Output 2

Training Programme for First Stage Researchers

CORE Project Team

2024

CORE PROJECT TEAM

Ceyda Cer Karabulut

Dana Rad

Ece Yağcı Akgündüz

Ezgi Güney Uygun

Ivana Marinković

José Manuel Carvalho Vieira

Ljubica Diković

Mustafa Özgenel

Vesna Vasović

This book is an output of the Competence-Oriented Methodologies for R&D in Social and Educational Sciences (CORE) Project (2021-1-TR01-KA220-HED-000031999), implemented under the Erasmus+ Strategic Partnerships Programme.

Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

TÜRKİYE ULUSAL AJANSI
TURKISH NATIONAL AGENCY

Funded by the European Union

CONTENTS

QUANTITATIVE RESEARCH DESIGNS

Part 1. Fundamentals of Quantitative Research Designs

The purpose of this chapter is to provide a comprehensive examination of quantitative research designs. These designs are crucial in empirical studies as they provide a systematic way of collecting and evaluating data. The chapter explores the different types of quantitative research designs commonly used in social sciences research. It analyses each design's characteristics, advantages, and drawbacks, allowing researchers to make informed decisions about which design is best suited for their study. Additionally, it discusses the critical factors to consider when selecting and implementing quantitative designs. It also offers practical advice on conducting research using quantitative methods.

The term research design pertains to the techniques and tactics employed to collect relevant data for investigating a specific query. Depending on the focus of the investigation, some queries may be more effectively answered by analysing data through quantitative research, which involves statistical analysis. Conversely, others may be better addressed by identifying patterns, features, or themes in the data, known as qualitative research.

A firm understanding of research designs is crucial, as even peer-reviewed research can be unreliable. To properly evaluate research, one must assess how a study was designed and executed to determine its validity for one's practice or study. Familiarity with different study designs enables one to distinguish between sound and flawed research, thereby applying critical thinking when evaluating evidence in activities such as literature reviews. Research can be classified into three main groups based on the application of the research study, its objectives in undertaking the research and how the information is sought (Fig. 1).

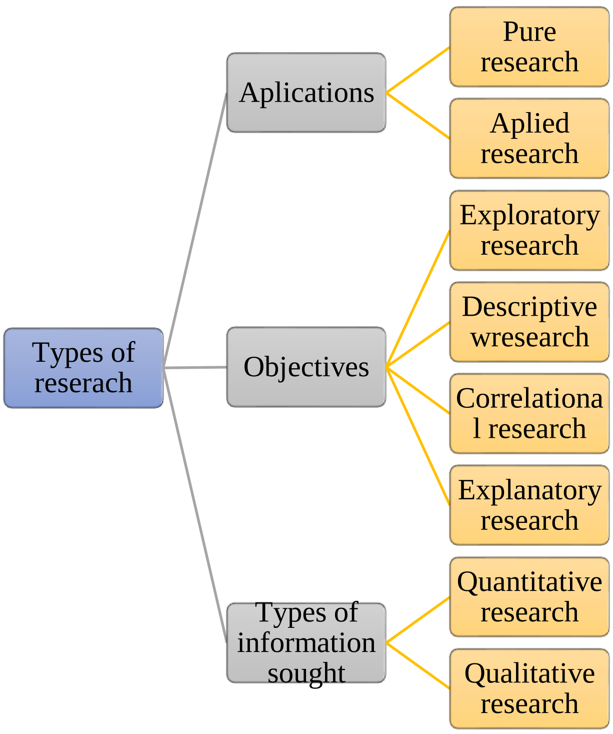


Figure 1. *General Types of Research* (Sukamolson, 2007).

What is quantitative research? Quantitative research is a type of research that involves collecting numerical data and analysing it using mathematical methods, particularly statistics. The definition of quantitative research may vary between researchers and educators, but it is generally agreed upon that this approach aims to explain phenomena through numerical data. For instance, (Creswell, 2014; 2018), a proponent of mixed methods, defines quantitative research as a method that explains phenomena by collecting numerical data and analysing it using mathematically based methods.

Quantitative research designs are more prevalent than qualitative research designs. Quantitative designs are structured, tested for validity and reliability, and can be easily defined and replicated. They provide enough details about a study design to ensure it can be verified and trusted. However, good quantitative research requires combining quantitative and qualitative skills to ascertain the nature and extent of diversity and variation in a given phenomenon (Tab. 1.).

Table 1. *Qualitative vs. Quantitative Methods*

Qualitative methods	Quantitative methods
Flexible investigation	Non-flexible investigation
Seeking an understanding of the frame of reference	Search results
Closer approach to reason (why and how)	Representative
Richer and deeper information	Generalizable
Based on intensive/ structural techniques	Based on extensive techniques

Quantitative and qualitative research methods are often perceived as distinct approaches, yet they exist on a continuum of research methodologies. Quantitative research tends to prioritize generalizability, reliability, and validity, whereas qualitative research emphasizes dependability, credibility, and confirmability. While both methodologies have inherent strengths and limitations, researchers must meticulously consider their research questions and context to determine which approach is most suitable for their research (Fryer et al., 2018).

The qualitative-quantitative-qualitative approach to research is the most comprehensive and worth considering, involving starting with qualitative methods to determine diversity, using quantitative methods to quantify the spread, and then going back to qualitative methods to explain the observed patterns (Kumar, 2011). Quantitative studies use different types of designs that can be classified based on (1) the number of contacts with the study population, (2) the reference period of the study, and (3) the nature of the investigation.

1. *Study designs based on the number of contacts.*

There are three study designs based on the number of contacts with the population: cross-sectional, before-and-after, and longitudinal studies. Cross-sectional studies are the most common and enable researchers to obtain an overall picture of a phenomenon or issue at one point in time. Before-and-after studies measure change in a phenomenon by comparing data collected before and after an intervention. Longitudinal studies study the pattern of change over time and involve multiple contacts with the study population. However, frequent contact with respondents may lead to the conditioning effect, where they respond with little thought or lose interest.

2. *Study designs based on the reference period.*

In research, the design of studies often centres around a specific reference period which examines a situation, event, problem, or phenomenon. There are two main types of studies - retrospective and prospective. Retrospective studies analyse past events using data collected from that time or people's memories, while prospective studies aim to predict future outcomes or the potential prevalence of a phenomenon. Experiments fall under the category of prospective studies, as the researcher has to wait for an intervention to impact the study population. Retrospective-prospective studies combine both approaches by examining past trends in a phenomenon and then tracking the study population to determine the impact of an intervention.

3. *Study designs based on the nature of the investigation.*

Quantitative research designs can be categorized as experimental, non-experimental, quasi-, or semi-experimental based on the nature of the investigation (Cash et al., 2016; Miller et al., 2020; White & Sabarwal, 2014). The experimental study designs can be classified as follows:

a. *The after-only experimental design refers to a scenario where the researcher intends to study the impact of an intervention on a population that is being or has been exposed to the intervention. In this case, the change in the dependent variable is measured by comparing the "before" (baseline) and "after" data sets. However, this design needs to be revised as it does not provide a proper baseline for comparison, and the two data sets are not comparable. Some of the changes in the dependent variable may be due to differences in how the data sets were compiled.*
b. *The before-and-after experimental design overcomes the problem of retrospectively constructing the "before" observation by establishing it before introducing the intervention to the study population. While this design addresses the comparability issue of the after-only design, it does not necessarily attribute any change to the intervention. To address this, a control group is introduced.*
c. *In a study utilizing the control group design, the researcher selects two population groups, a control group and an experimental group, to be as comparable as possible except for the intervention. The "before" observations are made on both groups simultaneously, and the experimental group is exposed to the intervention. When it is assumed that the intervention has had an impact, an "after" observation is made on both groups and the difference in the dependent variable(s) between the groups is attributed to the intervention.*
d. *The double-control design goes a step further than the control design in quantifying the impact attributed to extraneous variables. In this design, two control groups are used instead of one to separate other effects that may be due to the research instrument or respondents.*
e. *In a comparative design, a study can be carried out either as an experiment or a non-experiment. In the comparative experimental design, the study population is divided into the same number of groups as the number of treatments to be tested. The baseline concerning the dependent variable is established for each group, and the different treatment models are introduced to the other groups. After a certain period, when the treatment models have had their effect, the "after" observation is carried out to ascertain any change in the dependent variable. The study compares the effectiveness of interventions by analysing the degree of change in the dependent variable among different population groups.*
f. *In a matched control experimental design, comparability is determined individually-by-individually. Two individuals from the study population who are almost identical concerning a selected characteristic and condition are matched and allocated to a separate group. Once the groups are formed, the researcher decides which group is to be considered control and which experimental.*
g. *A placebo design attempts to determine the extent of the placebo effect, a patient's belief that they are receiving treatment, even if it is ineffective. In this design, two or three groups are used, depending on whether or not the researcher wants to have a control group.*

h. In the cross-over comparative experimental design, also known as the ABAB design, two groups are formed, and the intervention is introduced to one of them. After a certain period, the impact of this intervention is measured, and the interventions are crossed over.

Numerous researchers take a pragmatic approach to their research and utilize quantitative methods to investigate extensive data sets, test hypotheses or examine subjects that can be quantified. Nevertheless, selecting the appropriate research design and data collection instruments is more fundamental than using the proper data analysis tools. This remains a critical component of all research, regardless of its quantitative or qualitative nature (Sukamolson, 2007).

Despite the inherent challenges in measuring qualitative information, we can still glean meaningful insights by utilizing specialized research tools designed to convert attitudes, beliefs, and other intangible concepts into quantifiable data. This approach enables us to investigate various phenomena using quantitative methods, providing valuable insights into the complexities of human behaviour and experience.

Grimes and Schulz (2002) formulated a comprehensive framework of alternative research methodologies that researchers can employ when selecting an appropriate approach for their study, depending on their research question and the challenges inherent in their research design (Fig. 2).

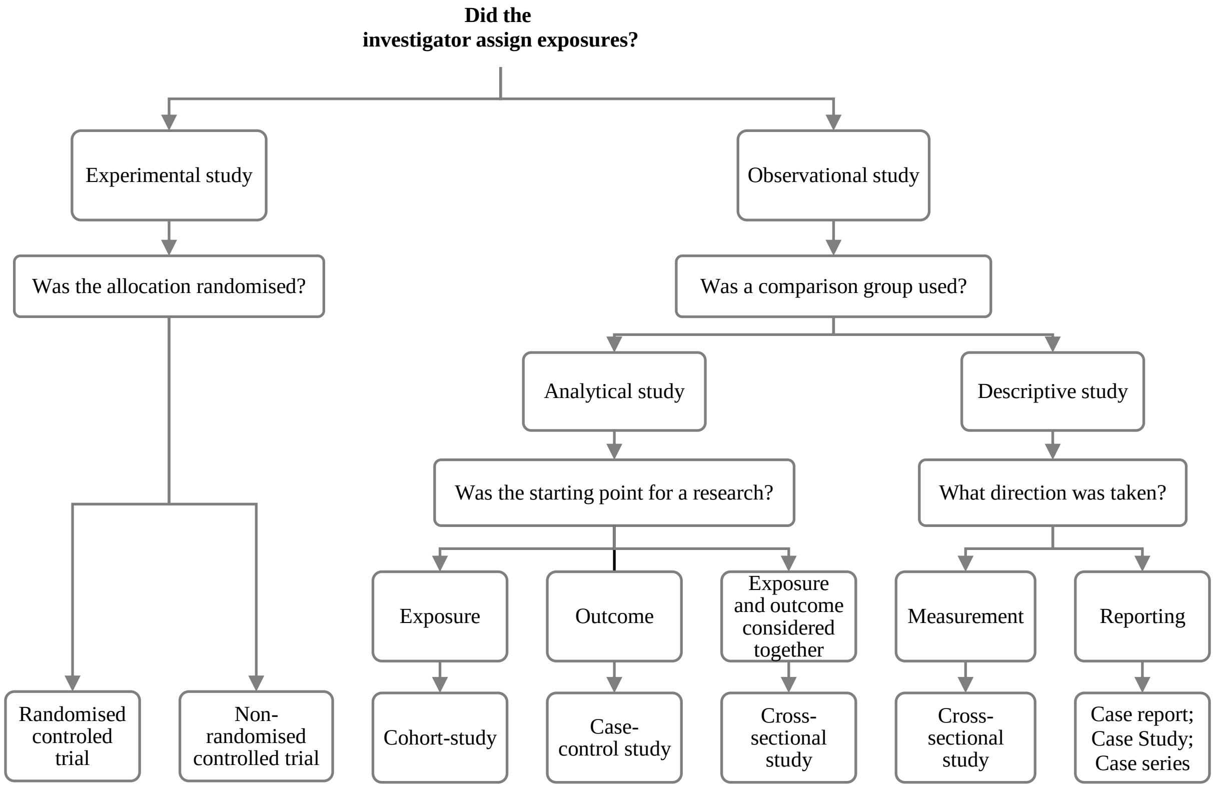


Figure 2. *Alternative Research Methodologies* (Grimes & Schulz, 2002).

1.1.Theoretical Delimitation of Quantitative Studies

Research methods are the various tools and techniques that researchers use to gather information and data. Research methodology, on the other hand, is a systematic approach researchers take to address research problems thoroughly. The systematic study of research methods encompasses defining research questions, selecting appropriate methods, collecting and analysing data, and drawing valid conclusions. A well-designed research methodology ensures that research is conducted accurately, efficiently, and ethically (Bhushan Mishra & Alok, 2019).

Research design, as the plan or proposal to conduct research, involves the intersection of philosophy, inquiry strategies, and specific methods (Creswell, 2014; Creswell & Clark, 2018). It is the blueprint that provides the rationale from which researchers derive meaning at the end of their study and guides the entire research process, enabling researchers to address their research questions and objectives effectively (Mwansa et al., 2022). Quantitative research designs offer a structured and objective data collection and analysis approach. This section provides an overview of the importance of research design and introduces the key concepts related to quantitative research designs.

When conducting research, it is essential to recognise that philosophical ideas can significantly impact the practices used. These ideas should still be acknowledged, even if they are not always apparent. Individuals must explicitly state their most significant philosophical ideas in research proposals or plans to ensure transparency and clarity. This will help us understand why specific approaches, such as qualitative, quantitative, or mixed methods, were chosen (Dawadi et al., 2021; Schoonenboom & Johnson, 2017). A proposal may include a section that discusses the proposed philosophical worldview of the study, a definition of the primary considerations of that worldview, and how the worldview influenced the researcher's approach to their study. The term worldview refers to a fundamental set of beliefs that guide action, which can be understood through paradigms, epistemologies, ontologies, or broadly conceived research methodologies. Worldviews serve as a general orientation about the world and the nature of research that a researcher holds. The primary worldviews are post-positivism, constructivism, advocacy/participatory, and pragmatism.

Social constructivism, often paired with interpretivism, is a commonly used approach to qualitative research, such as advocacy/participatory research.

The Post-positivist Worldview aligns with traditional research practices and is commonly associated with quantitative research. It is also called the scientific method, positivist/post-positivist research, empirical science, and post-positivism.

The pragmatic worldview emphasises actions, situations, and consequences rather than antecedent conditions, as post-positivism does. Its primary focus is on discovering practical solutions to problems and implementing them effectively. Researchers prioritise solving the research problem by utilising all available approaches to comprehend it rather than emphasising specific methods (Kotari, 2004).

Exploring paradigm alternatives can provide a more profound comprehension of research methodologies (Tab. 2). Quantitative research depends on statistically analysed numerical data, whereas qualitative research utilises non-numerical data. These differences have sparked debates, known as 'paradigm wars', as the perceived incompatibility between quantitative and qualitative research has led to disagreements.

Quantitative and qualitative research concepts are deeply rooted in researchers' philosophies and worldviews, also known as epistemologies or assumptions. Quantitative research is commonly referred to as "realist" or "positivist," while qualitative research is associated with a "subjectivist" perspective.

Realist researchers believe that their work reveals an objective reality that already exists. To uncover this truth, they maintain that objective research methods must be employed, often drawing on techniques derived from natural sciences and adapted to social sciences.

Table 2. Worldviews Used in Research Methods

Post-positivism	Constructivism	Transformative	Pragmatism
Determination	Understanding	Political and Activist	Consequences of action
Reductionism	Multiple participants meaning	Empowerment, human rights, social justice oriented	Problem oriented
Empirical observation and measurement	Social and historical construction	Collaborative	Pluralistic
Theory verification	Theory generation	Change, emancipatory oriented	Real-world practice oriented

Source: Adapted from Creswell (2014) and Creswell and Clark (2018).

Positivism represents the most radical manifestation of this worldview. Its adherents assert that the universe operates under unchanging principles of cause and effect. Nonetheless, we cannot entirely distance ourselves from what we are studying since we are all integral components of the world we are examining. Those who subscribe to subjectivism emphasize the significance of human subjectivity in the research process. They recognize that observing reality has a transformative effect on it, leading subjectivists to adopt a more relativistic stance.

Quantitative and qualitative research methods may be fundamentally incompatible when examining the two perspectives we have presented. Qualitative research encompasses a wide range of distinct methodologies, including participant observation, interviews, case studies, and ethnographic research.

1.1.1. Post-positivism, Experiential Realism and Pragmatism

Post-positivist thinkers understand the limitations of viewing the world as unbiased observers and acknowledge that natural science cannot govern all social research. Rather than pursuing absolute truth, post-positivists aim to present the most precise representation of reality possible. The principle of falsifiability, introduced by Popper, asserts that a theory's potential to be proven false is crucial in driving scientific inquiry. This implies that a theory's testability, whether through quantitative or qualitative means and its subsequent ability to be refuted or revised, are the fundamental components that propel investigation forward. Consequently, scientific inquiry is more concerned with combating errors and fallacies rather than establishing absolute truths (Burkholder et al., 2019).

Post-positivist social science focuses on our findings' dependability and capacity to anticipate results. Whether they specialize in quantitative or qualitative methods, researchers frequently take a practical research approach, using a range of techniques tailored to the issue at hand (Haig, 2017). Occasionally, a mixed-methods approach incorporating both quantitative and qualitative methods is the most suitable course of action. Nevertheless, quantitative methods are the only means to analyse numerical changes accurately.

1.2. When do we use quantitative methods?

Among the five primary research approaches, namely quantitative, qualitative, mixed methods research, arts-based research, and community-based participatory research, quantitative research stands out for its deductive approach to the research process that aims to prove, disprove, or lend credence to existing theories. Researchers measure variables and test relationships between them in this method to reveal patterns, correlations, or causal relationships. Quantitative research values neutrality, objectivity, and acquiring a comprehensive range of knowledge, such as a statistical overview from a large sample. This approach is typically suitable when the primary goal is to explain or evaluate (Leavy, 2022).
Research literature suggests that six primary types of research questions are particularly well-suited for quantitative research. These include audience segmentation, quantifying results obtained in a qualitative study, verifying data from a qualitative study, quantifying opinions, attitudes, and behaviours, explaining phenomena, and testing hypotheses. The first four types of research are referred to as "descriptive research," while the remaining two are known as "inferential research." Descriptive research relies on descriptive statistics, while inferential research employs inferential statistics.

During the initial stages of developing quantitative designs, researchers often conduct descriptive research to identify the characteristics of individuals, groups, or situations. This type of research aims to uncover new insights, describe current conditions, determine the frequency of occurrences, and classify data.

Based on the positivist philosophy of human inquiry, the quantitative research method is often regarded as the quintessential scientific approach to conducting research. Positivist research is characterised by a rigorous, systematic process emphasising rationality, objectivity, predictability, and control. Advocates of the quantitative approach are generally viewed as unbiased scientists who uncover factual data (Walker, 2005).

In quantitative research, statistics, mathematics, and numerical data processing are used to systematically and empirically study phenomena. This involves breaking down reality into smaller, more manageable pieces through the analysis of numerical data. Statistically oriented approaches are used to test and reproduce hypotheses about variable relationships by collecting and generalising data across groups of people. When designing a quantitative research study, the type of research philosophy, approach to theory development, characteristics, and research strategy are key factors to consider (Mwansa, et al., 2022). While positivism is commonly associated with quantitative research, it can adopt realistic and pragmatic philosophies. The deductive approach is strongly associated with quantitative research, while the inductive approach is rarely used (Tab. 3).

Table 3. Quantitative and Qualitative Paradigm Assumptions

Assumptions	Questions	Quantitative	Qualitative
Ontological assumption	What is the nature of reality?	Reality is objective and singular apart from the researcher.	Reality is subjective and multiple as seen by participant in a study.
Epistemological assumption	What is the relationship of the researcher to that researched?	Researcher is independent from that being researched.	Researchers interact with that being researched.
Axiological assumption	What is the role of values?	Value-free and unbiased	Value-laden and biased
Rhetorical assumption	What is the language of research?	Formal, based on set definitions, impersonal voice, and use of accepted quantitative word.	Informal, evolving decisions, personal voice, and accepted qualitative words
Methodological assumption	What is the process of research?	Deductive process. Cause and effect, Static design-categories isolated before, study, Generalizations leading to prediction, explanation, and understanding, and Accurate and reliable through validity and reliability.	Inductive process, mutual simultaneous shaping of factors, emerging design-categories identified during research process, context-bound, patterns, theories developed for understanding, and accurate and reliable through verification.

Source: Adapted from Sukamolson (2007) and Wilson et al. (2021).

Typically categorised into three tiers: descriptive, correlational, and causal, with the latter utilising experimental designs, it aims to generalise the findings from a sample to the whole population. A variety of statistical techniques can be used to predict scores on one variable based on scores from another variable. Qualitative research, on the other hand, aims to achieve a deeper understanding of a particular phenomenon (Sukamolson, 2007).

1.3.Theory Construction and Testing

Constructing a theory entail utilizing inductive reasoning to create a theory grounded in observations and account for various phenomena. Conversely, the theory-testing method commences with a theory that directs the selection of which observations to conduct, progressing from the general to the specific. The theory's accuracy is assessed through observations, utilizing deductive reasoning to deduce a set of propositions from it (Fig. 3).

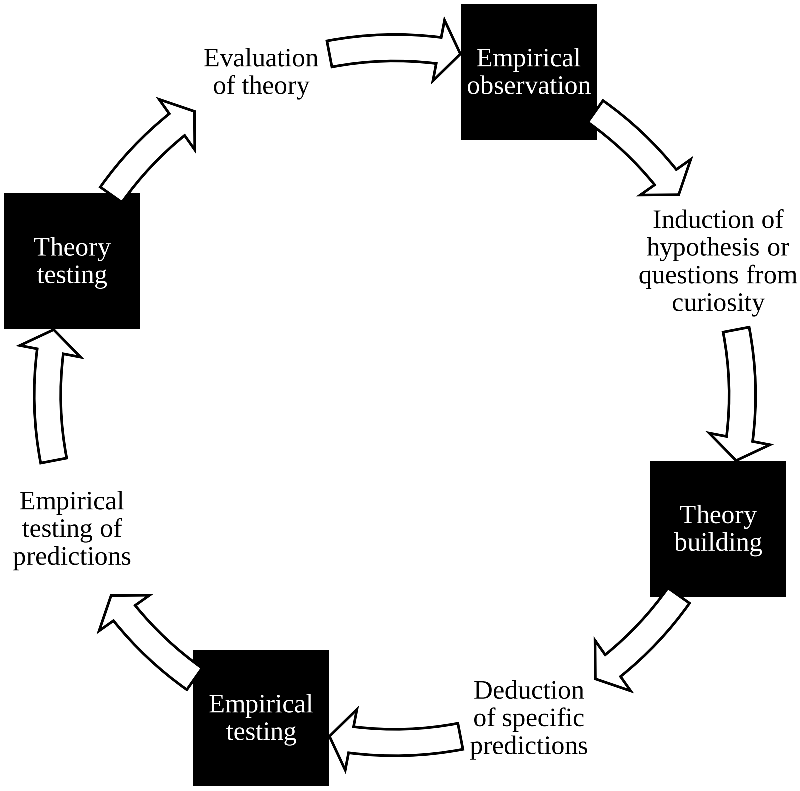


Figure 3. *Theory Building and Testing* (Cash et al., 2016).

The selection of inductive or deductive approaches depends on the research objectives: exploration, description, verification (explanation), and modelling. These approaches can be taken separately (diachronically) or together (synchronically) to perform distinct yet complementary functions throughout the research process (Fig. 4).

The exploratory (inductive) approach is useful when the research investigation aims to garner comprehensive and generic indications about the Problem, potential decision alternatives, and relevant variables that must be considered for descriptive and verification studies/phases. On the other hand, the deductive approach is instrumental in causal studies, where the objective is to demonstrate that a variable influences or determines the values of other variables, allowing researchers to test the explanations proposed by the set of hypotheses formulated.

While descriptive studies provide an objective "fixation" of some relevant aspect that may be useful in the subsequent interpretation of the results, they do not allow us to explain or anticipate the dynamics of any phenomenon, which warrants a more complex approach. Similarly, verification (explanatory) studies become imperative when it is essential to demonstrate that a variable has a causal impact on other variables.

Finally, modelling objectives are achieved when the theory underlying the structure of hypotheses is sustainable in light of the results, leading to a more complex and valuable level of Knowledge. Such investigations make a sensitive contribution to the body of Knowledge they are part of, to the extent that they now have a forecasting and management tool.

In summary, while the exploration phase identifies all potential reasons for a problem, the study's descriptive and explanatory phases aim to estimate its possible causes.

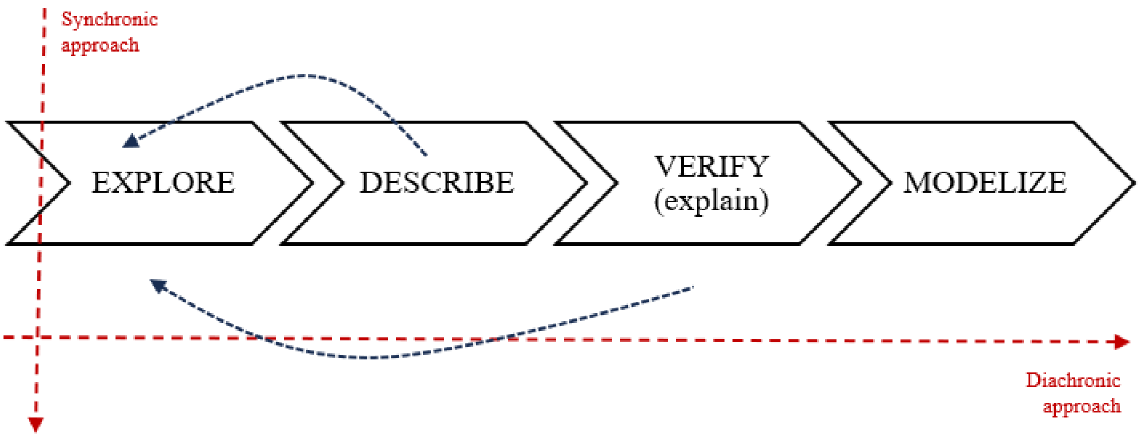


Figure 4. *Research Objectives*

1.4.Operational Framework of Research Projects

In research, there are two primary quantitative strategies: surveys and experiments.

Surveys examine a sample to obtain numerical representations of it's mine a population's trends, attitudes, or opinions. Researchers can conduct cross-sectional or longitudinal studies using questionnaires or structured interviews to collect data and generalize the results to a larger population.

On the other hand, experimental research aims to determine whether a specific treatment affects an outcome. This is achieved by administering the treatment to one group while withholding it from another and then measuring the outcomes of both groups. Examples of experiments include assigning subjects to treatment conditions at random and non-randomized quasi-experiments, which may involve single-subject designs.

