

Research Design and Methods for Developing Individual Research Projects

Maria Sassi

University of Pavia – Italy
maria.sassi@unipv.it



2026

Introduction

The development of strong research capacities requires not only theoretical knowledge but also the effective transfer of methodological skills that allow researchers to design, implement, and communicate rigorous research projects. Knowledge transfer plays a crucial role in strengthening the research process because it enables researchers and doctoral students to acquire practical competencies in research design, literature review, data collection, and data analysis. Through structured training activities, methodological concepts are translated into practical tools that researchers can apply directly to their own research projects, thereby improving the quality, coherence, and feasibility of their studies.

The training programme *Research Design and Methods for Developing Individual Research Projects* is designed precisely to support this transfer of knowledge by guiding participants through the main stages of the research process in a structured and practice-oriented way. The programme provides a step-by-step learning experience in which participants progressively transform an initial research idea into a coherent research proposal by developing a research problem, formulating research questions, identifying a research gap through the literature review, selecting an appropriate research design, and planning data collection and analysis strategies

Within the overall research process, the knowledge presented in the attached document corresponds to the stage that follows the definition of the research design and the identification of data sources. At this stage, researchers must operationalise their conceptual framework and transform theoretical constructs into measurable variables through appropriate data collection instruments. The design of questionnaires and other data collection tools, therefore, represents a critical phase of the research process because it determines how empirical information will be gathered and how accurately the research variables will be measured.

The material used in the training also addresses the methodological phase dedicated to data collection design, with particular emphasis on questionnaire development, the relationship between research variables and measurement instruments, and the principles of validity, reliability, clarity, and logical sequencing of questions. This stage is essential because poorly designed instruments can lead to measurement errors, biased responses, and unreliable results, ultimately compromising the validity of the entire research project.

The topics addressed are:

- How to Formulate a Strong Research Design;
- Literature Review for Research Gap and Statement Definition;
- Definition of a Research Question;
- Selection of a Research Approach;
- Design a Questionnaire;
- Data Cleaning.

By integrating conceptual explanations with practical guidelines, the training supports participants in translating their research objectives and conceptual frameworks into well-structured data collection tools that can generate high-quality empirical data. In this way, the transfer of methodological knowledge contributes directly to improving the rigour, transparency, and credibility of the research process and supports researchers in developing robust and policy-relevant research outputs.

Research Design: A Descriptive Overview

Introduction to Research Design

Research design is a structured plan used to conduct research in a systematic and organised way. It provides a clear framework that guides researchers through all stages of the research process, from defining the research problem to reporting the final results. A well-designed research plan ensures that the research question is addressed logically and that the methods used are appropriate for collecting and analysing data.

When researchers use a structured research design, they can reduce errors, control bias, and increase the reliability and validity of their results. This systematic approach allows the research process to produce meaningful and trustworthy conclusions.

Relevance of Research Design

Research design plays a crucial role in ensuring that research findings are accurate, reliable, and meaningful. It helps researchers organise their study, select appropriate methods, and ensure that the collected data effectively answers the research question.

Three key principles determine the quality of a research design:

- **Validity**
- **Reliability**
- **Objectivity**

These principles ensure that the research findings are accurate, consistent, and free from bias.

Validity in Research

Validity refers to how accurately a study measures what it is intended to measure. A study with high validity produces results that truly reflect the phenomenon being investigated.

There are several types of validity.

Internal Validity

Internal validity refers to the degree to which the results of a study are trustworthy and not influenced by external factors. It ensures that the relationship observed between variables is genuine and not caused by other hidden factors.

For example, a researcher testing a new teaching method may compare two groups of students. One group uses the new teaching method while the other uses the traditional method. If both groups study the same material, receive the same amount of instruction, and take the same exam, any difference in performance can be attributed to the teaching method itself. This control of variables helps ensure internal validity.

External Validity

External validity refers to whether the results of a study can be generalized to other populations, settings, or time periods.

For instance, if a study tests a learning strategy using students from several universities and finds positive results, the researcher may conclude that the strategy could also work in other universities or educational contexts.

Construct Validity

Construct validity refers to how well a research instrument measures the theoretical concept it is intended to measure.

For example, if a researcher wants to measure employee stress levels, they may use a validated psychological questionnaire that measures anxiety, pressure, and emotional exhaustion. Because the instrument measures aspects directly related to stress, the study demonstrates construct validity.

Content Validity

Content validity refers to whether a measurement instrument covers all aspects of the concept being studied.

For example, a mathematics exam that includes questions about algebra, calculus, statistics, and problem-solving covers the full content of the course. Because the test represents the entire subject area, it has high content validity.

Bias in Research

Bias occurs when the research process favours certain outcomes or groups, which can distort the results and lead to inaccurate conclusions. Bias can reduce the validity of a study and undermine the credibility of research findings.

For example, if a researcher wants to know whether students prefer online learning but surveys only students who are already enrolled in online courses, the results will likely be biased. The sample does not represent all students, including those who prefer traditional classrooms.

Several types of bias can occur in research:

- **Sampling bias** occurs when the selected sample does not represent the entire population.
- **Researcher bias** occurs when the researcher's expectations influence the interpretation of results.
- **Response bias** occurs when participants provide answers they believe are socially acceptable rather than truthful.
- **Measurement bias** occurs when the tools used to collect data are unclear or inaccurate.

Confounding Variables

A confounding variable is an external factor that influences both the independent and dependent variables in a study. This variable can distort the relationship between the variables being studied.

For example, a researcher may investigate whether drinking coffee improves student concentration. The results might show that students who drink coffee perform better. However, another factor, such as sleep quality or study habits, may also influence concentration. If these factors are not controlled, they become confounding variables that affect the results.

