, the literature review section is based on secondary data sources. Thus, secondary data is an essential part of research that can help to get information from past studies as basis conduction for

implementing a research or as the required background information. It can also help to design a study and provide a baseline to compare primary results. However, it should be noted that researchers need to re-examine the validity and reliability of these backgrounds to gain authentic results.



FIGURE I. DATA COLLECTION METHODS

There are different sources of secondary data such as records, books, research articles, and internet articles. Although they are not valid as much as primary data sources, these sources are still important for scientific studies as primary data collection is sometimes hard or impossible to obtain. Also, in some cases, the respondents do not allow to reveal data and the study should be conducted based on secondary data. Generally, secondary data is cheaper and easier to obtain in comparison to primary data and there is

no responsibility for the quality of data as it is just reported in the study. It also brings some disadvantages as it may not be reliable or accurate. It also cannot be used for a different situation for example environmental factors can affect the data and the time factor is important since sometimes it is required to use up-to-date data. Besides, copyright issues may happen (Hox & Boeije, 2005; Kabir, 2016).

### **3.I. PRIMARY DATA COLLECTION METHODS**

Primary data collection is based on the processes by which you gather data yourself for your purpose of study and no one has access to use this data until it is published and both qualitative and quantitative approaches are used for this purpose. The main primary data collection is discussed here, considering 14 different types are listed in figure 1. The most common types are initially explained including questionnaires, interviews, focus groups, observation, survey, case studies, and experimental methods in detail. Then, other methods are reviewed shortly.

#### **3.I.I. Questionnaire Method**

The questionnaire is one of the common devices for collecting information and a form or instrument including a set of questions and secure answers that respondents (from a specific population) fill to give the researcher information needed for the study. The data given from a questionnaire cannot be achieved from the secondary resources (Pandey & Pandey, 2015). These forms are suitable to gather both qualitative and quantitative data. Although they are not the most common methods used in qualitative research, they are useful in case of facing a large sample in a study.

Sir Francis Galton designed a questionnaire for the first time. A questionnaire is utilized for different purposes, although it is commonly used to gather statistical data. It can be designed for measuring separate variables such as behaviors, preferences, and facts (Kabir, 2016). Although the preparation and administration of a questionnaire are not hard, specific points in these processes should be required. This form is used, normally, when it is not possible to discuss each participant personally (Pandey & Pandey, 2015). Thus, it helps to gather data from different individuals, groups, and companies easily. Questionnaires can be categorized based on different aspects such as types of questions and administration modes.

##### **A) Types of Questions**

First, questions can be designed to measure variables for example in a survey. On the other hand, questions can be based on aggregating into indexes or scales, for instance in tests. Second, question types can be categorized into closed-ended and open-ended questions. In close-ended questions, the respondents face a specific range of answers to choose from, but the respondent is asked to provide

formulated answers using open-ended questions. Qualitative questions are open-ended (Taherdoost, 2019). In this type, then, the answers should be coded into a response scale. Therefore, in comparison to the open-ended questions, close-ended ones are pre-coded to make the work quickly be implemented (Olsen, 2012).

For close-ended questions, there are four types of options to respond the questions:

- You can have a two-option as the responses possibilities which are known as dichotomous scales.
- If you add more than two options for the respondents, the scale is known as nominal-polychromous.
- In ordinal-polytomous scales, you prepare more than two options which are also ordinal.
- Finally, you can use continuous or bounded types which use a continuous scale as a possible response case.

### **B) The Mode of Administration**

Questionnaires can be implemented in different ways. A face-to-face questionnaire mode can be used which provides the chance of presenting the questions orally, paper-and-pencil types can be utilized with the items presented in the paper or computerized questionnaires for data collection (Kabir, 2016). Questionnaires can be also utilized through telephone, online, or even posting. An online questionnaire is a cost-efficient option; however, you should consider the possibility of missing samples due to problems with internet access. In these types, different online survey services can be used which provide questionnaires for the purpose of study, and then the collected data can be easily added to the analyzing software (Taherdoost, 2021). In all these choices, it is important to secure ethical concerns such as the confidentiality of participants. On the other hand, participants should try to answer the questions politely and clearly (Kabir, 2016).

### **C) General Rules for Constructing a Questionnaire**

- Use simple and short questions as much as possible;
- Navigate respondents clearly to avoid any difficulty and motivate participants through answering questions (Olsen, 2012);
- Use understandable, simple, and clear statements for all respondents with different educational levels;
- Utilize positive sentences;
- Do not use more than one question (double-barreled) in one item;
- Add an open-answer possibility after providing the listed answers and where possible;
- Avoid making assumptions for the respondents (Kabir, 2016);
- Try to increase reliability by appropriate word selection;
- Avoid directing the respondent to any answer using objective questions including clues, suggestions, and hints) (Taherdoost, 2021);

- Explain the importance of the questionnaire in its content as well as its cover letter (Pandey & Pandey, 2015);
- Use sensitive questions after some innocuous or concrete ones. The order of questions is an important point. For example, those questions which are about sexual activities, drugs, incomes, etc. should be added to the final parts of questions to help the researcher gain participants' trust and avoid their embarrassment (Kabir, 2016; Taherdoost, 2016a).

There are also specific challenges and concerns that may be faced through designing an appropriate questionnaire. First, the maximum respondents' rates should be guaranteed together with securing maximum reliability and validity as much as possible. The respondent rates can be maximized when you:

- Can convince them you secure their information and keep their side;
- Can reward their cooperation.

On the other hand, you can gain an accurate data set considering two points:

- Prepare a suitable set of questions;
- Select appropriate sample size and type which can avoid biases and non-responded questions (Taherdoost, 2021).

The procedure of designing a questionnaire is shown in Figure 2.

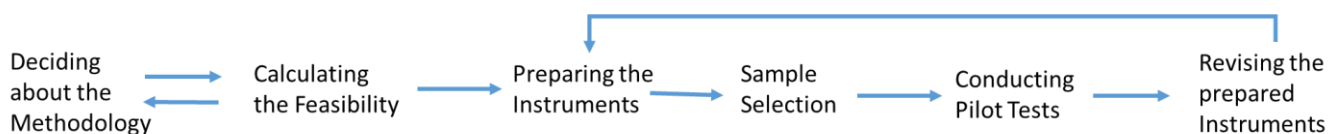


FIGURE 2. QUESTIONNAIRE DESIGN STEPS (TAHERDOOST, 2016A)

#### D) Advantages of Questionnaires

Questionnaires provide several merits in comparison to other survey methods as listed in the following:

- Collecting a large amount of data from a large sample size;
- Time saver;
- Cost-effective options;
- Highly structured;
- The possibility of gaining high accurate data;
- The possibility of being carried out by other people instead of the researcher regardless of affecting the reliability and validity term, and the possibility of group administrations;
- Analyzing the results easily by entering the achieved data to the software quickly in the majority of cases;



- The opportunity of more objective and scientific analysis;
- The achieved quantitative data can be used to compare and contrast the results of the study with others to measure the changes;
- The possibility of achieving comprehensive design and tests, and administrating the research with required details;
- Creating novel theories or/ and testing an existing hypothesis using the achieved quantitative data (Frechtling, 2002; Kabir, 2016; Taherdoost, 2021);
- Suitable in a wide range of study fields;
- Suitable and reliable in special cases (Pandey & Pandey, 2015).

