**Data Collection Methods in Social Sciences: A Primer for Novice Researchers and Students**

**Abstract**

Many early researchers and students sometimes face the challenges with the type of data they need and the methods that can help them collect data for their academic research. Some of them often struggle with selecting appropriate data types and methods, leading to delays in research progress. Scientific data collection is one of the most important stages of academic research. Poor data leads to poor and unreliable estimated results that cannot be used for shaping public policy, and helping governments and other stakeholders make informed decisions. Thus, this review aims to equip beginners with tools to design robust studies and complete their research efficiently. This study reviewed the relevant literature on the types of data and the data collection methods in social sciences. The narrative literature review focused on the studies that are relevant to the topic and are published in English. The study used published articles from databases such as ResearchGate, Scopus, Google Scholar, and Web of Science. The findings depicted that data can be categorized based on three main dimensions. These include the nature of data (Qualitative, quantitative data or Mixed methods) and (Categorical or Continuous), data source (Primary or Secondary data), data collection (Cross-sectional, Longitudinal or Time series data). The literature search also showed that the widely used methods of data collection in social sciences include interviews, focus group discussions, key informant interviews, observations and secondary data sources such as archives, records and published documents. In general, researchers should choose the right type of data and data collection method that can answer their research objectives. This will ensure that they collect valid and reliable data that can guide policymakers and the government.

Key words: Data types, Data collection methods, Social sciences, Research process, Novice researchers

**1.0 INTRODUCTION**

The term “Research” can be described as a methodical and scientific pursuit of new insights or the expansion of existing understanding on a particular subject within a specified timeframe (Koh, 2000). It is a craft of scientific exploration that entails posing questions and the discipline of uncovering dependable answers. It involves seeking to know the unknown about a given object, people, or situation, among others. Research in the social sciences offers numerous advantages, such as gaining insights into human behavior within intricate settings. It aids in clarifying the reasons behind people's thoughts, emotions, and actions across various cultures and communities. Social science research provides evidence that shapes government policies across various timeframes, influencing areas such as education reform, criminal justice, healthcare accessibility, food production enhancement, and poverty alleviation. It also uncovers the underlying causes of significant challenges like gender inequality, discrimination, mental health crises, economic development, inflation, and unemployment, offering policy recommendations to improve decision-making and enhance quality of life. Engaging in research contributes to the enhancement of skills and knowledge by closing the knowledge gap.

Prior to embarking on research in fields such as social science, agronomy, soil science, or medicine, it is important to grasp the entire research process, including the types of data and methods of data collection. This understanding aids in identifying the necessary data and selecting the most suitable data collection techniques to obtain accurate information for informed decision-making. Students and novice researchers are sometimes confused about which data type they need and the data collection methods to achieve their goals (Strangman & Knowles, 2012). Some individuals resort to using structured questions that result in all categorical variables, which may not be suitable for analysis through other methods, such as calculating means. Conversely, other students and novice researchers opt for entirely open-ended questions, leading to inadequate analysis. Additionally, some struggle to identify the type of data necessary to draw conclusions on a specific topic (Ameen et al., 2019). They also find it challenging to understand the research process steps. For instance, a novice researcher might be uncertain about which chapter to begin writing. This has caused numerous students and novice researchers to become stuck in the research proposal development phase for extended periods, while others find themselves going in circles (Fauzan et al., 2022; Zafar et al., 2021). To assist them in learning and applying the appropriate data types, there is a need for updated reviews of data types and data collection methods. This mini-review, therefore, intends to examine the current data types and data collection methods in scientific research. This is particularly important for novice researchers aiming to enhance their research skills. This review therefore equips beginners with data collection tools and methods in order to design robust studies and complete research efficiently.

**2.0 METHODOLOGY**

This study conducted a concise literature review to verify the specific findings. This process included examining published materials such as reports, journal articles, books, and book chapters, among other sources. Data was gathered from reputable databases like Scopus, ResearchGate, Google Scholar, and Web of Science. The key criteria for inclusion include relevance, year of publication and language. Only the relevant, papers published in English were used in this work. The review process encompassed reviewing reliable databases, curating data, analyzing the data, and presenting the results (Figure 1).



*Figure 1: Review process*

**3.0 RESULTS AND DISCUSSION**

**3.1 Research process**

To conduct a successful research study, researchers must adhere to a series of steps, as illustrated in Figure 2. Consequently, research is a sequential process, with each phase highlighting important elements that cannot be overlooked. The entire research procedure is elaborated upon below.