Kumar (2011) organizes these “decisions” based on three criteria (Fig. 5):

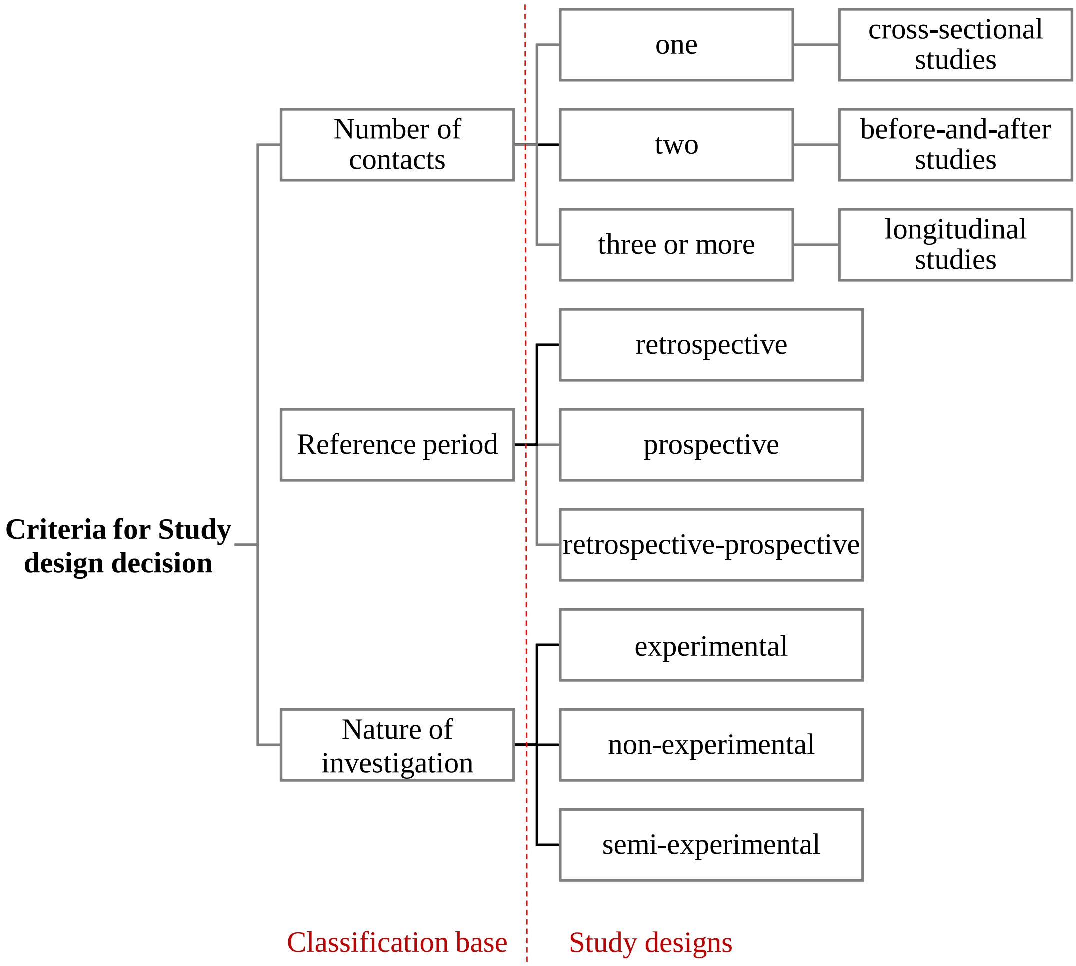


Figure 5. *Types of Study Design* (Kumar, 2011).

Part 2. Components of Quantitative Research Designs

Comprehending the role and significance of research designs is crucial for effective research. The design encompasses the entire research process, from framing the question to analysing and reporting data.

Two fundamental research questions exist: descriptive research, which explores what is happening, and explanatory research, which focuses on why things are happening.

Descriptive research can be advantageous, especially when exploring new areas, as it can provoke "why" questions for explanatory research. Explanatory research involves developing causal explanations that argue that a specific factor affects a particular phenomenon. For instance, gender may affect income levels. However, the complexity of causal explanations may vary, and hidden or unmeasured variables may be at play.

It is important to note that people often mistake correlation for causation. When two events are linked, it does not necessarily imply that one causes the other. The link between them may be coincidental rather than causal. Therefore, it is crucial to understand the distinction between correlation and causation to conduct effective research.

Aaker et al. (2013) organise the process/design of a study as shown in Fig. 6. It all starts with specifying the Research Question, that is, the problem that the project will try to solve and the knowledge to which it will contribute or initiate.

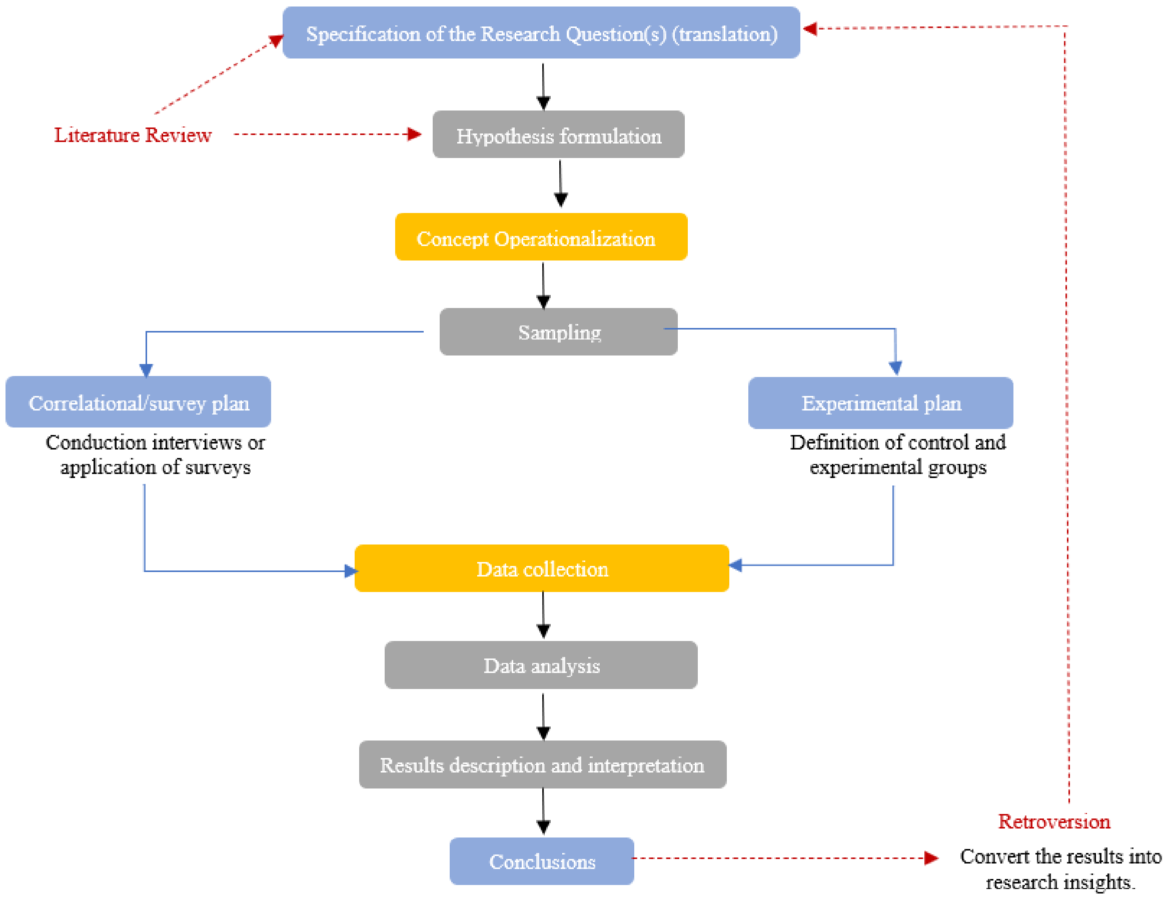


Figure 6. *Research Blueprint*

Directly resulting from the literature review, it must immediately be "translated" into research questions, that is, hypotheses that will determine what will be measured, from what sources of information and with what methodologies. Research hypotheses are systems of variables whose sets, although only partially exhaustive, cover the main dimensions of the phenomenon under analysis. They also clarify the relationships proposed between such variables that need testing. With this, the following very relevant and demanding task is operationalising (making measurable) the variables whose relationships will be tested (scales).

Once the research question has been specified, the concepts (variables), latent or directly observable, have been defined, whose relationships will be tested, and the measures with which each will be captured have been specified, it is essential to define which units of information will contain the information required (secondary or primary).

Quantitative studies (experimental/non-experimental) must also define the sampling method (random/non-random) that will be applied to this "population" and the size and characteristics of the groups (non-experimental; experimental; control) that will be heard. With this knowledge, the researcher must decide which concrete information collection plan should be adopted: correlational/survey (cross-sectional; longitudinal) or experimental.

Collecting information (questionnaire) is complex, susceptible to adding "errors," and dependent on the researcher's experience. For all these reasons, it is advisable to use scales that have already been validated in previous studies whenever possible, reinforcing their reliability and validity.

Once organised information is available, the data will be subjected to adjusted and planned analyses to test the research hypotheses (descriptive, univariate, multivariate, inferential). The results obtained must then be described and interpreted to, in conclusion, be "converted" into an answer(s) to the initial Research Question that triggered the entire process.

2.1. Research Questions

Understanding whether a research question is descriptive, or explanatory is crucial as it significantly affects the research design and the information collected. Researchers must develop causal explanations when answering the 'why' questions. Causal explanations attempt to prove that a particular factor X, such as gender, affects a phenomenon Y, such as income level. While some causal explanations can be simple, others can be more complex.

When dealing with predictions, researchers must distinguish between correlation and causation. It is a common mistake to assume that two events are causally related just because they occur together, or one follows the other. The correlation is likely coincidental and does not indicate a causal relationship.

Distinguishing between causation and correlation is essential to accurately understanding prediction, causation, and explanation. It is important to note that an accurate prediction does not always require a causal relationship, and the ability to make a prediction does not necessarily prove a causal relationship. Confusing these concepts can lead to a lack of understanding and incorrect conclusions.

Recognizing the difference between correlation and causation is essential because we can observe correlation but directly observe causation. Therefore, we must infer the cause, making avoiding invalid inferences a primary goal of explanatory research design.

There are two approaches to causation: deterministic and probabilistic. In deterministic causation, variable X causes Y without exception if it reliably produces Y. This approach aims to establish causal laws, such as the rule that water boils at 100ºC.

However, most causal thinking in social sciences is probabilistic rather than deterministic. We can enhance probabilistic explanations by specifying the conditions under which one factor is more or less likely to affect another. However, we will never achieve complete or deterministic explanations. Two events are causally related because they occur together or follow the other. The correlation is likely coincidental and does not indicate a causal relationship.

Mistaking causation for correlation can lead to understanding prediction, causation, and explanation. Accurate prediction does not necessarily require a causal relationship, and the ability to predict does not prove causality.

Research objectives can be single or multiple and can be covered synchronously or diachronically (from the 1st to the 4th).

2.2. Variables

Quantitative research focuses on measurement and assumes that the subject of inquiry can be quantified. Its main aim is to provide comprehensive data through measurement, analyse the data for patterns and connections, and verify its accuracy. The scope of quantitative research ranges from easily quantifiable attributes such as height and weight to more intangible elements such as human emotions and thoughts.

The quantitative research approach is highly precise and logical, utilizing statistical analysis to its fullest extent. Its ability to test theories through hypothesis formulation and formal statistical analysis sets it apart as a methodology. It is especially useful in measuring variables such as height, weight, attitude, and well-being, differentiating between independent and dependent classifications and capturing the influence of the former on the latter. Multiple hierarchical measurement theories are also employed to acquire diverse measurement types (Tab. 4).

Table 4. *Levels of Measurement*

Measurement level	Attributes	Examples
Ratio	A zero value is meaningful permitting direct comparisons between measurements.	Height, weight, length
Interval	Distance between measured variables is meaningful	Temperatures scales where the zero point is arbitrary, but set intervals are meaningful (for example, Centigrade or Fahrenheit)
Ordinal	Attributes can be ordered.	Opinion is measured by asking if you: “strongly agree”, agree”, “do not know”, “disagree”, and “strongly disagree”
Nominal	Hair colour, gender, nationality	Hair colour, gender, nationality

Source: Watson (2015)

At its most basic level, nominal classification categorises data without quantitative analysis. As we move towards ordinal measurement, we introduce a hierarchical structure to the data, although this method may require more precision. We rely on interval and ratio-level measurements for increased precision, although generating a ratio can be challenging when studying social phenomena. Ordinal and interval measurements are the most commonly used techniques in quantitative research.

Regardless of the method of measurement, errors are bound to occur. These errors can stem from various sources, including instrument, human, and random errors.

Although it is possible to reduce instrument and human errors, it is impossible to control random errors. Therefore, it is essential to consider random errors when designing and using any instrument. Instrumental and human errors can manifest in two ways: within the instrument (or within the human operator), which means that the same instrument may produce varying results in different settings, or inter-instrument (or human-to-human), which means that two seemingly identical instruments may yield different results.

Similarly, human errors imply that individuals using the same instrument may obtain divergent results with different advantages. On the other hand, instrument errors imply that two people using the same instrument may obtain different depths simultaneously. While errors cannot be eliminated, they can be minimised.

Effective instruments must be designed to minimise instrument errors. In social research, this means ensuring that observational questionnaires and checklists are easily comprehended and that questions are answered precisely.

When designing instruments, it is crucial to balance "authenticity" and "directivity." An authentic instrument measures as much as possible about a phenomenon but risks becoming indirect, while a direct instrument focuses only on items directly related to the phenomenon, potentially losing some authenticity (Watson, 2015).

2.3. Hypotheses

A hypothesis is a preliminary explanation that considers a collection of facts and is subject to further examination. In quantitative research, experiments are formulated to evaluate these hypotheses. We collect pertinent data and employ statistical methods to ascertain whether the hypothesis should be tentatively accepted or rejected. It is crucial to recognize that accepting a hypothesis is never absolute, as additional data may surface in the future which could prompt its rejection (Sukamolson, 2007).

Experiments are conducted to test how introducing an intervention, also known as a variable, affects what happens. Hypothesis testing is used to test variable relationships. It is necessary to control all other factors to ensure that you are measuring the impact of the intervention you have introduced.

Experiments are used in explanatory research based on causal logic, which identifies causal relationships between variables. For instance, A causes B or A causes B under C circumstances. Certain necessary conditions must be present to support the presence of a causal relationship. The cause must precede the effect (temporal order), the reason must be related to the effect, and there must be no alternative explanation.

Explained in terms of variables (Leavy, 2022):

- *The independent variable must precede the dependent variable, and a relationship must exist between the two.*
- *No extraneous variable can provide an alternative explanation for the dependent variable.*
- *Experimental groups receive the experimental intervention (the experimental stimulus), while control groups do not.*
- *In some cases, the control group may receive a placebo.*
- *All experiments have at least one experimental group, but not all experiments have control groups.*
- *Using control groups is necessary to accurately compare the results of the experimental group whose members received the intervention with those of a similar group whose members did not.*
- *Depending on the type of experiment, there may be one, two, or four groups in total.*
- *Some experiments involve pretests and/or post-tests in addition to the experimental intervention.*

- *A pretest determines a subject's baseline before introducing the experimental intervention.*
- *A post-test is given after the experimental intervention to assess the impact of the intervention.*

When forming a hypothesis, it is essential to identify independent and dependent variables. The hypothesis should be a plausible statement of how the independent variable interacts with the dependent variable. Additionally, potential control variables must be identified.

The next step involves determining how to measure the independent, dependent, and control variables. During the operationalisation process, ensuring high content validity between the numerical representation and the conceptual definition of any given concept is crucial.

Once the variables are defined and operationalised, the researcher must consider sampling. Which empirical referents will be used to test the hypothesis?

Stockermer (2019) point out that measurement and sampling are typically done simultaneously because the empirical referents that the researcher studies may affect the choice of operationalisation of an indicator over another.

After collecting the data, the researcher can conduct statistical tests to evaluate the research question and hypothesis. Ideally, the study's results will influence theory.

After constructing a set of hypotheses to test the initial theory, the researcher must also pinpoint other variables potentially impacting the phenomenon under investigation. These variables, such as socio-demographic, psychographic, and behavioural factors, should be controlled for in the study. With hypotheses and control variables in place, the researcher can then identify the best methods for measuring both the main variables of interest and the control variables before selecting an appropriate sample for the study.

2.4. Causality

The term "causation" refers to the idea that a change in one variable will result in another change. In this case, the definition of causation is expanded to include the idea that a precondition can influence a variable of interest. For example, one can imagine that a person's gender influences credit card usage. This means that gender could be seen as having a causal relationship with credit card usage, even though it is impossible to change a person's gender to observe whether credit card usage would change. The term "influence" is sometimes used instead of "cause" if it is more appropriate, but the logic of the analysis remains the same. If two variables are causally linked, then it is reasonable to assume they will be associated. If an association provides evidence of causation, then the lack of association suggests that causation is not present. Therefore, an association between attitude and behaviour is evidence of a causal relationship: Attitude --> behaviour (A. Aaker et al., 2013).

Researchers need to remember the vast array of causal relationships when attempting to establish causality in their studies. This requires implementing various methods and analysis techniques of varying complexity.

A comprehensive set of causal relationships can be observed in both experimental and non-experimental quantitative studies. Such relationships may be direct, mediated, or mutual, with some being more complex than others. These complexities can range from simple linear regression to structural equation models (SEMs).

In Figure 7, eight types of causal relationships are depicted:

- *Direct linear causal relationship in which Y is a function only of A.*
- *Mediated causal relationship in which the influence of A on Y is mediated by B.*
- *Direct causal relationship in which it is possible to estimate the total effect (direct and indirect) of A on Y.*
- *Direct mutual linear causal relationship in which the influence of A on Y is reciprocal.*
- *Indirect mutual linear causal relationship in which Y reciprocally influences the influence of A on Y (mediated by B).*
- *Multi-mediated causal relationship (domino) in which A generates a sequential unfolding of effects over time on Y.*
- *Moderate direct causal relationship in which the influence of A on Y is contingent on the conditions of C.*
- *Causal relationship in which A (exogenous variable) initiates a complex structure (path) of influences (direct and mediated) on Y.*
- *Apparent or spurious correlation refers to a statistical association between two variables that do not have a causal link. This type of correlation can arise due to mere chance or the influence of a third variable. It is important to be aware of the existence of spurious correlations to avoid drawing incorrect conclusions or making erroneous predictions based on statistical data.*

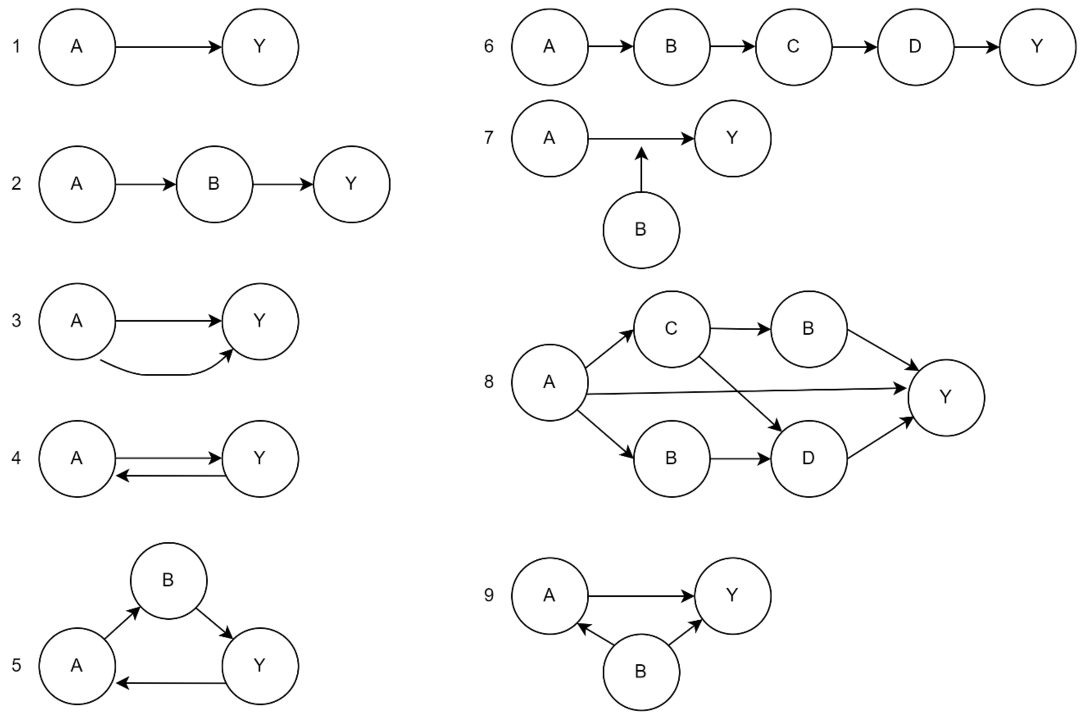


Figure 7. *Causal Hypotheses*

2.5. Surveys: Population and Sample

Quantitative research is a widely used survey research method that aims to systematically gather information from respondents to predict and understand various aspects of the population's behaviour. This involves sampling, designing questionnaires, administering surveys, and analysing data. Different types of surveys include in-person and telephone interviews, omnibus surveys, and self-administered questionnaires.

Despite its effectiveness, quantitative research has some limitations. It primarily collects naturally quantitative data and analyses it using mathematical methods. However, researchers can employ measuring instruments to convert non-quantitative phenomena, such as beliefs, into quantitative data. Many researchers adopt a pragmatic approach and use quantitative methods to gain a broad understanding, test hypotheses, or study quantitative phenomena.

Survey Research has become a major, if not the primary, technique for studying and asking participants one or more questions about their attitudes, perceptions, or behaviours (Stockemer, 2019).

2.5.1. Types of Surveys

Survey research is a widely utilized quantitative method in social sciences, education, and healthcare. It entails gathering standardized data through a series of questions posed to participants, which can then be analysed statistically. Through this statistical analysis, researchers can conclude the broader population from which the sample was taken. Surveys are commonly used better to understand individuals' beliefs, attitudes, and opinions and to report their experiences and behaviours. While the data collected from surveys is subjective, more objective data such as age and place of birth can be obtained through other means (Leavy, 2022).

There are two primary methodological designs in survey research.

Cross-Sectional Survey is a method of collecting data about a group of individuals at a specific point in time (Mtshweni, 2019). Such surveys are often legitimate from a theoretical standpoint, as they allow researchers to conclude the relationship between independent and dependent variables. However, since only one data set is available for independent and dependent variables, cross-sectional surveys cannot prove causality. Researchers must rely on theory, logic, and intuition to support the findings and conclusions drawn from cross-sectional studies. In other words, researchers should only use cross-sectional data to test theories if there is a clear temporal relationship between the independent and dependent variables. A cross-sectional survey can be a powerful tool for testing hypotheses when there are explicit theoretical assumptions about the relationship. Nonetheless, empirical relationships are only sometimes clear-cut, and it can be challenging to derive causal explanations from cross-sectional studies.

Longitudinal Surveys are a unique type of research that differs from cross-sectional studies in that they are conducted repeatedly over an extended period of time. Each survey consists of the same set of questions, allowing researchers to gain invaluable insights into the evolution of attitudes and behaviours within a population. These surveys can be classified into three categories: trend, cohort, and panel (Leavy, 2022; Stockemer, 2019; Watson, 2015).

- *A Trend study, often called a repeated cross-sectional survey, involves conducting multiple surveys with different groups of individuals over time. The surveys include the same questions in each wave, fully or partially. This method enables researchers to identify significant shifts in attitudes and behaviours over time.*
- *Cohort studies have a narrower focus, as they focus on a particular group rather than the entire population. Like trend studies, cohort studies involve recurring inquiries directed*

solely at a selected group of people with a shared characteristic. Each iteration involves a new sample from the same population, indicating that the population remains constant while the individuals sampled vary (Price & Lovell, 2018).

- *Finally, Panel studies are known for asking the same set of questions to the same individuals in multiple waves. While these surveys can be quite costly and challenging to conduct, they are considered the most effective way to detect causal relationships or changes in individual behaviour. As a result, panel studies serve as a powerful tool in identifying such relationships.*

2.6. Sampling

Sampling is a crucial aspect of research that enables researchers to collect data about a population of interest. The primary objective of sampling is to acquire data representative of the entire population, which can be challenging due to the population's size and diversity. To achieve a representative sample, researchers must ensure that the sample consists of individuals with the same characteristics as the population of interest. This can be achieved through various sampling techniques, such as random, purposive, and quota.

A representative sample consists of individuals with the same characteristics as the population. For example, suppose a researcher knows that 55% of the population he intends to study is male, 18% are African-American, 7% are homeless, and 23% earn more than 100,000 euros. In that case, he/she should try to match these characteristics in the sample to represent the population.

Random sampling is used when researchers cannot match the population's characteristics in the sample. Randomization helps to offset confounding effects by randomly selecting cases. Fig. 8. displays a graphical representation of a population (people, events, households, institutions, or something else) that is the subject of research, a sample frame (set of units from which a sample will be drawn: in the case of a simple random sample, all units in the sampling frame have an equal chance of being drawn and of occurring in the sample), and a sample (the subset of the population chosen for the research or survey).

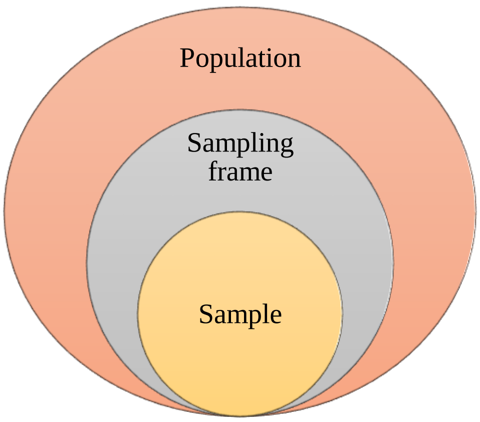


Figure 8. *Sample* (Stockemer, 2019).

A biased sample is neither representative nor random. Its answers do not reflect those obtained from the entire population. Survey responses can suffer from different biases, such as selection bias, non-response bias, and response bias. Sampling error is always present due to statistical imprecision.

Convenience sampling is a non-probabilistic sampling technique where people are chosen because they are readily available. In purposive sampling, subjects are selected based on predetermined characteristics. Volunteer and snowball sampling are other non-probabilistic sampling techniques employed in populations that are difficult to access. *Quota sampling* is a technique used in online surveys where sampling is done based on pre-established criteria. For instance, many polls have an implicit quota, such as customer satisfaction.

2.6.1. Sample Size Determination in Quantitative Research Design

Selecting the optimal sample size is crucial for quantitative researchers seeking precise and accurate results from significance tests. Determining the sample size involves various methods, such as statistical formulas and electronic sample size calculators. However, researchers must first establish details about the target population to identify the appropriate sample size. They must consider important factors such as the population size, error level, confidence interval, and confidence level. These determinants aid in deciding how much the sample mean can deviate from the population mean and how sure researchers want to be that the actual mean will fall within their confidence range. The confidence interval is usually set at 90%, 95%, or 99% confidence.

Furthermore, researchers must also account for the standard deviation to anticipate the variation between responses. The size of the quantitative sample is assessed based on the strength of a hypothesis test and the quality of produced estimates (Mwansa et al., 2022). Five important study design parameters, such as the significance criterion, minimum expected difference, estimated measurement variability, desired statistical power, and one or two-tailed statistical analysis, typically determine the primary determinant of sample size in a quantitative design.

2.6.2. Sampling Procedures in Quantitative Research Design

Probability sampling is a method employed in quantitative research designs that aims to ensure that each member of the population has an equal chance of being included in the sample (Mwansa et al., 2022). The primary objective of this method is to ensure that researchers draw valid conclusions from their findings and that their results represent the entire population. Researchers employ four primary sampling techniques to achieve this, which we will discuss below.

Simple random sampling involves using completely random techniques or tools, such as random number generators, to give each individual in the population an equal chance of being selected.

Systematic sampling is like simple sampling, but specific individuals are chosen regularly. However, it is important to ensure that the list does not contain hidden patterns that could skew the sample (Mwansa et al., 2022).

When sampling the population, stratified sampling involves dividing it into different subpopulations that are significantly different from one another. Each subgroup is well-represented in the sample, and researchers must divide the population into subgroups based on related characteristics, such as gender, age group, income class, or function. Then, they randomly or systematically select samples from each subgroup.

Cluster sampling involves dividing a population into subgroups with characteristics comparable to those of the sample as a whole and randomly selecting entire subgroups. This method is proper when dealing with large and dispersed populations, but it is more likely to introduce sampling errors as there can be significant differences between clusters.

On the other hand, non-probability sampling is a subjective approach to selecting units from a population, making it a fast, easy, and inexpensive way of obtaining data. However, it assumes that the sample is representative of the population, which can be a risky assumption to make. Additionally, elements are chosen arbitrarily, making it impossible to estimate the probability of any element being included in the sample or identify possible bias.