Reliability in Research

Reliability refers to the consistency or repeatability of research results. If a study is reliable, it produces similar results when repeated under the same conditions.

There are several types of reliability.

Test–Retest Reliability

Test–retest reliability examines whether the same test produces similar results over time.

For example, a researcher may measure students' stress levels using a questionnaire. If the same questionnaire is given again after a short period and produces similar results, the instrument demonstrates high test–retest reliability.

Inter-Rater Reliability

Inter-rater reliability refers to the consistency of results when different observers evaluate the same situation.

For example, two trained observers may watch the same classroom discussion and evaluate student participation using the same checklist. If both observers give similar ratings, the measurement method shows strong inter-rater reliability.

Internal Consistency

Internal consistency refers to how well the items in a questionnaire measure the same concept.

For example, a survey designed to measure job satisfaction may include several statements about workplace happiness, motivation, and willingness to recommend the workplace. If participants who report high satisfaction give similar positive responses across these items, the survey demonstrates internal consistency.

Objectivity in Research

Objectivity means that research findings are based on evidence rather than personal opinions or biases. An objective study uses standardised procedures and transparent methods so that the results can be verified or replicated by other researchers.

Indicators of objectivity include:

- Use of standardised procedures
- Transparent research methodology
- Reproducible results

For example, when studying whether remote work improves employee productivity, a researcher might measure productivity using objective indicators such as completed tasks or meeting deadlines. Anonymous surveys may also be used to collect honest responses from employees. By applying clear and standardised methods, the research remains objective and unbiased.

How Research Design Helps Answer Research Questions

A well-structured research design helps answer research questions in a systematic and logical way. It organises the research process by defining how data will be collected, analysed, and interpreted.

For example, if a researcher wants to determine whether online learning improves student performance, the research design will specify:

- who will participate in the study
- how performance will be measured
- what data will be collected
- how the results will be analysed

This organised structure ensures that the study directly addresses the research question.

Key Benefits of Research Design

A strong research design provides several important benefits.

Provides a Clear Structure

Research design outlines what data will be collected, how it will be collected, and how it will be analysed. This structure helps researchers follow logical steps rather than collecting data randomly.

Aligns Methods with the Research Question

The research design ensures that the chosen methods, such as surveys, interviews, or experiments, are appropriate for answering the research question. The collected data, therefore, becomes relevant and meaningful.

Reduces Bias and Errors

By defining procedures and controlling variables, research design helps minimise bias and mistakes. For example, randomly assigning participants to groups ensures that results are not influenced by the researcher's preferences.

Ensures Consistency

Research design standardises procedures so that all participants are treated equally and measured using the same tools. This consistency improves the reliability and validity of the study.

Makes Results Easier to Interpret

When research is carefully structured, the results become easier to analyse and interpret. Researchers can clearly determine whether the observed outcomes are caused by the factors being studied.

Steps in the Research Design Process

The research design process follows a series of organised steps that guide the researcher from the initial idea to the final conclusions.

1. Identify the research problem
2. Review the existing literature
3. Formulate research questions or hypotheses
4. Choose the research design (qualitative, quantitative, or mixed methods)
5. Select the population and sampling method
6. Choose data collection methods such as surveys, interviews, or observations
7. Plan the method of data analysis
8. Address ethical considerations such as consent and confidentiality
9. Conduct a pilot study if necessary
10. Collect the data
11. Clean, analyze, and interpret the data
12. Report the research findings
13. Draw conclusions and make recommendations

These steps ensure that the research is conducted systematically and that the results are credible and meaningful.

Conclusion

Research design is a fundamental component of the research process. It provides a structured plan that guides researchers through each stage of their study. By ensuring validity, reliability, and objectivity, research design helps produce accurate and trustworthy results.

A well-designed study also reduces bias, improves consistency, and makes findings easier to interpret. Following a clear research design process allows researchers to answer research questions effectively and contribute valuable knowledge to their field.

Literature Review in Research: A Descriptive Overview

Introduction

A literature review is an essential component of a research paper. It is the section where researchers examine, analyse, and summarise existing studies related to their research topic. By reviewing previous research, scholars can understand what has already been studied, identify the main theories and findings in the field, and recognise areas that require further investigation.

A well-written literature review does more than simply summarise previous studies. It critically evaluates existing knowledge and demonstrates how the new research contributes to the academic conversation. Through this process, researchers build a strong foundation for their study and justify the need for their research.

Importance of a Literature Review

The literature review plays a crucial role in the research process for several reasons.

First, it helps show what is already known about the research topic. By examining previous studies, theories, and empirical findings, researchers can present the current state of knowledge in a particular field. This helps readers understand how the topic has been studied and what conclusions have been reached.

Second, a literature review helps identify research gaps. By carefully analysing existing studies, researchers can determine which questions remain unanswered or which areas have not been sufficiently explored. Identifying these gaps allows researchers to design studies that contribute new knowledge.

Third, the literature review provides context for the research. It places the study within the broader academic discussion and shows how the research relates to previous work. This contextualization demonstrates the relevance and contribution of the study.

Another important role of the literature review is to justify the need for the research. By highlighting limitations or unanswered questions in previous studies, researchers can explain why further investigation is necessary.

Additionally, the literature review helps develop the research framework. It often guides the formulation of research questions, hypotheses, and theoretical frameworks by building upon ideas proposed by earlier researchers.

Finally, reviewing previous studies helps prevent duplication of research. By understanding what has already been done, researchers can avoid repeating existing work and instead focus on contributing new insights to the field.