1. **Research problem identification**

The first step in research involves the identification of the research problem to be studied. The research problem represents the central issue or question that the researcher aims to tackle. It pinpoints the particular problem within the study area and explains why it warrants investigation. This problem should be succinct and embody the SMART criteria: Specific, Measurable, Achievable, Reliable, and Time-bound. Identifying the research problem is a pivotal step in the research process, as it forms the foundation of the entire endeavor. Authors must articulate the research problem with clarity. During this stage, the issue should be distinctly outlined (Arthur & Hancock, 2009). Researchers in fields such as social sciences, education, humanities, natural sciences, and formal sciences are required to specify what is already understood and what remains unknown. They should present the knowledge gap in a straightforward manner that is accessible to all readers.

**Example of the research problem**: “*In recent years, Sub-Saharan Africa has witnessed numerous instances of teenage pregnancies. These occurrences have resulted in early marriages, school dropouts, diminished academic achievements, and social stigma for those affected. Several factors may be contributing to the prevalence of teenage pregnancies in Sub-Saharan African nations. Despite the consequences of teenage pregnancies, the underlying causes have not been thoroughly investigated. This study, therefore, seeks to examine the factors leading to teenage pregnancies in Country ABC*”.

This problem statement clearly identifies the burning issue (impact of teenage pregnancies in Sub-Saharan Africa), the context (teenagers), importance (growing concern), and the research focus (marriage, school dropouts, and poor academic performance). A research gap is another component within the research problem identification that most young researchers ignore when discussing the research problem formulation.

A research gap refers to areas or questions within the field that have not been fully explored or studied in sufficient detail and depth. Identifying a research gap is crucial in research because it informs the researcher where there is a lack of knowledge or where existing research is insufficient or outdated.

**How to identify research gaps:**

1. Conducting a literature review: Thoroughly evaluate and scrutinize existing publications in your field of interest to achieve a detailed understanding of the current knowledge landscape and pinpoint any gaps or areas that warrant further exploration. Essentially, examine recent research in your domain to determine if there are aspects that have been overlooked or remain unexplored by other scholars. Identify inconsistencies or contradictions: This involves highlighting instances where various studies, theories, or sources offer findings, interpretations, or conclusions that are not in agreement, sometimes even directly conflicting with one another. Use these discrepancies to critically evaluate the current state of knowledge in your field and suggest potential directions for future research.

One crucial element in identifying research problems that emerging researchers often overlook is the originality of the research. Originality in research signifies the introduction of new, unique, or distinct ideas within a specific academic field. It is a fundamental criterion for high-quality research, as it ensures that a study offers valuable and previously unexplored contributions to academic knowledge. It drives progress, attracts attention from journals, funding agencies and inspires further research. Research is deemed original if it tackles new issues, introduces novel theories or models, employs innovative methodologies, or presents findings that differ from earlier studies. Merely replicating what has already been discovered or examined does not qualify as original. Research originality can be demonstrated by introducing novel theories or models, using existing methods uniquely or applying them to different issues. It involves uncovering new phenomena or providing fresh interpretations of data. Additionally, it includes studying unexamined populations or settings or creating innovative solutions for applications.

1. **Development of research objectives and research questions**

The study's direction is determined by its research objectives and questions. Research objectives are succinct and precise declarations that outline the goals a researcher intends to accomplish through their study. These objectives act as a guiding structure for the research process, aiding in the concentration of efforts, methodologies, and analysis. Effective research objectives should adhere to the SMART criteria. They can be divided into general objectives and specific objectives. The general objective typically represents the broad aim of the study, while the specific objectives are more detailed and targeted goals that elaborate on the general objective. Consequently, researchers must formulate research objectives that are pertinent to their study (Farrugia et al., 2010). These objectives must be straightforward, achievable, and quantifiable. A research question is generally a clear, focused, and specific question that a researcher aims to answer in a particular study. It defines the scope and direction of a study, and it is typically developed from a research problem or gap in existing knowledge. Characteristics of a good research question include focused, researchable, clear and concise, and relevant. It should be noted that the subsequent sections, such as the Literature Review, Methodology, Results, and Discussion, are heavily reliant on the research objectives. These sections are crafted based on the established research objectives. Examples of research objectives based on teenage pregnancy study include;