Convenience sampling, also known as random sampling, utilises individuals most easily accessible as study participants. Snowball sampling, also known as chain sampling or network sampling, asks early sample members to find and refer additional people who meet eligibility requirements. Quota sampling involves the researcher determining the necessary number of participants from each population stratum and identifying population strata.

Lastly, Purposive sampling, also known as judgmental sampling, is based on the idea that the researcher's understanding of the population can select individuals for the sample.

2.6. Data Collection Methods

Questionnaires are the primary data collection tool used in survey research. They are also referred to as survey instruments. This phase is considered the upfront work in survey research and determines everything else. To produce a valid measure instrument (questionnaire), clear and justifiable links must be between your indicators (questions) and the concepts you say you are measuring (Leavy, 2022).

There are many preexisting surveys available on a wide range of topics. Therefore, it is advisable to consult published research on your subject and available online databases to determine whether or not there are preexisting surveys you can use or draw from to answer your research questions. Often, you do not need to design a brand-new survey instrument.

Survey items, which are the questions in the questionnaire, are designed to help you test your hypotheses or answer your research questions. These hypotheses or research questions have been constructed regarding the variables you want to measure. Survey questions are designed to measure the concept you are interested in as precisely as possible. The questions you create around each concept (construct) in the study are how you operationalize your variables. They are the indicators that a variable is present or not.

The more multidimensional a variable concept is, the more questions you will likely ask regarding that specific variable. For example, it could take many questionnaire items to address one concept (e.g., ECSI). Question construction is at the heart of survey research. Keep sight of your goal to measure the phenomenon of interest as precisely as possible.

There are some general dos and don'ts for creating practical survey questions. Beginning with the dos, it is vital to use clear, understandable, and, whenever possible, highly specific language. There is a long list of things to avoid when constructing survey questions, including double-barrelled questions, double-negative questions, negatively phrased questions, biased or leading questions, questions with built-in assumptions, abbreviations, slang and contractions or ambiguous phrases, and questions that ask respondents to recall information from an unrealistic time frame.

Whether you create open-ended or forced-choice questions plays a role in determining the nature of your inquiries. Forced-choice or fixed-choice questions provide respondents with a range of response options to choose from. This type of question design enables you to collect a wide range of data, produce easily quantifiable data, and have high generalizability when large samples are used. Multiple choice, dichotomous, checklists, and rating and Likert scales are examples of different forced-choice questions.

2.6.1. Measurement Scales

Measurement is the process of systematically characterizing or quantifying information about people, events, ideas, or objects of interest. The social sciences have made significant progress in measuring people's underlying beliefs, knowledge, attitudes, and values through scientific scaling approaches and well-designed questionnaires and scales. These tools help researchers access the thoughts and views of respondents and understand the views of the person or group under observation. Rating scales are formalized versions of questionnaires that use multiple items to triangulate or help define a concept. Scales are more appropriately used to measure attitudes, values, or personality dispositions because they use multiple items to reflect the view that people's attitudes or beliefs are not singularly defined (Crano et al., 2014; Hair et al., 2021; Leavy, 2022).

Constructing effective questionnaire items applies to scale development as well. Several things to avoid when constructing survey questions include double-barrelled questions, double-negative questions, negatively phrased questions, biased or leading questions, questions with built-in assumptions, abbreviations, slang and contractions, ambiguous phrases, and questions that ask respondents to recall information from an unrealistic time frame.

Scale measurement involves assigning a set of scale descriptors to represent the range of possible responses to a question about a particular object or construct. Scale measurement assigns degrees of intensity to the responses, commonly called scale points. There are four basic scale levels: nominal, ordinal, interval, and ratio (Tab. 5).

Table 5. *Relationships between Scale Levels and Measures*

	Basic Levels of Scales			
Measurement	**Nominal**	**Ordinal**	**Interval**	**Ratio**
Central Tendency				
Mode	**Appropriate**	Appropriate	Appropriate	Appropriate
Median	Inappropriate	**More Appropriate**	Appropriate	Appropriate
Mean	Inappropriate	Inappropriate	**Most Appropriate**	**Most Appropriate**
Dispersion				
Frequency distribution	**Appropriate**	Appropriate	Appropriate	Appropriate
Range	Inappropriate	**More Appropriate**	Appropriate	Appropriate
Estimated standard deviation	Inappropriate	Inappropriate	**Most Appropriate**	**Most Appropriate**

Source: Adapted from Hair et al. (2017, p. 187).

Nominal scales are the most basic and least powerful scale designs, requiring respondents only to provide some descriptor as the response. Responses do not contain a level of intensity, so ranking the responses is impossible. Nominal scales only allow the researcher to categorize the responses into mutually exclusive subsets that do not have distances between them.

Ordinal scales enable respondents to express relative magnitude between the answers to a question, and responses can be rank ordered in a hierarchical pattern. The researcher can determine relationships between reactions, such as "greater than/less than," "higher than/lower than", "more often/less often," "more important/less important," or "more favourable/less favourable." Mathematical calculations with ordinal scales include mode, median, frequency distributions, and ranges but cannot determine the absolute difference between rankings.

Interval scales can measure the absolute difference between scale points. The intervals between scale numbers tell us how far apart the measured objects are on a particular attribute. This approach enables us to compare different levels of any attribute. Beyond the mode and median, researchers can calculate the mean and standard deviation of respondents' responses for interval scales. Researchers can report findings not only about hierarchical differences (better than or worse than) but also the absolute differences between the data.

Ratio scales are the highest-level scale because they enable the researcher to identify the absolute differences between each scale point and make absolute comparisons between the responses. Ratio scales are designed to enable a "true natural zero" or "true state of nothing" response to be a valid response to a question. Generally, ratio scales ask respondents to provide a specific numerical value as their response, regardless of whether a set of scale points is used. In addition to the mode, median, mean, and standard deviation, one can compare levels.

It is important to consider the nature of the variables being studied when selecting the appropriate level of measurement for research. Figure 9 provides a helpful guide for researchers to determine the level of measurement that is most appropriate for their particular study. By carefully considering the characteristics of the variables at hand, researchers can ensure that their data is accurately and appropriately measured, leading to more valid and reliable findings.

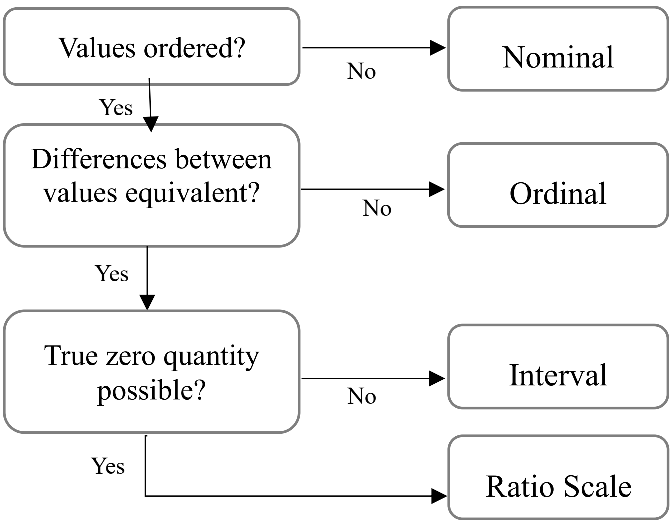


Figure 9. *Decision Tree To Determine Level of Measurement (Crano et al., 2014).*

Part 3. Types of Quantitative Research Designs

Research design is divided into three groups: quantitative, qualitative, and mixed method. The researcher must decide on the most appropriate design for the type of research work (Fig. 10). The field of social sciences and education research employs a variety of quantitative research designs and methods, including experimental, correlational, survey, and quasi-experimental designs.

Quantitative research encompasses various experimental designs, including true experiments, quasi-experiments, applied behavioural analysis, and single-subject experiments. On the other hand, non-experimental quantitative research involves causal-comparative research, where the investigator compares multiple groups based on an independent variable that has already occurred, and correlational design, where investigators use correlational statistics to measure the association between variables or sets of scores.

Furthermore, these designs have evolved into more intricate relationships among variables, such as structural equation modelling, hierarchical linear modelling, and logistic regression techniques. Quantitative strategies have become more complex in recent years, with experiments featuring numerous variables and treatments, such as factorial designs and repeated measure designs. Elaborate structural equation models have also been developed to identify the collective strength of multiple variables and causal paths.

Each design has distinct characteristics and goals. In this section, we will explore these designs in depth, providing a thorough understanding of their suitability for various research inquiries and highlighting their distinctive approaches.

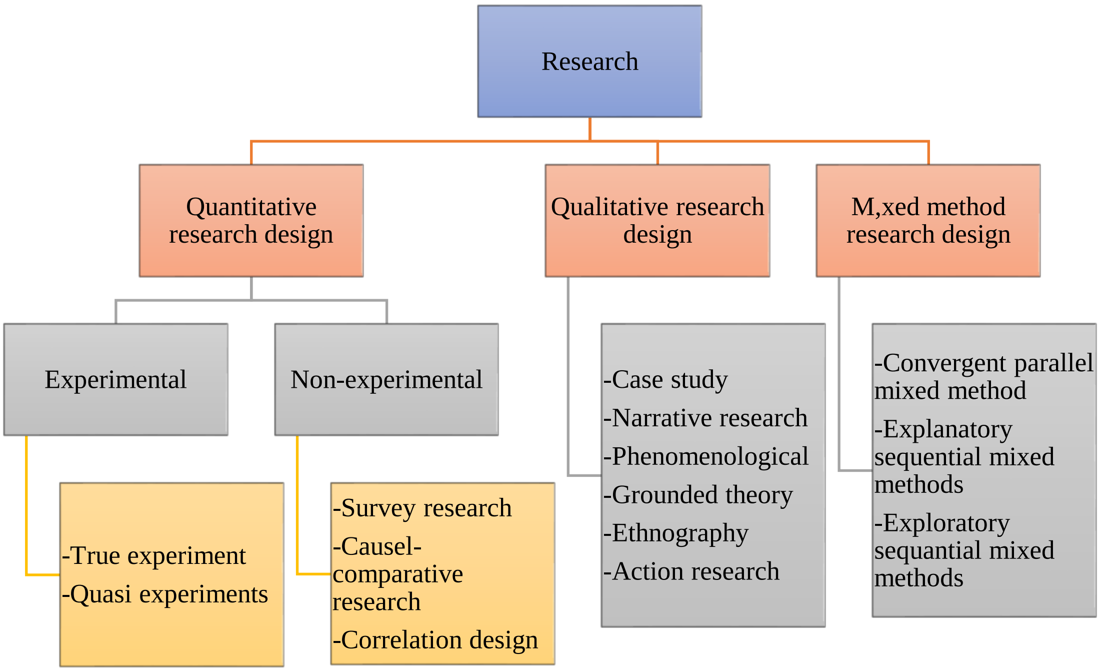


Figure 10. *Summarized form of Research Design Types* (Asenahabi, 2019).

3.1. Assumptions, Objectives, Structure, Methods, and Techniques of Quantitative Research

The methodology employed by a researcher can significantly affect the outcome and accuracy of their study. The research design encompasses the framework, structure, and strategy researchers use to address a research question. By carefully considering these elements, researchers can establish their hypothesis, conduct their study, and interpret their data (Leavy, 2022). In quantitative research, it is pivotal to maintain control and mitigate any factors that could sway or skew the findings. Reviewing the key characteristics and assumptions that underpin quantitative research is vital to appreciate how control plays a crucial role in quantitative research design (Asenahabi, 2019; Bloomfield & Fisher, 2019).

Quantitative research is a formal and systematic process for describing variables, testing their relationships, and examining cause-and-effect associations between variables. Quantitative research generates numerical data, predominantly informed by positivist or post-positivist paradigms and underpinned by various assumptions, such as the belief in a single truth or reality, objectivity, and deduction (Hair et al., 2021; Sukamolson, 2007).

Quantitative research is a scientific method that employs objective and impartial techniques to test hypotheses and find the correct answer. The research process involves drawing a representative sample of participants from a known population and measuring the variables of interest. Typically, researchers begin by testing the null hypothesis, which posits no connection between independent and dependent variables. The hypothesis is scrutinised, and the findings are assessed through statistical analysis. Ultimately, the null hypothesis is deemed acceptable or unacceptable based on the results of the statistical evaluations. Once the null hypothesis is accepted or rejected, inferences or generalisations can be applied to the population of interest. The research design must be reliable and have internal and external validity, collectively known as rigour, to allow confident generalisations about a population (Watson, 2015).

Rigour in quantitative research can be described as the amount of control the researcher exerts to prevent the effects of extraneous or confounding variables on the dependent (test or outcome) variable.

To assess the effect of a single factor on a particular outcome, the researcher must account for any other variables or external factors that could potentially affect the outcome. This can help isolate the specific impact of the independent variable being studied.

For example, to create a fall risk profile, a researcher needs to compare the characteristics of a sample of patients who have experienced a fall with those of a sample of patients who have not.

In this situation, if there is an error in the sample selection of the non-falling patient group and this group accidentally consists of a higher mean for age, then the difference – or not – between the two groups may be due to age due to the sampling error.

Quantitative research uses a range of different study designs. These may vary in how they are categorised according to four significant types of quantitative research: descriptive, correlational, quasi-experimental, and experimental (Tab. 6).

Table 6. *Type of Quantitative Research Design*

Descriptive	It helps describe a phenomenon in a real context. It quantifies and monitors the dynamics of variables without proposing any explanation, as it does not involve manipulating variables.
Correlational	In addition to describing, it determines the degree and direction of the relationship between variables through correlational statistics without determining any cause-and-effect relationship.
Quasi-experimental	It describes and examines relationships between variables, evaluating one variable's influence (effect) on another, but without being able to implement any controls (separation; anteriority), which are only within the reach of experimental studies.
Experimental	Describe and evaluate causal relationships between dependent and independent variables under well-controlled conditions. With the manipulation of independent variables, the random distribution of subjects between groups (experimental and control) exposes the first to at least one intervention to which the second will not be subject.

Source: Adapted from Bloomfield and Fisher (2019).

3.2.Descriptive Research Design

A descriptive quantitative study aims to examine variables in a single sample and systematically measure, describe, and interpret them. It is commonly used to gather data about a specific phenomenon or attribute of interest within a known sample or population in its natural environment without controlling or manipulating variables.

Research studies that describe and examine variables in two or more groups are comparative descriptive designs. The variable/s of interest are measured and described in both groups and are then compared. For example, researchers might conduct a comparative descriptive study to describe the differences in educational qualifications between male and female nurses employed at a local hospital.

Findings from descriptive research studies are most valuable in determining the frequency to which something exists. They are also useful for describing a new or little-known phenomenon. However, while findings cannot be used to establish cause and effect, they may help develop hypotheses that can be tested in future studies.

Researchers must use methods to ensure the data collected is reliable and valid. This includes using a probability sampling technique to select a sample of adequate size and accurately represent the target population.

The instruments and methods used to collect data in descriptive studies include surveys, checklists, observations, interviews, and equipment to measure physiological variables, such as weight scales and thermometers. These must also be calibrated, standardised, and piloted prior to use to ensure internal validity.

3.3.Correlational Research Design

The primary objective of correlational research is to establish a relationship's existence, strength, and direction between two or more variables. That is the degree to which changes in one variable correspond to changes in another. By analysing the correlation coefficient, researchers can determine the nature and direction of the relationship between the variables, which is crucial in making informed decisions based on the data collected.

Like descriptive research, correlational studies do not manipulate the variables investigated and do not seek to determine cause or effect. Instead, they can describe or predict relationships or test theoretical models of relationships. Causal inferences regarding the relationships between independent and dependent variables are not drawn without random selection or manipulation of the independent variable. There is no attempt to make such inferences without following the above-mentioned procedure (Rumrill, 2004). The findings from correlational studies can be statistically explained in three ways: positive, negative, and no correlation.

In statistics, a positive correlation refers to a connection or association between two variables such that when one variable increases, the other variable also increases, or when one variable decreases, the other variable also decreases (Bloomfield & Fisher, 2019). This means that the two variables move in the same direction. For example, the amount of food a person consumes might correlate positively with weight.

A negative correlation between variables occurs when an increase in one variable results in a decrease in another and vice versa. For example, the more food a person consumes, the lower their hunger levels will be. Two variables are said to be uncorrelated when an alteration in one does not lead to an alternation in the other and vice versa.

Researchers typically rely on a statistical measure called the correlation coefficient when reporting the results of correlational studies. This value ranges from +1 to -1, with a figure near +1 indicating a robust positive correlation and a value close to -1 signifying a significant negative correlation. A value close to zero denotes that the variables are not correlated.

3.4.Quasi-experimental research design

The third category of quantitative research design is quasi-experimental studies. These are similar to experimental studies in that they aim to test the effectiveness of interventions and, therefore, involve the manipulation of an independent variable.

However, unlike a proper experimental study (e.g. a randomised controlled trial), they lack the random allocation of participants to certain conditions, such as an intervention/experimental or control group. This may have considerable implications because factors other than those being investigated may affect the findings. These are known as confounding or extraneous variables.

When conducting a randomised controlled trial is not feasible or ethical, quasi-experiments are usually performed. Different types of quasi-experimental studies include, among others (Tab. 7.):

- *non-equivalent control pre-test post-test design*
- *non-equivalent control post-test only design*
- *one group pre-test–post-test design and*
- *time-interrupted series.*

Table 7. *Features of Quasi-Experimental Research Design*

Type	Feature		Test groups
Non-equivalent control pre-test - post-test	Non-randomly assigned control	Experimental group Pre-test → intervention →Post-test	Control Pre-test → Placebo/normal care → post-test
Non-equivalent control post-test only design	Non-randomly assigned control	Experimental group	Control

		Intervention → Post-test	Placebo/normal care → Post-test
One group pre-test – post-test design	No control group	Experimental group Intervention → Post-test	Control Placebo/normal care → Post-test
Time-interrupted series	Non-randomly assigned control or no control group	Experimental group Pre-test → Intervention → Post-test → Folow-up post-test	Control Pre-test → Placebo/normal care → Post-test → Folow-up post-test

Source: Bloomfield and Fisher (2019).

3.5.Experimental designs

The experimental study design has the greatest level of control and has frequently been identified as the gold standard of quantitative research due to its ability to determine a cause-and-effect relationship between an intervention (the cause) and the study outcome (the effect) (Rogers & Révész, 2020).

Scientific research widely recognizes experimental designs as the gold standard. This method, known as true experimentation, establishes a cause-and-effect relationship between variables within a study. Despite common misconceptions, true experimentation is not exclusive to laboratory settings.

Experimental research provides a structured approach to establishing causal relationships between variables. Using this approach, the researcher is actively involved in deducing and testing hypotheses. The researcher manipulates an independent variable (cause) and observes its effect on a dependent variable while attempting to control for extraneous variables. This is achieved by administering the treatment to one group while withholding it from another and then analysing the resulting scores of both groups.

In the realm of research, an experiment involves selecting participants randomly and exposing them to different levels of one or more variables, known as independent variables. The researcher then observes the impact of this exposure on one or more outcome variables called dependent variables. The aim of conducting an experiment is to establish a correlation between the independent and dependent variables and deduce conclusions regarding the effectiveness of the intervention and its causal connection. A key aspect of this process is controlling for extraneous variables. This control is essential as it ensures that any observed effects are solely due to manipulating the independent variables, enhancing the study's validity. Experiments are a potent tool for investigating cause-and-effect relationships in diverse fields, including psychology, medicine, physics, and engineering (Mizik & Hanssens, 2018).

True experiments randomly allocate subjects to treatment conditions, while quasi-experiments utilize nonrandomized assignments. To ensure the similarity of subjects, cases are matched on various characteristics and randomly allocated to control and experimental groups. Only observable facts are considered, and inferential statistics produce precise numerical results. Various experimental designs have been developed, from simple before-after to complex multivariate factorial designs, including:

- *Parallel design - In a parallel design, participants are randomly assigned to either the intervention or control group.*

- *Crossover design - In a crossover design, participants are initially assigned to either the intervention or control group and then switch over to the other group after a certain period of time. This design helps negate bias from individual differences since each participant becomes their own control.*
- *Cluster design - In many research contexts, it is not always possible to randomise individuals to receive different interventions. To overcome this, groups or clusters of individuals (for example, wards, units or hospitals) can be randomly assigned to either the control or intervention and all cluster members will receive the allocation.*

The randomised controlled trial (RCT) is a highly valued research approach that embodies these virtues (Styles & Torgerson, 2018) (Fig. 11).

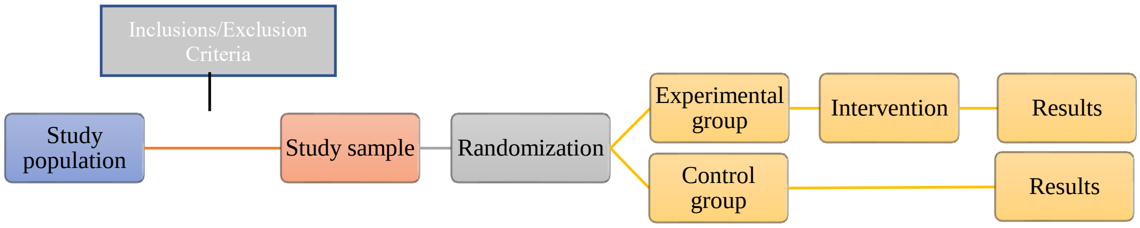


Figure 11. *RTC-Randomised Controlled Trial* **(Crano et al., 2014).**

All variables are identified and controlled in this type of experiment except for one. The --- independent variable is manipulated to observe its effects on dependent variables. Additionally, participants are randomly assigned to experimental treatments instead of being selected from naturally occurring groups. This ensures the validity of the research.

The fundamental principles of experimental designs include random assignment, variable manipulation, and control groups. While experimental designs effectively establish causal relationships, they also have limitations, such as ethical considerations and practical constraints.

The fundamental framework of a quantitative design is rooted in the scientific method, utilising deductive reasoning. This involves the researcher developing a hypothesis, investigating to gather data on the problem, and then analysing and sharing the conclusions to demonstrate that the hypotheses are not untrue.

To follow this procedure, one should:

- *Observe an unknown, unexplained, or new phenomenon and research the current theories relating to the issue.*
- *Create a hypothesis to explain the observations made.*
- *Predict outcomes based on these hypotheses and create a plan to test the prediction.*
- *Collect and process data. If the prediction is accurate, proceed to the next step. If not, create a new hypothesis based on the available knowledge.*
- *Verify the findings, draw your conclusions, and present the results in a suitable format.*

Crano et al. (2014) outline the steps of a classic true experimental research design, which involve gathering a group of participants, conducting a pretest on the dependent variable, randomly assigning participants to either the experimental or control group, closely controlling the application of the experimental treatment between the two groups, and measuring both groups again on the dependent variable after the experimental manipulation (Fig. 12). Variations exist, such as removing pretesting, including multiple experimental treatments, or using the same participants across all experimental conditions.

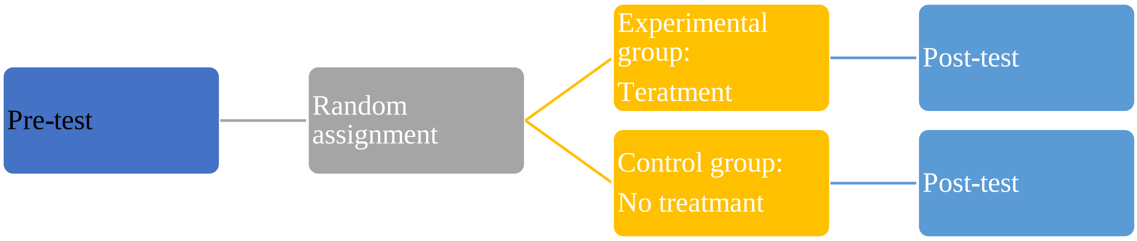


Figure 12. *Pre-test - Post-test Control Group Design* (Crano et al., 2014).

Following crucial steps is essential to designing and conducting experiments effectively. These steps include the following (Kuçuksayraç, 2007):

- *Sampling participants for the study.*
- *Randomly assigning participants to groups.*
- *Randomly assigning groups to experimental or control conditions.*
- *Defining the independent variable, which refers to the aspect of the environment being studied that varies between the groups.*
- *Defining the dependent variable, which measures any resulting behavioural changes.*
- *Controlling all other variables that may affect the dependent variable while keeping the independent variable consistent.*
- *Conducting statistical tests to confirm or refute the hypothesis to determine whether there are any differences between the two groups regarding the dependent variable measurements.*
- *If the hypothesis is confirmed, explain and generalise the findings.*
- *Finally, predict how the findings may apply to other situations, potentially through replication.*

In conclusion, quantitative research is an important method for measuring variables and evaluating the effectiveness of interventions. Unlike qualitative research, quantitative research is conducted objectively, emphasizing reducing bias. Researchers who seek to adopt evidence-based practices must have a strong understanding of quantitative research design. This knowledge allows them to comprehend and evaluate research literature better and potentially integrate study outcomes and recommendations into their work.

Table 8. summarizes the alternatives for implementing quantitative and experimental research projects adjusted to different research conditions.

Table 8. *Quantitative Experimental Design Options*

	Design	Groups	Action
Pre-experimental design	On-shot case study	Single group	One group experimental intervention
	One-group pre-test – Post-test design	Single group	Pre-test, experimental intervention, post-test
	Static-group comparison	Two groups	One group experimental intervention, both groups post-test
True Experimental designs	Pre-test – post-test control group design	Two groups randomly assigned	Both groups pre-test, one group experimental intervention, both groups post-test
	Solomon four groups design	Four groups randomly assigned	One group pre-test, experimental intervention, post-test
			One group pre-test and post-test
			One group experimental, intervention and post-test
			One group post-test only
	Post-test only control group design	Two groups randomly assigned	One group experimental intervention, post-test
			One group post-test only
Quasi-experimental designs	Time-series experiment	Single group	Measures taken overtime, experimental intervention, measures taken
	Multiple time-series experiment	Two groups	Measures of both taken over time, one group experimental intervention, measures of both groups taken
	Non-equivalent control group design	Two groups	One group pre-test, experimental intervention, post-test, one group pre-test and post-test only
Single subject designs	Single subject	One individual	Record multiple observations to determine baseline and then introduce the experimental intervention and record multiple observations

Source: Adapted from Leavy (2022).

3.6.Non-experimental Designs

3.6.1. Correlational Designs

Correlational designs are a valuable tool for exploring the connections between variables without manipulating them. This type of research involves intentionally investigating the associations between variables without any intervention. The goal is establishing relationships that can be further examined through experimental research. However, it's important to note that correlational designs have limitations, including the potential for third-variable effects and a lack of causality. These limitations are important to keep in mind when interpreting the results of correlational studies (Asenahabi, 2019).

3.6.2. Survey Designs

Survey research is a method of collecting data that provides a numerical or quantitative description of a population's trends, attitudes, or opinions. The research involves studying a population sample, to generalize the results to the whole population. The data is collected through questionnaires or structured interviews using cross-sectional and longitudinal studies.

Survey research is a valuable tool for providing a numerical description of trends, attitudes, or opinions within a population. By studying a sample of that population, researchers can conduct cross-sectional or longitudinal studies using questionnaires or structured interviews for data collection (Hu & Chang, 2017). The ultimate goal is to generalize the findings from the sample to the larger population.

Survey designs involve collecting data through structured questionnaires or interviews. This section discusses the key components of survey designs, including sampling techniques, questionnaire design, and data analysis. It explores the strengths of survey designs in capturing large amounts of data and providing insights into attitudes, opinions, and behaviours. The limitations, such as response biases and reliance on self-reported data, are also addressed.

Survey research uses well-designed questionnaires and scientific sampling to accurately measure a population's characteristics. This allows for precise comparisons between groups and provides estimates representative of the entire population. Survey research requires random sampling techniques such as random digit dialling and sampling procedures to ensure scientific accuracy.

Survey research utilizes scientifically sound sampling techniques and well-designed questionnaires to accurately measure the characteristics of a population, allowing for confident estimates that can be compared across different groups. To ensure scientific accuracy, specific techniques such as random digit dialling and sampling procedures are employed to randomly sample respondents (Leavy, 2022).