The Concept of a Research Gap

A research gap refers to a question or issue that has not been adequately addressed in existing research. In other words, it represents missing or insufficient information in the current literature.

Identifying a research gap is a crucial step in developing a research study. When researchers discover areas that have not been fully explored, they can design studies aimed at addressing these gaps and advancing knowledge in the field.

A research gap statement typically highlights three important elements. First, it acknowledges existing research and demonstrates familiarity with the literature. Second, it identifies what is missing or insufficient in the current knowledge. Third, it explains why addressing the gap is important for advancing understanding or solving practical problems.

Through this process, researchers can formulate research questions that directly address the identified gap.

Types of Research Gaps

Research gaps can take different forms depending on the nature of the missing knowledge. Several common types of research gaps are recognised in academic research.

Evidence Gap

An evidence gap occurs when existing studies produce conflicting or inconsistent findings. When research results contradict one another, it becomes difficult to draw clear conclusions. In such situations, further research is required to clarify the relationship between variables and determine which findings are most accurate.

Evidence gaps highlight uncertainty in the literature and indicate the need for additional investigation to resolve contradictions.

Knowledge Gap

A knowledge gap exists when little or no research has been conducted on a particular topic. These gaps often occur in emerging fields or interdisciplinary areas where new technologies, theories, or social phenomena are being explored.

Knowledge gaps provide opportunities for innovative research because they involve topics that have not yet been thoroughly examined.

Practical Knowledge Gap

A practical knowledge gap refers to a difference between what research suggests and what occurs in practice. In some cases, research findings recommend certain practices or solutions, but these recommendations are not widely adopted in real-world situations.

This type of gap highlights the need to understand why theoretical knowledge is not applied in practice and what barriers prevent implementation.

Methodological Gap

A methodological gap occurs when existing studies rely heavily on certain research methods that may not fully capture the complexity of the problem being studied. This gap suggests the need for new or improved research methods that can provide deeper insights or more accurate data.

Identifying methodological gaps encourages researchers to develop innovative approaches to data collection and analysis.

Empirical Gap

An empirical gap exists when theoretical ideas or assumptions have not yet been supported by sufficient empirical evidence. Although theories may suggest certain relationships or outcomes, these claims may not have been tested with real-world data.

Research addressing empirical gaps focuses on collecting and analysing data to validate or challenge theoretical assumptions.

Theoretical Gap

A theoretical gap arises when existing theories do not adequately explain a particular phenomenon. This situation often occurs when theories developed in one context are applied to different contexts where they may not fully account for local conditions or social factors.

Addressing theoretical gaps may involve refining existing theories or developing new theoretical frameworks.

Population Gap

A population gap occurs when certain groups are underrepresented or excluded in research studies. When research focuses primarily on a specific group, such as a particular demographic or geographic population, the findings may not reflect the experiences of other groups.

Research addressing population gaps aims to include underrepresented populations to provide a more comprehensive understanding of the research topic.

Characteristics of a Good Research Gap

A strong research gap should possess several important characteristics that ensure the research contributes meaningful knowledge.

Relevance

A research gap should address issues that are relevant to the field of study and important to researchers, practitioners, or policymakers. Research that focuses on relevant problems is more likely to influence policy, practice, or future research.

Novelty

A research gap should involve a new or insufficiently explored topic or examine an existing issue from a new perspective. Novel research contributes fresh insights and expands the existing body of knowledge.

Significance

The research gap should be significant enough to justify further investigation. It should address an issue that can advance understanding, influence policy, or improve professional practice.

Feasibility

A research gap must also be feasible to investigate. Researchers must consider whether the study can realistically be conducted using available resources, time, data, and research methods.

Clarity

A research gap should be clearly defined and easy to understand. It should not be too broad or too vague, as this can make the research difficult to focus and manage.

Specificity

The research gap should be specific enough to guide the development of clear research questions and methods. A well-defined gap helps researchers design focused and targeted studies.

Timeliness

Research gaps should address current or emerging issues. Studying timely topics ensures that the research remains relevant and contributes to ongoing academic discussions.

Originality

A good research gap should represent an original contribution to knowledge. The research should explore new problems, contexts, or relationships rather than repeating previous studies.

Researchability

The research gap must be researchable, meaning that it can be investigated using appropriate research methods and that relevant data can be collected and analysed.

Importance

Finally, the research gap should be important enough to justify the effort and resources required to conduct the study. Research should focus on problems that have meaningful implications for theory, policy, or practice.

Poor Research Gaps

A poor research gap occurs when the proposed research does not add meaningful value to the existing body of knowledge. This may happen when the gap is too vague, too broad, or lacks clear direction.

For example, a research gap that simply states that there is a lack of research on a topic without identifying a specific problem or research question is insufficient. Similarly, a gap that claims that no research exists on a topic is often unrealistic and fails to demonstrate familiarity with existing literature.

A poorly defined research gap makes it difficult to formulate clear research questions and weakens the justification for the study.

Identifying a Research Gap

Identifying a research gap requires a systematic approach. The first step is to define a broad area of interest. Researchers should select a topic that is relevant to their field while ensuring that it is focused enough to allow detailed investigation.

The next step involves conducting a systematic literature review. By examining existing research, scholars can identify patterns, debates, and areas where knowledge is incomplete.

Researchers should look for signals that indicate potential gaps in the literature. These signals may include underexplored topics, contradictory findings, outdated data, limited research scope, or a lack of diversity in research samples.

Another useful strategy is to map existing knowledge. Creating conceptual maps or charts can help researchers visualise areas where research is dense and areas where information is limited.

Engaging with experts such as professors, industry professionals, or experienced researchers can also help identify emerging issues and unresolved questions in the field.