1. *“To evaluate the rate of teenage pregnancies in country ABC”*
2. *“To determine the factors affecting teenage pregnancies in Country ABC”*
3. *“To assess the effects of teenage pregnancies on education in Country ABC”*

**Table 1. Difference between research objectives and research questions**

|  |  |  |
| --- | --- | --- |
| ***Aspect*** | ***Research Objectives*** | ***Research Questions*** |
| *Purpose* | *States what the research aims to achieve at the end of the study* | *Defines the questions the research seeks to answer at the end of the study.* |
| *Form* | *Declarative (e.g., “To evaluate”, “To determine,” and “To assess”)* | *Interrogative (e.g., “How”, “Why”, and “What”)* |
| *Example* | *To evaluate the rate of teenage pregnancies in the country ABC in Sub-Saharan Africa.* | *What is the rate of teenage pregnancies in the country ABC in Sub-Saharan Africa?* |

1. **Literature review**

In academic research, conducting a literature review is typically the second step. Most academic works, such as Master's and PhD theses, include the literature review as the second chapter. This part involves examining the pertinent literature related to the study's topic. The literature review is crucial because it enables the researcher to pinpoint gaps and highlight them clearly (Hair, 2007). It also aids the researcher in identifying suitable methodological approaches for their work. When crafting the literature review, authors should align it with the study's objectives. This section also encompasses the conceptual and theoretical frameworks. Within the conceptual framework, authors must establish the relationships between various variables in the study, such as the link between dependent and independent variables. Conversely, the theoretical framework pertains to the theories that support the study.

1. **Developing study methodology**

A research methodology is a carefully organized framework that directs the planning, execution, and analysis of research. It includes the comprehensive approach, tactics, and methods employed to tackle research questions or issues, ensuring that the results are trustworthy, dependable, and accurate. The methodology of the study outlines the approach for research design, study area, sample size determination, sampling techniques, data instrument and collection, data analysis, and specification of econometric models (Goundar, 2022). The authors need to clearly describe the data collection methods (such as Surveys, Observations, Focus Group Discussions, Key Informant Interviews, etc.) and the tools, like data collection questionnaires, that will be utilized.

1. **Data collection**

According to Taherdoost (2021), this section of research holds significant importance. Once the study's methodology is established, the researcher proceeds to the field for data collection. It is highly imperative for researchers to follow ethical guidelines during this process, which include obtaining consent from participants before conducting interviews and ensuring the privacy and confidentiality of the information collected, among other considerations (Akaranga & Makau, 2016; Fleming & Zegwaard, 2018; Gajjar, 2013). This will ensure that the study respondents are protected from unethical practices (Midamba et al., 2023).

1. **Data analysis**

Data analysis involves a structured approach to examining, refining, altering, and interpreting raw data to derive valuable insights, identify patterns, and aid in decision-making or research conclusions. This process converts raw data into actionable insights using a variety of quantitative and qualitative methods. The selection of techniques is guided by the research goals, data type, and intended results. It plays an important role in comprehending data, validating theories, predicting trends, and making informed choices across various domains. Researchers need to apply appropriate data analysis methods. Popular software for data analysis includes R, Python, SAS, Stata, SPSS, E-Views, Microsoft Power BI, Tableau, and NVivo, among others, depending on the data type and analysis needs. Interesting models for data analysis encompass linear regression, multiple linear regression, multivariate probit, multinomial logistic regression, Poisson regression, binary regression (probit or logit), ordered probit/logit model, Tobit model, Seemingly Unrelated Regression (SUR), and time series model: Autoregressive Distributed Lag model (ARDL), Vector Autoregressive (VAR), Vector Error Correction Model (VECM) Panel data models (Fixed Effects and Random Effects models), Panel Ordinary Least Squares (POLS) etc. For instance, they might utilize descriptive statistics to condense data, apply t-tests and chi-square tests for comparing two groups, use regression analysis like Ordinary Least Squares to evaluate the impact of independent variables on dependent ones, and conduct correlation analysis to examine the relationship between two variables. It is important to ensure that the data being analyzed is devoid of errors and outliers, which can be removed during the data cleaning process. Additionally, data analysis should align with the research objectives.