### **E) Disadvantages of Questionnaires**

However, there are also several demerits that are not negligible. There are several difficulties researchers may face using questionnaires as the following:

- Hard or inadequate to perceive gathered data in some cases such as emotional, feelings, and behavioral changes;
- Human errors for example if the respondent is forgetful and cannot consider the whole concept truly;
- Determining the reliability of answers is not possible;
- The possibility of misunderstanding the questions which can overshadow the answers;
- The effects of differences in human beliefs on their answers in some cases since even a standard subject can be considered good for one group and bad for others (Kabir, 2016);
- Facing difficulties when participants need clarifications for particular questions in impersonal administrations and the possibility of failing to answer those questions (Taherdoost, 2021);
- Low response rates if respondents' low interests cannot be addressed to answer questions (Frechtling, 2002);
- The possibility of illegible answers;
- Useless and wrong answers are prevalent (Pandey & Pandey, 2015).

### **3.1.2. Interviews**

In interviews, as a fundamental way of social interaction, questions are asked and data is collected using provided answers and it is in contrast to the questionnaire with indirectly collected data methodology. Thus, the chance of getting confidential data from interviewees is also possible; however, it requires special skills which are not necessary for questionnaires. Researchers can employ different methods to conduct an interview (Pandey & Pandey, 2015; Taherdoost, 2021) and perform them in individual, or group face-to-face interviews, as well as not personally for example using telephone, computer, etc. (Kabir, 2016). Both forms bring along merits and demerits which can be listed in Table.I.



TABLE. I. ADVANTAGES AND DISADVANTAGES OF DIFFERENT INTERVIEW TYPES (KABIR, 2016)

Type	Merit	Demerit
Face-to-face Interviews	-Asking detailed questions. -Obtaining rich data. -Literacy requirements are not an issue. -Possibility of clarifying questions. -High response rate. -Exploring complex and sensitive issues.	-Expensive. -Effects of interviewers' biases. -Possibility of facing some challenges for sensitive issues. -Training interviewers is necessary.
Telephone Interviews	-Cheap and accurate data. -Quick. -Possibility of clarifying questions. -Literacy requirements are not an issue. -Using fewer resources than face-to-face.	-Possibility of not accessing to the participants easily and the first time. -Only possible for interviewees who access telephone. -Not possible to discover sensitive issues.

There are different ways to record interviews including paper-based, self-reports, and tape-record interviews, etc. It can be also noted that they are useful to gain both qualitative and quantitative data using open-ended and close-ended questions. Researchers rely on interviews, generally, when they face complex or sensitive concepts and need detailed and high-status information (Frechtling, 2002).

#### A) Types of Interviews

In addition to the administration mode of interviews which was discussed before, interview types can be structured, semi-structured, and unstructured.

- **Structured interviews:**

In these kinds of interviews, interviewees face the same set of standardized questions which are pre-prepared before the interview session. The possible responses are limited, and participants may just face a few open-ended questions. In case of gaining a good understanding of the research topic, it is possible to provide a questionnaire, or an interview guide based on the structural interview characteristics. A suitable place to use these kinds is when there is a comprehensive literature review about the topic of study or to utilize it following a less structural method.

- **Semi-structured Interviews**

Semi-structured interviews are formal and are conducted based on a guide. The interviewers ask questions considering the guidance; however, when researchers or interviewers need extra information, they can continue the conversation based on the questions provided ahead of time. For example, it is an appropriate choice when you do not have the chance of

interviewing with individuals more than once and you should gather clear, comparable, and reliable qualitative data in the first meeting. Thus, training of the interviewers and the design of open-end questions are important points. It can also be used when you aim to develop a subject such as a culture, setting, or experience which is not completely understood, and as an initial step for developing another more-structured method like a survey or an interview.

▪ **Unstructured interviews**

Unstructured interviews are informal methods of interviewing without using a specific structure. There is no guide in this type, and they just conduct casual conversations. Interviewers collect data using brief notes and try to memorize the responses. This process is a part of field observation and can be a good choice when there is limited literature about the field of study.

In these forms of interviews as they are "on the fly", the chance of tape-recording or writing detailed notes is slim, thus, the researcher should try to participate in the setting directly as soon as possible and should approve the notes and achieve the final data quickly after the observation to avoid losing information. Like semi-structured methods, informal ones can be also a fundamental step for preparing structured interviews. These methods can enhance the trust of the interviewees for the next levels as well, as both sides speak openly and freely to each other. Achieving qualitative data using informal interview methods can help the researcher to understand other individuals' experiences. These types also allow asking questions to the interviewees as a follow-up response to the interview questions (Kabir, 2016).

## **B) Important Notes in Conducting Interviews**

As running interviews can be both rewarding and challenging in a study, different points should be considered to conduct the process as far as possible. Some important points are as the following:

- The interviewer plays an important role as they are responsible to motivate participants to answer questions and judge whether the answers cover the expected quality or not. They conduct and guide interviewees when needed. Thus, training the interviewers is essential in these kinds of data collection types.
- The setting should be completely ready for the interviewers. They should address all their needs to the materials (known as interviewer's kit) including identifications, phone numbers for later contacts, maps, copies of guides, cover letters for example provided by the sponsors and main researcher, notebooks, etc.
- Another important point is that interviewers' biases can affect the findings of the study. Thus, this bias and judgmental thoughts should be minimized to gain the best result. This factor can be also addressed by interviewers' training. These problems can mainly happen through job interviews.

The process of executing an interview is shown in Figure 3.