Part 4: Strengths and Limitations of Quantitative Research Designs

The most trusted way to gather dependable data on treatment or intervention effects is through experimental research, widely recognised as the top quantitative methodology. Amongst all research techniques, the randomised controlled trial is regarded as the epitome of evidence-based research. This is because experimental research is empowered by the ability to exercise control, which entails adhering to standardised protocols and minimising potential biases to eliminate inaccurate findings.

There exist various techniques for exerting control, includingl6 but not limited to

- *the random selection of a sample,*
- *the application of inclusion/exclusion criteria,*
- *the use of a comparison group,*
- *the matching of subjects across groups,*
- *the manipulation of the independent variable,*
- *the implementation of single, double, or treble blinding procedures and*
- *the utilisation of precise measuring tools and standardised statistical tests in the final data analysis.*

Quantitative research provides valuable insights into the larger population, allowing us to uncover the prevalence of attitudes held by individuals. It generates data that can be easily summarized into statistical figures, enabling comparisons between groups using statistics. It is highly accurate, conclusive, and consistent, quantifying the frequency of events, actions, and trends. This type of research is particularly useful when answering questions such as "How many?" and "How often?"(Walker, 2005).

Different research designs are available for quantitative measures, each with advantages and disadvantages. Experimental research may have certain methodological limitations that affect the precision and applicability of the research outcomes. These drawbacks can hinder the research's real-world significance.

4.1. Objectivity, Reliability, Validity, Generalizability

One benefit of conducting experimental research is that it enables researchers to attribute the results to the experiment's effects confidently. This differs from descriptive and correlational research, which uses fewer rigid approaches to explore and describe phenomena and may not yield clear conclusions. Experimental research's scientific and statistical rigor maximizes internal validity and increases the likelihood of generalising findings beyond the study sample. However, it is essential to note that the literature highlights methodological limitations that could impact research results' internal and external validity, ultimately limiting their practical applications (Walker, 2005).

Achieving sample randomisation in a study is essential for ensuring that the sample accurately reflects the parent population and that the findings are generalisable. Without randomisation, it can be challenging for researchers to match participants based on critical traits that could impact the study's outcomes. Even with systematic and protocolised procedures in place, external factors can still influence the results of experiments with humans, especially in longitudinal research designs.

Challenges to validity, both internal and external, as well as the Hawthorne effect, pose limitations to experimental research. Participants may alter their behaviour simply because they are observed, prompting researchers to adopt a 'double-blind' technique. Additionally, reliable, and validated outcome measures are crucial for meaningful results. Quantitative research designs can contribute significantly to the evidence base, while descriptive and correlational designs are essential for generating questions and hypotheses. Despite rigorous procedures to control error and bias, ethical and methodological concerns remain for experimental research.

A sound research design aims to reduce bias and increase the dependability of the data gathered and analysed. The design that produces the least experimental error is typically considered the optimal approach in scientific inquiry. Likewise, a suitable and effective design results in the acquisition of comprehensive information and enables consideration of various facets of a given problem. The traits of a robust research design include:

- *Accuracy is crucial regarding the methods used to collect data and evaluate responses. To ensure Objectivity, it is essential to implement unbiased measuring tools that deliver consistent results, regardless of who performs the evaluation.*
- *Reliability, in turn, refers to the degree of consistency in responses obtained from multiple measurements. Essentially, if a participant offers a specific response to a particular question, they are expected to provide the same response if the question is posed again. Any fluctuations in their answers may weaken the Reliability of the data gathered. As such,*

researchers must design their surveys to ensure Reliability and consistency in the responses obtained.

- *Quantitative Research Design faces a significant obstacle in accurately measuring intended variables. The validity of a measuring device or instrument must be deemed valid by measuring only what it intends to, and this quality can be demonstrated. For example, an IQ test should only measure intelligence and use appropriately framed questions. To estimate the validity of measures, including facial, convergent, discriminant, and predictive validity, agreement with available literature and quantitative analyses of consistency and available predictions are necessary despite the constantly evolving nature of measures.*
- *Generalizability refers to applying data collected from a sample to a larger population. It is crucial to design the research study carefully to ensure that a researcher's findings can be generalized. This involves accurately defining the population, selecting an appropriate sample, conducting proper statistical analysis, and ensuring that the study's conclusions can be applied to the population.*

Sufficient information is essential to analyse the research problem from a broader perspective, create an effective research design, and identify important factors such as research objectives, methods for obtaining information, skilled human and financial resources availability, minimizing bias, and maximizing reliability and generalization. An ideal research design should be flexible, adaptable, efficient, and cost-effective.

The research design is the foundation for data collection and analysis. It outlines the methods and procedures for measuring and analysing data, allowing investigators to investigate research questions and establish the conditions for data collection and analysis so that the findings of a sample can be generalized to the larger population (Kuçuksayraç, 2007).

4.2. Considerations in Choosing the Design

A research design is a foundation for data collection and analysis in a study. It outlines the methods and procedures, serving as a blueprint for measuring and analysing data. This strategy enables investigators to investigate research questions and arrange the conditions for data collection and analysis in a manner that permits the findings of a sample to be generalised to the larger population (Pandey & Pandey, 2015).

Choosing the appropriate research design is crucial while conducting research. It involves considering various factors, from broader assumptions to specific data collection and analysis techniques. It is essential to note that there is no fixed order for making these decisions, and they should depend on what is most relevant to your research objectives.

Research approaches consist of many plans and procedures that guide the steps from broader assumptions to detailed data collection, analysis, and interpretation methods. When selecting an approach to study a topic, you need to consider the philosophical assumptions, procedures of inquiry (research designs), and specific research methods of data collection, analysis, and interpretation. Your choice of a research approach should be informed by the nature of the research problem or issue, your personal experiences, and the intended audience. These three elements - research approaches, research designs, and research methods - provide a framework for understanding the perspective of research.

Ultimately, your research design will depend on various factors, such as the problem or issue you are investigating, your personal experiences, and the audience you intend to reach. For instance, quantitative research may be the most suitable approach if you aim to test objective theories by examining the relationship between variables. This method involves measuring variables using instruments, which generates numerical data that can be analysed using statistical methods (Creswell, 2014).

A design or structure must be established to begin data collection or analysis in social research. However, a research design is more than just a work plan. While a work plan outlines the steps necessary to complete a project, it is based on the research design. Essentially, the function of a research design is to ensure that the evidence collected allows us to clearly answer the initial question.

It is important to note that research design is a logical problem rather than a logistical one. In social research, factors such as sampling, data collection method, and question design are all subsidiary to the question of what evidence needs to be collected. Unfortunately, many researchers jump into designing questionnaires or conducting interviews before considering what information they need to answer their research questions.

It is also essential to distinguish between the research design and the data collection method. The data collection method has no bearing on the logic of the design. Explanatory research seeks to develop and evaluate causal theories, but in social sciences, causation is probabilistic rather than deterministic. This is why the role of research design is so crucial - it minimizes the likelihood of drawing incorrect causal inferences from data. When designing research, the type of evidence needed to answer the research question convincingly must be identified.

4.3. Research Problem and Purpose, Research Question and Hypothesis

The research purpose serves as a statement explaining why a study is being conducted and what the study aims to achieve. It typically involves identifying, describing, explaining, or predicting a specific concept, situation or solution. The purpose statement is important because it clarifies the variables, population, and setting for the study, and it is usually derived from the research problem. The purpose statement must be objective and free from any researcher's biases or values to ensure the research is credible and trustworthy. Typically, the purpose statement comes after the research problem and supports it by clarifying the knowledge generated through the study (Durand, 2023).

Quantitative research is rooted in deduction. This type of research is guided by theory, meaning it begins and ends with theory. Before embarking on any research project, researchers need to become acquainted with the relevant literature. This involves understanding the established theories and explanations about the phenomenon being studied and identifying gaps or controversies in the existing knowledge (Fig. 13.). By utilising this knowledge, researchers can develop hypotheses to address these gaps or controversies. The primary aim of quantitative research is to test existing theories using new quantitative data to establish the limits of a theory or determine the conditions under which it applies. Regardless of the research question, it must always be derived from the theory. A well-formulated research question should be pertinent and have the potential to make a theoretical contribution to the literature. It should either add to, modify, challenge, or fill gaps in the existing theory or controversies (Stockemer, 2019).

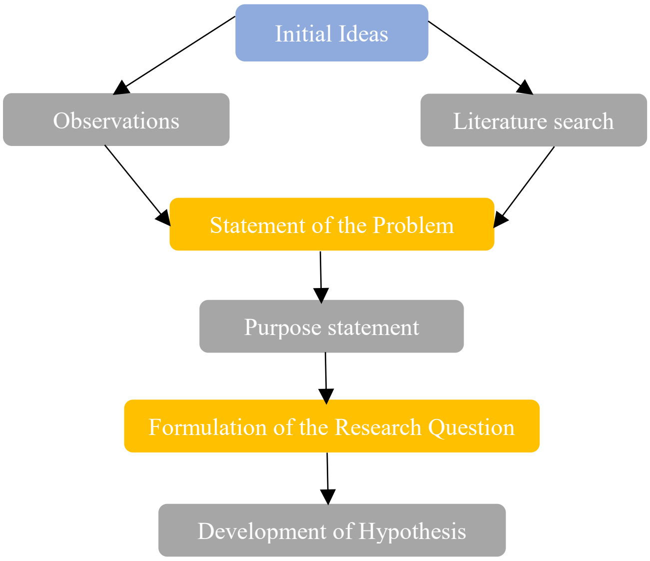


Figure 13. ***Research Purpose, Research Question and Hypothesis***

A well-crafted research design is the key to efficient and cost-effective research without compromising the accuracy of findings. It lays the foundation for the entire research journey and ensures that all aspects of the project are meticulously planned to yield valuable insights. With a solid Research Design in place, scaling becomes effortless, and resources are allocated judiciously, thus making it an indispensable element in accomplishing research objectives.

Proper research design is also crucial when mapping out methodologies for collecting pertinent data and analysing techniques while considering research objectives, staff time, and budgetary constraints. Failing to prepare a research design properly can harm the entire project (Pandey & Pandey, 2015).

Research design is like a blueprint for seamless research operations, much like a blueprint or map is necessary for a house's efficient and aesthetically pleasing construction.

Furthermore, research design offers an overview of the research process, allowing field experts to provide valuable input. It also assists investigators in organising their thoughts and identifying any potential flaws.

Ultimately, research design provides a clear trajectory for all parties involved in the research project, guaranteeing a coordinated and successful outcome.

References

Aaker, D., Kumar, V., Leone, R. P., & Day, G. S. (2013). *Marketing research* (11th ed.). Wiley.

Asenahabi, B. M. (2019). Basics of research design: A guide to selecting appropriate research design. *International Journal of Contemporary Applied Researches*, *6*(5), 76-89.

Bhushan Mishra, S., & Alok, S. (2019). *Handbook of research methodology.* EDUCREATION PUBLISHING. www.educreation.in

Bloomfield, J., & Fisher, M. J. (2019). Quantitative research design. *Journal of the Australasian Rehabilitation Nurses Association*, *22*(2), 27-30.

Burkholder, G. J., Cox, K. A., Crawford, L. M., & Hitchcock, J. H. (2019). *Research design and methods: An applied guide for the scholar-practitioner.* Sage.

Cash, P., Stanković, T., & Štorga, M. (2016). *Experimental design research: Approaches, perspectives, applications*. Springer.

Crano, W. D., Brewer, M. B., & Lac, A. (2014). *Principles and methods of social research* (Third edition). Routledge.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed). Sage.

Creswell, J. W., & Clark, V. L. P. (2018*). Designing and conducting mixed methods research*. Sage.

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

Fryer, L., Larson-Hall, J., & Stewart, J. (2018). Quantitative methodology: Experimental and longitudinal advances in language learning research. In A. Phakiti, P. D. Costa, L. Plonsky, & S. Starfield (Eds.), *Palgrave handbook of applied linguistics research methodology.* Palgrave.

Haig, B. D. (2017). *The philosophy of quantitative methods*. Oxford University Press.

Hair, J. F., Ortinau, D. J., & Harrison, D. E. (2017). *Essentials of marketing research*. McGraw-Hill.

Hair, J. F., Ortinau, D. J., & Harrison, D. E. (2021). *Essentials of marketing research*. McGraw-Hill Education.

Hu, C.-P., & Chang, Y.-Y. (2017). John W. Creswell, Research design: Qualitative, quantitative, and mixed methods approaches. *Journal of Social and Administrative Sciences, 4*(2), 205–207. https://doi.org/10.1453/jsas.v4i2.1313

Kotari, C. R. (2004). *Research methodology: Methods and techniques* (2nd ed.). New Age International.

Kuçuksayraç, B.Sc., E. (2007). *Exploring the term "experiment" in industrial design* [İstanbul Technical University]. https://core.ac.uk/download/62730442.pdf

Kumar, R. (2011). *Research Methodology: A step-by-step guide for beginners* (3rd ed.). Sage.

Leavy, P. (2022). *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches*. Guilford.

Miller, C. J., Smith, S. N., & Pugatch, M. (2020). Experimental and quasi-experimental designs in implementation research. *Psychiatry Research, 283,* 112452. https://doi.org/10.1016/j.psychres.2019.06.027

Mizik, N., & Hanssens, D. (2018). *Handbook of marketing analytics*. Edward Elgar. https://doi.org/10.4337/9781784716752

Mtshweni, V. B. (2019). *The effect of sense of belonging and adjustment on undergraduate students' intention to dropout of university* [Doctoral dissertation]. University of South Africa. https://core.ac.uk/download/pdf/286383953.pdf

Mwansa, B., Austin Mwange, D. B. A., Windu Matoka, D. B. A., Joseph, C. I., Chibawe, O., Robbson Manda, M. P. A., & Mashiri, G. (2022). Theoretical review: An explanation of data types, statistical tests, and factors that influence presentation of findings. *Research on Humanities and Social Sciences, 12*(22), 52-67. https://doi.org/10.7176/JMCR/87-06

Pandey, D. P., & Pandey, D. M. M. (2015). *Research methodology: Tools and techniques.* Bridge Center.

Price, O., & Lovell, K. (2018). Quantitative research design. In *A research handbook for patient and public involvement researchers* (pp. 40–50). Manchester University Press.https://www.manchesteropenhive.com/display/9781526136527/9781526136527.00008.xml

Durand, Rodolphe (2023) From the boardroom: Making purpose research relevant for practice. *Strategy Science, 8*(2),149-158. https://doi.org/10.1287/stsc.2023.0182

Rogers, J., & Révész, A. (2020). Experimental and quasi-experimental designs. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 133-143). Routledge.

Rumrill P. D., Jr (2004). Non-manipulation quantitative designs. *Work (Reading, Mass.), 22*(3), 255–260.

Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie, 69*(S2), 107–131. https://doi.org/10.1007/s11577-017-0454-1

Styles, B., & Torgerson, C. (2018). Randomised controlled trials (RCTs) in education research –methodological debates, questions, challenges. Educational Research, 60(3), 255–264. https://doi.org/10.1080/00131881.2018.1500194

Stockemer, D. (2019). *Quantitative methods for the social sciences: A practical introduction with examples in SPSS and Stata*. Springer.

Sukamolson, S. (2007). Fundamentals of quantitative research. *Language Institute Chulalongkorn University*, *1*(3), 1-20.

White, H., & Sabarwal, S. (2014). *Quasi-experimental design and methods. Methodological briefs: Impact evaluation*. UNICEF Office of Research.

Wilson, B., Austria, M.-J., & Casucci, T. (2021). *Understanding quantitative and qualitative approaches*. Health University of UTAH. https://rb.gy/9zrvuh.

QUANTITATIVE ANALYSIS

1. INTRODUCTION

Although Generalized Linear Models (GLM) allows for easy execution of regression models, in practice, univariate GLM is primarily used for analysis of variance (ANOVA) and analysis of covariance (ANCOVA) models (Rutherford, 2011:1-2). On the other hand, multivariate GLM is primarily utilized for multiple analysis of variance (MANOVA) and multiple analysis of covariance (MANCOVA) models (Huberty & Petoskey, 2000). In SPSS, multivariate GLM is a separate module, while in SAS, it is implemented within PROC GLM using the MANOVA statement.

ANOVA is a statistical technique used to investigate the main and interaction effects of categorical independent variables (referred to as "factors") on a continuous dependent variable (West etal., 1996). It examines whether the means of the groups formed by different values of the independent variable(s) differ significantly. ANOVA allows for the identification of both main effects (the direct effect of an independent variable on the dependent variable) and interaction effects (the combined effect of two or more independent variables). In contrast to regression models, which require the explicit addition of interaction terms, ANOVA inherently detects interaction effects (Jaccard, 1998). In the case of multiple dependent variables, multivariate GLM implements MANOVA, which can also incorporate control variables as covariates (MANCOVA).

The key statistic in ANOVA is the F-test, which evaluates whether the differences in group means are significant enough to suggest that they did not occur by chance (Tian etal., 2018: 61) If the group means do not differ significantly, it implies that the independent variable(s) did not have a significant effect on the dependent variable. However, if the F-test indicates a significant relationship between the independent variable(s) and the dependent variable, multiple comparison tests can be conducted to determine which specific values of the independent variable(s) contribute the most to this relationship.

It is important to note that ANOVA tests the null hypothesis that group means are equal, not that variances are equal. However, ANOVA assumes relative homogeneity of variances, meaning that the groups formed by the independent variable(s) have similar variances on the dependent variable. Homogeneity of variances can be assessed using tests such as Levene's test (Levene, 1960). Similar to regression, ANOVA is a parametric procedure that assumes the multivariate normality of the dependent variable for each value category of the independent variable(s) (Dattalo, 2013: 14).

ANCOVA, on the other hand, is used to test the main and interaction effects of categorical variables on a continuous dependent variable while controlling for the effects of selected continuous variables that covary with the dependent variable (Ankarali etal, 2018: 283). These covariates, also known as control variables, can be used to predict the dependent variable through regression analysis. ANCOVA then performs an ANOVA on the residuals (the predicted minus actual dependent variables) to determine if the factors are still significantly related to the dependent variable after accounting for the variation explained by the covariates. ANCOVA serves three purposes: 1) in quasi-experimental designs, it helps remove the effects of variables that modify the relationship between categorical independent variables and the interval dependent variable; 2) in experimental designs, it controls for factors that cannot be randomized but can be measured on an interval scale; and 3) in regression models, it accommodates the presence of both categorical and interval independent variables.

All three purposes of ANCOVA aim to reduce the error term in the model. ANCOVA can be seen as a type of "what if" analysis, examining what would happen if all cases had equal scores on the covariates, allowing for the isolation of the effects of factors beyond the influence of the covariates. The use of ANCOVA is applicable in various ANOVA designs, and the same assumptions regarding homogeneity of variances and multivariate normality still apply.

It is important to distinguish GLM from other types of models, such as generalized linear models (GZLM) that incorporate nonlinear link functions, linear mixed models (LMM) that handle multilevel data, and generalized linear mixed models (GLMM) that combine nonlinear link functions with LMM. SPSS also offers analysis of variance components (VC), which is a subset of LMM and serves similar functions as ANOVA under GLM. A comparison between GLM, LMM, and VC, along with data illustrations, can be found in the section on linear mixed models. While both GLM and LMM allow for the inclusion of random effects in models, LMM is generally preferred when random effects are present, as explained in the comparison.

1.1 Normality Assumption

Normality is one of the most important assumptions in ANOVA type analysis. So, it is important to check whether each variable in the analysis has a normal distribution.

There are several measures and indicators that you can use to check the normality assumption.

- *You can read skewness and kurtosis statistics, values and z-test results.*
- *You can use Kolmogorov-Smirnov (KS Test) and Shapiro-Wilk Tests (Razali & Wah, 2011).*
- *You can examine the histogram or any other graphs.*

Skewness & Kurtosis
Samples < 50 use the z value -1.96 and +1.96 (SPSS does not report Z-test results. However, it is possible to calculate it manually. Z-Test value: divide skewness and kurtosis statistics by their Standard Error Values)
Samples 50 < N < 300, you can use a more wide range for examining the Z-test results: -3.29 and +3.29
Z value is extremely sensitive to sample size. So it is not efficient to use it for larger sample sizes.
For larger samples:
Absolute skewness value will be between -2 and +2
Absolute kurtosis value will be between -7 and +7 (Westfall & Henning, 2013: 249)
Kolmogorov-Smirnov (KS Test) and Shapiro-Wilk Tests
Use with samples below 300
Samples larger than 300, these test may be unreliable. (Kim, 2013, p. 52-54)

Null hypothesis of the both test are that the data is normally distributed. So, p-values should be higher than 0.05, so we can accept the null hypothesis. However, if samples are more than 300, skewness and kurtosis values should be considered.

Let's practice the normality test!

Select cross_sell.sav

Click on Analyze button on top menu. Then go to Descriptive Statistics and click on Explore button.

Select following variables and put them on to dependent list:

Special offer purchases [buyoff]

CD purchases [buycd]

Book purchases [buybk]

CD club discount [disccd]

Book club discount [discbk]

Log of CD club discount [lndisccd]

Log of Book club discount [lndiscbk]

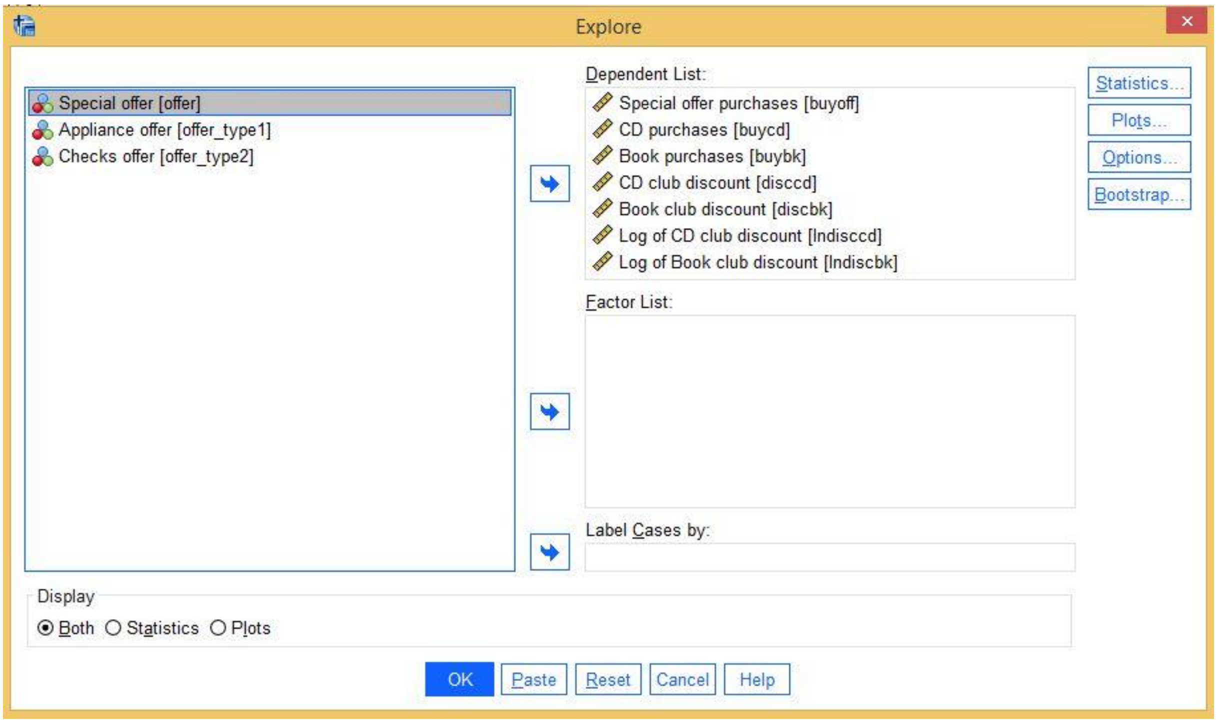


Figure 1. Variable Selection

After that click on Plots button on the right menu.

Click Histogram under the Descriptive title and also select Normality plots with tests. After that click on Continue button.

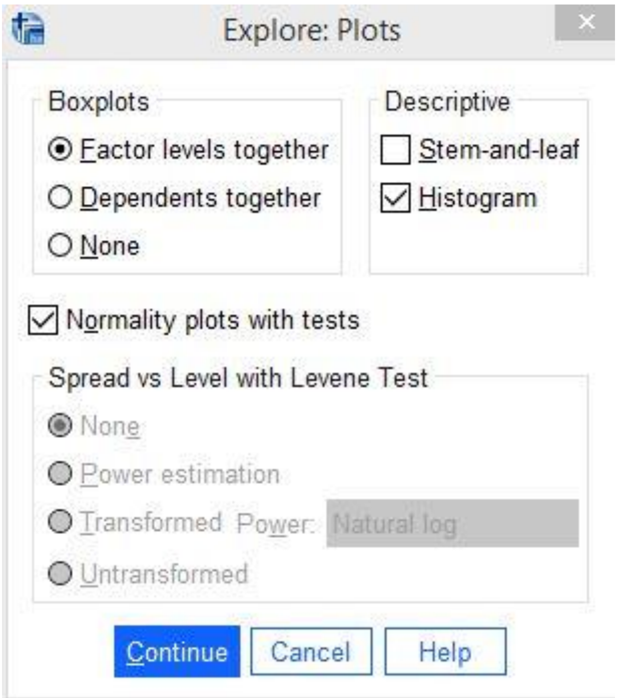


Figure 2. ***Normality Plots with Tests***

On the main menu click OK to undertake the tests and see the results.