1. **Reporting and Discussing the Results**

Subsequently, the researcher should present the outcomes derived from the data analysis. This can be achieved through the use of tables and figures. It is important to ensure that the results reported are both valid and reliable. Following this, a comprehensive discussion should take place. In discussing the results, it is crucial to address the following questions: What were your findings? What do these findings imply? What have other researchers reported? Make sure to support your results with references to 3 to 5 pieces of literature.

1. **Conclusion and Recommendation**

The final phase of research involves drawing conclusions and suggesting policies for implementation. When crafting the conclusion, it is important to do so impartially, aligning each conclusion with its corresponding objective. Therefore, if there are three objectives, three separate conclusions are necessary. Recommendations should be derived from the findings rather than the literature. Authors should propose policies based on their results and suggest any further research if needed.

|  |
| --- |
| **Research Process** |

*Figure 2: Summary of the research process*

**3.2 Types of data in research**

Data is described as "facts or information, particularly when analyzed and utilized to discover insights or make decisions” (Olson, 2021). Data can be categorized in various ways, but there are three primary domains for classifying data. These domains include classification by source, classification by the nature of the data, and classification by data collection methods.

1. **Classification based on nature of data**

In this type of classification, data can be classified as either quantitative or qualitative, categorical or continuous data.

1. **Quantitative data** refers to the numerical data such as the number of infants, height of buildings, income of households, number of people, etc. Quantitative data can be analyze using quantitative software such as Stata, SPSS, SAS, E-views. It is possible to carry out descriptive analysis such as mean, standard deviation, median and mode. Moreover, one can run correlation and regression analysis when dealing with quantitative data. The majority of researchers in social science collect quantitative data since it is easy to analyze, compare and make conclusions. Examples of quantitative data include 30 people, 2500 acres, 234 girls, 20% death rates, 20 kilograms of fertilizer, 200 bags of maize, etc.
2. **Qualitative data encompasses** descriptive, non-numerical details that focus on exploring ideas, feelings, perceptions and traits rather than numerical values (Wesely, 2021). Examples of this type of data include gender, perceptions, emotions, and accessibility. However, qualitative data can be transformed into a quantitative format through coding. For instance, assigning the number 1 to males and 2 to females allows for the analysis of the male-to-female ratio. Qualitative data can be analyzed using methods such as content analysis, narrative analysis, and thematic analysis, among others. Both qualitative and quantitative data hold significance.
3. **Mixed methods** involves combining both quantitative and qualitative data in a study(Dawadi et al., 2021)**.** Mixed methods yields the most effective outcomes and has been widely used and recommended in social sciences (Dawadi et al., 2021; Morgan, 2017)**.** Many researchers such as Wasti et al. (2022) have argued that “By integrating both qualitative and quantitative data, researchers can gain a more in-depth understanding of the research question. Qualitative data provides rich contextual insights and explores the "why" behind observed patterns, while quantitative data offers broad, generalizable insights and identifies statistical trends, according to experts”. This results in more in-depth analysis of a give topic.
4. **Continuous data** refers to countable information that can take on numerical values, including whole numbers, decimals, and fractions. This type of data is suitable for analysis using descriptive statistics such as mean, median, mode, and standard deviation. Additionally, regression analysis can be employed to evaluate the impact of one variable on another, and correlation analysis can also be conducted. Examples of continuous data include 20% of the population, 250 patients, and a temperature of 38.8 degrees Celsius**.**
5. **Categorical data;** This kind of data is organized into categories, making it challenging to perform certain analyses, such as calculating the means of categorical variables. To conduct more sophisticated analyses, this data often needs to be transformed, for instance, by creating dummy variables. Examples of categorical variables include Education (Primary, Secondary, Tertiary), Gender (Male, Female), and Color (Blue, Yellow, Indigo, Red). Categorical data can also be classified as ordinal or nominal data.

*Ordinal data* refers to the categorical data that can be ranked. An example include education level, which can be ranked as primary (lowest level), Secondary (Middle level) and tertiary (Upper level). Another example include ranking the adopters of a given technology as Lower adopters, Middle adopters and Higher adopters.

*Nominal data* on the other hand is the type of categorical data which is not ranked. Examples include Gender of a person which can be Male or Female; Color, which can be Yellow, White, Red; Membership (Member, Non-member), among others.

1. **Classification based on the data source**

The time the data was collected plays a key role when classifying data. Data can be classified as primary or secondary data based on the source of data.