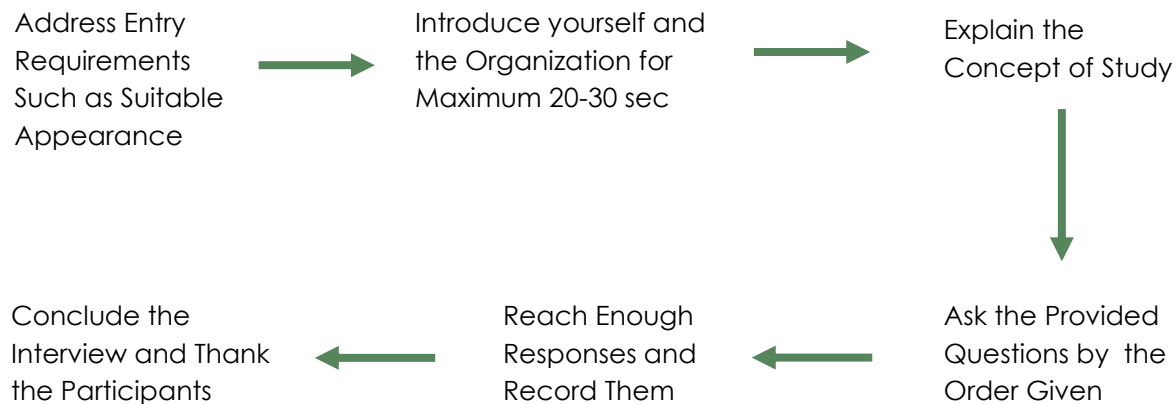


FIGURE. 3. INTERVIEW PROCESS

### C) Merits and Demerits

The main advantages and disadvantages of conducting interviews as a method of collecting data are provided in the following. The merits can be listed as:

- Gathering rich, in-depth, and detailed data directly
- The opportunity of obtaining knowledge about past and future for particular events and features
- The flexibility of administration of interviews
- The possibility of an explanation of the questions to the interviewees to clarify the questions

On the other hand, there are some difficulties that may be faced when conducting an interview method including:

- Hiring and training interviewers
- Complex process
- Scheduling where and when to meet people and the possibility of changing plans at the last minutes
- The possibility of missing information
- The coding process can be difficult and long
- Being expensive (Frechtling, 2002; Kabir, 2016; Pandey & Pandey, 2015)

### 3.1.3. Focus Groups Discussion (FGD)

This method, simply, is a mixture of interviewing and observation. This method is used to discover human behavior, attitudes, and respondents facing a particular concept. This in-depth field method gathers a group of individuals, normally between 6-12 people in each group, commonly with a shared

characteristic such as sex, age, and educational status to discuss a specific study field. The mediators aim to stimulate and discover the behavior of the participants and the reasons of each behavior using social dynamic of the group (Frechtling, 2002) (Kabir, 2016).

For this purpose, data regarding a particular subject is collected using a semi-structured interview and utilizing a specific guide in a relaxed and friendly atmosphere which helps the interviewees to easily trust the facilitators of the program and answer them honestly. As in interviews which interviewers should be trained to be supportive and good listeners, the training of the mediators is essential also here to gain especial skills such as analyzing the participants' behavior, facilitating, controlling, and managing the meeting sessions, listening and observing the events. Furthermore, to record data and provide detailed notes an observer often helps the mediator (Kabir, 2016).

This method is used in times of aiming to pretest a concept or an idea, generate an idea, discover a topic further, discover new areas, clarify the findings of other methods, and achieve a full understanding of a commonly social concept using small groups by stimulating discussions. It can also assist researchers to develop the interviews and surveys. When a concept is hard to observe, it can also be a useful method to obtain a concentrated observation set in a short period of time. When the results of a concept cannot be discussed statistically, this method helps to gain meaning for the findings of a survey (Frechtling, 2002) (Kabir, 2016).

#### **A) General Rules**

After identifying the suitable target group considering shared required characteristics, a short explanation about the concept should be provided to help the participants get familiar with the background and the meaning of the concept. Then, the standardized questions should be discussed based on the protocol and the responses of the group to issues should be written. Discussions about the concept should be also facilitated to gain more than just normal question-answer interactions. To obtain a comprehensive investigation, more than one focus group (minimum three) should be managed. After summarizing the results of the discussions, the meeting can be finalized, although sometimes more than one session is needed. General steps of conducting an FGD are shown in figure 4.

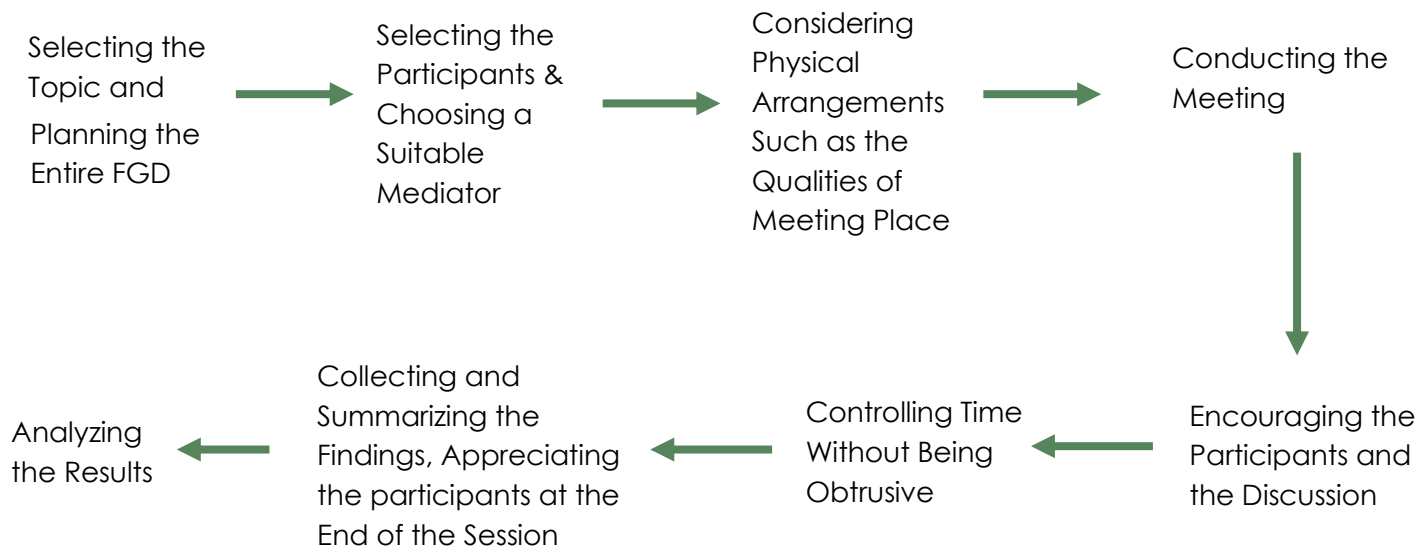


FIGURE. 4. PROCESS OF CONDUCTING AN FGD

### **B) Interview or Focus Group**

There are general circumstances to decide which method can be effective. First, when the interaction between members is not productive or may negatively affect the findings, an interview is a good choice. On the other hand, if the group pressure can positively overshadow the challenges of the participants' thoughts, a focus group can be conducted. For covering a large volume of issues, interviews are more appropriate choices. In focus groups, a single behavior is commonly examined regardless of considering connections between behaviors and attitudes. Additionally, the participants do not have a depth understanding of the concept and it is assumed that individuals can discuss all their known relevant knowledge in less than 10 minutes (Frechtling, 2002).