Table 1. ***Case Processing Summary***

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Special offer purchases	99	100,0%	0	0,0%	99	100,0%
CD purchases	99	100,0%	0	0,0%	99	100,0%
Book purchases	99	100,0%	0	0,0%	99	100,0%
CD club discount	99	100,0%	0	0,0%	99	100,0%
Book club discount	99	100,0%	0	0,0%	99	100,0%
Log of CD club discount	99	100,0%	0	0,0%	99	100,0%
Log of Book club discount	99	100,0%	0	0,0%	99	100,0%

Table 2. ***Descriptives***

			Statistic	Std. Error
Special offer purchases	Mean		2,5507	,04892
	95% Confidence Interval for Mean	Lower Bound	2,4536	
		Upper Bound	2,6478	
	5% Trimmed Mean		2,5553	
	Median		2,4800	
	Variance		,237	
	Std. Deviation		,48670	

	Minimum		1,35	
	Maximum		3,80	
	Range		2,45	
	Interquartile Range		,74	
	Skewness		,010	,243
	Kurtosis		-,313	,481
CD purchases	Mean		7,8790	,15059
	95% Confidence Interval for Mean	Lower Bound	7,5801	
		Upper Bound	8,1778	
	5% Trimmed Mean		7,8596	
	Median		7,6700	
	Variance		2,245	
	Std. Deviation		1,49835	
	Minimum		3,61	
	Maximum		11,35	
	Range		7,74	
	Interquartile Range		2,06	
	Skewness		,237	,243
	Kurtosis		,099	,481
Book purchases	Mean		6,0407	,14023
	95% Confidence Interval for Mean	Lower Bound	5,7624	
		Upper Bound	6,3190	
	5% Trimmed Mean		6,0591	
	Median		6,1400	
	Variance		1,947	
	Std. Deviation		1,39532	
	Minimum		2,26	
	Maximum		9,20	
	Range		6,94	
	Interquartile Range		1,81	
	Skewness		-,194	,243
	Kurtosis		-,177	,481
CD club discount	Mean		26,92	1,904
	95% Confidence Interval for Mean	Lower Bound	23,14	
		Upper Bound	30,70	

	5% Trimmed Mean		25,91	
	Median		20,00	
	Variance		359,034	
	Std. Deviation		18,948	
	Minimum		5	
	Maximum		70	
	Range		65	
	Interquartile Range		30	
	Skewness		,615	,243
	Kurtosis		-,778	,481
Book club discount	Mean		28,54	1,948
	95% Confidence Interval for Mean	Lower Bound	24,67	
		Upper Bound	32,40	
	5% Trimmed Mean		27,42	
	Median		25,00	
	Variance		375,639	
	Std. Deviation		19,381	
	Minimum		5	
	Maximum		80	
	Range		75	
	Interquartile Range		35	
	Skewness		,682	,243
	Kurtosis		-,425	,481
Log of CD club discount	Mean		2,9912	,08427
	95% Confidence Interval for Mean	Lower Bound	2,8240	
		Upper Bound	3,1584	
	5% Trimmed Mean		3,0006	
	Median		2,9957	
	Variance		,703	
	Std. Deviation		,83846	
	Minimum		1,61	
	Maximum		4,25	
	Range		2,64	
	Interquartile Range		1,39	

	Skewness		-,332	,243
	Kurtosis		-1,095	,481
Log of Book club discount	Mean		3,0750	,08075
	95% Confidence Interval for Mean	Lower Bound	2,9147	
		Upper Bound	3,2352	
	5% Trimmed Mean		3,0897	
	Median		3,2189	
	Variance		,645	
	Std. Deviation		,80341	
	Minimum		1,61	
	Maximum		4,38	
	Range		2,77	
	Interquartile Range		1,50	
	Skewness		-,407	,243
	Kurtosis		-,870	,481

When we examine the descriptive statistics for variables:

Variable: Special offer purchases

Skewness: Statistic: 0.01 Standard Error: 0.243 – Z-Test value: 0.01 / 0.243 = **0.041**

Kurtosis Statistic: -0.313 Standard Error: 0.481 – Z-Test value: -0.313 / 0.481 = **-0.65**

Variable: CD purchases

Skewness: Statistic: 0.237 Standard Error: 0.243 – Z-Test value: 0.237 / 0.243 = **0.975**

Kurtosis: Statistic: 0.099 Standard Error: 0.481 – Z-Test value: 0.099 / 0.481 = **0.203**

Variable: Book purchases

Skewness: Statistic: -0.194 Standard Error: 0.243 – Z-Test value: -0.194 / 0.243 = **-0.798**

Kurtosis: Statistic: -0.177 Standard Error: 0.481 – Z-Test value: -0.177 / 0.481 = **-0.368**

Variable: CD club discount

Skewness: Statistic: 0.615 Standard Error: 0.243 – Z-Test value: 0.615 / 0.243 = **2.53**

Kurtosis: Statistic: -0.778 Standard Error: 0.481 – Z-Test value: -0.778 / 0.481 = **-1.617**

Variable: Book club discount

Skewness: Statistic: 0.682 Standard Error: 0.243 – Z-Test value: 0.682 / 0.243 = **2.81**

Kurtosis: Statistic: -0.425 Standard Error: 0.481 – Z-Test value: -0.425 / 0.481 = **-0.88**

Variable: Log of CD club discount

Skewness: Statistic: -0.332 Standard Error: 0.243 – Z-Test value: -0.332 / 0.243 = **-1.37**

Kurtosis: Statistic: -1.095 Standard Error: 0.481 – Z-Test value: -1.095 / 0.481 = **-2.28**

Variable: Log of Book club discount

Skewness: Statistic: -0.407 Standard Error: 0.243 – Z-Test value: -0.407 / 0.243 = **-1.674**

Kurtosis: Statistic: -0.870 Standard Error: 0.481 – Z-Test value: -0.870 / 0.481 = **-1.808**

Since the number of N of each variable is 99. It is possible to check the Z-test values in the range of -3.29 and +3.29. Therefore, it can be said that all of the variables are normally distributed.

However, we also need to look for normality test results. Kolmogorov-Smirnov and Shapiro-Wilk tests results show that Special offer purchases, CD purchases and Book purchases are normally distributed since their Sig. (p-value) is bigger than 0.05. For the rest of the variables, we have to reject the null hypothesis. When both of these tests are examined, even though the p-values differ, they yield consistant results.

Table 3. ***Tests of Normality***

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Special offer purchases	,063	99	,200*	,988	99	,493
CD purchases	,072	99	,200*	,982	99	,180
Book purchases	,058	99	,200*	,993	99	,881
CD club discount	,148	99	,000	,907	99	,000
Book club discount	,148	99	,000	,921	99	,000
Log of CD club discount	,117	99	,002	,918	99	,000
Log of Book club discount	,117	99	,002	,934	99	,000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

It is also possible to check the distribution from histogram of the variables. Here is the an example of perfect normal distribution:

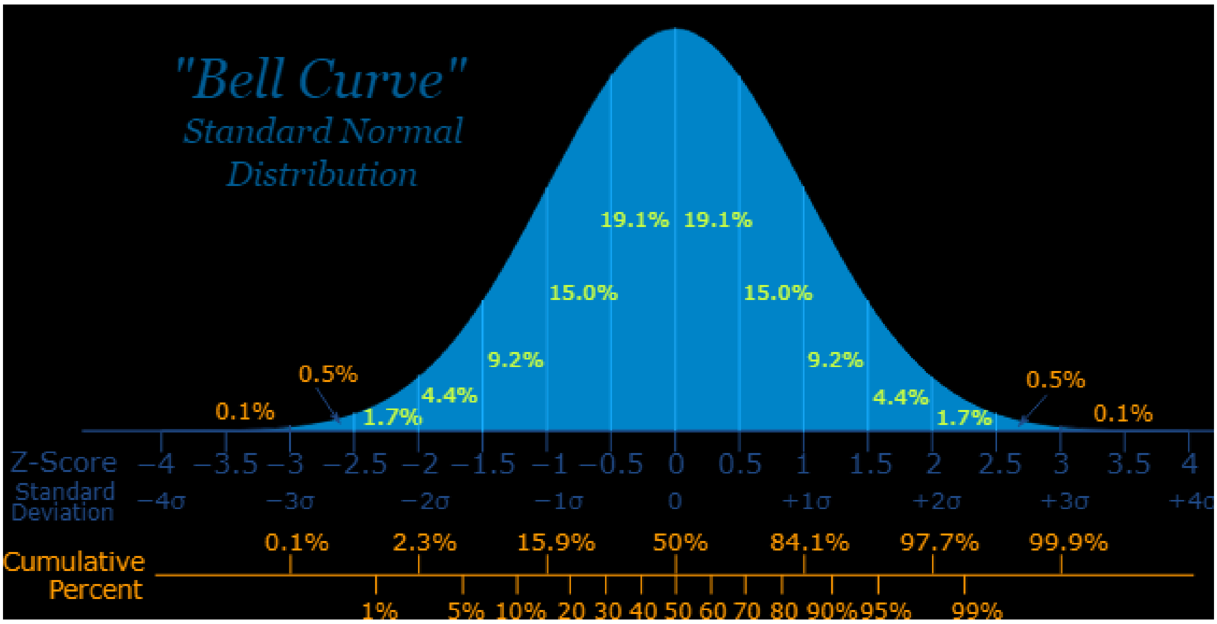
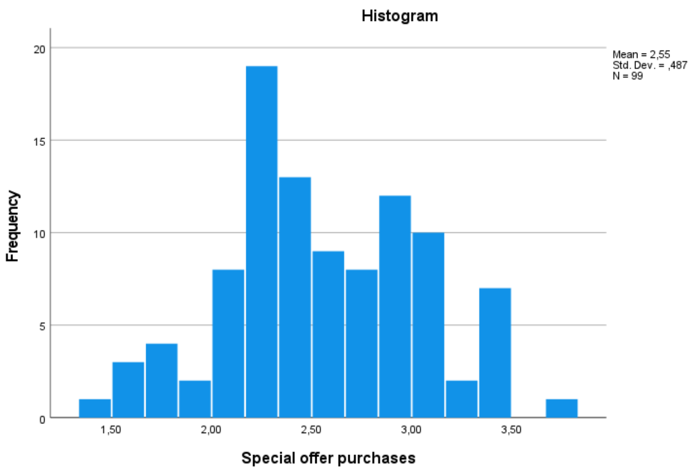
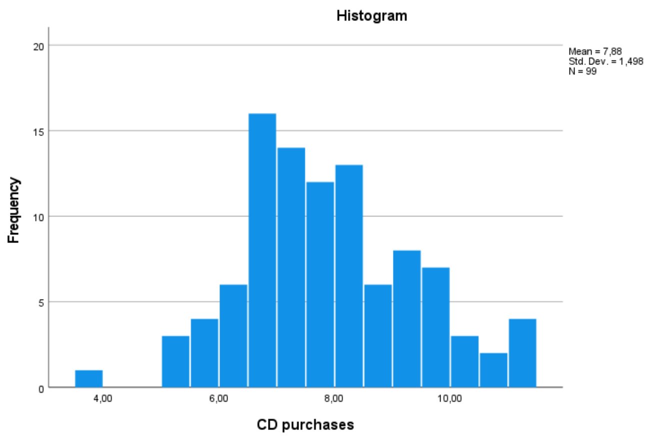


Figure 3. ***Standard Normal Distribution***

Source: https://www.mathsisfun.com/data/standard-normal-distribution.html (Accessed: 01.06.2023)





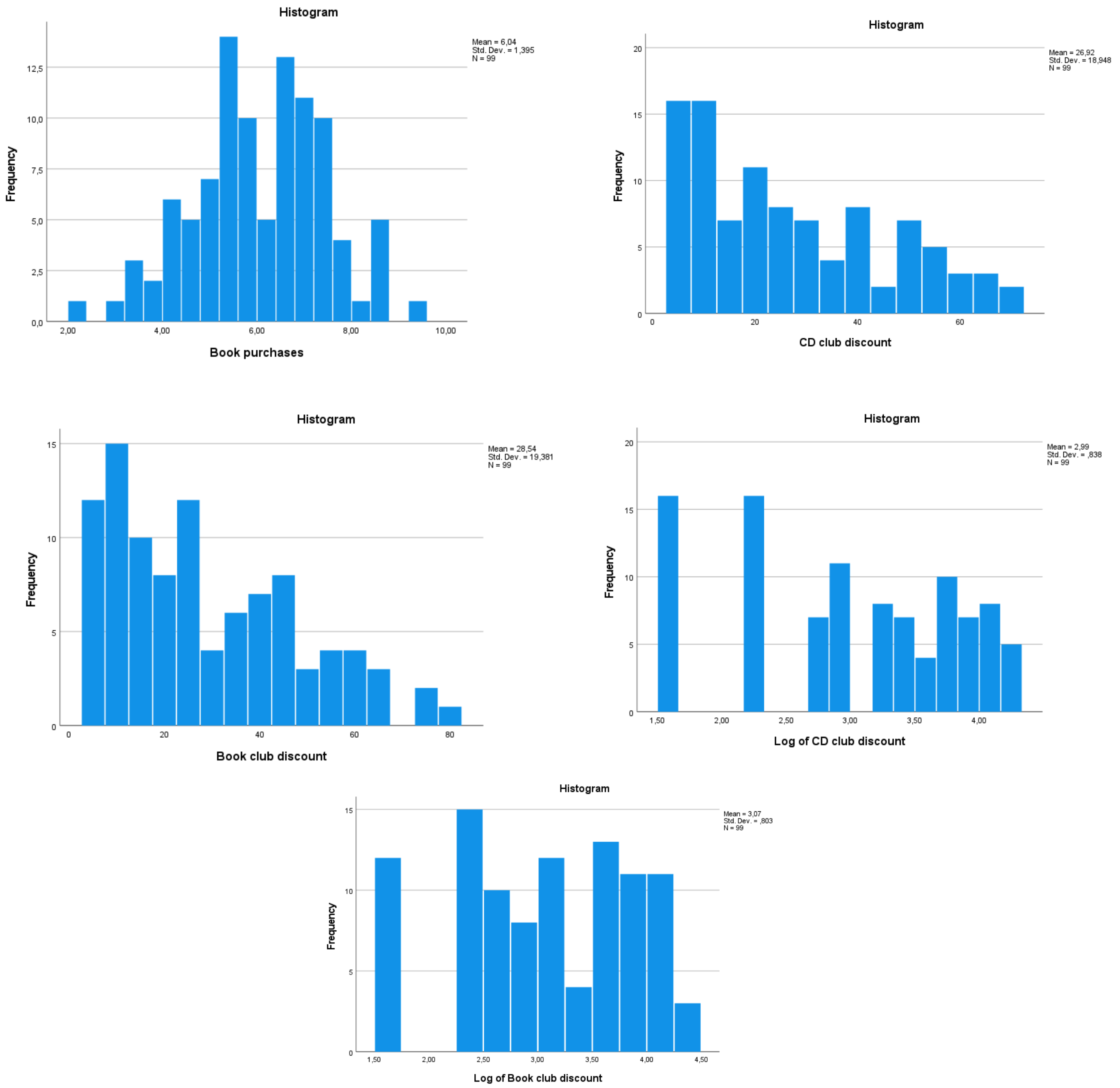


Figure 4. ***Variable Histograms***

When the histograms are examined, it can be seen that histograms of the first 3 variables which are found to be normally distributed according to normality test results, are more similar with the perfect example of normal distribution. Histograms of the rest of the variables start with a high frequency which decreases gradually and/or by fluctuating.

1.2 ANOVA (Analysis of Variance)

ANOVA, which stands for "Analysis of Variance," is a statistical method used to determine if there is a significant difference between the means of three or more independent groups. There are two common types of ANOVA: one-way ANOVA and two-way ANOVA.

One-way ANOVA is used when comparing two groups to assess whether there is a difference between them. It examines two means from unrelated groups using the F-distribution. The null hypothesis assumes that the means are equal, and a significant result suggests that the means are unequal. However, a one-way ANOVA does not identify which specific groups differ from each other.

On the other hand, two-way ANOVA is used to assess how two factors affect a response variable and determine if there is an interaction between the two factors on the response variable. It extends the concept of one-way ANOVA. In a two-way ANOVA, you have one measurement variable (quantitative) and two nominal variables.

A two-way ANOVA provides results for both main effects (considering the effects of each factor separately) and interaction effects (considering the simultaneous effects of all factors). Interaction effects are easier to test when there is more than one observation in each cell. The null hypotheses for a two-way ANOVA include testing the equality of means for each factor and the absence of an interaction effect.

It is important to consider certain assumptions when conducting a two-way ANOVA. These include the assumption of normal distribution in the population, independence of samples, equality of population variances (homoscedasticity), and equal sample sizes in the groups (Leech etal., 2013: 129).

Example 1: Examining the Effectiveness of Three Different Fertilizers

Suppose you are a researcher studying the effect of three different fertilizers (Fertilizer A, B, and C) on the growth of tomato plants. You randomly assign 30 tomato plants to three groups: Group 1 receives Fertilizer A, Group 2 receives Fertilizer B, and Group 3 receives Fertilizer C. After a month, you measure the height of each tomato plant.

To analyze the data using ANOVA, you would calculate the mean height of the tomato plants in each group (Group 1 mean height, Group 2 mean height, and Group 3 mean height). ANOVA allows you to determine whether there is a statistically significant difference in the mean heights of the tomato plants among the three fertilizer groups. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that there is a significant difference in the mean heights, suggesting that the type of fertilizer used has an effect on plant growth.

Example 2: Evaluating the Impact of Different Teaching Methods on Test Scores

Let's say you are interested in comparing the effectiveness of three teaching methods (Method A, B, and C) on student test scores in a math class. You randomly assign 100 students to three groups: Group 1 receives teaching Method A, Group 2 receives Method B, and Group 3 receives Method C. After completing a unit, you administer the same test to all students and record their scores.

To analyze the data using ANOVA, you would calculate the mean test scores for each group (Group 1 mean score, Group 2 mean score, and Group 3 mean score). ANOVA helps determine whether there is a significant difference in the mean test scores among the three teaching methods. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that there is a significant difference in the mean scores, indicating that the teaching method employed has an impact on student performance.

In both examples, ANOVA allows you to compare multiple groups and assess whether there are significant differences among them. It helps you understand the effect of different variables on an outcome of interest by analyzing the variability within and between groups.

1.3 ANCOVA (Analysis of Covariance)

ANCOVA (Analysis of Covariance) is a statistical method used to determine if there is a significant difference between the means of three or more independent groups, similar to ANOVA. However, ANCOVA incorporates one or more covariates, which help in understanding how a factor influences a response variable while accounting for the covariate(s).

ANCOVA is commonly employed when there are baseline group differences, as well as in pretest/posttest analyses where regression to the mean affects the posttest measurement. It is also utilized in non-experimental research, such as surveys, and in quasi-experimental designs where random assignment of study participants is not possible. However, the latter application of ANCOVA is not universally recommended.

Similar to regression analysis, ANCOVA allows examination of how an independent variable acts on a dependent variable. It removes the effects of covariates, which are variables not of primary interest in the study. For instance, if the aim is to investigate how different levels of teaching skills affect student performance in math, it may not be feasible to randomly assign students to classrooms. In this case, systematic differences between students in different classes, such as varying initial math skill levels between gifted and mainstream students, need to be taken into account.

As an extension of ANOVA, ANCOVA can be used in two ways:

- *To control for covariates that are not the main focus of the study, typically continuous or variables on a specific scale.*
- *To study combinations of categorical and continuous variables or variables on a scale as predictors, where the covariate of interest is a variable of interest rather than a control variable.*

The assumptions for ANCOVA are essentially the same as those for ANOVA. Before conducting the test, it is necessary to ensure the following (Leech etal, 2013: 141)

- *Independent variables (minimum of two) should be categorical variables.*
- *The dependent variable and covariate should be continuous variables measured on an interval or ratio scale.*
- *Observations should be independent, with individuals not assigned to more than one group.*

Software tools can typically verify the following assumptions:

- *Normality: The dependent variable should exhibit approximate normality for each category of independent variables.*
- *Homogeneity of variance: The data should demonstrate similar variance across groups.*

- *Linear relationship: The covariate and dependent variable (at each level of the independent variable) should exhibit a linear relationship.*
- *Homoscedasticity: The data should display consistent spread of the dependent variable for each value of the independent variable.*
- *Absence of interaction: The covariate and independent variable should not interact, indicating homogeneity of regression slopes.*

Example: Consider the previous example of splitting a class of 90 students into three groups, each using a different studying technique for one month to prepare for an exam. To account for the students' current grade in the class, their grade is used as a covariate in an ANCOVA. The aim is to determine if there is a significant difference in mean exam scores between the three groups. By conducting the ANCOVA, it becomes possible to examine whether the studying technique has an impact on exam scores after removing the influence of the covariate. Thus, if a statistically significant difference in exam scores is found among the three studying techniques, it can be concluded that this difference exists even after considering the students' current grade in the class.

Example 1: Assessing the Effect of a Teaching Intervention on Test Scores while Controlling for a Covariate

Suppose you are conducting a study to evaluate the effectiveness of a teaching intervention designed to improve student test scores in a mathematics class. However, you suspect that the students' prior mathematical ability, as measured by a pre-test score, may influence their post-test scores. To account for this potential confounding factor, you collect data on both the pre-test score and the post-test score for each student.

To analyze the data using ANCOVA, you would consider the post-test score as the dependent variable, the teaching intervention as the independent variable, and the pre-test score as the covariate. ANCOVA allows you to determine whether there is a significant difference in the post-test scores among the different teaching intervention groups, while adjusting for the influence of the pre-test scores. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that the teaching intervention has a significant effect on the post-test scores, even after accounting for the influence of the pre-test scores.

Example 2: Examining the Impact of a Drug Treatment on Blood Pressure while Controlling for a Covariate

Let's say you are interested in studying the effect of a new drug treatment on blood pressure in patients with a specific medical condition. However, you suspect that age may be a confounding factor, as it is known to be associated with blood pressure. Therefore, you collect data on both the patients' blood pressure measurements and their age.

To analyze the data using ANCOVA, you would consider the blood pressure measurement as the dependent variable, the drug treatment as the independent variable, and age as the covariate. ANCOVA allows you to determine whether there is a significant difference in blood pressure among the different drug treatment groups, while adjusting for the influence of age. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that the drug treatment has a significant effect on blood pressure, even after accounting for the influence of age.

In both examples, ANCOVA enables you to assess the relationship between an independent variable and a dependent variable, while controlling for the influence of a covariate. It helps you understand the effect of the independent variable on the dependent variable, while taking into account the potential confounding effect of the covariate.

1.4 MANOVA (Multivariate Analysis of Variance)

MANOVA (Multivariate Analysis of Variance) is a statistical technique that is similar to ANOVA but involves two or more response variables (Huberty & Olejnik, 2006:7) Like ANOVA, MANOVA can be conducted as a one-way or two-way analysis.

In a MANOVA, the purpose is to determine if the independent variable(s) affect the response variable(s), similar to other tests and experiments. For example, if the goal is to investigate if different textbooks have an impact on students' scores in math and science, where improvements in math and science are the two dependent variables, a MANOVA would be appropriate.

While ANOVA provides a single f-value for analysis of one dependent variable, MANOVA produces a multivariate F value to assess multiple dependent variables. MANOVA tests the combined effect of the dependent variables by creating new artificial dependent variables that maximize the differences between groups. These new dependent variables are linear combinations of the original measured dependent variables.

Assumptions for the MANOVA analysis are having independent observations, multivariate normality homogeneity of variance and/or covariance matrices. If groups have nearly equal size, MANOVA is robost for violations of normality and homogeneity (Leech etal, 2013: 162).

Example 1: Assessing the Effect of Exercise on Multiple Health Parameters

Suppose you are conducting a study to investigate the effect of exercise on multiple health parameters in a group of individuals. You randomly assign 50 participants to two groups: Group 1 undergoes an exercise program for 12 weeks, while Group 2 serves as a control and does not participate in any exercise program. At the end of the 12-week period, you measure several health parameters, including blood pressure, cholesterol levels, and body fat percentage.

To analyze the data using MANOVA, you would organize the health parameters into a multivariate outcome variable. In this case, the outcome variable would include blood pressure, cholesterol levels, and body fat percentage. MANOVA allows you to determine whether there is a significant difference in the combined multivariate outcome between the exercise group and the control group. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that there is a significant difference in the overall health parameters, suggesting that exercise has an effect on multiple health variables simultaneously.

Example 2: Evaluating the Impact of Different Advertising Channels on Consumer Behavior

Let's say you work for a marketing agency and you want to assess the impact of different advertising channels (TV, radio, and online) on consumer behavior. You randomly select 100 participants and expose them to advertisements through one of the three channels. After exposure, you measure multiple consumer behavior variables, such as brand preference, purchase intention, and recall.

To analyze the data using MANOVA, you would create a multivariate outcome variable that includes the consumer behavior variables (brand preference, purchase intention, and recall). MANOVA allows you to determine whether there is a significant difference in the combined multivariate outcome across the different advertising channels. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that there is a significant difference in consumer behavior among the advertising channels, indicating that the choice of advertising medium has an impact on multiple consumer variables simultaneously.

In both examples, MANOVA allows you to analyze the relationship between multiple dependent variables and one or more independent variables. It helps you understand whether there are significant differences among groups when considering multiple outcome variables simultaneously, providing a more comprehensive understanding of the relationship between the variables of interest.

1.5 MANCOVA (Multivariate Analysis of Covariance)

MANCOVA (Multivariate Analysis of Covariance) is a statistical technique that is similar to MANOVA but includes one or more covariates. It is the multivariate counterpart of ANCOVA. MANCOVA is used to determine if there are statistically significant mean differences among groups while taking into account the effects of covariates (Dattalo, 2013: 63).

By removing the effects of covariates from the model, MANCOVA allows for the examination of the true effects of independent variables on dependent variables without unwanted interference. However, it is important to note that MANCOVA typically requires larger sample sizes compared to other tests. Therefore, the decision to use MANCOVA should consider the trade-off between the additional time and expense required and the potential benefits. In many cases, a simpler MANOVA without considering covariates may be more powerful.

Similar to MANOVA, MANCOVA can be conducted as a one-way or two-way analysis. Covariance refers to the measure of how two random variables vary together. A covariate is a variable that affects how independent variables act upon dependent variables. It is typically a variable that needs to be controlled for in the analysis, such as confounding variables.

The assumptions for MANCOVA are similar to those for MANOVA, with the addition of a couple of assumptions specific to covariance (Dattalo, 2013: 64). These assumptions include the continuous and ratio/ordinal nature of covariates and dependent variables, equality of covariance matrices (to reduce Type I error), categorical independent variables, independence of variables, random sampling, normality of dependent variables for each group, absence of multicollinearity, and homogeneity of variance between groups.

Prior to their inclusion in MANCOVA, it is important for the chosen covariates to be correlated with the dependent variables, which can be assessed using correlation analysis. Additionally, the dependent variables should ideally not be significantly correlated with each other. Statistical software is often used to assess these assumptions before conducting MANCOVA.

Example 1: Assessing the Effect of a Drug Treatment on Multiple Outcome Variables while Controlling for Covariates

Suppose you are conducting a clinical trial to evaluate the effectiveness of a new drug treatment on multiple outcome variables, such as pain relief, quality of life, and mobility, in patients with a specific medical condition. However, you suspect that age and baseline symptom severity may influence the outcome variables. To account for these potential confounding factors, you collect data on the participants' age and baseline symptom severity.

To analyze the data using MANCOVA, you would consider the outcome variables (pain relief, quality of life, and mobility) as the multivariate dependent variable and the drug treatment as the independent variable. Additionally, you would include the covariates (age and baseline symptom severity) in the analysis to control for their potential effects. MANCOVA allows you to determine whether there is a significant difference in the combined multivariate outcome across the different drug treatment groups while accounting for the covariates. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that the drug treatment has a significant effect on the outcome variables, after controlling for the influence of age and baseline symptom severity.

Example 2: Examining the Influence of Socioeconomic Status on Multiple Academic Achievement Variables with Adjustment for Covariates

Let's say you are interested in investigating the relationship between socioeconomic status (SES) and multiple academic achievement variables, such as test scores in math, reading, and science, in a group of students. However, you suspect that factors like parental education level and the presence of learning disabilities might also impact academic achievement. Therefore, you collect data on SES, parental education level, and learning disability status.

To analyze the data using MANCOVA, you would consider the academic achievement variables (math scores, reading scores, and science scores) as the multivariate dependent variable, and SES as the independent variable. Additionally, you would include the covariates (parental education level and learning disability status) in the analysis to control for their potential effects. MANCOVA allows you to determine whether there is a significant relationship between SES and the combined multivariate academic achievement variables, while adjusting for the influence of the covariates. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that SES has a significant impact on academic achievement, even after accounting for the effects of parental education level and learning disability status. In both examples, MANCOVA enables you to assess the relationship between multiple dependent variables and an independent variable, while controlling for the influence of covariates. It helps you understand the joint effect of the independent variable on the multivariate outcome, while taking into account the effects of other variables that might confound the relationship.

1.6 T-Test

The t-test is a statistical test used to determine if there is a significant difference between the means of two groups or samples (Kim, 2015: 540). It is based on the t-distribution, which is similar to the normal distribution but has slightly heavier tails.

The theoretical background of the t-test is rooted in the concept of sampling distributions and the Central Limit Theorem (Livingston 2004: 59-60). The Central Limit Theorem states that when independent random samples are drawn from a population with a finite mean and standard deviation, the distribution of the sample means approaches a normal distribution as the sample size increases, regardless of the shape of the population distribution. This is the case even if the population itself does not follow a normal distribution.

The t-test uses the t-distribution to compare the means of two groups by estimating the standard error of the difference between the sample means. The formula for the t-statistic is:

$$t = (x1 - x2) / SE \tag{1}$$

where x1 and x2 are the sample means of the two groups, and SE is the standard error of the difference between the means. The standard error takes into account the variability within each group and the sample sizes. The t-statistic measures the difference between the sample means in terms of standard error units.

The t-distribution is used because, in practice, we often do not know the population standard deviation and need to estimate it from the sample data. The t-distribution accounts for the added uncertainty in the estimation of the standard deviation by incorporating the degrees of freedom, which is based on the sample sizes of the two groups.

The degrees of freedom determine the shape of the t-distribution and affect the critical values used to determine statistical significance. For independent samples t-tests, the degrees of freedom are calculated as the sum of the sample sizes minus two.

By comparing the calculated t-value to the critical values from the t-distribution, we can determine if the observed difference in means is statistically significant. If the calculated t-value exceeds the critical value at a chosen significance level (e.g., 0.05), we reject the null hypothesis and conclude that there is a significant difference between the means of the two groups (Liu & Wang, 2021: 266).

Overall, the t-test provides a statistical framework for comparing means and evaluating the significance of differences between two groups, taking into account the sample sizes and the inherent variability within the data.

Example 1: Comparing the Mean Heights of Two Groups

Suppose you are interested in comparing the mean heights of two groups: Group A and Group B. You collect height measurements from a sample of individuals from each group. Group A consists of 30 participants, and Group B consists of 35 participants.