1. **Primary data** refers to the actual data collected from the field. When a researcher collects data from the field, such data is known as primary data. Majority of the researchers and students in social sciences prefer primary data since it has several benefits such as relevance, accuracy, reliability, among others. Unlike secondary data, primary data also offers wealth of variables since the researcher decides the variables to include in their study.
2. **Secondary data** involves data that has already been collected and stored. These include magazines, government archives, newspapers, published work, and institution repositories, among others. The main challenge with secondary fata is that some of them might be outdated and may not provide all the data that the researchers need for their studies. This limits the researchers study scope.

**Advantages of secondary data**

* Secondary data are readily available, this implies that it is easy to access them
* Secondary data are cost effective
* They can be used to compare findings.
* They may also be used to make conclusions

**Disadvantages of secondary data**

* Secondary data may not contain all the variables that one requires
* Some of them are old and outdated.
* It is hard to understand some of the secondary data
* Secondary may contain irrelevant data that might not be useful
* **Classification based on data coding**
1. **Classification based on data collection**

In this type of classification, data can classified as cross-sectional, time series or longitudinal data.

**Cross-sectional data**

Cross-sectional data is defined as "data collected at a single point in time" (Wang &amp; Cheng, 2020). When researchers gather information on a specific subject, such as academic data, at one particular moment, this is referred to as cross-sectional data(Sanchez et al., 2023). The primary benefit of this type of data is its cost-effectiveness, while still yielding valid and reliable results that can be used for drawing conclusions. However, a significant drawback is that cross-sectional data cannot be utilized for making comparisons over time**.**

**Longitudinal data**

Longitudinal data refers to information gathered from the same subjects over a span of time, such as every two or three years. Unlike time series data that collect data on the same unit, longitudinal data involves collecting data from various study units. For instance, it might include assessing the crop yields of numerous farmers before and after they adopt a particular technology, possibly over three growing seasons. This would measure the impact of adoption of a given technology on yields over years. Another example is collecting data on teenage pregnancies rates from different locations before and after community trainings on teenage pregnancies. This would measure the impacts of teenage pregnancies trainings on the rate of teenage pregnancies. The primary benefit of longitudinal data is its ability to facilitate comparisons and assess the impact of newly implemented technologies. However, because it is collected at different intervals, longitudinal data tends to be more costly than cross-sectional data.

**Time series data**

Time series data shares several characteristics with longitudinal data. Both types are gathered at specific intervals, such as every five years or two seasons. However, while longitudinal data examines multiple study units, time series data focuses on a single unit observed over time. For instance, tracking the unemployment rate of a particular country every year is an example of time series data. Another example of time series data is measuring carbon dioxide emission rates in a given country over a span of time. The other example include collecting data on pollution rate of a certain country over 5 years’ time span. Nonetheless, it holds significant value and relevance for impact studies. The main difference between time series and longitudinal data that in time series, measurement is done on the same study units while in longitudinal data, the data is collected from different units of study.

**3.3 Methods of data collection**

Researchers have various methods for gathering data, but this study focused on the five primary methods discussed below.

**3.3.1 Primary data collection methods**

**Interviews**

In the field of social sciences, interviews are a commonly employed method for gathering data. This approach involves a researcher selecting individuals to participate in interviews. The researcher then conducts one-on-one sessions, during which they pose questions directly to the participants and document their responses.

**Types of interviews**

There are two main types of interviews in research.

1. **Phone or video interviews**, where the interview is done on the phone or video. This is mainly applicable to participants who are not easy to access. These may include those in foreign countries.
2. **One-on-one interview** is the second type of interview. This involves sitting down with the respondent and conducting the interview.

**Interview tools**

The primary instrument for carrying out interviews is the questionnaire. This document contains all the questions that the respondent aims to answer. There are three types of questionnaires, which include;

**Structured questionnaire**: This tool features questions with predetermined answers. Respondents choose from the given options, as illustrated in the example below.

|  |
| --- |
| **Questions**1. What is your age range?1. 18-25
2. 26-35
3. 36-46
4. Above 46

2. What is your household size?1. 1-5
2. 6-10
3. Above 10

3. What is your gender?1. Male
2. Female

How much do you earn per month?1. Less than 250$
2. 250-500$
3. 500-750$
4. Above 750$
 |

**Unstructured questionnaire**: This type of questionnaire uses open-ended questions to capture data from the respondents. This type of questionnaire is mainly used to collect qualitative data. An Example is shown below.