### **C) Advantages and Disadvantages**

The merits and demerits of the focus group method are focused on in this section as these types also possess both. They have several merits as:

- They can help to discover social, health, and cultural concepts for example by considering humans' behavior facing different situations;
- Literacy of individuals is not an issue;
- They are suitable to explore complex subjects;
- They are useful for the development of hypotheses.

On the contrary, they are also:

- Can be expensive and time-consuming;
- Face privacy lacks;
- Need trained facilitators;
- Can face the issues due to the domination of limited individuals in the focus groups (Frechtling, 2002; Kabir, 2016).

### **3.1.4. Observational Methods**

In these techniques, first-hand data is gathered through the observation of events, behaviors, interactions, processes, etc. directly to obtain an understanding of the concepts. For example, observation is an appropriate technique to evaluate teaching methods in the classes. It can be used when focus groups and interviews cannot help to gather data due to the different reasons including times that participants:

- Are not aware of the concept;
- Are not able to talk about the concept;
- Do not prefer to discuss the concept.

It can be also utilized to explore whether a study is progressing as planned, or whether the study has been successful or not. In the evaluation of studies, these two phases are known as formative and summative, respectively (Frechtling, 2002). It also can be helpful when the concept is unexplored or not well-known. If it is required to explore a subject in the natural setting and the reported information can be different from the findings of the real setting, an observational technique should be used.

This method can collect both qualitative and quantitative data. The qualitative data is gathered as a description of events in the setting. The quantitative data can be obtained by using the duration or frequency of the particular subjects. During this kind of systematic observation, formal and structured instruments and protocols nominal, ordinal, ratio, and interval scales are utilized. Thus, it can be used to record the findings template coding sheets with specific guides if the observer is not the main researcher. On the other hand, data achieved through this method can be used in conjunction with the quantitative findings of other methods.

Generally, observation helps the researcher to find out what is going on (Kabir, 2016) in the surrounding environment; however, as a data collection method, it is further than just listening and looking. This method includes an engagement with the setting, a clear expression of the events, technical improvisations, high attention, and good recording.

### A) Advantages and Disadvantages

The observational method also possesses several pros and cons. In this section, the most important ones are listed. The advantages are as the following:

- Gathering direct information;
- The participation of evaluators in the natural setting;
- Flexible and natural atmosphere;
- Free from biases;
- Can be generalized as large samples can be covered in the studies;
- High reliable and precise data can be achieved.

These techniques also provide some difficulties as:

- They can be time-consuming and not economical
- The training of observers is effective
- Observers can be selective and distort data
- It can be sometimes unreliable due to the misrepresenting of the qualitative data measurement
- It does not consider processes and the changes during them and cannot be appropriate for fresh concepts (Frechtling, 2002; Kabir, 2016).

### B) Special Notes

After conducting an observation, the researcher should first analyze achieved data. For this purpose, data is summarized in a process known as data reduction and it is coded based on particular criteria to specific categories. The reliability of data according to the agreements of independent observers should be also considered to show how the behaviors are measured accurately. It should be noted that the participants can act differently when they are in the research setting. These acts should be controlled using techniques of controlling reactivity such as indirect observation, the adaptation of participants, and unobtrusive measurements.

Biases of the observers are another important point to consider what happens when an observer's bias can affect what behaviors to choose and record; however, it can be minimized by keeping them unaware about the aims (Kabir, 2016).

### 3.1.5. Survey Methods

A survey simply is an appropriate method to determine feelings, opinions, and thoughts. The aim of the survey can be both globally and specifically. They can provide a large volume of data using telephone calls, emails, or face-to-face interviews.



On the other hand, data can be collected in self-completion surveys or by the interviewer. A survey can be used to explore social behaviors such as measuring the behavior of political candidates and professional people in educational institutions. However, it is not useful when evaluating people for government programs since in these programs, all members of the population should be studied (Kabir, 2016). Overall, in both formative and summative phases of a study, surveys are useful when it is required to collect information from a large target population, and detailed and in-depth data are not necessary in the project (Frechtling, 2002).

In a survey, a set of questions are provided to give a sample that is chosen from a specific target population. This sample presents the characteristics and behaviors of the population. Surveys are conducted to explore the populations' attitudes, the differences between different populations' behavior and discover the possible changes over time by repeating surveys in regular time intervals. Thus, the sample selection is an important stage in this process which can highly affect the findings. Sample sizes should be chosen based on the possibility of selecting every participant with a non-zero chance. Therefore, samples need to be chosen using a non-volunteer and non-haphazard selection technique (Kabir, 2016).

Sampling process steps can be simply listed as the following:

- Defining target population such as the number of individuals that are living in the country;
- Selecting a frame for sampling as the actual cases that we select a sample from them;
- Choosing the method of sampling which can be either a random or non-random technique;
- Measuring the appropriate sample size to avoid biases and sampling errors using the related formula (Taherdoost, 2016c).

Biases of participants can commonly happen, especially when they need to answer about sensitive subjects, or when the individuals need to trust the team before giving the right answer.

Questions can be designed in different ways as the following:

- Open-ended questions in which participants answer questions in their own ways;
- Close-ended questions mostly based on yes/no or true /false answers;
- Multiple choices that provide the opportunity to choose a favorite topic by participants (Kabir, 2016).

Here, as discussed in the questionnaires, the questions should be written considering several aspects ranging from their language to their length and their presenting order. For example, sensitive questions should be added among final questions as well. Cover letter and introduction should be also provided as discussed in other types.

## A) Survey Process Steps

Figure 5 summarizes main process steps in surveys; however, it should be noted that although some steps such as pilot testing are essential, some steps are not necessary in conducting all surveys, and as the surveys are complex, the process does not need to follow the order and these sequential steps are provided as a common example. First, the design and plan of the study should be provided. It includes defining the purpose of study, sample selection method, its size, and designing and providing questionnaires.