To analyze the data using a t-test, you would calculate the mean height of each group (Group A mean height and Group B mean height) and also calculate the standard deviation for each group. Then, you would perform a two-sample t-test to determine whether there is a significant difference in the mean heights between the two groups. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that there is a significant difference in the mean heights, indicating that the two groups differ in average height.

Example 2: Evaluating the Effectiveness of a New Treatment

Let's say you are evaluating the effectiveness of a new treatment for a specific medical condition. You randomly assign 50 patients to two groups: Group A receives the new treatment, and Group B receives a placebo. After a specified treatment period, you measure a relevant outcome variable, such as pain intensity.

To analyze the data using a t-test, you would calculate the mean pain intensity for each group (Group A mean pain intensity and Group B mean pain intensity). Then, you would perform an independent samples t-test to determine whether there is a significant difference in the mean pain intensity between the two groups. If the p-value is below a predetermined significance level (e.g., 0.05), you can conclude that there is a significant difference in pain intensity, suggesting that the new treatment has an effect on reducing pain compared to the placebo.

In both examples, the t-test allows you to compare the means of two groups and determine whether there is a significant difference between them. It is commonly used when you have two independent groups and want to evaluate whether there is evidence to support a meaningful difference between their means.

1.7 Regression Analysis

Regression analysis is a statistical method used to examine the relationship between a dependent variable and one or more independent variables (Uyanık & Güler, 2013: 234). It is based on the concept of fitting a regression model to the data and estimating the coefficients that represent the relationship between the variables.

The theoretical background of regression analysis is grounded in the concept of a linear relationship between variables. Linear regression assumes that there is a linear, additive relationship between the independent variables and the dependent variable. This means that the effect of the independent variables on the dependent variable can be represented by a straight line in a scatterplot.

The goal of regression analysis is to estimate the parameters (coefficients) of the linear equation that best fits the data. The most common form of linear regression is called simple linear regression, which involves one dependent variable and one independent variable. The equation for simple linear regression is:

$$Y = \beta 0 + \beta 1X + \varepsilon \qquad (2)$$

where Y is the dependent variable, X is the independent variable, β0 is the y-intercept (the value of Y when X is 0), β1 is the slope (the change in Y for a one-unit change in X), and ε is the error term (representing the variability or randomness not explained by the model).

The coefficients β0 and β1 are estimated using a method called Ordinary Least Squares (OLS), which minimizes the sum of the squared differences between the observed values of the dependent variable and the predicted values based on the regression equation (Rawlings etal., 1998: 2-4).

Multiple linear regression extends the concept of simple linear regression to include more than one independent variable. The equation becomes:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta nXn + \varepsilon \quad (3)$$

where X1, X2, ..., Xn are the independent variables, and β1, β2, ..., βn are the corresponding coefficients.

The premise is that the data points of the dependent variable, denoted as Y, are considered as random samples from populations of random variables, where the average of each population is represented by Y. To incorporate the difference between an observation Y and its population average Y, a random error is introduced into the statistical model (Rawlings etal., 1998: 2).

Regression analysis aims to estimate the coefficients (β0, β1, β2, ..., βn) that provide the best fit to the data and allow for predicting the dependent variable based on the independent variables. These coefficients indicate the direction and magnitude of the relationship between the variables. A positive coefficient suggests a positive relationship (as the independent variable increases, the dependent variable tends to increase), while a negative coefficient suggests a negative relationship.

Additionally, regression analysis allows for hypothesis testing and evaluating the statistical significance of the coefficients. Hypothesis tests such as t-tests or F-tests are used to assess whether the coefficients are significantly different from zero, indicating a meaningful relationship between the variables.

Overall, regression analysis provides a statistical framework for understanding and quantifying the relationship between variables, estimating coefficients, and making predictions based on the regression equation. It enables the identification of key factors that influence the dependent variable and aids in uncovering patterns and insights within the data.

Example 1: Predicting House Prices based on Features

Suppose you are a real estate agent and want to predict house prices based on various features such as the size of the house, the number of bedrooms, the location, and the age of the property. You collect data on recently sold houses, including information about these features and their corresponding sale prices.

To analyze the data using regression analysis, you would use a multiple linear regression model. You would treat the house price as the dependent variable and the house features (size, number of bedrooms, location, age) as independent variables. Regression analysis allows you to estimate the relationship between the independent variables and the dependent variable, providing insights into how each feature contributes to the variation in house prices. You can interpret the regression coefficients to understand the direction and magnitude of the effect of each independent variable on house prices.

Example 2: Examining the Relationship between Study Time and Exam Scores

Let's say you want to investigate the relationship between the amount of time students spend studying and their exam scores. You collect data from a group of students, recording the number of hours they spend studying and their corresponding exam scores.

To analyze the data using regression analysis, you would use a simple linear regression model. You would treat the exam score as the dependent variable and the study time as the independent variable. Regression analysis allows you to estimate the slope and intercept of the regression line, which represents the average change in the exam score associated with each additional hour of study time. By examining the coefficient of determination (R-squared value), you can determine the proportion of the variability in the exam scores that can be explained by the study time variable.

In both examples, regression analysis allows you to understand the relationship between a dependent variable and one or more independent variables. It helps you estimate the coefficients and assess the significance of the relationships, enabling predictions, and understanding the impact of the independent variables on the dependent variable.

1.8 Correlation Analysis

Correlation analysis is a statistical method used to measure and assess the strength and direction of the relationship between two variables. It is based on the concept of covariance and the idea of quantifying the degree of linear association between variables.

The theoretical background of correlation analysis is rooted in the concept of correlation coefficient, which provides a numerical measure of the relationship between variables. The most commonly used correlation coefficient is the Pearson correlation coefficient (r), which measures the linear relationship between two continuous variables (Gogtay & Thatte, 2017: 80).

The Pearson correlation coefficient ranges between -1 and 1. A correlation coefficient of +1 indicates a perfect positive linear relationship, meaning that as one variable increases, the other variable increases proportionally. A correlation coefficient of -1 indicates a perfect negative linear relationship, meaning that as one variable increases, the other variable decreases proportionally. A correlation coefficient of 0 suggests no linear relationship between the variables.

The formula for calculating the Pearson correlation coefficient is:

$$r = (\Sigma[(X - \bar{X})(Y - \bar{Y})]) / [\mathrm{sqrt}(\Sigma(X - \bar{X})^2) * \mathrm{sqrt}(\Sigma(Y - \bar{Y})^2)] \tag{4}$$

where X and Y are the values of the two variables, $\bar{X}$ and $\bar{Y}$ are their respective means, and Σ denotes the sum across the data points.

Correlation analysis allows us to determine the direction and strength of the relationship between variables. The magnitude of the correlation coefficient indicates the strength, with values closer to 1 or -1 representing a stronger linear relationship. The sign of the coefficient (+ or -) indicates the direction of the relationship.

It's important to note that correlation does not imply causation. A high correlation between two variables does not necessarily mean that one variable causes the other to change. Correlation analysis only quantifies the association between variables.

In addition to the Pearson correlation coefficient, there are other correlation coefficients that are used for specific types of data, such as Spearman's rank correlation coefficient for ranked or ordinal data, and Kendall's tau for ranked data with ties.

Overall, correlation analysis provides a quantitative measure of the strength and direction of the linear relationship between variables. It helps in understanding the degree to which changes in one variable are associated with changes in another, but it does not provide information about causality or the presence of other types of relationships between variables.

Example 1: Examining the Relationship between Age and Blood Pressure

Suppose you are interested in understanding the relationship between age and blood pressure. You collect data from a sample of individuals, recording their age (in years) and their corresponding blood pressure measurements (e.g., systolic or diastolic pressure).

To analyze the data using correlation analysis, you would calculate the correlation coefficient between age and blood pressure. The correlation coefficient measures the strength and direction of the linear relationship between two variables. A positive correlation coefficient indicates a positive linear relationship (both variables increase or decrease together), while a negative correlation coefficient indicates a negative linear relationship (as one variable increases, the other decreases). The magnitude of the correlation coefficient represents the strength of the relationship, with values closer to 1 or -1 indicating a stronger relationship.

Example 2: Assessing the Relationship between Advertising Spending and Sales Revenue

Let's say you want to examine the relationship between advertising spending and sales revenue for a company. You collect data on the amount of money spent on advertising (e.g., in dollars) and the corresponding sales revenue (e.g., in dollars) generated during specific periods.

To analyze the data using correlation analysis, you would calculate the correlation coefficient between advertising spending and sales revenue. The correlation coefficient provides insights into the strength and direction of the relationship between the two variables. A positive correlation coefficient suggests that higher advertising spending is associated with higher sales revenue, while a negative correlation coefficient suggests an inverse relationship. By examining the magnitude of the correlation coefficient, you can assess the strength of the relationship, with values closer to 1 or -1 indicating a stronger association.

In both examples, correlation analysis allows you to quantify the relationship between two variables. It helps you understand the direction and strength of the association, providing insights into how changes in one variable are related to changes in another. However, it's important to note that correlation does not imply causation, and additional analysis and consideration of other factors are often necessary to establish causal relationships.

2. PRACTICES IN SPSS

a. Practices For ANOVA

b. One-Way Anova (Practice)

Click on the sample files button.

Select bankloan.sav and open it.

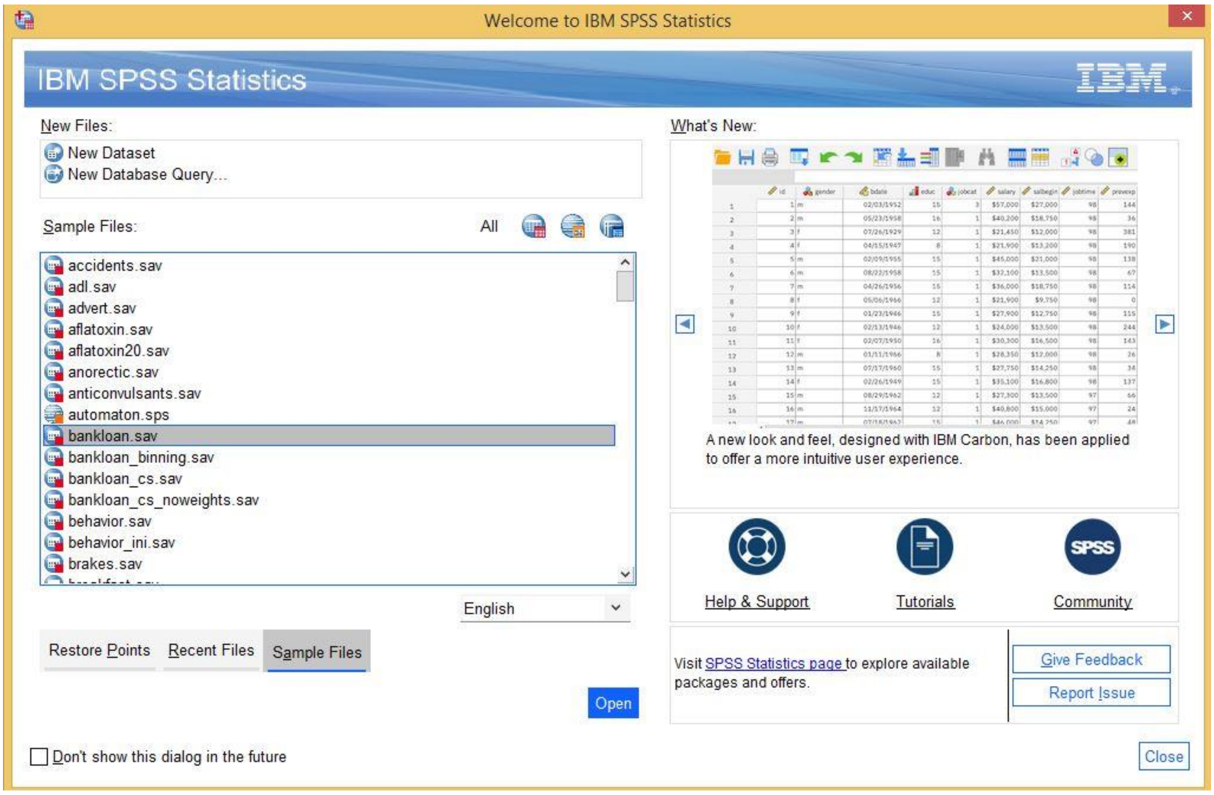


Figure 5. ***Selection of Variables***

From the menu, click on analyze, select compare means and click on One-Way Anova button.

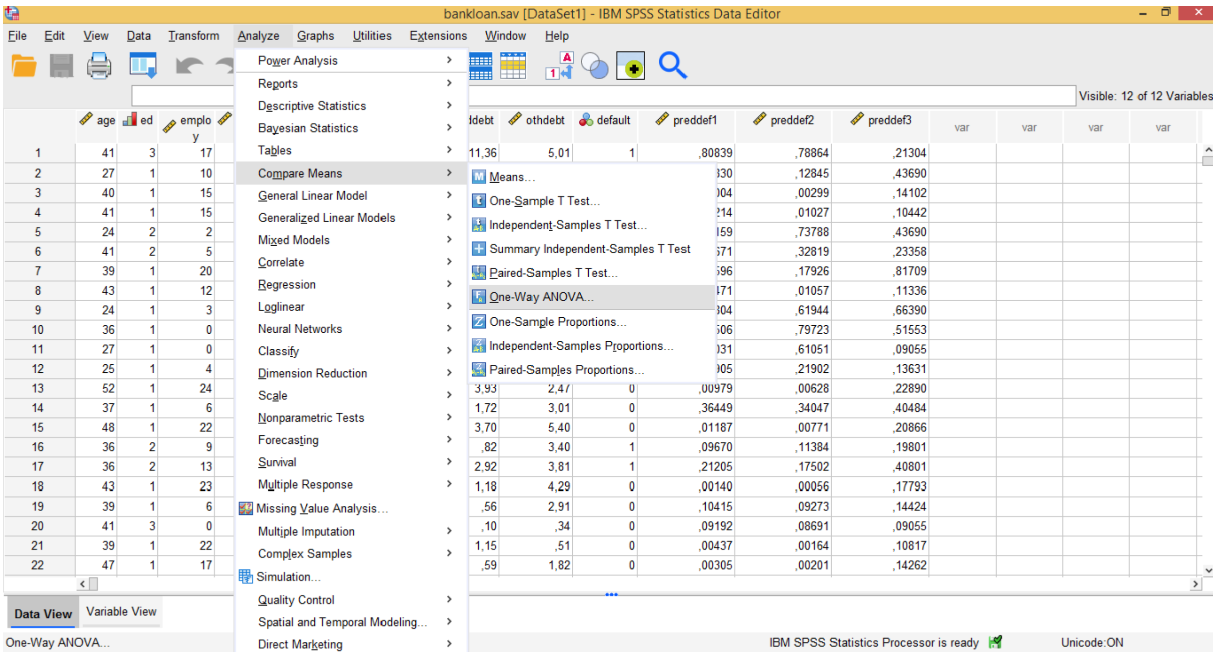


Figure 6. ***Selection of Test (One-Way ANOVA)***

You can select your dependent variable and factor variable from the list. In this example, we select household income as dependent variable while choosing level of education as factor variable.

This means that we will examine the relationship between education and household income.

Click on the household income variable on the left, then click on the arrow to put it as a dependent variable and do the same for factor variable which is level of education.

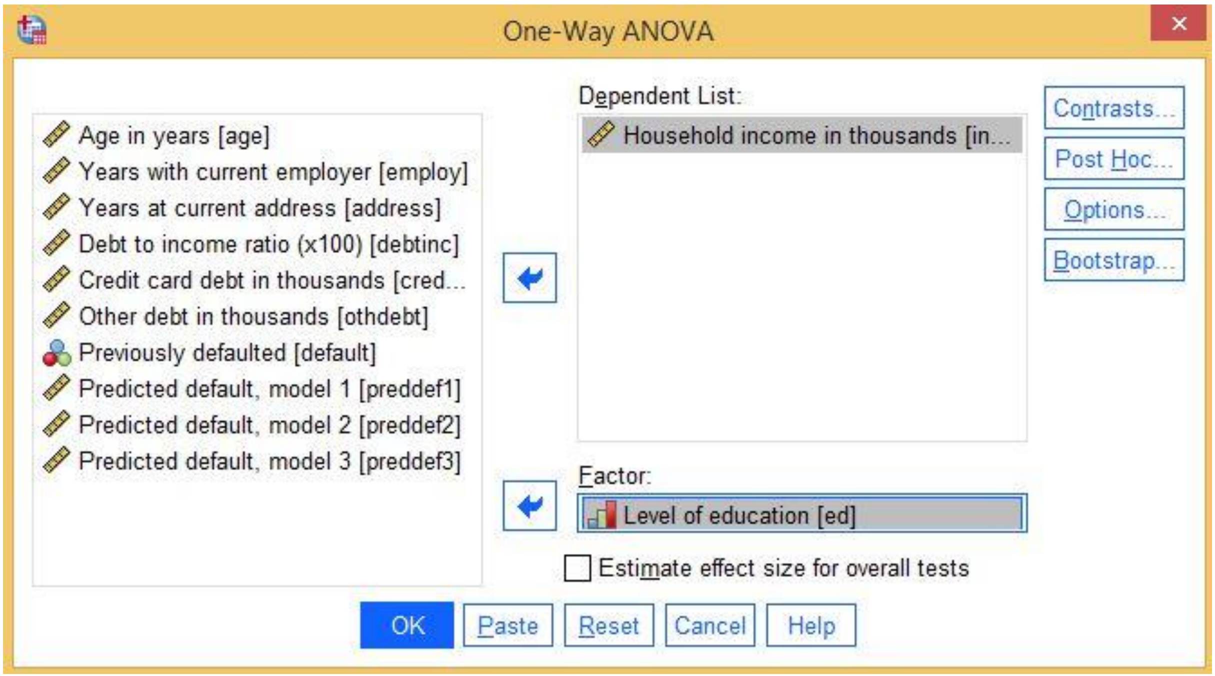


Figure 7. ***Selection of Variables for ANOVA***

Once you do this, click on the options button on the right side of the menu. Then you will see the following menu.

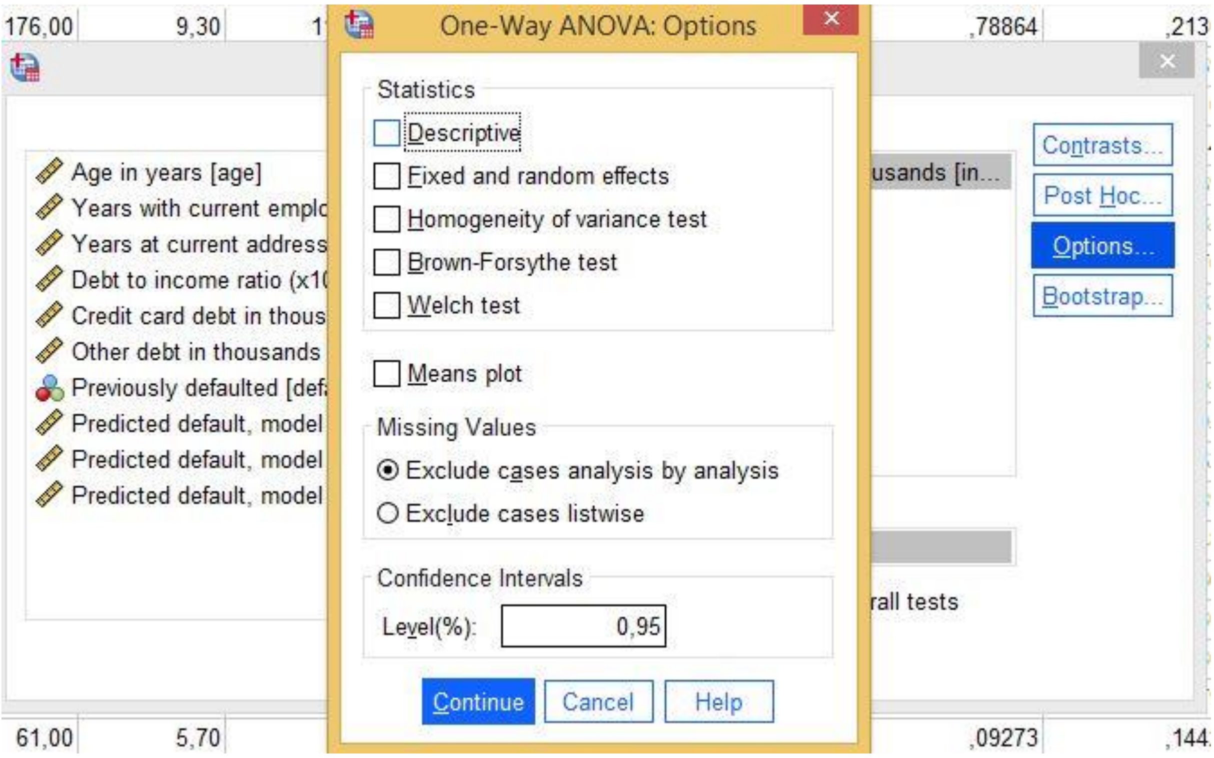


Figure 8. ***Selection of One-Way ANOVA Options***

Select descriptive and homogeneity of variance test. Also, you can choose the confidence intervals. In practice, 0.05 is usually the ideal number. So in this practice, we will change it as 0.05. Once you finished it, click on the continue button. After that click on the OK button to implement the analysis. Once we run the analysis, based on the results regarding the homogeneity of variances, we will carry out a post hoc test.

Table 4. ***Descriptive Statistics***

Descriptives

Household income in thousands

	N	Mean	Std. Deviation	Std. Error	5% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Did not complete high school	460	40,8370	32,19679	1,50118	40,7428	40,9311	13,00	324,00
High school degree	235	46,9532	32,34843	2,11018	46,8207	47,0857	14,00	249,00
Some college	101	63,0198	45,50999	4,52841	62,7351	63,3045	15,00	266,00
College degree	49	59,3265	70,39424	10,05632	58,6926	59,9604	18,00	446,00
Post-undergraduate degree	5	116,6000	71,81086	32,11479	114,4570	118,7430	20,00	190,00
Total	850	46,6753	38,54305	1,32202	46,5924	46,7582	13,00	446,00

Table 5. ***Tests of Homogeneity of Variances***

		Levene Statistic	df1	df2	Sig.
Household income in thousands	Based on Mean	10,239	4	845	,000
	Based on Median	6,044	4	845	,000
	Based on Median and with adjusted df	6,044	4	484,302	,000
	Based on trimmed mean	7,933	4	845	,000

Table 6. ***ANOVA Results***

Household income in thousands

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	74969,188	4	18742,297	13,350	,000
Within Groups	1186277,193	845	1403,878		
Total	1261246,381	849			

From the results as it can be seen above, Sig. (p-value) value of Tests of Homogeneity (Levene Test) is below 0.0001. This means that we have to reject the null hypothesis (H0) which is the variances are homogeneously distributed. In other words, there is a heterogeneity issue with the variances. This will violate one of the assumptions of ANOVA test. Therefore we will select compatible post hoc test.

ANOVA shows us that there is a significant relationship between variables which allows us to further delve into the analysis by using post hoc test.

Results of the post hoc tests show us that there is a significant difference between household incomes of people who did not complete high school and people with some college degree. There is also another significant difference between high school graduates and some college degree in terms of household income.

Table 7. ***One-Way ANOVA Comparison Results***

Multiple Comparisons

Dependent Variable: Household income in thousands

	(I) Level of education	(J) Level of education	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tamhane	Did not complete high school	High school degree	-6,11623	2,58967	,171	-13,4006	1,1681
		Some college	-22,18285*	4,77075	<,001	-35,7834	-8,5823
		College degree	-18,48957	10,16775	,541	-48,2623	11,2832
		Post-undergraduate degree	-75,76304	32,14986	,554	-253,9509	102,4248
	High school degree	Did not complete high school	6,11623	2,58967	,171	-1,1681	13,4006
		Some college	-16,06661*	4,99594	,016	-30,2692	-1,8640
		College degree	-12,37334	10,27533	,930	-42,4049	17,6582
		Post-undergraduate degree	-69,64681	32,18405	,635	-247,4114	108,1178
	Some college	Did not complete high school	22,18285*	4,77075	<,001	8,5823	35,7834
		High school degree	16,06661*	4,99594	,016	1,8640	30,2692
		College degree	3,69327	11,02888	1,000	-28,2162	35,6028
		Post-undergraduate degree	-53,58020	32,43249	,847	-228,4107	121,2503
	College degree	Did not complete high school	18,48957	10,16775	,541	-11,2832	48,2623
		High school degree	12,37334	10,27533	,930	-17,6582	42,4049
		Some college	-3,69327	11,02888	1,000	-35,6028	28,2162
		Post-undergraduate degree	-57,27347	33,65248	,807	-220,7786	106,2317
	Post-undergraduate degree	Did not complete high school	75,76304	32,14986	,554	-102,4248	253,9509
		High school degree	69,64681	32,18405	,635	-108,1178	247,4114
		Some college	53,58020	32,43249	,847	-121,2503	228,4107
		College degree	57,27347	33,65248	,807	-106,2317	220,7786
Dunnett T3	Did not complete high school	High school degree	-6,11623	2,58967	,171	-13,3990	1,1665
		Some college	-22,18285*	4,77075	<,001	-35,7709	-8,5948
		College degree	-18,48957	10,16775	,522	-48,1835	11,2043
		Post-undergraduate degree	-75,76304	32,14986	,371	-230,2471	78,7210
	High school degree	Did not complete high school	6,11623	2,58967	,171	-1,1665	13,3990
		Some college	-16,06661*	4,99594	,016	-30,2583	-1,8749
		College degree	-12,37334	10,27533	,919	-42,3295	17,5828
		Post-undergraduate degree	-69,64681	32,18405	,440	-223,9316	84,6380
	Some college	Did not complete high school	22,18285*	4,77075	<,001	8,5948	35,7709
		High school degree	16,06661*	4,99594	,016	1,8749	30,2583
		College degree	3,69327	11,02888	1,000	-28,1582	35,5447
		Post-undergraduate degree	-53,58020	32,43249	,668	-206,4800	99,3196
	College degree	Did not complete high school	18,48957	10,16775	,522	-11,2043	48,1835
		High school degree	12,37334	10,27533	,919	-17,5828	42,3295
		Some college	-3,69327	11,02888	1,000	-35,5447	28,1582
		Post-undergraduate degree	-57,27347	33,65248	,639	-204,7929	90,2460
	Post-undergraduate degree	Did not complete high school	75,76304	32,14986	,371	-78,7210	230,2471
		High school degree	69,64681	32,18405	,440	-84,6380	223,9316
		Some college	53,58020	32,43249	,668	-99,3196	206,4800
		College degree	57,27347	33,65248	,639	-90,2460	204,7929

*. The mean difference is significant at the 0.05 level.

c. Two-Way Anova (Practice)

For this example we will use another dataset from SPSS samples: customer_dbase.sav

Select customer_dbase.sav from the same menu in One-way ANOVA test example.

Click on Analyze section from the top menu.

This time find General Linear Model section under Analyze. Then click on Univariate… button.