|  |
| --- |
| **Questions**What is your age range?……………………….What is your household size?……………………….What is your gender?……………………….How many years did you spend in school?………………………..How much do you earn per month in USD?……………………… |

**Semi-structured questionnaire:** This is the recommended tool for those in social sciences. It incorporates both open and closed questions to capture in-depth information from the respondents as shown in the example below.

|  |
| --- |
| **Questions**1. What is your age range? (Open)

……………………………1. How many people live in your house? (Open)

………………………….1. What is your gender? (Closed)
2. Male
3. Female
4. What is your education level (Closed)
5. Primary
6. Secondary
7. Tertiary
8. Do you have access to credit? (Closed)
9. Yes
10. No
11. If Yes, how much do you earn per month? (Closed)

………………………………… |

**Key considerations when conducting interviews**

1. Make sure your questionnaire includes all the necessary questions to address your research objectives.
2. Follow the suggested sampling method to ensure you have a sufficient number of participants.
3. Arrive early and discuss the scope of your study with the respondent beforehand.
4. Obtain consent prior to conducting the interview.
5. Keep the participants' responses as confidential as possible.
6. Avoid lengthy interviews; it is not advisable for them to exceed two hours.

**Focus Group Discussions**

Focus group discussions (FGDs) are extensively employed to gather qualitative data. This method involves assembling a group of 6 to 12 participants to engage in a brainstorming session on a specific topic, facilitated by a moderator. The primary researcher formulates a series of guiding questions to steer the conversation with the selected participants. The primary objective is to collect detailed insights from the group and utilize this information to draw conclusions on the subject at hand. FGDs are particularly valuable for capturing diverse perspectives from the participants. It is imperative to recognize that FGDs are significant and favored in qualitative research.

**Advantages of FGDs**

* The FGDs help to gather in-depth information
* FGDs helps to generate varied information. This may not be possible with other methods such as interviews
* It is easy to receive instant feedback from the team members
* FGDs allows for flexibility and easy follow ups

**Considerations when conducting FGDs**

* Apply the recommended sampling to select the participants
* Create a free environment where the respondents can share their opinions easily
* Maintain time and do not spend too much time
* Adhere to the research ethics and principles
* Engage all the team members
* Respect each one’s opinions
* Document all the relevant findings
* Make it as interactive as possible

Example of Focus Group Discussion guiding questions on teenage pregnancies

1. “Has your community experienced teenage pregnancies over the last three years?”
2. “In your opinion, what are the causes of teenage pregnancies in your locality?”
3. “What are the possible impacts of teenage pregnancies in your area?”
4. “What can be done to reduce the rate of teenage pregnancies in your area?”
5. “In your opinion, do girls have equal access to education resources as boys?”

**Key informant interviews**

Within any society or community, there are individuals who hold knowledge on specific subjects. These individuals have extensive experience in particular areas and are referred to as key informants. Consequently, key informant interviews involve speaking with people who have specialized insights into a certain topic. Such individuals might include local chiefs, village elders, County Commissioners, and Community Health Practitioners, among others. It is important to note that Key Informant Interviews (KII) are primarily used to collect qualitative data from these key informants. Unlike Focus Group Discussions (FGDs), which are conducted with groups, KIIs are conducted on a one-on-one basis.

**Example of KII:** “*Over the last decades, the rate of insecurity has been on the rise in ABC district. The government has been keen on protecting the lives and properties of its citizens. To recommend policies to reduce the escalating rate of insecurity, the County government of ABC aimed at conducting a study in district ABC among village elders and chiefs to assess the possible causes of insecurity in their district. This study will be done on village elders and chiefs since they handle insecurity cases in their areas”*

**Advantages of Key Informant Interviews**

1. Key Informant Interviews (KIIs) play a key role in obtaining specialized knowledge that can only be acquired from experts. Researchers can gather comprehensive insights through KIIs. Similar to Focus Group Discussions (FGDs), KIIs offer flexibility. They also allow for the collection of sensitive data. Additionally, KIIs are both cost-effective and simple to conduct.
2. KII helps to gather additional information such as local history, people’s beliefs, historical practices, etc.
3. Sine it is done on few people with specialized skills and expertise, KII are generally cost effective, yet it provides first-hand information.