At the next level, you go through the testing and modifying processes which include examining pilot tests and pre-tests if possible, analyzing the tests, and modifying them if needed, and repeating the whole process from the first step after all modifications, then you should conduct the survey, finalize the documents and questionnaire forms, select your sample, train the interviewers and conduct the survey. Finally, the gathered data should be analyzed by entering data, coding it, preparing the result reports and lastly, finalizing them as technical reports (Kabir, 2016).

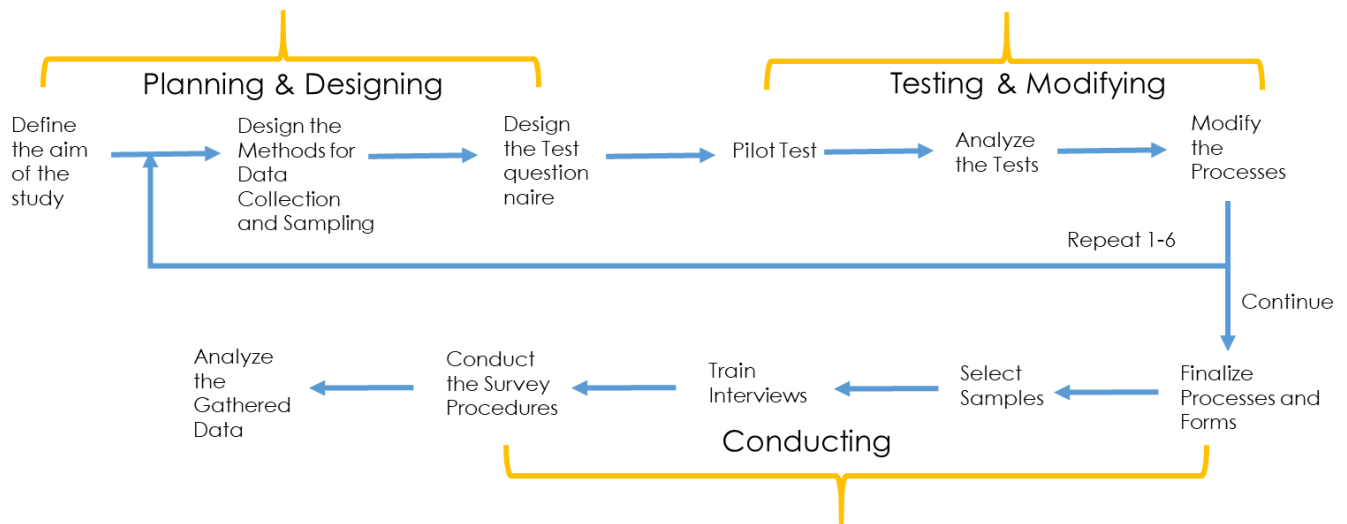


FIGURE. 5. SURVEY PROCESS

## B) Common Issues and Errors

Different issues and errors may happen during the survey process due to the several possible resources. These errors can stem from the following origins:

- Inappropriate planning due to the lack in well-understanding of the study concept;
- Selecting unappropriated survey method including emails, face-to-face, etc.;
- Inappropriate questionnaire design;
- Interviewers issues such as misunderstanding the answers, biases, etc.;
- Sample selection issues such as biased selection, small size;
- Problems of the respondents, for example, memorial mistakes;

Errors can also happen during the processing of the gathered data. For example, in coding processes and also when researchers estimate the results of the survey. To prevent the errors or minimize them as much as possible, researchers should consider different points considering error types, they need to:

- Gain a comprehensive definition about all concepts;
- Use sample selection methods;
- Test the surveys using suitable methods;
- Consider the points of designing a questionnaire;
- Train the interviewers;
- Use introductions and guides for the interviewees;
- Use trained staff for data analysis;
- Use professional statisticians for estimation processes.

### C) Advantages and Disadvantage

As discussed in other proposed methods, this type also possesses several advantages and disadvantages. The important merits and demerits of the survey method can be summarized in Table 2.

TABLE. 2. SURVEY PROS AND CONS (FRECHTLING, 2002; KABIR, 2016)

Merits	Demerits
Covering wide ranges of subjects	Possibility of biased reports
Not expensive	Lack of in-depth comprehensive data
Easy to analyze utilizing different available software	Lack of comprehensive data
Obtaining descriptive information	Possibility of different sources of error such as sampling errors and interviewer errors

### 3.1.6. Case-Studies

Case studies give you the opportunity to investigate issues deeply and descriptively. It covers different concepts ranging from individual(s) to organizations and sectors when the number of sites is small. As they develop a real-world context by observing what happens there and also interviewing the participants, they can help the researcher to study the site precisely and obtain deep explorations. It is basically an empirical investigation that can study different phenomena in a real-life context when the boundaries between them are not clear and also needs a comprehensive investigation. Thus, brief and occasional visits cannot be helpful to achieve this complex task. For example, detailed data gathered through these types are mainly used by a psychologist who needs to gather information about their patients' lives to gain the best treatments for their diseases (Frechtling, 2002; Taherdoost, 2021). Although case studies are not considered as research methods by themselves, researchers conduct different data collection and

analysis methods to gain the material that they need for their study. For example, to gather data in qualitative methods, interviews, and diaries can be used. Furthermore, different official and personally provided notes can be utilized to collect information. Methods such as grounded theory can be used to analyze the achieved data (Kabir, 2016). The common process steps in a case-study method are shown in Figure 6.

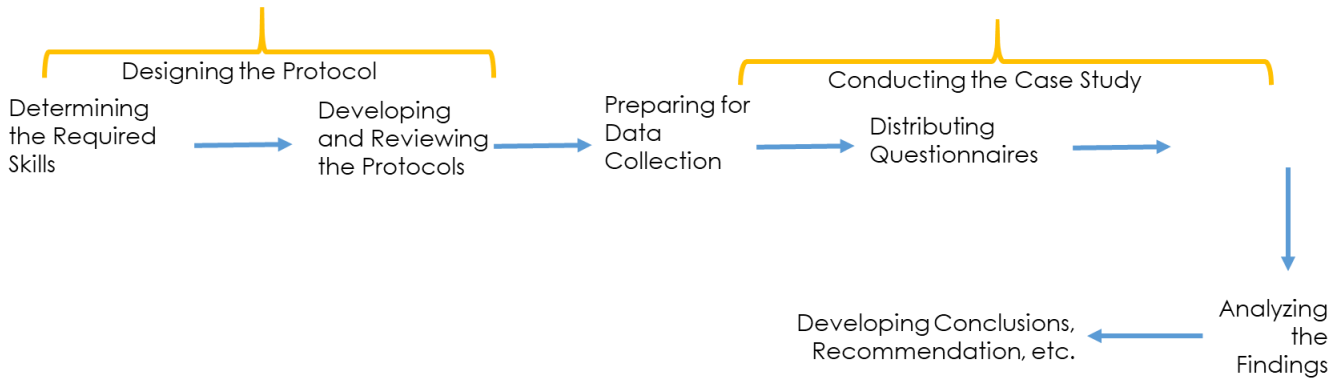


FIGURE. 6. CASE STUDY PROCEDURE

### A) Types of Case Studies

Case studies generally can be categorized into single and multiple ones. Although, these two types can be either narrow or broad. Single case studies are single experiments that can be utilized to analyze a concept deeply, gain an initial result for the next extended multiple groups of cases, and study long-standing theories. Multiple case studies cover different units, and the number of units can positively affect the results of the study. The results of the considered units can be identical or opposite. Holistic cases analyze one unit and embedded analyze two or more sub-units and, in these methods, the number of units is an important factor. Furthermore, considering these four discussed types, case studies can be broad or narrow-based on Table 3 (Taherdoost, 2021).