Once you clicked you will see the following menu:

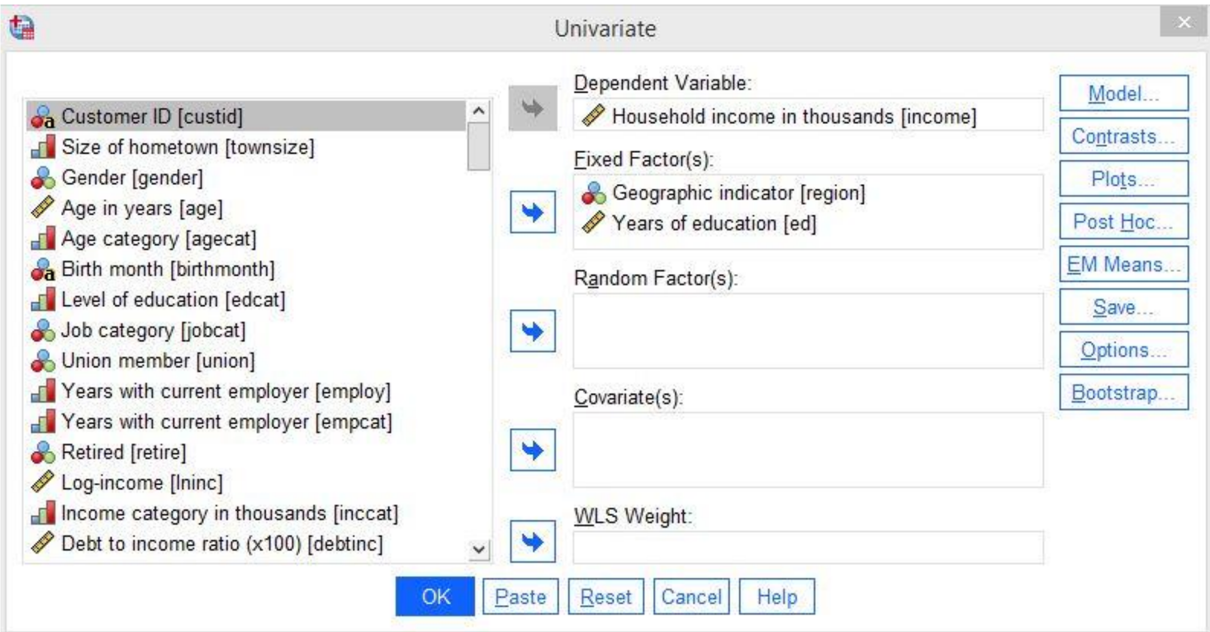


Figure 9. ***Variable Selection***

As you can see from the image above, we selected Household income in thousands (income) as dependent variable and geographic indicator (region) and years of education (ed) as factor variables.

Click on the model button on the right.

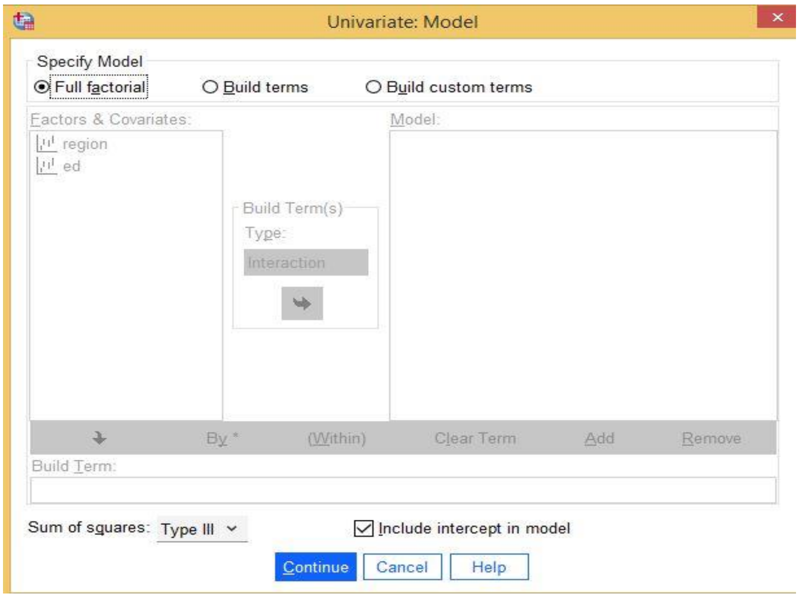


Figure 10. ***Specifying the Model***

You can customize variables for the model. The default option is full factorial. This means that all variables including interaction of region and ed will be in the model. This interaction between variables show the effects of both variables at the same time. This will provide information the effects of years of education and living in a specific zone on the household income.

Click on the Continue button and then on the Post Hoc button.

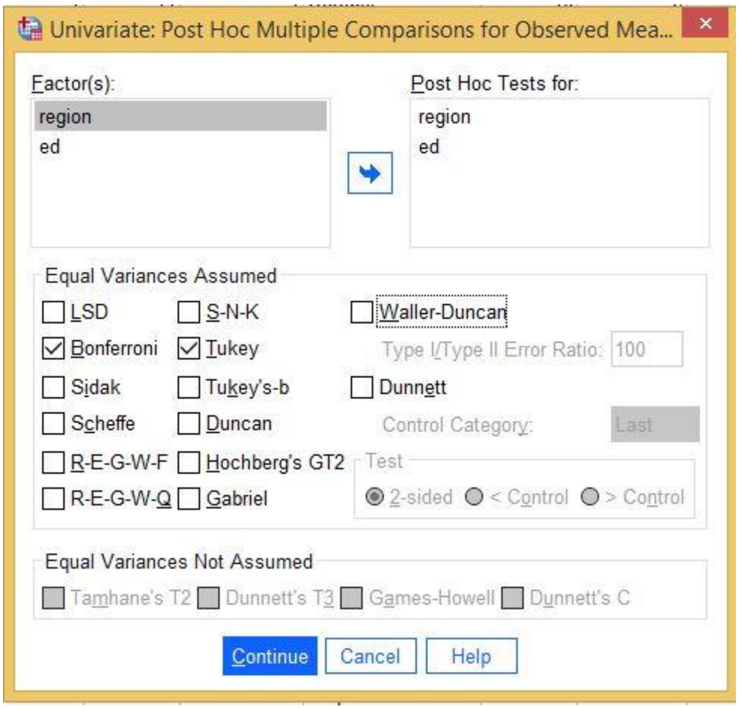


Figure 11. ***Post Hoc Tests***

Select Factors on the left for Post Hoc analysis. Note that if you select a variable as Random Factor, you cannot use that variable for the post-hoc test. Since we selected our variables as Fixed Factor, it is possible for us to select that.

Note that Two Way ANOVA analysis is robust to Unequal variance distribution. This means that you can use test that requires Equal Variance Assumption. As you can see from the image above, it is not possible to select the tests that you can use for the situations for unequal variance distribution.

In this example we will select Bonferroni and Tukey tests.

Click on Continue button and then click on Options button.

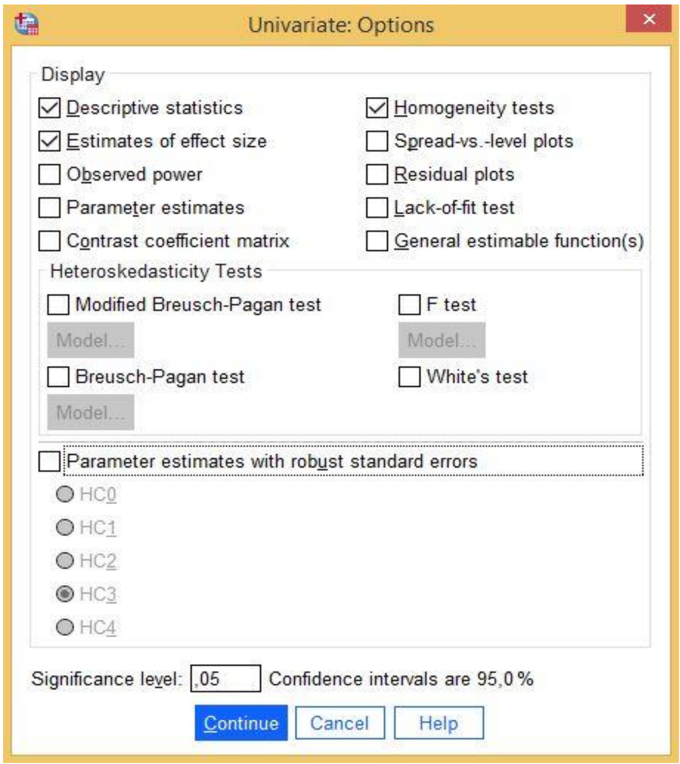


Figure 12. ***Descriptives, Estimates and Test Selection***

For this example, we use Descriptive statistics, Estimates of effect size and Homogeneity tests. So select the relevant sections.

Click on Continue and after that click OK button from the main menu. This will run the selected tests:

Table 8. ***General Descriptives***

Between-Subjects Factors

		Value Label	N
Geographic indicator	1	Zone 1	1019
	2	Zone 2	1005
	3	Zone 3	981
	4	Zone 4	943
	5	Zone 5	1052
Years of education	6		8
	7		31
	8		112
	9		175
	10		264
	11		363
	12		472
	13		527
	14		572
	15		535
	16		467
	17		452
	18		353
	19		308
	20		214
	21		110
	22		32
	23		5

Between Subjects Factors shows general descriptive statistics for factors.

Table 9. ***Descriptive Statistics***

Dependent Variable: Household income in thousands

Geographic indicator	Years of education	Mean	Std. Deviation	N
Zone 1	7	24,0000	7,07107	2
	8	60,2258	45,13071	31
	9	39,5897	29,66454	39
	10	41,9275	43,40992	69
	11	41,0395	28,06917	76
	12	48,7981	48,57611	104
	13	43,2018	39,66591	109
	14	45,5856	38,97189	111
	15	49,8544	51,82331	103
	16	46,6818	35,29174	88
	17	58,4405	45,09063	84
	18	64,0167	58,48294	60
	19	69,7867	61,05343	75
	20	74,1220	49,92304	41
	21	58,9524	56,22764	21
	22	57,2500	27,32978	4
	23	63,0000	7,07107	2
	Total	51,2012	46,01240	1019
Zone 2	6	17,7500	8,30161	4
	7	22,0000	7,54983	3
	8	39,7727	24,90349	22
	9	32,7059	25,25690	34
	10	48,8571	45,44664	56
	11	41,7470	28,52720	83
	12	44,2469	33,56990	81
	13	39,9583	31,15797	96
	14	58,7899	55,69999	119
	15	47,8362	35,39318	116
	16	52,5904	50,83259	83
	17	58,6139	48,19024	101
	18	61,4079	48,49005	76
	19	56,5862	40,65172	58
	20	78,7568	65,84806	37
	21	84,3214	70,90790	28
	22	63,5000	40,97386	8
	Total	52,0468	45,22640	1005
Zone 3	6	45,0000	.	1

	7	25,6667	18,68333	6
	8	42,0833	37,44319	24
	9	46,6452	39,41789	31
	10	48,4898	35,63538	49
	11	48,9286	48,27647	70
	12	49,9608	35,19955	102
	13	48,3362	45,79784	116
	14	50,2315	40,84358	108
	15	53,0326	45,04856	92
	16	59,9368	53,61264	95
	17	61,7558	69,80459	86
	18	73,7313	60,30744	67
	19	65,2857	47,12695	63
	20	82,7826	62,00319	46
	21	79,7895	41,94782	19
	22	200,6667	238,92230	6
	Total	57,1346	53,61579	981
Zone 4	7	27,0000	12,63473	12
	8	57,0000	45,67106	14
	9	48,3793	53,44718	29
	10	46,4054	45,17341	37
	11	49,1159	54,23764	69
	12	48,3146	44,45953	89
	13	42,4712	35,96977	104
	14	47,5842	31,81172	101
	15	53,8571	50,33607	105
	16	65,8679	64,49837	106
	17	64,3780	68,21076	82
	18	61,1912	54,23505	68
	19	76,2778	146,30666	54
	20	71,5918	55,60198	49
	21	106,2381	171,51848	21
	22	33,0000	.	1
	23	128,5000	99,70206	2
	Total	56,6055	66,07229	943
Zone 5	6	38,3333	25,10644	3
	7	46,7500	32,75777	8
	8	38,2381	30,64621	21
	9	48,1190	30,37024	42
	10	49,1698	39,88432	53

	11	51,0769	44,60357	65
	12	48,7083	46,57646	96
	13	48,2647	50,84532	102
	14	50,4812	56,71321	133
	15	64,4958	59,50215	119
	16	54,1895	60,88501	95
	17	60,6162	49,42665	99
	18	76,9024	118,69502	82
	19	81,0000	81,52752	58
	20	71,2683	59,15531	41
	21	99,2857	103,50031	21
	22	63,3077	56,08979	13
	23	84,0000	.	1
	Total	58,2643	63,53879	1052
Total	6	28,8750	18,86370	8
	7	31,1613	21,49744	31
	8	47,7946	38,24604	112
	9	43,0057	35,98475	175
	10	46,6970	41,87569	264
	11	46,0551	41,39537	363
	12	48,1589	42,23122	472
	13	44,5769	41,44433	527
	14	50,7010	46,77187	572
	15	54,0056	49,40335	535
	16	56,3105	54,74004	467
	17	60,6637	56,37845	452
	18	67,7479	74,85806	353
	19	69,6299	81,55727	308
	20	75,6589	58,20242	214
	21	85,7364	98,69568	110
	22	87,4063	118,16496	32
	23	93,4000	59,98166	5
	Total	55,0406	55,54475	5000

Descriptive statistics privde detailed information such as the relationship between factor variables. In our example, you can see how many people with different years of education live in which zones. These statistics also show standart deviations and means.

Table 10. ***Levene's Test Results***

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
Household income in thousands	Based on Mean	4,173	82	4914	<,001
	Based on Median	2,244	82	4914	<,001
	Based on Median and with adjusted df	2,244	82	1534,375	<,001
	Based on trimmed mean	3,203	82	4914	<,001

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Household income in thousands

b. Design: Intercept + region + ed + region * ed

Since Sig. (p-value) of the Levene's test is lower than 0,05, we reject the null hypothesis which is error variance of the dependent variable is equal across groups. However, as it was mentioned, two way ANOVA is robust to unequal variance problem. So, we can continue with our analysis. However, in your articles, studies, works or papers, it is advised that you mention about your result and state it as your constraint in the analysis.

Table 11. ***General Results***

Tests of Between-Subjects Effects

Dependent Variable: Household income in thousands

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	826592,252[a]	85	9724,615	3,274	<,001	,054
Intercept	2125968,747	1	2125968,747	715,724	<,001	,127
region	33110,071	4	8277,518	2,787	,025	,002
ed	531800,717	17	31282,395	10,531	<,001	,035
region * ed	219810,899	64	3434,545	1,156	,186	,015
Error	14596418.51	4914	2970,374			
Total	30570349.00	5000				
Corrected Total	15423010.76	4999				

a. R Squared = ,054 (Adjusted R Squared = ,037)

The image above shows that the model is significant when you checked the Sig. (p-value). The only variable that is insignificant is interaction of region and education. So, living a specific region and education doesn't have any effect on the household income. On the other hand, living a specific region and education have an impact on the the household income, seperately.

The last statistic test is post-hoc test. But due to the size of the years of education (ed) variable, we will only show the results of the region varible. But the test is basically same with the one way ANOVA analysis. The analysis of both variables are done seperately by SPSS.

Table 12. ***Comparison Results***

Multiple Comparisons

Dependent Variable: Household income in thousands

	(I) Geographic indicator	(J) Geographic indicator	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Zone 1	Zone 2	-,8456	2,42293	,997	-7,4572	5,7661
		Zone 3	-5,9334	2,43781	,107	-12,5856	,7189
		Zone 4	-5,4043	2,46270	,182	-12,1245	1,3159
		Zone 5	-7,0631*	2,39552	,027	-13,6000	-,5262
	Zone 2	Zone 1	,8456	2,42293	,997	-5,7661	7,4572
		Zone 3	-5,0878	2,44612	,229	-11,7627	1,5872
		Zone 4	-4,5587	2,47093	,348	-11,3014	2,1839
		Zone 5	-6,2175	2,40399	,073	-12,7775	,3425
	Zone 3	Zone 1	5,9334	2,43781	,107	-,7189	12,5856
		Zone 2	5,0878	2,44612	,229	-1,5872	11,7627
		Zone 4	,5290	2,48552	1,000	-6,2534	7,3115
		Zone 5	-1,1297	2,41898	,990	-7,7306	5,4712
	Zone 4	Zone 1	5,4043	2,46270	,182	-1,3159	12,1245
		Zone 2	4,5587	2,47093	,348	-2,1839	11,3014
		Zone 3	-,5290	2,48552	1,000	-7,3115	6,2534
		Zone 5	-1,6587	2,44407	,961	-8,3281	5,0106
	Zone 5	Zone 1	7,0631*	2,39552	,027	,5262	13,6000
		Zone 2	6,2175	2,40399	,073	-,3425	12,7775
		Zone 3	1,1297	2,41898	,990	-5,4712	7,7306
		Zone 4	1,6587	2,44407	,961	-5,0106	8,3281
Bonferroni	Zone 1	Zone 2	-,8456	2,42293	1,000	-7,6499	5,9587
		Zone 3	-5,9334	2,43781	,150	-12,7795	,9127
		Zone 4	-5,4043	2,46270	,282	-12,3204	1,5117
		Zone 5	-7,0631*	2,39552	,032	-13,7904	-,3357
	Zone 2	Zone 1	,8456	2,42293	1,000	-5,9587	7,6499
		Zone 3	-5,0878	2,44612	,376	-11,9572	1,7817
		Zone 4	-4,5587	2,47093	,651	-11,4979	2,3804
		Zone 5	-6,2175	2,40399	,097	-12,9686	,5336
	Zone 3	Zone 1	5,9334	2,43781	,150	-,9127	12,7795
		Zone 2	5,0878	2,44612	,376	-1,7817	11,9572
		Zone 4	,5290	2,48552	1,000	-6,4511	7,5091
		Zone 5	-1,1297	2,41898	1,000	-7,9229	5,6635
	Zone 4	Zone 1	5,4043	2,46270	,282	-1,5117	12,3204
		Zone 2	4,5587	2,47093	,651	-2,3804	11,4979
		Zone 3	-,5290	2,48552	1,000	-7,5091	6,4511
		Zone 5	-1,6587	2,44407	1,000	-8,5224	5,2049
	Zone 5	Zone 1	7,0631*	2,39552	,032	,3357	13,7904
		Zone 2	6,2175	2,40399	,097	-,5336	12,9686
		Zone 3	1,1297	2,41898	1,000	-5,6635	7,9229
		Zone 4	1,6587	2,44407	1,000	-5,2049	8,5224

Based on observed means.
The error term is Mean Square(Error) = 2970,374.
*. The mean difference is significant at the ,05 level.

Tukey and Bonferoni tests show basically the same results. There is only one significant difference between people living in Zone 1 and Zone 5 regarding the household income. Analysis shows that people living in Zone 5 have better household income compared with the people in Zone 1.

d. ANCOVA (Practice)

For this example we will use dataset from SPSS samples: customer_dbase.sav

Select customer_dbase.sav.

Click on Analyze section from the top menu.

Find General Linear Model section under Analyze. Then click on Univariate… button.

Once you clicked you will see the following menu:

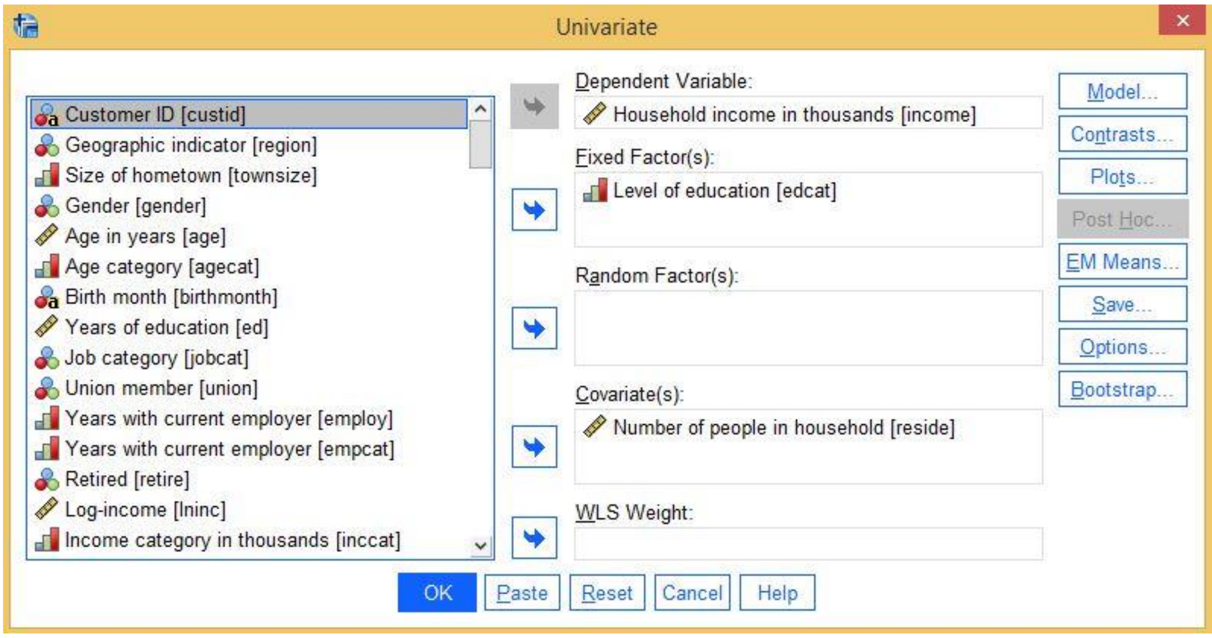


Figure 13. ***Variable Selection***

In ANCOVA, you will have a dependent variable, factor variables and covariates.

In this example, we will use one factor variable and covariate.

We selected Houshold income in thousands (income) as dependent variable, Level of education (edcat) as categorical / factor variable and Number of people in household (reside) as covariate / control variable.

In ANCOVA analysis, there is one additional assumption: Homogeneity of regression slopes.

In order to test this assumption, click on the Model button on the right.

Click on build terms or custom model.

Select each of the factor and covariate. Then select both of them on the right than click on the arrow button. This way, you will be able to analyze factor variable, covariate and their interaction term.

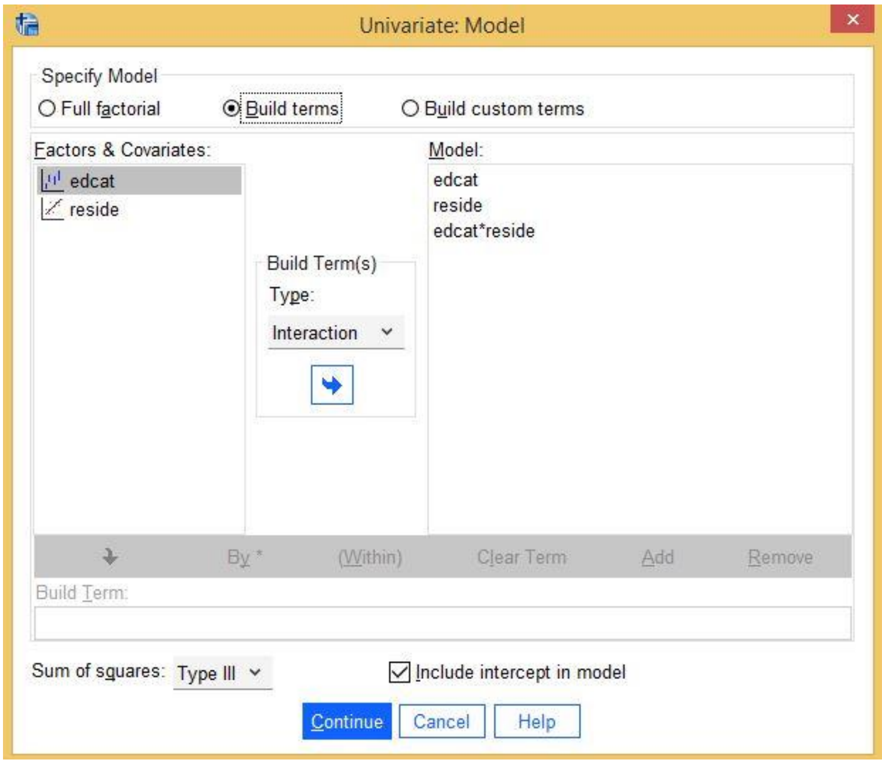


Figure 14. ***Specifying the Model***

Once you are done, click on Continue button. Then click on the OK button in the main menu.

Table 13. ***General Results***

Tests of Between-Subjects Effects

Dependent Variable: Household income in thousands

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	609364,395[a]	9	67707,155	22,807	<,001
Intercept	4423818,963	1	4423818,963	1490,170	<,001
edcat	113984,936	4	28496,234	9,599	<,001
reside	31371,013	1	31371,013	10,567	,001
edcat * reside	21362,023	4	5340,506	1,799	,126
Error	14813646.36	4990	2968,667		
Total	30570349.00	5000			
Corrected Total	15423010.76	4999			

a. R Squared = ,040 (Adjusted R Squared = ,038)

What you need to check in the table of Test of Between-Subjects Effects is the Sig. (p-value) of interaction term which is edcat*reside. If p-value is bigger than 0.05, in other words insignificant, then your model doesn't violate the assumption of homogeneity of regression slopes. In this example, the assumption is not violated (since the p-value of the interaction term is 0.126 which is bigger than 0.05), so we can continue the analysis.

So, you need to click Analyze -> General Linear Model -> Univariate again.

Now, you need to click Model… button on the menu at the right side. Then select Full factorial and continue.

After that click on the Options… button and select Descriptive Statistics, Estimates of effect size, Homogeneity tests and click on Continue… button.

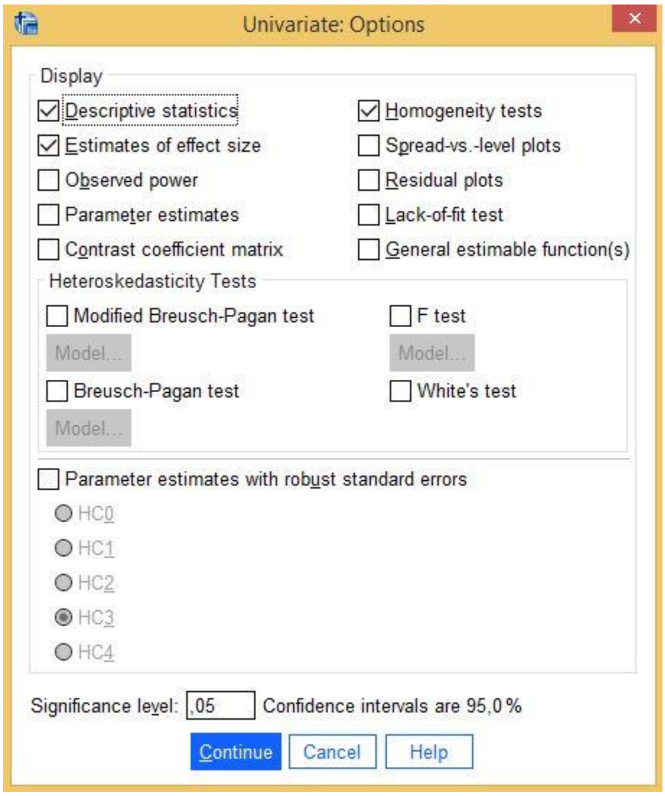


Figure 15. ***Descriptives, Estimates and Test Selection***

Now in the main menu click OK… to see the final results.

Table 14. ***General Results***

Tests of Between-Subjects Effects

Dependent Variable: Household income in thousands

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	588002,373[a]	5	117600,475	39,589	<,001	,038
Intercept	5541820,816	1	5541820,816	1865,577	,000	,272
reside	71692,842	1	71692,842	24,134	<,001	,005
edcat	516598,759	4	129149,690	43,476	<,001	,034
Error	14835008.39	4994	2970,566			
Total	30570349.00	5000				
Corrected Total	15423010.76	4999				

a. R Squared = ,038 (Adjusted R Squared = ,037)

From the results, we can clearly say that Level of education (edcat) and Number of people in household (reside) have significant impact on Household income in thousands.

e. MANOVA (Practice)

In the Manova, there are at least two dependent variables in the model. It is possible to have more than one categorical variable (not covariate) in the analysis.

For this example we will use dataset from SPSS samples: customer_dbase.sav

Select customer_dbase.sav.

Click on Analyze section from the top menu.

Find General Linear Model section under Analyze. Then click on Multivariate… button.

