Factors to consider when choosing data collection methods

Kelvin M Mwita

Lecturer, Department of Public Services and Human Resource Management, Mzumbe University, P. O Box 1, Mzumbe-Tanzania

ARTICLE INFO

Received 03 February 2022
Received in rev. form 12 May 2022
Accepted 18 June 2022

Keywords:
Data collection methods, research tools, research methodology, qualitative research, quantitative research

JEL Classification:
C8

ABSTRACT

Novice and experienced researchers may encounter numerous methodological challenges when it comes to data collection. Data have to be collected from the right sources using appropriate data collection tools to make research studies impactful. There are numerous data collection methods that researchers may use but not all methods are appropriate for all research studies. Choosing inappropriate research methods may result in invalid research findings and lead to erroneous conclusions. This means researchers have to carefully choose data collection methods to make their studies achieve their respective goals. This study intended to identify factors that researchers have to consider in choosing data collection methods. A systematic literature review was used to analyze 28 journal articles published between 2018 and 2022. From the analysis the study identified ten (10) factors that researchers have to consider in choosing data collection methods. These factors are: the research goal, the scope of the study, sample size, type of data, time and user-friendliness to the subjects. Other factors are research approach, safety and security of a researcher, the need for triangulation and theoretical framework. Before choosing methods for data collection, it is important to put into consideration these factors to increase the chances of achieving research objectives and answering research questions of research projects.

© 2022 by the authors. Licensee SSBFNET, Istanbul, Turkey. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Conducting a research study requires special skills to make it meaningful and impactful (Carberry, et al. 2021; Hendriarto et al., 2021). Novice and experienced researchers may encounter numerous methodological challenges specifically on what data should be collected, from whom to collect and by which means to collect them in order to achieve research objectives (Mwita, 2022; Brown et al., 2018; Barrett & Twycross, 2018).

Admittedly, data collection is a sensitive area in research studies therefore it has to be done in a right way, from right sources with right means (Mkandawire, 2019; Heath et al., 2018). As it is defined, data collection is a systematic process of gathering information needed to answer research questions, solve a particular research problem or/and provide a basis for accepting or rejecting research hypotheses. This gives an implication that when it is not done in a proper way it may fail to answer research questions and ultimately fail to solve research problems. Consequently, one of the most important decisions that each researcher has to make is selection of a method or tool for data collection (Heath et al., 2018).

Data collection methods are means, tools or techniques that researchers use to collect data in their respective research projects. Data collection is always done in the field but there are some cases in which it is done in other places such as libraries, historical archives, and on other online sources (Yin, 2011).

There are numerous data collection methods but not all methods are appropriate in all circumstances (Busetto et al., 2020; Hennink et al, 2019). Each research method has its own strengths and weaknesses. Failure to choose appropriate methods may affect the whole study project and make it less effective. Using innapropriate data collection methods may also result into invalid research findings and erroneous conclusions. Choosing data collection methods is a multi-dimensional phenomenon that needs to put into
consideration different issues and factors to make it less costly and more effective. This study focused on assessing factors that researchers have to consider in choosing data collection methods.

This paper aims to identify factors that researchers have to consider in choosing data collection methods. A systematic literature review was used to analyse 28 journal articles published between 2018 and 2022. From the analysis the study identified ten (10) factors that researchers have to consider in choosing data collection methods. These factors are; research goal, scope of the study, sample size, type of data, time and user-friendliness to the subjects. Other factors are research approach, safety and security of a researcher, the need for triangulation and theoretical framework. Before choosing methods for data collection, it is important to put into consideration these factors to increase chances of achieving research objectives and answer research questions of research projects.

**Conceptual Background**

**Data collection methods**

Before discussing factors that researchers have to consider while choosing data collection methods it is important to briefly explain each data collection tool. The following are data tools that one may use for data collection.