Finally, researchers should evaluate whether the identified gap is relevant, researchable, and capable of contributing new knowledge.

Purpose of the Literature Review

The literature review serves several important purposes in a research study.

It provides a background for the research by summarising relevant studies and academic discussions. It also classifies research into different categories and examines how knowledge in the field has evolved over time.

Additionally, the literature review highlights areas of agreement and disagreement among researchers. By evaluating previous studies, researchers can identify knowledge gaps and justify the need for their research.

Another important purpose is to help avoid unnecessary duplication of research. By understanding existing work, researchers can focus on contributing new insights.

The literature review also supports conceptual clarity and helps researchers develop theoretical frameworks and research questions.

Types of Literature Reviews

There are several types of literature reviews commonly used in academic research.

Narrative Literature Review

A narrative review provides a broad overview of a research topic from a theoretical or contextual perspective. In this type of review, the author summarises and discusses the most important studies related to the topic.

However, narrative reviews may be subjective because they do not always follow a strict methodological procedure for selecting and analysing studies.

Systematic Literature Review

A systematic literature review follows a structured and transparent process to identify, evaluate, and synthesise all relevant studies on a particular research question.

Systematic reviews use predefined search strategies, databases, and inclusion criteria to ensure that the review is comprehensive and unbiased.

Meta-Analysis

Meta-analysis is a statistical method used to combine results from multiple quantitative studies. It allows researchers to calculate overall effect sizes and identify general trends across different studies.

Unlike traditional literature reviews, meta-analysis uses statistical techniques to produce objective summaries of research findings.

Meta-Synthesis

Meta-synthesis is a method used to combine findings from multiple qualitative studies. Instead of statistical analysis, this approach focuses on interpreting and integrating qualitative insights to develop a deeper understanding of the research topic.

The PRISMA Protocol in Systematic Reviews

PRISMA, which stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses, is a guideline used to ensure transparency and quality in systematic reviews.

The PRISMA protocol outlines a structured process for identifying, selecting, and evaluating research studies.

The PRISMA flow diagram describes four main stages of the systematic review process:

1. Identification of relevant studies
2. Screening of titles and abstracts
3. Assessment of full-text articles for eligibility
4. Inclusion of studies in the final review

Following the PRISMA guidelines improves the credibility and replicability of literature reviews.

Conclusion

The literature review is a fundamental component of academic research. It provides a comprehensive understanding of existing knowledge, identifies research gaps, and establishes the foundation for new studies.

By examining previous research, scholars can position their work within the broader academic context and justify the importance of their research questions. A well-structured literature review also helps prevent duplication of work and supports the development of theoretical frameworks.

Ultimately, conducting a thorough literature review allows researchers to contribute meaningful insights that advance knowledge and address important problems within their field.

Developing Research Questions and Research Approaches: Concepts and Methods

Introduction

A research question is a central element of any research project. It defines the main issue that the study aims to investigate and guides the entire research process. A well-formulated research question helps researchers focus their investigation, select appropriate methods, and interpret results effectively.

Developing a strong research question requires careful consideration of the research problem, the availability of data, and the methodological approach. The process also involves identifying research gaps in existing literature and selecting suitable research strategies for collecting and analysing data.

What is a Research Question?

A research question is a clear and focused inquiry that a study aims to answer through systematic investigation. It represents the central objective of the research and provides direction for the research design, data collection, and analysis.

A good research question is not simply a general topic or idea. Instead, it must clearly define the problem being studied and outline the specific issue that the researcher intends to explore. By doing so, it ensures that the research remains structured and coherent.

Characteristics of a Good Research Question

All research questions should possess several key characteristics that make them suitable for academic investigation.

First, a research question should be **focused on a single problem or issue**. This helps maintain clarity and ensures that the research does not become too broad or unfocused.

Second, it should be **researchable** using either primary or secondary data sources. The researcher must be able to collect relevant information through surveys, interviews, experiments, datasets, or existing scholarly literature.

Third, the research question should be **feasible** to answer within the time, resources, and practical constraints of the study.

Fourth, it should be **specific enough** to allow for thorough investigation and analysis. Vague or overly broad questions make it difficult to design effective research methods.

Fifth, a good research question should be **complex enough** to allow for discussion, interpretation, and critical analysis rather than a simple factual answer.

Finally, it should be **relevant** to the researcher's field of study and ideally contribute to solving a broader academic or societal problem.

Focused and Researchable Questions

A strong research question must remain focused on a single topic or problem. The main research question should work closely with the research problem to maintain clarity and direction in the study.

When multiple questions are used in a study, they should all relate directly to the central objective of the research. This ensures that each part of the research contributes to answering the main question.

Another important characteristic of a research question is that it must be answerable using credible sources. Credible sources are reliable and unbiased sources of information that are supported by evidence and produced by trustworthy authors or institutions.

Researchers must also avoid research questions based on subjective judgments. Words such as "better," "good," or "bad" introduce value judgments that cannot be objectively measured or analysed.

Feasibility and Specificity of Research Questions

Feasibility refers to the practicality of answering the research question within the available time, data, and resources. If the required data is difficult to obtain or the scope of the research is too large, the question should be refined or narrowed.

Specificity is another essential feature of a research question. The terms used in the question must be clearly defined and easily understood. Vague language or broad concepts make it difficult to design appropriate research methods and collect relevant data.

In addition, a strong research question should not require a direct policy recommendation or solution. Instead, it should focus on understanding relationships, patterns, or explanations related to the research problem.

Complex and Arguable Research Questions

Research questions should be complex enough to require analysis and interpretation. Questions that can be answered with a simple “yes” or “no” do not provide sufficient scope for academic research.