TABLE. 3. CASE STUDY TYPES

	Case Designs- Scope		
	Type	Narrow	Broad
Analysis	Narrow	Single case Designs (Holistic Analysis)	Multiple Case Designs (Holistic Analysis)
	Broad	Single Case Designs (Embedded Analysis)	Multiple Case Designs (Embedded Analysis)

In addition to the above categories which are based on the number of cases and units of the study, case studies can be categorized as the following:

- ✓ Explanatory: they study the surface and deep levels of data to gain a comprehensive explanation of the phenomenon in data. They are used to match a pattern for a phenomenon in complex casual cases.
- ✓ Exploratory: this type of case study is used to obtain an introduction for further complete studies. It does not consider in-depth factors and helps to collect data to achieve a protocol before developing the research question and hypothesis of the study. Pilot studies are considered as an example of this case study type.
- ✓ Descriptive: the researcher in this case study type aims to achieve a description of data as occurring or in other words the occurred phenomena within the data. Thus, the researcher needs a description basis theory to gain the description of the phenomenon as a result. This can make challenges as the theory can fail from the first step and it can bring different problems during the next steps of the study. Journalistic descriptions are considered descriptive case studies.
- ✓ Prospective: the outcome of the process is the aim of this type of case study. For example, if researchers observe a group of patients to discover what happens to them after a particular period of time, it is a prospective case study type.
- ✓ Retrospective: in this type, researchers consider the historical information for their study case. For example, they gather historical data regarding the patients to collect the risk factors when studying the outcomes of a specific illness.
- ✓ Intrinsic: In this form of case study, the subject of the study is the researcher's interest.
- ✓ Instrumental: the individuals provide observers with more information than just what happens in the setting.

## **B) Information Sources**

Several methods are used to collect data in a case study. Researchers can use a direct observation method in which they observe participants in the natural setting and gather information regarding the reactions and interactions of the individuals. In this method, a single or a group of observers can participate. Another method is participant observation in which the researcher participates in the setting like other under-study people and observes the happenings from a closer perspective.

Interviews are also among the most common methods that can be conducted based on survey-type questions which are structured or based on more open-ended question sets. Finally, researchers can use different census and survey records, newspapers, letters, instruments, tools, etc. (Kabir, 2016).

### C) Advantages and Disadvantages

Several merits can be noted in case-study methods as they:

- Assist specific human sources cases by gathering in-depth information to gain a hypothesis;
- Can utilize different research methods based on researchers' preferences;
- Consider past, present and future aspects together (Kabir, 2016);
- Provide an explanation about the changes and impacting factors that are not always driven from the data. This can bring positive consequences in social units;
- Consider both individuals and their interactions (Taherdoost, 2021).

However, they also provide some challenges as:

- They are complex processes, time consuming and expensive;
- There is no limit for collecting data and it is a challenge for the researcher to realize when it is time to stop the data collecting process;
- The taken basic assumptions are not always realistic and eventually, the gathered data should be tested in these situations;
- The personal documents and letters which are used by the researcher can face distortion problems (Kabir, 2016);
- Case studies need to use expert and trained conducting teams;
- Over-interpreting and over-generalizing issues can also happen in these types (Taherdoost, 2021).

### 3.1.7. Experimental Methods

Experiments are also one of the main methods and even the prime method of data collection. In this method, the researcher can create a research situation, choose the participants and manipulate the independent variables (often limited and small numbers) which can affect a dependent one and explore how they overshadow the dependent variable (Hox & Boeijs, 2005). Briefly, it discovers a causal relationship between them known as effect and causes relationships (Taherdoost, 2021). These methods are objective, and the thoughts of the researchers do not affect the results. Therefore, their validity is high and bias effects are minimized. However, the possible biases can still stem from different aspects including:

- Selection, when you can recognize the differences of the groups at the beginning of the research;
- The placebo effect, when a belief about a study subject can affect the findings, for example, you think the understudy medicines can help you to be better;
- Experimenter, when you as the researcher have a specific belief about the subject (Kabir, 2016).

### A) Types of Experiments with Their Merits and Demerits

Experiments can be conducted in three ways:

- ✓ Laboratory/controlled settings: This provides you with high control over study design and processes. You can gain precise and accurate data although it may cause some problems. For example, in the term of ecological validity (how much the results of our study can be generalized to real-life?) in laboratory settings, according to the artificial degree of the studies. Thus, these methods are better choices when a phenomenon is manageable in the laboratory, not about the variables which are seen in everyday activities in humans' real lives (Hox & Boeije, 2005).
- ✓ Field Experiments: you can conduct it where your participants are in their real-life situations. Although you can still manipulate the variables, the ecological validity of your research is higher as your control is less than the first category.
- ✓ Natural Experiments: These types are done in the participants' natural setting; however, in contrast to two other types, here you have no control over the setting. They can assure a very high ecological validity as it is less likely that different parameters can affect the results. However, on the other hand, the effects of biases are possible due to the control of less extraneous variables which are not independent but can overshadow the result). In addition, other researchers cannot conduct the same study with the exact parameters as the study as well (Kabir, 2016).

#### 3.1.8. Diaries

these self-administered questionnaires ask participants to fill out the forms in two different situations. First, event-based ones ask to fill in when something happens. The second type is known as time-based diaries which should be filled at a particular time frame or interval. On the other hand, diaries can also be elicitation and feedback studies. Elicitation types let the participants utilize media to help their memory. Feedback studies use predesigned questions to get immediate responses.

This method can be used in the following situations:

- Tracking people over time;
- Making a cross-sectional comparison of individuals;
- Investigating processes among families and people.