**Questionnaires**

These are documents containing questions or statements that research subjects or respondents have to respond to in order to offer information needed to achieve research objectives. These questions may either be open-ended or close-ended. Open ended questions tend to seek for opinions or clarifications from the respondents therefore respondents have to write down their responses in their own words. Questions like, “How do you think organisational culture affects employee performance?”, “why do employees leave an organisation you are working for?” are examples of open-ended questions. On the other hand, close ended questions always limit respondents on what they are expected to answer. These questions give a respondent a number of options in each question or statement from which the respondent is expected to select the most appropriate or correct option.

**Interviews**

An interview is a method that is used to collect data through words of mouth between a researcher and a respondent. Interviews can be done face-to-face or through the use of technological devices such as telephones and video conferencing. Interviews may either be structured, semi-structured or unstructured. Structured interviews involve asking questions from a list of questions that have been prepared in advance and each respondent is asked the same questions. Semi-structured interviews are more flexible, on top of predetermined questions a researcher is allowed to ask more questions. Unstructured interview offers a room for a researcher to ask various questions which were not prepared in advance. In unstructured interviews a researcher asks questions which he/she finds important for a study without referring to any list of pre-determined questions.

**Observations**

Observations involve data collection by using sense organs. A researcher collects data through seeing, touching, listening, tasting or smelling. Researchers normally, take record of what is being observed. Observations are categorized into participatory observation and non-participatory observation. Through participatory observation a researcher observes a particular phenomenon while he/she is part of the research subjects. If a researcher intends to study on how decision making is done in meetings, he/she has to be part of meeting sessions. In non-participatory observations, a researcher observes a particular phenomenon without being part of the subjects he/she is studying. By considering the previous example, here a researcher observes how meetings are done from a distance; without participating in a meeting in anyhow.

**Focus group discussion**

Focus group discussions involve collection of data through discussions that take place in small manageable groups. For efficiency and effectiveness, each group normally has a person who facilitates the process. In these groups, discussions are always done in form of dialogue through which group members discuss different issues relating to a researched topic as presented by the facilitator. The facilitator decides what issue should be discussed and who should speak.

**Document analysis/documentary review**

Document analysis is a data collection tool that is used to gather information from published work that contains relevant information for a particular study. A researcher assembles various written materials and collects the needed information. Document analysis normally has a limited scope to help reducing the amount of unnecessary and irrelevant information that would be collected. It is recommended to have an observation checklist to achieve that goal.

**Methods**

The study used systematic literature review to collect the needed data from journal articles. Articles were searched from *google scholar* database. A total of 67 articles were obtained from the data base. A total of 28 articles published between 2018 and 2022.
were included for analysis and 39 were excluded. The major criterion of inclusion was justification(s) of using data collection methods. Articles that did not offer reason(s) why they chose particular data collection methods were excluded. Articles included in the study underwent thematic analysis through which the most repeated themes were identified to offer insights for factors that journal articles generally consider in choosing data collection methods.

**Results**

Based on the analysis conducted by this study the following are the factors one has to consider when choosing data collection methods. The factors are also summarized in table 1 with corresponding authors.

**Research goal**

Every research has a particular goal it intends to achieve. Research goal is a major determining factor of what research tool has to be employed in data observation. Observation method for instance is appropriate for a research topic focusing on observing characteristics or behaviors of a subject(s). In the study of Corwin (2018) observation was used since the study intended to observe social interactions among the study subjects. In the studies of Sun et al. (2021); Humphreys et al. (2021); Weber et al., 2019; Radovic et al. (2018); Arifin et al., 2018; Bellhouse et al. (2018); and Leisterer & Jekauc (2019) interviews were used to understand respondents’ experiences. Interviews can also be used to enhance in-depth understanding of researched issues (Konig & Drebler, 2021). Conversely, questionnaires can be used to create numerical values for establishing relationship between variables (Haand & Shuwang, 2020; Palicarpo et al., 2019; Neufeld & Malin, 2019). The danger of not considering research goal in choosing a research method is failure to achieve the set goal. The reason is obvious, data that is collected by inappropriate method may fail to offer any relevant insights to a research problem at hand. Compatibility between a research goal and data collection methods is of most importance.