Similarly, questions that can be answered quickly using easily accessible facts are not suitable for research. A strong research question requires the examination of multiple sources of information and may involve the analysis of original data.

Such complexity encourages critical thinking and allows the researcher to develop well-supported arguments based on evidence.

The Role of Sub-Questions

In many research projects, the main research question is supported by several sub-questions. These sub-questions break down the main question into smaller and more manageable parts.

Sub-questions should be less complex than the main research question and should focus on specific aspects of the problem. They should also be logically organised so that answering them helps build a complete answer to the main question.

Typically, researchers limit the number of sub-questions to four or five. If too many sub-questions are required, it may indicate that the main research question is too broad and needs refinement.

Developing a Research Question

Developing a research question involves several important steps.

First, researchers must choose a topic of interest within their field of study. This topic should be broad enough to explore but also relevant to existing academic discussions.

Second, researchers should conduct preliminary reading to understand the current state of knowledge in the field.

Third, the focus should be narrowed to a specific issue or niche area within the broader topic.

Finally, researchers identify a research problem that needs to be addressed. The research question is then formulated to investigate this problem.

Research Approaches

Research approaches refer to the overall strategies used to conduct research and answer research questions. Different approaches are used depending on the research objectives and the type of data required.

Two common approaches are **inductive** and **deductive** reasoning.

Inductive research begins with observations or data and develops broader generalisations or theories based on these observations.

Deductive research starts with an existing theory or hypothesis and tests it using data and empirical evidence.

Qualitative and Quantitative Research

Research methods can also be categorised as qualitative or quantitative depending on the type of data collected.

Quantitative Research

Quantitative research involves collecting and analysing numerical data. It is commonly used to measure variables, identify patterns, test hypotheses, and make predictions.

Quantitative research often uses statistical analysis to examine relationships between variables and to generalise findings to larger populations.

Some advantages of quantitative research include the ability to replicate studies, compare results across different settings, and analyse large datasets. The standardised nature of quantitative data collection also makes it possible to test hypotheses using established statistical methods.

However, quantitative research also has limitations. It may oversimplify complex concepts, ignore contextual factors, or introduce structural biases due to measurement errors or sampling problems.

Qualitative Research

Qualitative research focuses on understanding concepts, experiences, and social processes through non-numerical data such as interviews, observations, or textual analysis.

This approach is often used when researchers aim to gain deeper insights into human behaviour, perceptions, and social interactions.

Qualitative methods are flexible and allow researchers to adapt their research design as new patterns emerge during the study. Data is often collected in natural settings, providing detailed descriptions of participants' experiences.

Despite these strengths, qualitative research also has some limitations. Results may be difficult to replicate because interpretations depend on the researcher's perspective. Additionally, smaller sample sizes may limit the ability to generalise findings to broader populations.

Mixed Methods Research

Mixed methods research combines both qualitative and quantitative approaches within a single study. This approach allows researchers to benefit from the strengths of both methods.

Quantitative data provides measurable evidence and statistical analysis, while qualitative data offers deeper insights into underlying experiences and explanations.

By integrating both types of data, mixed methods research can provide a more comprehensive understanding of complex research problems.

Data Collection and Sources

Data used in research can be obtained from two main sources: primary data and secondary data.

Primary data refers to new information collected directly by the researcher through surveys, experiments, interviews, or observations. Primary data allows researchers to design data collection specifically to answer their research question.

Secondary data refers to existing information collected by other organisations or researchers. This includes public datasets, government statistics, and previously published research.

Primary data offers advantages such as relevance to the research question and control over data collection procedures, while secondary data allows researchers to access large datasets without conducting their own data collection.

Population and Sampling

In research, the population refers to the entire group that the researcher wants to study. Because it is usually impossible to collect data from every individual in the population, researchers select a smaller group known as a sample.

Sampling methods determine how individuals are selected from the population. The choice of sampling method affects the reliability of the results and the ability to generalise findings to the broader population.

Two main types of sampling methods are probability sampling and non-probability sampling. Probability sampling is generally considered more statistically reliable because each individual in the population has a known chance of being selected.

Case Selection in Qualitative Research

In qualitative research designs, traditional sampling methods may not always be necessary. Instead, researchers often select cases that provide meaningful insights into the research problem.

For example, in ethnographic studies or case studies, researchers may focus on a specific community, organisation, or social setting. The goal is not necessarily to generalise findings to a larger population but to gain a deep and detailed understanding of a particular context.

The choice of case must be carefully justified and should be directly relevant to the research question.

Conclusion

Developing a strong research question is a critical step in the research process. A well-designed question provides direction for the study, guides the selection of research methods, and helps ensure that the research produces meaningful results.

Researchers must carefully consider the characteristics of effective research questions, including focus, feasibility, complexity, and relevance. They must also select appropriate research approaches and methods based on the type of data required to answer the question.

By combining well-defined research questions with suitable research methods, researchers can generate reliable evidence and contribute valuable knowledge to their field of study.

Designing a Questionnaire for Research: Principles, Methods, and Data Quality

Introduction

Questionnaires are one of the most widely used tools for collecting data in social science and development research. They allow researchers to gather information directly from respondents in a systematic and efficient manner. A well-designed questionnaire enables researchers to collect reliable and relevant data that can be used to answer research questions and test hypotheses.

Because questionnaires often involve collecting information from a large number of individuals within a limited time frame, the design of the questionnaire plays a critical role in ensuring the accuracy and usefulness of the collected data. Poorly designed questionnaires may produce ambiguous responses, incomplete information, or biased results. Therefore, careful planning and methodological rigour are essential when designing questionnaires.