**Observations**

Yeradkar & Jaywant (2020) describe observation as a method of gathering data where the researcher watches and notes behaviors, trends, and patterns in a natural setting without conducting interviews. For instance, this approach can be used to study the behavior of mentally challenged patients without engaging them in interviews.

**Types of observations**

1. Structured observation involves the researcher creating a checklist to guide data collection through observation, with responses recorded according to this list. In contrast, unstructured observation allows the researcher to observe behaviors, trends, and patterns freely, without relying on a checklist.

**Advantages of observations**

1. Observations provide accurate and reliable data
2. Observations provide rich information
3. It is less expensive since it does not need respondents
4. It is time-saving since the researcher does everything without spending time on the respondents.

**3.3.2 Secondary data collection methods**

**Documents, Records and Archives**

These methods pertain to the collection of data from secondary sources. This includes gathering information from sources like archived government documents, institutional records, libraries, newspapers, published journals, historical documents, books, reports, and social media. Researchers can utilize secondary data sources to support primary data.

**Advantages of secondary data**

1. Cost-efficient; numerous secondary data sources, like published articles, are readily accessible and can be downloaded and utilized at no cost.
2. Time-efficient; secondary data saves time as it is easily obtainable.
3. Secondary data can be employed to compare outcomes.
4. Researchers can access extensive data in a brief amount of time.
5. Secondary data can be used to develop research gaps, eg published papers.

**4. CONCLUSION**

The novice researchers including post-graduate do not possess the advanced research methods skills. Sometimes they are not sure of the type of data they need and the data collection methods that can generate valid and reliable results, yet data collection is one of the important stages of research. This makes some of them stuck hence delaying the research process. Strikingly, some end up abandoning their studies for other certification programs that do not require intensive research. This work therefore reviewed the data types and data collection techniques in social sciences. The findings from literature showed that there are different types of data based on different categories. These include the nature of data (Qualitative, Quantitative data or Mixed methods) and (Categorical or Continuous data), data source (Primary or Secondary data), data collection (Cross-sectional, Longitudinal, or Time series data. The literature search also showed that the widely used primary methods of data collection include interviews, focus group discussions, key informant interviews and observations while the main secondary data sources include archives, records and published documents.

**4.1 Key recommendations**

**4.1.1 Recommendation on data collection methods**

* Consider your research objective when choosing data collection method: Some data collection methods such as Focus Group Discussion are suitable for qualitative studies while other like interviews are better for quantitative or mixed studies.
* Consider your budget: Some data collection methods such as interviews require large sums of money. Others like secondary data source such as published work may be cheap.
* Consider the types of respondents you need; methods such as key informant interviews require respondents with certain skills. On the other hand, Focus Group Discussions require not special skills.
* Take into account the depth of research; when you need in-depth information and different opinions from the respondent, Focus Group Discussions will be effective.
* Consider sample size; surveys require large sample size to draw conclusions while methods like focus group discussions does not need large sample size.

**4.1.2 Recommendation on data types**

* Before collecting data, ensure that your data collection tool is complete and contain all the relevant questions that would provide answers to your research questions.
* It is recommended to use semi-structured data collection tool since it involves in-depth analysis, structured tool may be limiting.
* Ensure that you collect the right data you need, this can be qualitative or quantitative data. Some studies require a blend of both qualitative and quantitative for in-depth analysis.
* Consider the type of analysis you will run before collecting data. If you collect all categorical data, it becomes hard to compute means, mode, and medians. Correlation and regression analysis also work best for continuous data.
* If you are interested in comparative analysis, you may been to use longitudinal or time series data since they can compare data across time.
* A blend of secondary and primary data is best. Some secondary data do not contain all the variables you will need. Some are also outdated and cannot be used for decision making.

**4.1.3 Ares for further studies**

This work focused on data types and data collection methods, yet there are other important components of research. Future research should therefore focus on sampling techniques for the novice researchers and students. This will help the students and novice researchers to design the best sampling methods for their work.

**COMPETING INTERESTS DISCLAIMER:**

Authors have declared that they have no known competing financial interests OR non-financial

interests OR personal relationships that could have appeared to influence the work reported in

this paper.

**Disclaimer (Artificial intelligence)**

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**References**

Akaranga, S. I., & Makau, B. K. (2016). Ethical Considerations and their Applications to Research : a Case of the University of Nairobi. *Educational Policy and Entrepreneurial Research*, *3*(12), 1–9.