**Scope of the study**

The extent to which the study will explore in the research area is another important factor to consider in choosing data collection methods. This includes geographical coverage of the study. In the study of Dehghani (2019) semi-structured interviews and focus group discussions were used since the study was conducted in a single organisation. Conversely, studies that cover relatively big geographical area and perhaps involve big number of respondents cannot be easily reached by methods other than questionnaires (Policarpo et al., 2019; Okumah et al., 2019). It should be noted that some studies cover big areas but do not involve big number of subjects. In a situation like this, interviews can be used and if it may take long time to reach all subjects by travelling to interview the respondents in their respective locations, telephone interviews may be appropriate. The study of Humphrey et al. (2021) involved 17 participants from UK and 1 from USA yet data were collected using telephone interviews. Duration is another important aspect of scope of a study which needs to be put into consideration. In some cases, respondents take time to respond to questionnaires. If the study intends to take a short period of time questionnaires will not be appropriate. In most cases interviews, focus group discussion and observations take a very short time to collect data if data collection planning is appropriately done (Sun et al., 2021; Humphreys et al., 2022; Bellhouse et al., 2018). In a situation where a study covers data which are documented from different written sources document analysis method is the most appropriate method (Ruiz-Parrado et al., 2022; Kalogirou et al., 2021; Spitzman, 2019; Sommerhoff et al., 2018).

**Sample size**

Sample size entails number of people or subjects that the study is expected to collect data from. Number of a study’s respondents largely affects selection of data collection tools. Some of research tools are appropriate for relatively large samples and some are not. Interviews and focus group discussions for instance are considered appropriate when a study involves a relatively small sample size. In the study of Dehghani (2019) semi-structured interviews and focus groups were used because the sample size of 32 participants used in the study was relatively small. However, when the study involves a large sample size interviews may not be appropriate. Questionnaires are suitable for large samples since they can easier reach many people at once (Konig & Drebler, 2021; Agha, 2019; Gilakjani et al., 2019; Wunderlich & Veit, 2019). This implies that sample sizes have cost implication as well. Small sample sizes are comparatively less expensive to cover than large sample sizes.

**Type of data (primary or secondary)**

Data may be categorized into primary and secondary data. Primary data is that which is collected for the first time by a researcher from main sources. On the other hand, secondary data is that which has already been collected by other people and is readily available for researchers to use in their studies. Primary data may only be collected by research methods such as interviews, questionnaires, focus group discussion and observation while secondary data may be collected by document analysis. Decision on what tool has to be used for data collection by considering type of data may be done only by using common sense. It is unexpected that one could use document review to collect data from a group of people. The need for primary data in the studies of Sun et al. (2021); Sony et al. (2020); Bellhouse et al. (2018); Humphreys et al.; Radovic et al. (2018) dictated the use of interviews. Same applies to the studies of Konig & Drebler (2021); Agha (2019); Gilakjani et al. (2019); Wunderlich & Veit (2019); Gilakjani et al. (2019) where questionnaires were used. Conversely, the use of secondary data determined the use of document analysis in the studies of Ruiz-Parrado et al. (2022); Kalogirou et al. (2021); Spitzman (2019); & Sommerhoff et al. (2018).
Time factor

In deciding what tool to use, one of important questions that a researcher has to ask him/herself is when am I expecting to collect data and how long to expect the exercise to last. Because of urgency of time a researcher may use interviews to collect data in order to achieve a certain goal in a limited scope of time. Just imagine, an eruption of a not-well-known disease that needs to be immediately addressed occurs, a researcher may opt for interviews to quickly collect data from the victims. In a situation like, questionnaires may be less effective since they may take longer time to be filled and returned while the issue has to be addressed immediately. Moreover, the focus of the study in a particular period of time may also dictate a method to be used. Document analysis tends to collect data which have been collected and documented. A researcher may decide to make analysis of a particular duration and make document analysis the most appropriate method for that (Ruiz-Parrado et al., 2022; Kalogirou et al., 2021; Spitzman, 2019; Sommerhoff et al., 2018). This study also opted for document analysis to get its findings from articles published between 2018 and 2022; a time duration of five years.