Purpose of a Questionnaire

A questionnaire is primarily used to collect information from respondents in order to answer specific research questions. It provides a structured set of questions that respondents must answer, allowing researchers to gather standardised information across a large group of participants.

The questionnaire method is particularly useful when researchers need to collect data from many individuals within a relatively short period of time. It is widely used in surveys, policy evaluations, market research, and development studies.

However, the usefulness of a questionnaire depends largely on its design. A poorly designed questionnaire can lead to incorrect conclusions, as respondents may misunderstand the questions or provide inaccurate answers.

Characteristics of a Good Questionnaire

A good questionnaire should possess several essential qualities that ensure the quality of the data collected. These include validity, reliability, clarity, interest, and succinctness.

Validity

A valid questionnaire measures exactly what it is intended to measure. The questions must be clearly phrased so that respondents understand the purpose of each question. If a question is vague or confusing, respondents may interpret it differently, leading to inaccurate data.

To ensure validity, the questionnaire should be reviewed by experts during the pilot testing phase. Experts can verify whether the questions accurately reflect the research objectives and whether they measure the intended variables.

Reliability

Reliability refers to the consistency of responses obtained from the questionnaire. A reliable questionnaire should produce similar answers if the same respondent is asked the same question again within a short period of time.

One way to assess reliability is through the test–retest method, where the same questionnaire is administered to respondents at two different times. If the responses remain consistent, the questionnaire can be considered reliable.

Reliable questions are typically clear, specific, and based on factual information rather than subjective opinions.

Interesting Questionnaires

A questionnaire should also be interesting to respondents in order to encourage participation and improve response rates. When respondents find the questions relevant and easy to answer, they are more likely to complete the questionnaire carefully.

Interest can be maintained by ensuring that the questions relate directly to the respondents' experiences and by arranging the questions in a logical sequence. Starting with general questions and gradually moving to more specific ones helps respondents remain engaged throughout the survey.

Succinct Questionnaires

A succinct questionnaire includes only the questions that are necessary to answer the research objectives. Questions that are not directly related to the research problem should be excluded.

Short and focused questionnaires reduce the burden on respondents and increase the likelihood that participants will complete the survey. They also help researchers collect relevant data without unnecessary complexity.

Designing a Questionnaire: Conceptual Framework

The first step in designing a questionnaire is to develop a conceptual framework. A conceptual framework outlines the relationships between the variables being studied and helps researchers determine what information must be collected.

The conceptual framework should be developed based on existing literature, theoretical models, and discussions with subject experts. It provides a structured representation of how different variables interact in the research study.

Constructing the conceptual framework before data collection ensures that the questionnaire is aligned with the research objectives.

Types of Variables in a Conceptual Framework

Several types of variables are typically included in a conceptual framework.

Independent Variables

Independent variables represent the factors that influence or explain an outcome. They are also known as explanatory or predictor variables.

Dependent Variables

Dependent variables represent the outcomes that researchers seek to explain. Their values depend on changes in the independent variables.

Mediating Variables

A mediating variable explains the mechanism through which the independent variable affects the dependent variable. It describes the process that links cause and effect.

Moderating Variables

Moderating variables influence the strength or direction of the relationship between independent and dependent variables.

Control Variables

Control variables are factors that are held constant in the study because they may also influence the outcome. Controlling these variables allows researchers to isolate the effect of the independent variable.

Confounding Variables

Confounding variables are external factors that are related to both the independent and dependent variables. If not properly controlled, they can create misleading relationships between variables.

Researchers must identify potential confounding variables during the research design stage and account for them through appropriate research methods or statistical controls.

Correlation and Causation

Understanding the difference between correlation and causation is essential in research design.

Correlation refers to a statistical association between two variables, meaning that they change together. However, correlation does not necessarily imply that one variable causes the other.

Causation, on the other hand, implies a direct cause-and-effect relationship between variables.

Two major problems can arise when interpreting correlations. The first is the third-variable problem, where a confounding variable influences both variables being studied. The second is the directionality problem, where it is unclear which variable causes the change in the other.

Appropriate research designs are necessary to distinguish between correlational relationships and causal relationships.

Types of Questions in a Questionnaire

When designing a questionnaire, researchers must carefully choose the form of the questions.

Closed-Ended Questions

Closed-ended questions provide respondents with a fixed set of response options. These questions are useful when the possible answers are known and limited.

Open-Ended Questions

Open-ended questions allow respondents to express their opinions or experiences in their own words. These questions are useful when researchers want to explore complex issues or gather detailed insights.

Designing Response Options

Response options should be as comprehensive as possible so that respondents can select an option that accurately reflects their situation.

Including an “Other: please specify” option allows respondents to provide answers that are not listed among the predefined choices.

When assessing knowledge-based questions, it is also useful to include a “Don’t know” option, as not all respondents may have the necessary information.

Filtering Questions

Filtering questions are used to guide respondents through the questionnaire and ensure that they only answer relevant questions.

For example, if a respondent indicates that they have not used a particular service, they may be directed to skip certain follow-up questions. While filtering can improve efficiency, excessive filtering may confuse respondents and make the questionnaire difficult to follow.

Question Order

The sequence of questions should follow a logical flow. Questionnaires should typically begin with simple and general questions before moving to more complex or sensitive questions.

This structure helps respondents feel comfortable and gradually engage with the survey.

Likert Scales

Likert scales are commonly used to measure attitudes, opinions, or perceptions. They provide a range of response options, usually on a five-point or seven-point scale, ranging from strong disagreement to strong agreement.

Likert scales allow researchers to measure the intensity of respondents' attitudes and to calculate average scores across responses.

Avoiding Question Design Errors

Several common errors must be avoided when designing questionnaire items.