#### 3.1.9. Activity Sampling Technique

This method utilizes consecutive observations on individuals, groups, and even machines in a specific time frame to record the happening events and achieve a rating if needed. Here, the main aim is to record delays and percentages of the works that are done during a specific period. It employs sampling techniques to define the activity of the understudy subjects and then analyze those using statistical



methods. The number of observations is a key factor since it can affect the accuracy of the study. However, as there are limits due to the time and costs, an optimum number should be utilized based on the approximate number of the occurrence of the factor and acceptable accuracy of the study.

### **3.I.10. Memo Motion Study**

This method is also known as spaced-shot photography, and uses a camera to analyze a long process. As a tool of time, it uses film material for study purposes. Nowadays, this technique is used commonly, as many of the current settings can address the required conditions.

### **3.I.11. Process Analysis**

This analysis is used to gain higher efficiency and decrease wastes by improving understanding of processes. In this method, the process is broken down into different steps including inputs, operational, and output steps. In this method, first, you need to define the process, design a flowchart, define steps' capacity, explore the lowest capacity as the bottleneck of the process, and quantify its impact on the process. Then, you can utilize the gathered data and findings from the sections to enhance the process.

### **3.I.12. Link Analysis**

This is a data analysis method that can be used to explore the links, connections, and relationships between different objects. This analysis can be based on relational, physical, and digital aspects. Thus, data collection in this method can be also based on digital data for example from websites. This method is also utilized in networking. For this, the data that pass the links are analyzed to explore the integrity of the links between each network node. Generally, this method uses known violated patterns to discover anomalies, considers linked objects to match the known patterns of interest, and finally discovers new patterns of interests.

### **3.I.13. Time and Motion Study**

This method uses different techniques and standards to simplify the industrial processes, improve employees' motivation, and gain better efficiency. For this purpose, you need to look at your activities well and find the areas which can improve to gain higher efficiency. Then, you need to make changes and see what the achieved results are. You can rinse and repeat the former steps to gain better results. Therefore, the next step is to make small changes that can bring important merits and consider what and how you conduct the process. Then, you should think about broad items and spot the points that can be achieved in enhancements. Finally, make the change and find out the results.

### 3.1.I4. Statistical Method

These methods are a combination of processes including data collection, summarizing the information, analyzing them, and reporting the findings as quantitative data. The data collection step includes deciding what should be observed and making observations to address the questions using samples that should represent the whole population and be adequate in number. The summarization process includes measuring suitable statistics and displaying them in terms of tables, graphs, etc. Then, using statistical analysis methods such as regression the gathered statistical data can be analyzed and related. Different models can also determine to what extent the gathered information can be utilized in similar projects (Kabir, 2016).

## 3.2. SECONDARY DATA COLLECTION METHODS

The secondary data is defined as the data gathered from published sources. Secondary data are gathered using secondary data methods. This gathered data can be from both qualitative sources such as interview reports and quantitative sources data such as the census. The secondary data collection methods, generally, can be categorized as Table 4.

TABLE. 4. TYPES OF SECONDARY DATA COLLECTION METHODS (KABIR, 2016)

Secondary Data Collection Methods	Description and Credibility Points
Published Printed Sources	In these types, the writer, publication, and publishing time are important.
Books	The most authentic ones among the secondary sources.
Journals/Periodicals	The most important methods as they are up-to-date and provide data and information about very specific subjects.
Magazines/Newspapers	Not highly reliable but can be effective in your research as well.
Published Electronic Sources	Very fast and accessible.
E-Journals	Very available compared to the printed ones. Nowadays these types are very credible as well
General Websites	Not very reliable.
Weblogs	Commonly used methods and provide reliable diaries.
Unpublished Personal Records	Useful in some cases.
Diaries	Useful in descriptive methods, although they can be not very accessible.
Letters	Must be checked in terms of reliability before using.
Government Records	Very useful in different subjects such as human research, social sciences, etc.
Census Data/Population Statistics	Include educational and health records.
Public Sector Records	Provided by private companies and NGO surveys.

#### IV. CHALLENGES OF THE DATA COLLECTION PROCESS

Specific challenges that can occur during each of the data collection methods have been discussed in the previous sections. Here, some other general challenges are explained that can occur due to the following conditions:

##### 4.1. Location of the Data Collection

Location is a critical part of data collection considerations. If you want to conduct a reliable process, you need to use a neutral location. Because the participants need to feel free to provide their responses. Otherwise, there is a high possibility of getting biased answers. For example, when you interview the students in the school atmosphere, they may feel they can face difficulties and troubles if they discuss the issue with their school staff.

##### 4.2. Literacy of the Participants and the Language of the Questions

You need to ensure that the design of the questions is appropriate considering the literacy level of the participants, or you need to select a target sample with a suitable level to respond to your questions. Pilot tests must be conducted to understand whether these cases are considered successful or not. Also, audio assistance methods can be used to address these kinds of issues.

##### 4.3. Timing

As discussed in other sections, different criteria can impact the required time for the data collection process. Here, it can be also added that a suitable schedule is necessary for research projects as the following:

- ✓ First, an adequate time can help participants to be engaged with the project easier and better without stress and other issues, and in contrast, their discomfort can bring different challenges.
- ✓ On the other hand, long study periods can be also problematic as there are possibilities in changing the circumstances for some participants if the length of study is so long.
- ✓ To provide an approximate required time, pilot tests can be also useful.

##### 4.4. Exhaustion of the Research Team

This factor is important as it can negatively affect the results. It can affect the ability of the researchers to:

- ✓ Conduct the data collection processes efficient;
- ✓ Manage the interviews;
- ✓ Control the sessions.

And as a result, the time taken may be wasted. To address these challenges, interviewers and researchers need to take adequate breaks between the processes and utilize a limited set of interviews during a particular day. If they can debrief with their colleagues after conducting a suitable number of interviews, it can also be helpful to face their fatigue.

#### 4.5. Sensitive Data

It is a vital point that participants should feel comfortable during interviews. They should not be distracted or feel embarrassed responding to some sensitive topics for example about their sexual experiences. In these situations, they can also refuse or hesitate to provide their personal information such as their addresses. To address these challenges, you can start interviews with icebreaking conversations, give them the rights of using alias names, and provide an approximate address instead of their exact locations (Rimando et al., 2015).

### V. ETHICAL COMMITMENTS

Ethical considerations should be noted during the data collection processes. It means, for example, the confidentiality of data must be considered. Thus, where it is needed, the data should be presented by an unknown respondent. In other words, there are different legal problems that can stem from research studies. Thus, you need to ensure you consider the necessary points to avoid them. The legal issues can be health risks happening for the volunteer and participants, risks against the reputation of the organizations, and injuring the team. These issues are affected by the laws of the country in which the study is conducted.