User-friendliness to the subjects

Research subjects vary from a place to a place and a project to a project. The differences may dictate what tool may be preferred by a particular group of respondents. For instance, questionnaires are always inappropriate for people who cannot read and write. This may influence a researcher to think of another method such as interviews or group discussions to collect data from the respondents. In some cases, researchers may find low return rate of questionnaires simply because the subjects do not prefer them. A researcher may decide to think of another method which can be more preferred by the subjects such as interviews or focus groups (Heath et al., 2018). In the study of Corwin (2018) participant observation and interviews were used since subjects were old people whose communication abilities using other methods such as questionnaires could not be effective. Additionally, conducting a pilot study may help a researcher to understand whether research methods to be used are preferred by the subjects or otherwise and make a better decision on what methods to be used.

Research approach

Researchers may use qualitative, quantitative or a mixed research approach. Data collection methods may be suitable for one approach and unsuitable for another. A researcher needs to carefully assess appropriateness of a method before embarking to data collection. Quantitative approach for instances matches with questionnaires in most cases since they are capable of providing research results in numerical values (Haand & Shuawng, 2020; Neufeld & Malin, 2019; Okumah et al., 2019; Fasoulis & Kurt, 2019; Palicarpo et al., 2019; Ramachandiran & Dhanapal, 2018). Conversely, Qualitative approach goes together with interviews, focus group discussion and observation (Konig & Drebler, 2021; Sun et al., 2021; Bellhouse et al., 2018; Humphreys et al., 2021; Radovic et al., 2018; Leisterer & Jekauc, 2019; Lewis-Smith et al., 2019). Document analysis may be used in both approaches depending on the nature of the subject. A researcher is expecting to collect. This is due to the fact that document analysis may collect both quantitative and qualitative data. For studies that use a mixed approach, researchers may opt to use data collection methods that are suitable for qualitative and those that are suitable for quantitative approach concurrently (Agha, 2019; Gilakjani et al., 2019; Wunderlich & Veit (2019): Konig & Drebler (2021).

Safety and security of a researcher

Data collection may involve numerous risks in the side of a researcher. Data collection method may expose the researcher to a more risk environment that may harm him/her in the course of data collection. Face-to-face interview or observation to respondents contracted with a pandemic disease such as COVID-19 may not appropriate because it may put the one collecting data in the risk of contracting the disease. Mwita (2020) conducted a study on the effects of COVID-19 on human resource management practices, the researcher used telephone interviews and online survey to avoid risks associated with the disease. Same applies to Sun et al. (2021) who used of telephone interviews to COVID patients. Moreover, in case direct contact with the subjects may lead to security threats a researcher is advised to use data collection methods that may not jeopardize his/her life like doing research relating to terrorism that involves collecting data from terrorists (Morgan, 2022).

The need for triangulation

Triangulation refers to the use of more than one data collection method in a study to offer insights from multiple data collection methods for various reasons including confirming whether the tools will offer similar findings or not. Studies that require more triangulation creates a need for various method. The studies of Konig & Drebler (2021); Califf et al., 2020; Agha (2019); Gilakjani et al. (2019); and Wunderlich & Veit (2019) used both questionnaires and interviews for triangulation purposes. In deciding which methods to be used, other factors narrated above may be taken into consideration.

Theoretical framework

A theory underpinning a particular study may form a basis for deciding what data collection tool or method to be used in that particular study (Collins & Stockton, 2018). For instance, in the study of Okumah et al. (2019) the Theory of Planned Behavior was used to operationalize a questionnaire used for data collection. Further, the grounded theory in most cases forms a basis for collecting
qualitative data through interviews, observations, focus groups and document analysis (Tie et al., 2019). This implies that, a selected theory in a particular study can offer insights of the right data collection tool to be used. This may also start with whether a theory used needs qualitative or quantitative data to validate its assumptions. As noted earlier there are specific methods suitable for qualitative studies and others that are suitable for quantitative studies.