One common problem is the use of double-barrelled questions, where two different issues are asked in a single question. This can create confusion and make responses difficult to interpret.

Another issue is ambiguous wording. Terms such as “frequently” or “often” may have different meanings for different respondents. Questions should therefore be precise and clearly defined.

Short questions, ideally containing fewer than twenty words, are easier for respondents to understand and answer accurately.

Designing the Questionnaire with Analysis in Mind

When designing questionnaire items, researchers must consider how the collected data will be analysed. The format of questions should align with the intended analytical methods.

For example, if the objective is to calculate the average age of respondents, the questionnaire should collect exact age values rather than broad age categories.

If researchers are unsure about the type of analysis they will perform, it is often better to collect raw data first and categorise it later during the analysis stage.

Translation of Questionnaires

In multilingual contexts, questionnaires should be translated into the language most familiar to respondents. Accurate translation is essential to ensure that all respondents interpret the questions in the same way.

The translate–back–translate method is commonly used to ensure translation accuracy. In this approach, the questionnaire is first translated into the local language and then independently translated back into the original language. Any discrepancies between the two versions are reviewed and corrected.

Pilot Testing

Pilot testing is a critical step in questionnaire design. Before conducting the full survey, the questionnaire should be tested on a small sample that represents the target population.

The pilot test helps identify potential problems such as unclear questions, missing response options, or difficulties in completing the survey. Respondents can also provide feedback on their experience with the questionnaire.

Based on this feedback, researchers can revise and improve the questionnaire before the main data collection begins.

Data Quality in Research

Data quality refers to the accuracy and reliability of the information collected through research instruments. High-quality data is essential for producing credible research results and making informed decisions.

One commonly used framework for assessing data quality is the VRIPT criteria, which include validity, reliability, integrity, precision, and timeliness.

Validity and Reliability in Data Quality

Validity ensures that the data collected truly measure the intended variables. Reliability ensures that data collection procedures produce consistent results across time, locations, and data sources.

Maintaining reliability requires standardized data collection procedures, consistent sampling methods, and clear documentation of research processes.

Integrity, Precision, and Timeliness

Integrity refers to the ethical collection and reporting of data. Researchers must avoid intentional falsification of data and ensure that data collection procedures are conducted honestly and transparently.

Precision refers to the level of detail and accuracy in the data collected. Data should be sufficiently detailed to support decision-making and should include appropriate disaggregation by factors such as gender, age, or location.

Timeliness refers to the prompt collection and reporting of data so that information remains relevant for research and policy decisions.

Conclusion

Designing a questionnaire is a critical step in the research process. A well-designed questionnaire ensures that the collected data are accurate, reliable, and relevant to the research objectives.

Researchers must carefully plan the structure of the questionnaire, ensure the validity and reliability of the questions, and align the design with the conceptual framework of the study. Proper attention must also be given to question wording, response options, translation, and pilot testing.

By following these principles, researchers can collect high-quality data that supports meaningful analysis and contributes to reliable research findings.

The Data Cleaning Process: Concepts, Importance, and Applications

Introduction

Data cleaning is a fundamental step in the data analysis process. Before any statistical analysis, modelling, or machine learning can be performed, the dataset must be examined to ensure that it is accurate, consistent, and reliable. The data cleaning process involves identifying, correcting, or removing errors, inconsistencies, duplicate records, and missing values that may exist in the dataset.

If these issues are not addressed, they can lead to incorrect statistical results, biased conclusions, and poor decision-making. Therefore, data cleaning is essential for ensuring that the data used in research and policy analysis accurately represents real-world conditions.

What is the Data Cleaning Process

The data cleaning process refers to the systematic procedure of identifying and correcting errors or inconsistencies within a dataset in order to improve its quality before performing analysis or building predictive models.

Datasets often contain several types of problems, such as duplicated observations, missing values, incorrect formats, extreme values, and logically impossible data entries. If these issues remain in the dataset, they can distort statistical calculations and lead to misleading results.

The goal of data cleaning is therefore to ensure that the dataset is accurate, complete, consistent, and ready for analysis.

1. Removing Duplicate Data

Duplicate data occurs when the same observation appears more than once in a dataset. These duplicates often arise from data entry errors, repeated survey submissions, or errors during data integration.

Removing duplicate records is an important step because it ensures that each observation is counted only once. If duplicates remain in the dataset, they may bias the analysis by giving extra weight to certain observations.

For example, in development economics research evaluating the impact of a cash transfer program on household poverty reduction, duplicate records may cause the number of beneficiaries to appear larger than it actually is. If 1,000 households are recorded as receiving assistance but 50 of these are

duplicates, the true number of beneficiaries is only 950. Such errors may lead researchers to overestimate the effectiveness of the program and could result in misguided policy decisions.

Removing duplicate records, therefore, improves the accuracy and reliability of the dataset.

2. Handling Missing Values

Missing data occurs when some observations lack information for one or more variables. Missing values may arise due to errors during data collection, incomplete survey responses, or difficulties in measuring certain variables.

Handling missing values is critical because many statistical methods and machine learning algorithms cannot process incomplete data. In addition, missing values can distort statistical estimates and reduce the representativeness of the dataset.

Several approaches can be used to address missing data. One option is to remove observations or variables that contain too many missing values. Another approach is imputation, where missing values are replaced with estimated values such as the mean, median, or predicted values generated by statistical models. In time-series datasets, forward or backward filling techniques can also be used to estimate missing observations.

For example, when studying the impact of microcredit programs on household income, some survey respondents may choose not to report their income. If these missing values are ignored or improperly handled, the calculated average income may be inaccurate. In some cases, poorer households are less likely to report income, meaning that deleting these observations may cause researchers to overestimate the program's impact on income.