Ethical clearance processes are used to assure a project is designed based on the required ethical rules. In this process, it must be examined what kinds of considerations are applied based on the provided guidelines. It can be a simple or complex process for example:

- A process can be simple if there is no participant, and the project is not experimental;
- A process can be complex if you need to pay the participants or provide other kinds of awards (Olsen, 2012).

There are different points that should be considered to conduct the data collection ethically. First, if you need to get personal identifications of the participants like their telephone number, postal address, and email address you should provide the reasons clearly and make them consent about this. Secondly, you need to make it clear whether there is any possibility that respondents can face a specific difficulty if they participate in the research or not. It means they must be aware of it; otherwise, it can put them in jeopardy in some ways. Furthermore, an identification method should be provided for the participants to be sure about who the researchers are. In addition, it must be noted that, if participating in that specific research project is voluntary, participants should feel free about answering the questions as it should not be compulsory.

For the under 14 groups of participants, the consent of their parents should be achieved before their participation (Kabir, 2016). Considering the above points, your provided statement for the committees should cover the subjects such as the following:

- A summary of the processes; helps the committee to discover the positive achievements of the project;
- If the data is provided for specific organizations for example universities, you need to provide the commitment of giving them full access, and you will lodge the data in them by considering the legal aspects for protecting data as well;
- An assessment document listing the risks which participants can face;
- A list of the considered activities can minimize the risks;
- The insurance benefits' possibilities;
- A list providing the participants' activities during the research;
- A list of activities can ensure the confidentiality of the participants for example for sensitive information;
- Finally, informed-consent forms ensure the participants are not forced to participate until the end of the data collection process and they can withdraw when they want.

After providing the documents for the committee, other considerations should be approved. An interview must be required to discuss the processes, methods, and convince them that your research is useful and can provide positive results. The other important point you need to convince them about is why putting participants in trouble or danger is rational and to what extent it can be beneficial.

Furthermore, there are some objects that can cause data-protection issues, and to avoid them some general points can be considered. For example, you should not hold information about the participants in the appointment stage before approving that you have this right. You also should not type the actual names on the questionnaires without their agreements. You also can not hold their MP3 files in your archive without getting their consent. These processes are time-consuming and can take months as discussed; however, there are specific circumstances that this period can be even more. For example, if you conduct a data collection for an organization and your co-researcher is from outside of this organization, the process can be longer, as:

- Differences in regulations can impact the risk factors.
- The other researcher is not conformed to the organization's rules and procedures, and it takes time to address this situation.

To sum, ethics are very complex, and you need to investigate not only the processes but also a variety of different factors which can overshadow them (Olsen, 2012).

## VI. CONCLUSION

This article provides an overview of different data collection methods, the challenges researchers can face conducting these processes, and finally, the ethical issues that must be considered in data collection processes. For this, we first discussed the most common methods including questionnaires, interviews, focus groups, observation, surveys, case studies, and experimental methods. The merits and demerits, as well as the situations each method can be utilized, were also reviewed. Then, the less common methods

were explained shortly. The most important points that were considered to investigate the advantages and disadvantages of these methods are required cost, time, and training factors as well as bias issues, reliability, and validity of the processes.

The general issues and challenges that may occur during data collection processes were explained then and some recommendations were also added to address the issues. Finally, the ethical commitments were studied including the issues that can happen as a result, how the process can be implemented ethically, and also the list of documents that should be provided for the committees to obtain the ethical commitments.

## VII. REFERENCES

- [1] Frechtling, J. (2002). An overview of quantitative and qualitative data collection methods *The 2002 user-friendly handbook for project evaluation* (pp. 43-62).
- [2] Hox, J. J., & Boeije, H. R. (2005). Data collection, primary versus secondary *Encyclopedia of social Measurement* (Vol. 1): Elsevier.
- [3] Data collection challenges (2005).
- [4] Kabir, S. M. S. (2016). Methods Of Data Collection *Basic Guidelines for Research: An Introductory Approach for All Disciplines* (first ed., pp. 201-275).
- [5] Olsen, W. (2012). *Data collection: Key debates and methods in social research* (Vol. 1): Sage.
- [6] Pandey, P., & Pandey, M. M. (2015). *Research Methodology: Tools and Techniques* (Vol. 1). Romania: Bridge Center.
- [7] Rimando, M., Brace, A. M., Namageyo-Funa, A., Parr, T. L., Sealy, D.-A., Davis, T. L., . . . Christiana, R. W. (2015). Data collection challenges and recommendations for early career researchers. *The Qualitative Report, 20*(12), 2025-2036.
- [8] Taherdoost, H. (2016a). How to design and create an effective survey/questionnaire; A step by step guide. *International Journal of Academic Research in Management (IJARM)*, 5(4), 37-41.
- [9] Taherdoost, H. (2016b). Measurement and scaling techniques in research methodology; survey/questionnaire development. *International Journal of Academic Research in Management (IJARM)*, 6(1), 1-5.
- [10] Taherdoost, H. (2016c). Sampling methods in research methodology; how to choose a sampling technique for research. *International Journal of Academic Research in Management (IJARM)*, 5(2), 18-27.
- [11] Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. *International Journal of Academic Research in Management (IJARM)*, 8(1), 1-10.
- [12] Taherdoost, H. (2021). *Handbook on Research Skills: The Essential Step-By-Step; Guide on How to Do a Research Project* (Kindle ed.): Amazon.

## AUTHORS' BIOGRAPHY



Hamed Taherdoost is an award-winning leader and R&D professional. Hamed is a founder of Hamta Group and Faculty member at University Canada West. He was lecturer at IAU and PNU universities, scientific researcher, R&D and Academic Manager at IAU, Research Club, MDTH, NAAS, Tablokar, and Hamta Academy. Hamed has authored over 130 scientific articles in peer-reviewed international journals and conference proceedings (h-index = 29; i10-index = 53), as well as eight book chapters and seven books in the field of technology and research methodology. He is the author of the Handbook of Research Skills and his current papers have been published in Behaviour & Information Technology, Electronics, Computers, Information and Computer Security, Electronics, Annual Data Science, Cogent Business & Management, Procedia Manufacturing, and International Journal of Intelligent Engineering Informatics. He has authored a series of articles in the field of Research Methodology.

He is a Senior Member of the IEEE, IAEEEE, IASED & IEDRC, Working group member of IFIP TC and Member of CSIAC, ACT-IAC, and many other professional bodies. Currently, he is involved in several multidisciplinary research projects among which includes studying innovation in information technology & web development, people's behavior, and technology acceptance.