Table 1: Summary of factors to consider and their corresponding authors

<table>
<thead>
<tr>
<th>S/N</th>
<th>Factors</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Research goal</td>
<td>Humphreys et al. (2021); Konig &amp; Drebler (2021); Sun et al. (2021); Sony et al. (2020); Haand &amp; Shuwang (2020); Palicarpo et al. (2019); Neufeld &amp; Malin (2019); Leisterer &amp; Jekauc (2019); Weber et al., 2019; Bellhouse et al. (2018); Arifin et al. (2018); Corwin (2018); Radovic et al. (2018)</td>
</tr>
<tr>
<td>2</td>
<td>Scope of the study</td>
<td>Ruiz-Parrado et al. (2022); Kalogirou et al. (2021); Humphrey et al. (2021); Dehghani (2019); Spitzman (2019); Policarpo et al. (2019); Okumah et al., 2019; Sommerhoff et al. (2018)</td>
</tr>
<tr>
<td>3</td>
<td>Sample size</td>
<td>Gilakjani et al. (2019); Dehghani (2019)</td>
</tr>
<tr>
<td>4</td>
<td>Type of data (primary or secondary)</td>
<td>Ruiz-Parrado et al. (2022); Kalogirou et al. (2021); Sun et al. (2021); Sony et al. (2020); Spitzman (2019); Gilakjani et al. (2019); Sommerhoff et al. (2018); Bellhouse et al. (2018); Humphreys et al., 2021; Radovic et al. (2018)</td>
</tr>
<tr>
<td>5</td>
<td>Time factor</td>
<td>Ruiz-Parrado et al., 2022; Kalogirou et al., 2021; Spitzman, 2019; Sommerhoff et al., 2018)</td>
</tr>
<tr>
<td>6</td>
<td>User-friendliness to the subjects</td>
<td>Bellhouse et al. (2018); Corwin (2018);</td>
</tr>
<tr>
<td>7</td>
<td>Research approach</td>
<td>Konig &amp; Drebler (2021); Sun et al. (2021); Humphreys et al. (2021); Haand &amp; Shuwang (2020); Neufeld &amp; Malin (2019); Okumah et al. (2019); Fasoulis &amp; Kurt (2019); Policarpo et al. (2019); Ramachandiran &amp; Dhanapal (2018); Leisterer &amp; Jekauc (2019); Lewis-Smith et al. (2019); Bellhouse et al. (2018); Radovic et al. (2018).</td>
</tr>
<tr>
<td>8</td>
<td>Safety and security of a researcher</td>
<td>Morgan, 2022; Sun et al. (2021); Mwita (2020)</td>
</tr>
<tr>
<td>9</td>
<td>The Need for triangulation</td>
<td>Konig &amp; Drebler (2021); Califf et al., 2020; Agha (2019); Gilakjani et al. (2019); Wunderlich &amp; Veit (2019)</td>
</tr>
<tr>
<td>10</td>
<td>Theoretical Framework</td>
<td>Okumah et al. (2019); Tie et al. (2019)</td>
</tr>
</tbody>
</table>

Source: Author, 2022

Conclusion

This study intended to identify factors to be considered when choosing data collection methods. Ten factors have been identified from a systematic literature review. These factors include research goal, scope of the study, sample size, type of data, time and user-friendliness to the subjects. Other factors are research approach, safety and security of a researcher, the need for triangulation and theoretical framework.

Given the fact that valid and reliable research findings find their roots in a well-designed and executed research methodology considering these factors increases the chances of making a research study more rigorous.

Acknowledgement

Heartfelt thanks to Dr. Issaya Lupogo of Mzumbe University for proofreading this paper. The author has read and agreed to the published version of the manuscript.

Author Contributions: The whole work was done by the author (K.M.M)

Funding: This research did not receive any funding

Informed Consent Statement: Nature of the study did not require informed consent

Data Availability Statement: The data presented in this study are available on request from the author.

Conflicts of Interest: The author declares no conflict of interest.

References


536


**Publisher’s Note:** SSBFNET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2022 by the authors. Licensee SSBFNET, Istanbul, Turkey. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of Research in Business and Social Science (2147-4478) by SSBFNET is licensed under a Creative Commons Attribution 4.0 International License.