Proper handling of missing values helps ensure that the dataset remains representative and suitable for analysis.

3. Fixing Structural Errors

Structural errors refer to inconsistencies in data entry, formatting, or categorisation. These errors often occur when data is collected from multiple sources or when different individuals record data using different formats.

Examples of structural errors include inconsistent spelling of categorical variables, mismatched measurement units, and inconsistent category labels. For instance, a country may appear as "Italy",

“italy”, or “ITALY” in the same dataset. Similarly, weight measurements might appear as “kg” in some cases and “kilograms” in others.

These inconsistencies can cause serious problems during analysis because data processing software may treat them as separate categories. As a result, the dataset may produce incorrect counts or misleading summaries.

For example, when analysing gender differences in access to agricultural credit among smallholder farmers, inconsistent gender labels such as “Male”, “M”, and “male” could be interpreted as three separate categories. This would lead to incorrect gender counts and inaccurate conclusions.

Standardising these categories into a single format ensures that the data is grouped correctly and that the results of the analysis are reliable.

4. Filtering or Handling Outliers

Outliers are extreme values that differ significantly from the rest of the data. These values may occur due to measurement errors, data entry mistakes, or unusual real-world observations.

Outliers are important because they can distort statistical measures such as averages, correlations, and regression results. Therefore, identifying and addressing outliers is a key step in the data cleaning process.

Outliers can be detected using statistical techniques such as the Interquartile Range (IQR), Z-score analysis, or graphical methods such as boxplots.

In development economics research, for example, when analysing household income to measure poverty levels, most households may report incomes between \$500 and \$2,000. However, if one observation shows an income of \$50,000 due to a data entry error, the average income of the sample will increase significantly. This can create a misleading picture of economic well-being and lead policymakers to underestimate the need for poverty reduction programs.

Once detected, researchers must decide whether to remove the outlier, cap it at a reasonable level, or investigate the observation further to determine whether it reflects a legitimate extreme value.

5. Standardising Data Formats

Another important step in data cleaning is ensuring that all data follows consistent formatting rules. Inconsistent data formats can create errors during analysis and make comparisons difficult.

Common examples of inconsistent formatting include differences in date formats, currency representations, and text capitalisation.

For instance, dates may appear in formats such as “06/03/2026” or “2026-03-06”. Currency values might appear as “\$100”, “100 USD”, or simply “100”. If these formats are not standardised, data analysis software may interpret them differently or produce incorrect calculations.

Consider a study evaluating government cash transfer payments across different regions. If some payment values are recorded in US dollars while others are recorded in local currency without conversion, the dataset may incorrectly treat these values as equivalent. This could lead to incorrect conclusions about the financial support provided to households.

Standardising the data format ensures that values are comparable and that calculations are performed correctly.

6. Validating Data

Data validation involves checking whether the dataset follows logical rules and real-world constraints. This step helps identify impossible or unrealistic values that may have been introduced during data entry.

Examples of invalid data include negative ages, unrealistic income values, or inconsistent dates, such as an order date occurring after the delivery date.

In development economics research, for instance, when studying participation in youth employment programs, the dataset might contain ages such as -5 or 150 due to input errors. These values are clearly unrealistic and can distort statistical results.

Similarly, when analysing agricultural production data, a dataset may contain records where harvest dates occur before planting dates. Such inconsistencies would produce unrealistic estimates of crop growth periods.

Applying validation rules—such as limiting age ranges or ensuring chronological consistency in dates—helps maintain data integrity and ensures that the dataset accurately represents real-world conditions.

7. Data Normalisation

Data normalisation, also known as scaling, is used when variables in a dataset have different measurement scales. Normalisation transforms these variables so that they become comparable and can be used effectively in statistical or machine learning models.

Normalisation is particularly important when variables have very different ranges. For example, in socioeconomic datasets, household income may range from 0 to 10,000 dollars, while years of education may range from 0 to 20. Without normalisation, the income variable would dominate the analysis because of its larger scale.

Normalisation is also essential in distance-based machine learning algorithms such as K-Nearest Neighbours, K-Means clustering, Support Vector Machines, and neural networks. In these models, variables with larger numerical values can bias the results.

Normalisation is also used when constructing composite indicators such as development indices. For example, indicators like income, education, and life expectancy must be normalised before being combined into a single index so that each indicator contributes equally.

Common normalisation methods include:

Min–Max normalisation, which rescales values to a fixed range, such as 0 to 1.

Z-score normalization, which standardizes values so that the dataset has a mean of zero and a standard deviation of one.

Log transformation, which is frequently applied to economic data such as income, GDP, or population, to reduce skewness and minimise the influence of extreme values.

Conclusion

The data cleaning process is an essential step in preparing datasets for analysis. By removing duplicate records, handling missing values, correcting structural errors, detecting outliers, standardising data formats, validating logical rules, and normalising variables when necessary, researchers can significantly improve the quality and reliability of their datasets.

High-quality data leads to more accurate statistical analysis, stronger research findings, and better-informed policy decisions. In fields such as development economics, where research often informs

public policies and development programs, ensuring data quality through proper data cleaning is particularly critical.



Co-Funded by the European Union under Grant No. 101237180
(ERASMUS -EDU-2025-CBHE-STRAND 1)

© 2026 Maria Sassi / University of Pavia

No part of this document may be reproduced without prior authorisation.

This publication was produced with the support of the ERASMUS-EDU-2025-CBHE-STRAND 1. The contents of this publication are the sole responsibility of Maria Sassi and do not necessarily reflect the views of the European Union.