Ameen, K., Batool, S. H., & Naveed, M. A. (2019). Difficulties novice LIS researchers face while formulating a research topic. *Information Development*, *35*(4), 592–600. https://doi.org/10.1177/0266666918774875

Arthur, A., & Hancock, B. (2009). Introduction to the Research Process Authors. *NIHR RDS for Yorkshire and the Humber*, 1–32.

Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-Methods Research: A Discussion on its Types, Challenges, and Criticisms. *Journal of Practical Studies in Education*, *2*(2), 25–36. https://doi.org/10.46809/jpse.v2i2.20

Farrugia, P., Petrisor, B. A., Farrokhyar, F., & Bhandari, M. (2010). Practical tips for surgical research: Research questions, hypotheses and objectives. *Canadian Journal of Surgery. Journal Canadien de Chirurgie*, *53*(4), 278–281. http://www.ncbi.nlm.nih.gov/pubmed/20646403%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2912019

Fauzan, U., Hasanah, N., & Hadijah, S. (2022). The Undergraduate Students’ Difficulties in Writing Thesis Proposal. *Indonesian Journal of EFL and Linguistics*, *7*(1), 175. https://doi.org/10.21462/ijefl.v7i1.515

Fleming, J., & Zegwaard, K. E. (2018). Methodologies, methods and ethical considerations for conducting research in work-integrated learning. *International Journal of Work-Integrated Learning*, *19*(3), 205–213.

Gajjar, N. B. (2013). Ethical Consideration in Research. *International Journal for Research in Education*, *2*(7), 8–15.

Goundar, S. (2022). Chapter 3 - Research Methodology and Research Method. Cloud computing. *Research Methodology and Research Method*, *Chapter3*, 84–193. https://www.researchgate.net/publication/333015026\_Chapter\_3\_-\_Research\_Methodology\_and\_Research\_Method/comments

Hair, J. F. (2007). Research Methods for Business. *Education + Training*, *49*(4), 336–337. https://doi.org/10.1108/et.2007.49.4.336.2

Koh, W. L. (2000). Meaning of Research. *Researchgate*, 1–10.

Midamba, D. C., Kwesiga, M., & Ouko, K. O. (2023). Determinants of adoption of sustainable agricultural practices among maize producers in Northern Uganda. *Cogent Social Sciences*, *10*(1). https://doi.org/10.1080/23311886.2023.2286034

Morgan, D. L. (2017). Mixed methods research. In *The Cambridge Handbook of Sociology* (Vol. 1, Issue January 2015). https://doi.org/10.1017/9781316418376.015

Olson, K. (2021). What Are Data? *Qualitative Health Research*, *31*(9), 1567–1569. https://doi.org/10.1177/10497323211015960

Sanchez, T. R., Inostroza-Nieves, Y., Hemal, K., & Chen, W. (2023). Cross-sectional study. *Handbook for Designing and Conducting Clinical and Translational Surgery*, 219–222. https://doi.org/10.1016/B978-0-323-90300-4.00030-6

Strangman, L., & Knowles, E. (2012). Improving the Development of Student’s Research Questions and Hypotheses in an Introductory Business Research Methods Course. *International Journal for the Scholarship of Teaching and Learning*, *6*(2). https://doi.org/10.20429/ijsotl.2012.060224

Wasti, S. P., Simkhada, P., van Teijlingen, E., Sathian, B., & Banerjee, I. (2022). The Growing Importance of Mixed-Methods Research in Health. *Nepal Journal of Epidemiology*, *12*(1), 1175–1178. https://doi.org/10.3126/nje.v12i1.43633

WESELY, P. M. (2011). Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences by TEDDLIE, CHARLES, & ABBAS TASHAKKORI . *The Modern Language Journal*, *95*(1), 152–153. https://doi.org/10.1111/j.1540-4781.2011.01158.x

Yeradkar, R., & Jaywant, S. (2020). Impact of Indian Stock Market Due to Crisis in March 2020 I nternational J ournal of M ultidisciplinary E ducational R esearch. *International Journal of Multidisciplinary Educational Research*, *7*(July), 89–103.

Zafar, S., Kamran, M., Raza, H., & Mehtab, A. (2021). Challenges Faced by Novice Research Students at Undergraduate Level: A Qualitative Study at Institute of Education and Research. *Review of Applied Management and Social Sciences*, *4*(2), 411–420. https://doi.org/10.47067/ramss.v4i2.142