Once you clicked you will see the following menu:

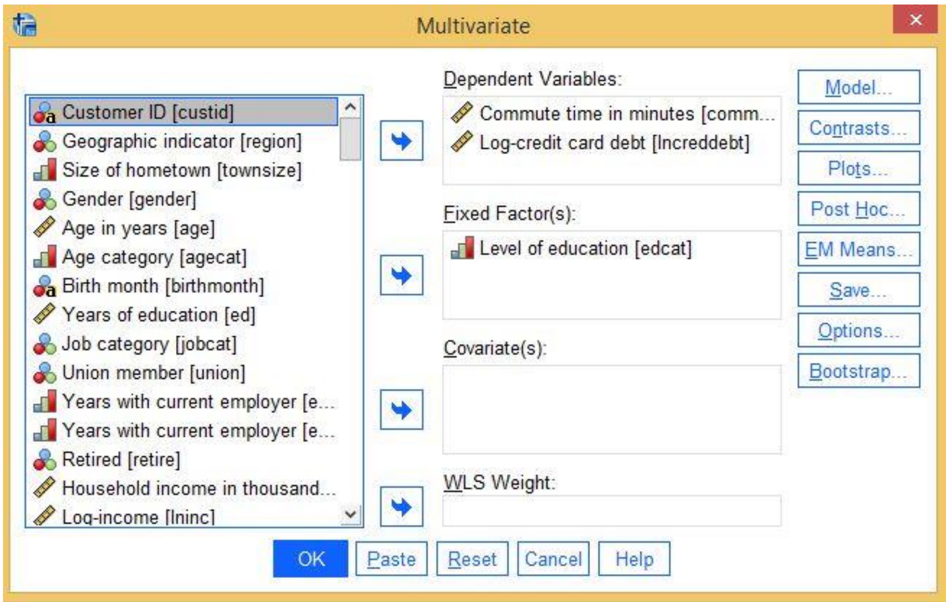


Figure 16. ***Variable Selection***

We use Commute time in minutes (commutetime) and Log-Credit Card Debt (lncreddebt) as dependent variables and level of education (edcat) as categorical independent variable.

As second step, click on the model button:

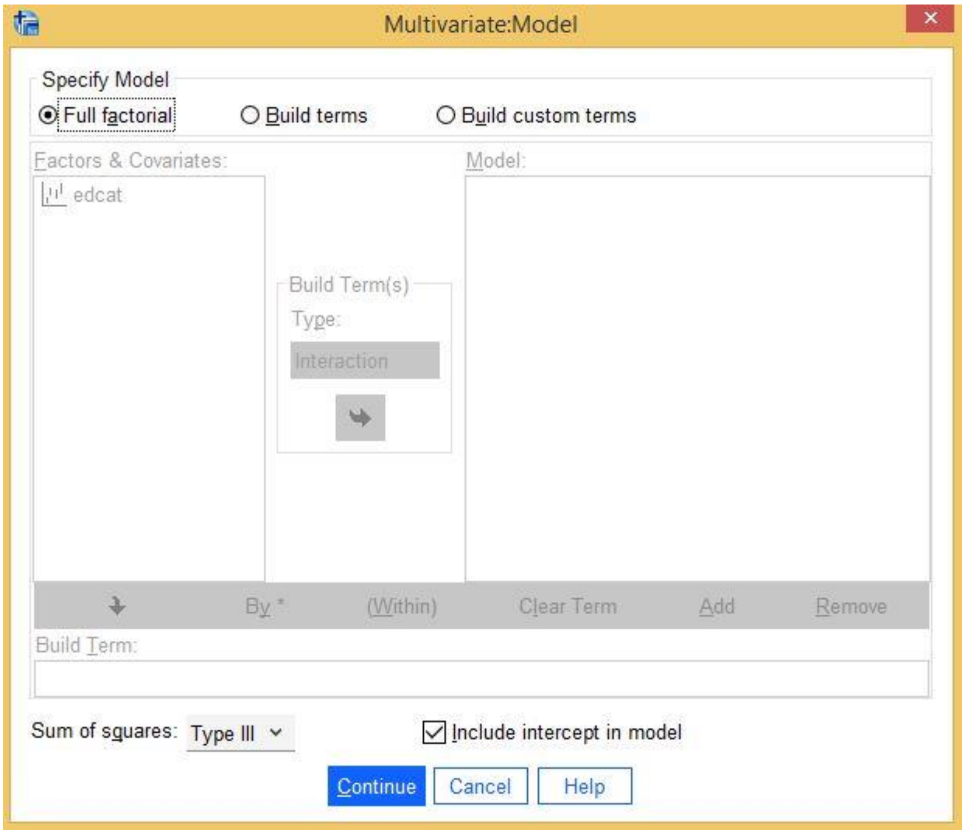


Figure 17. ***Specifying the Model***

Use full factorial model and click on Continue button and return to main menu.

Now click on the Post Hoc button and select edcat variable for the post hoc test.

Click on Turkey and Bonferroni tests. You may also click on Tamhane's and Dunnett's tests. If the results show that the variances are not equally distributed, you can use the latter tests.

Once you finished, click on Continue button.

Now on the main menu click on Options button.

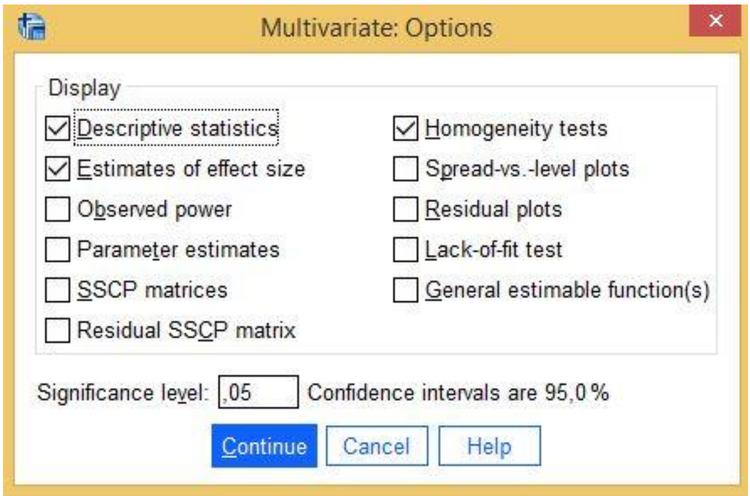


Figure 18. ***Descriptives, Estimates and Test Selection***

Select Descriptive Statistics, Estimates of effect size and homogeneity tests and click on continue button.

Table 15. ***General Descriptives***

Between-Subjects Factors

		Value Label	N
Level of education	1	Did not complete high school	952
	2	High school degree	1571
	3	Some college	1001
	4	College degree	1112
	5	Post-undergraduate degree	361

Between-subjects factors table shows how many samples are in each category.

Table 16. ***Descriptive Statistics***

Descriptive Statistics

	Level of education	Mean	Std. Deviation	N
Commute time in minutes	Did not complete high school	25,10	5,680	952
	High school degree	25,68	6,019	1571
	Some college	25,47	5,850	1001
	College degree	24,93	5,867	1112
	Post-undergraduate degree	25,45	5,992	361
	Total	25,35	5,891	4997
Log-credit card debt	Did not complete high school	-,3404	1,26471	952
	High school degree	-,2128	1,29089	1571
	Some college	-,1138	1,22975	1001
	College degree	,0725	1,24675	1112
	Post-undergraduate degree	,1872	1,34913	361
	Total	-,1249	1,27854	4997

Descriptive statistics table shows how many samples are in each category and their mean and standart deviation.

Table 17. ***Box's Test for Covariance Matrices Equality***

Box's Test of Equality of Covariance Matrices[a]

Box's M	14,979
F	1,247
df1	12
df2	23590520.03
Sig.	,243

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + edcat

One of the assumptions of MANOVA is the equality of covariance matrices. As you can see under the result table, the null hypothesis for the test is that covariance matrices of the dependent variables are equal across groups. As results show that the Sig. (p-value) is above 0.05 which means significant. Therefore we accept the null hypothesis and can continue with the analysis.

Table 18. ***Multivariate Tests***

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,936	36414,870[b]	2,000	4991,000	,000	,936
	Wilks' Lambda	,064	36414,870[b]	2,000	4991,000	,000	,936
	Hotelling's Trace	14,592	36414,870[b]	2,000	4991,000	,000	,936
	Roy's Largest Root	14,592	36414,870[b]	2,000	4991,000	,000	,936
edcat	Pillai's Trace	,019	11,976	8,000	9984,000	<,001	,010
	Wilks' Lambda	,981	12,007[b]	8,000	9982,000	<,001	,010
	Hotelling's Trace	,019	12,038	8,000	9980,000	<,001	,010
	Roy's Largest Root	,017	21,191[c]	4,000	4992,000	<,001	,017

a. Design: Intercept + edcat

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Multivariate tests show that the model is significant since Sig. (p-value) is under 0.05. So we can continue with the analysis.

Table 19. ***Levene's Test***

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
Commute time in minutes	Based on Mean	,865	4	4992	,484
	Based on Median	,820	4	4992	,512
	Based on Median and with adjusted df	,820	4	4973,814	,512
	Based on trimmed mean	,884	4	4992	,472
Log-credit card debt	Based on Mean	1,728	4	4992	,141
	Based on Median	1,585	4	4992	,175
	Based on Median and with adjusted df	1,585	4	4979,938	,175
	Based on trimmed mean	1,658	4	4992	,157

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + edcat

Another assumption of MANOVA is that error variances of the dependent variables are equal across groups. Levene's test show that Sig. (p-value) of both of the dependent variables are above 0.05. Therefore, we accept the null hypothesis and continue with the analysis.

Table 20. ***Multiple Comparisons***

Dependent Variable		(I) Level of education	(J) Level of education	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Commute time in minutes	Tukey HSD	Did not complete high school	High school degree	-,58	,242	,113	-1,24	,08
			Some college	-,37	,266	,624	-1,10	,35
			College degree	,17	,260	,969	-,54	,88
			Post-undergraduate degree	-,35	,364	,873	-1,34	,64
		High school degree	Did not complete high school	,58	,242	,113	-,08	1,24
			Some college	,21	,238	,907	-,44	,86
			College degree	,75*	,231	,010	,12	1,38
			Post-undergraduate degree	,23	,344	,961	-,70	1,17
		Some college	Did not complete high school	,37	,266	,624	-,35	1,10
			High school degree	-,21	,238	,907	-,86	,44
			College degree	,54	,256	,216	-,16	1,24
			Post-undergraduate degree	,03	,361	1,000	-,96	1,01
		College degree	Did not complete high school	-,17	,260	,969	-,88	,54
			High school degree	-,75*	,231	,010	-1,38	-,12
			Some college	-,54	,256	,216	-1,24	,16
			Post-undergraduate degree	-,52	,357	,598	-1,49	,46
			Did not complete high school	,35	,364	,873	-,64	1,34
			High school degree	-,23	,344	,961	-1,17	,70

		Post-undergraduate degree	Some college	-,03	,361	1,000	-1,01	,96
			College degree	,52	,357	,598	-,46	1,49
	Bonferroni	Did not complete high school	High school degree	-,58	,242	,160	-1,26	,10
			Some college	-,37	,266	1,000	-1,12	,37
			College degree	,17	,260	1,000	-,56	,90
			Post-undergraduate degree	-,35	,364	1,000	-1,37	,67
		High school degree	Did not complete high school	,58	,242	,160	-,10	1,26
			Some college	,21	,238	1,000	-,46	,88
			College degree	,75*	,231	,012	,10	1,40
			Post-undergraduate degree	,23	,344	1,000	-,73	1,20
		Some college	Did not complete high school	,37	,266	1,000	-,37	1,12
			High school degree	-,21	,238	1,000	-,88	,46
			College degree	,54	,256	,349	-,18	1,26
			Post-undergraduate degree	,03	,361	1,000	-,99	1,04
		College degree	Did not complete high school	-,17	,260	1,000	-,90	,56
			High school degree	-,75*	,231	,012	-1,40	-,10
			Some college	-,54	,256	,349	-1,26	,18
			Post-undergraduate degree	-,52	,357	1,000	-1,52	,49
		Post-undergraduate degree	Did not complete high school	,35	,364	1,000	-,67	1,37
			High school degree	-,23	,344	1,000	-1,20	,73
			Some college	-,03	,361	1,000	-1,04	,99
			College degree	,52	,357	1,000	-,49	1,52
Log-credit card debt	Tukey HSD	Did not complete high school	High school degree	-,1276	,05210	,103	-,2697	,0146
			Some college	-,2266*	,05742	,001	-,3833	-,0699
			College degree	-,4129*	,05601	,000	-,5658	-,2601
			Post-undergraduate degree	-,5276*	,07840	,000	-,7415	-,3137
		High school degree	Did not complete high school	,1276	,05210	,103	-,0146	,2697
			Some college	-,0990	,05130	,301	-,2390	,0409
			College degree	-,2854*	,04971	,000	-,4210	-,1497
			Post-undergraduate degree	-,4000*	,07403	,000	-,6021	-,1980
		Some college	Did not complete high school	,2266*	,05742	,001	,0699	,3833
			High school degree	,0990	,05130	,301	-,0409	,2390
			College degree	-,1863*	,05526	,007	-,3371	-,0355
			Post-undergraduate degree	-,3010*	,07787	,001	-,5135	-,0885
		College degree	Did not complete high school	,4129*	,05601	,000	,2601	,5658
			High school degree	,2854*	,04971	,000	,1497	,4210
			Some college	,1863*	,05526	,007	,0355	,3371
			Post-undergraduate degree	-,1147	,07684	,567	-,3243	,0950
			Did not complete high school	,5276*	,07840	,000	,3137	,7415
			High school degree	,4000*	,07403	,000	,1980	,6021

	Post-undergraduate degree	Some college	,3010*	,07787	,001	,0885	,5135
		College degree	,1147	,07684	,567	-,0950	,3243
Bonferroni	Did not complete high school	High school degree	-,1276	,05210	,144	-,2739	,0188
		Some college	-,2266*	,05742	,001	-,3879	-,0653
		College degree	-,4129*	,05601	,000	-,5702	-,2556
		Post-undergraduate degree	-,5276*	,07840	,000	-,7478	-,3074
	High school degree	Did not complete high school	,1276	,05210	,144	-,0188	,2739
		Some college	-,0990	,05130	,536	-,2431	,0450
		College degree	-,2854*	,04971	,000	-,4250	-,1458
		Post-undergraduate degree	-,4000*	,07403	,000	-,6080	-,1921
	Some college	Did not complete high school	,2266*	,05742	,001	,0653	,3879
		High school degree	,0990	,05130	,536	-,0450	,2431
		College degree	-,1863*	,05526	,008	-,3415	-,0311
		Post-undergraduate degree	-,3010*	,07787	,001	-,5197	-,0823
	College degree	Did not complete high school	,4129*	,05601	,000	,2556	,5702
		High school degree	,2854*	,04971	,000	,1458	,4250
		Some college	,1863*	,05526	,008	,0311	,3415
		Post-undergraduate degree	-,1147	,07684	1,000	-,3305	,1011
	Post-undergraduate degree	Did not complete high school	,5276*	,07840	,000	,3074	,7478
		High school degree	,4000*	,07403	,000	,1921	,6080
		Some college	,3010*	,07787	,001	,0823	,5197
		College degree	,1147	,07684	1,000	-,1011	,3305

Based on observed means.

The error term is Mean Square(Error) = 1,609.

*. The mean difference is significant at the ,05 level.

According to results, there is a significant difference between Commute time of high school and college graduates. Commute time of high school graduates are 75% higher than the college graduates. The main reason of this result can be rationalized as the bargaining power and/or chances of the college graduates are higher than the high school graduates in finding jobs closer where they live.

According to results for Credit Card Debt, there are several significant differences between education categories. In order to keep the example short, we will examine only the people with the College degree. Compared with the people with no high school graduation, high school graduation, some college degree, people with college degree have 41%, 29% and 19% more debt, respectively. There is no significant difference between post-undergraduate degree and college degree. The main reason that the people with college degree have more debt is that these people simply earn more Money, so they have a higher debt. Results show that there is a decrease in percentage when the graduation degree is higher.

f. MANCOVA (Practice)

In the Mancova, there are at least two dependent variables in the model. Your model has to have at least one covariate.

For this example we will use dataset from SPSS samples: customer_dbase.sav

Select customer_dbase.sav.

Click on Analyze section from the top menu.

Find General Linear Model section under Analyze. Then click on Multivariate… button.

Once you clicked you will see the following menu:

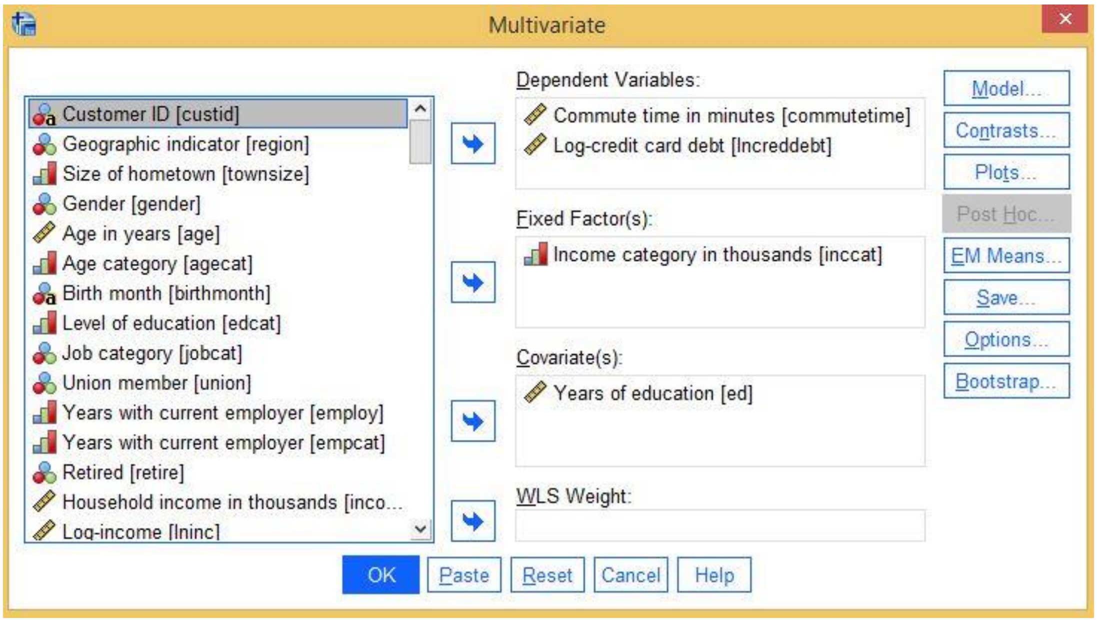


Figure 19. ***Variable Selection***

We use Commute time in minutes (commutetime) and Log-Credit Card Debt (lncreddebt) as dependent variables and Income category in thousands (inccat) as categorical independent variable and Years of education (ed) as covariate.

As second step, click on the model button:

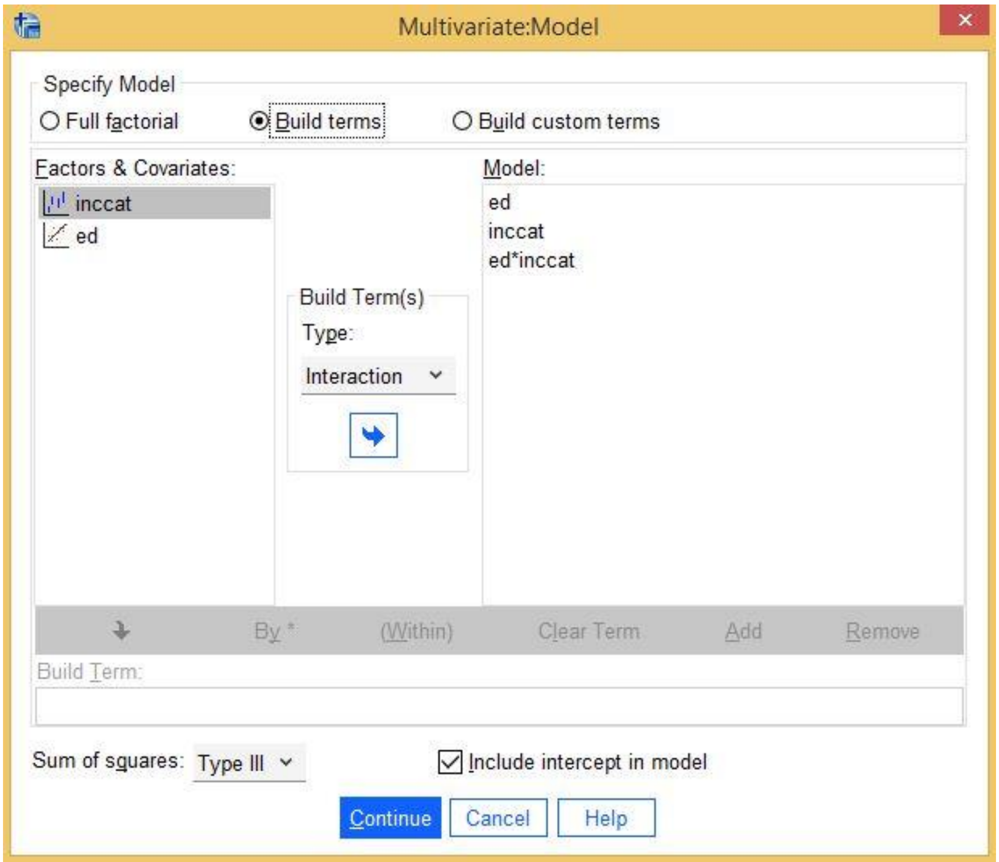


Figure 20. ***Specifying the Model***

In MANCOVA analysis (like in ANCOVA), there is one additional assumption: Homogeneity of regression slopes.

In order to test this assumption, click on the Model button on the right.

Click on build terms or custom model.

Select each of the factor and covariate. Then select both of them on the right than click on the arrow button. This way, you will be able to analyze factor variable, covariate and their interaction term.

Once you are done, click on Continue button. Then click on the OK button in the main menu.

Table 21. ***General Results***

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Commute time in minutes	262,724[a]	9	29,192	,841	,578	,002
	Log-credit card debt	2585,425[b]	9	287,269	256,677	,000	,317
Intercept	Commute time in minutes	110768,473	1	110768,473	3190,747	,000	,390
	Log-credit card debt	,619	1	,619	,553	,457	,000
ed	Commute time in minutes	,018	1	,018	,001	,982	,000
	Log-credit card debt	3,934	1	3,934	3,515	,061	,001
inccat	Commute time in minutes	106,033	4	26,508	,764	,549	,001
	Log-credit card debt	103,925	4	25,981	23,214	<,001	,018
inccat * ed	Commute time in minutes	76,992	4	19,248	,554	,696	,000
	Log-credit card debt	9,474	4	2,369	2,116	,076	,002
Error	Commute time in minutes	173126,337	4987	34,716			
	Log-credit card debt	5581,381	4987	1,119			
Total	Commute time in minutes	3383613,000	4997				
	Log-credit card debt	8244,757	4997				
Corrected Total	Commute time in minutes	173389,061	4996				
	Log-credit card debt	8166,806	4996				

a. R Squared = ,002 (Adjusted R Squared = ,000)

b. R Squared = ,317 (Adjusted R Squared = ,315)

What you need to check in the table of Test of Between-Subjects Effects is the Sig. (p-value) of interaction term which is inccat*ed. If p-value is bigger than 0.05, in other words insignificant, then your model doesn't violate the assumption of homogeneity of regression slopes. In this example, the assumption is not violated (since the p-value of the interaction term for both dependent variable are bigger than 0.05), so we can continue the analysis.

Table 22. ***Levene's Test***

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Commute time in minutes	1,140	4	4992	,335
Log-credit card debt	1,769	4	4992	,132

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + ed + inccat + inccat * ed

Another assumption for the equal distribution of the error variance of the dependent variable is tested by Levene's test. As it can be seen that both of the Sig. (p-value) are higher than 0.05. Therefore, we accept the null hypothesis and the model meets the assumption. So, we can continue with the analysis.

Since you checked the assumptions, you need to click Analyze -> General Linear Model -> Multivariate again.

Now, you need to click Model… button on the menu at the right side. Then select Full factorial and continue.

After that click on the Options… button and select Descriptive Statistics, Estimates of effect size, Homogeneity tests and click on Continue… button.

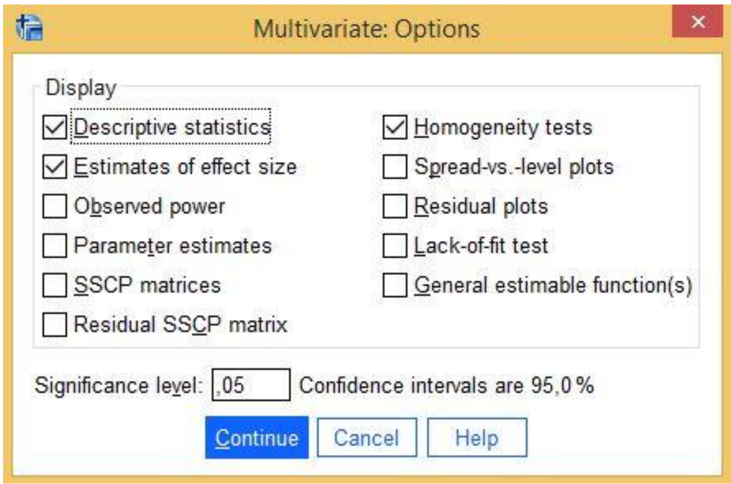


Figure 21. ***Descriptives, Estimates and Test Selection***

One of the assumptions of MANOVA is the equality of covariance matrices. As you can see under the result table, the null hypothesis for the test is that covariance matrices of the dependent variables are equal across groups. As results show that the Sig. (p-value) is above 0.05 which means significant. Therefore we accept the null hypothesis and can continue with the analysis.

Table 23. ***Box's Test for Covariance Matrices Equality***

Box's Test of Equality of Covariance Matrices[a]

Box's M	18,023
F	1,500
df1	12
df2	26350075.67
Sig.	,116

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + ed + inccat

Table 24. ***Multivariate Tests***

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,452	2058,416[b]	2,000	4990,000	,000	,452
	Wilks' Lambda	,548	2058,416[b]	2,000	4990,000	,000	,452
	Hotelling's Trace	,825	2058,416[b]	2,000	4990,000	,000	,452
	Roy's Largest Root	,825	2058,416[b]	2,000	4990,000	,000	,452
ed	Pillai's Trace	,001	1,605[b]	2,000	4990,000	,201	,001
	Wilks' Lambda	,999	1,605[b]	2,000	4990,000	,201	,001
	Hotelling's Trace	,001	1,605[b]	2,000	4990,000	,201	,001
	Roy's Largest Root	,001	1,605[b]	2,000	4990,000	,201	,001
inccat	Pillai's Trace	,304	223,608	8,000	9982,000	,000	,152
	Wilks' Lambda	,696	247,623[b]	8,000	9980,000	,000	,166
	Hotelling's Trace	,436	272,021	8,000	9978,000	,000	,179
	Roy's Largest Root	,436	543,693[c]	4,000	4991,000	,000	,303

a. Design: Intercept + ed + inccat

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Multivariate test results show that there is a significant relationship between income groups and dependent variables, on the other hand Years of Education variable has no significant effect on them.

Table 25. ***General Results***

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Commute time in minutes	185,731[a]	5	37,146	1,070	,375	,001
	Log-credit card debt	2575,951[b]	5	515,190	459,914	,000	,315
Intercept	Commute time in minutes	142746,681	1	142746,681	4113,366	,000	,452
	Log-credit card debt	1,655	1	1,655	1,477	,224	,000
ed	Commute time in minutes	,752	1	,752	,022	,883	,000
	Log-credit card debt	3,579	1	3,579	3,195	,074	,001
inccat	Commute time in minutes	177,014	4	44,254	1,275	,277	,001
	Log-credit card debt	2434,888	4	608,722	543,411	,000	,303
Error	Commute time in minutes	173203,329	4991	34,703			
	Log-credit card debt	5590,856	4991	1,120			
Total	Commute time in minutes	3383613,000	4997				
	Log-credit card debt	8244,757	4997				
Corrected Total	Commute time in minutes	173389,061	4996				
	Log-credit card debt	8166,806	4996				

a. R Squared = ,001 (Adjusted R Squared = ,000)

b. R Squared = ,315 (Adjusted R Squared = ,315)

Test of between-subjects effects shows that income groups variable has a significant relationship with credit card debt but insignificant relationship with commute time in minutes.

Since previous table showed us that there is no significant effect of Years of Education on both dependent variables, we cannot examine the between-subjects effect.

g. T-Test (Practice)

T test can be used for comparing the difference of means of 2 groups. If you have more than 2 groups to compare, you need to use ANOVA type analysis.

Normality and equality of variances assumptions should be hold for T-Test. For this example, we skip normality test since we already referred how to do it for each variable in the previous section.

For this example we will use dataset from SPSS samples: customer_dbase.sav

Select customer_dbase.sav.

Click on Analyze section from the top menu.

Find Compare Means section under Analyze. Then click on Independent Samples T-Test button.

Once you clicked you will see the following menu:

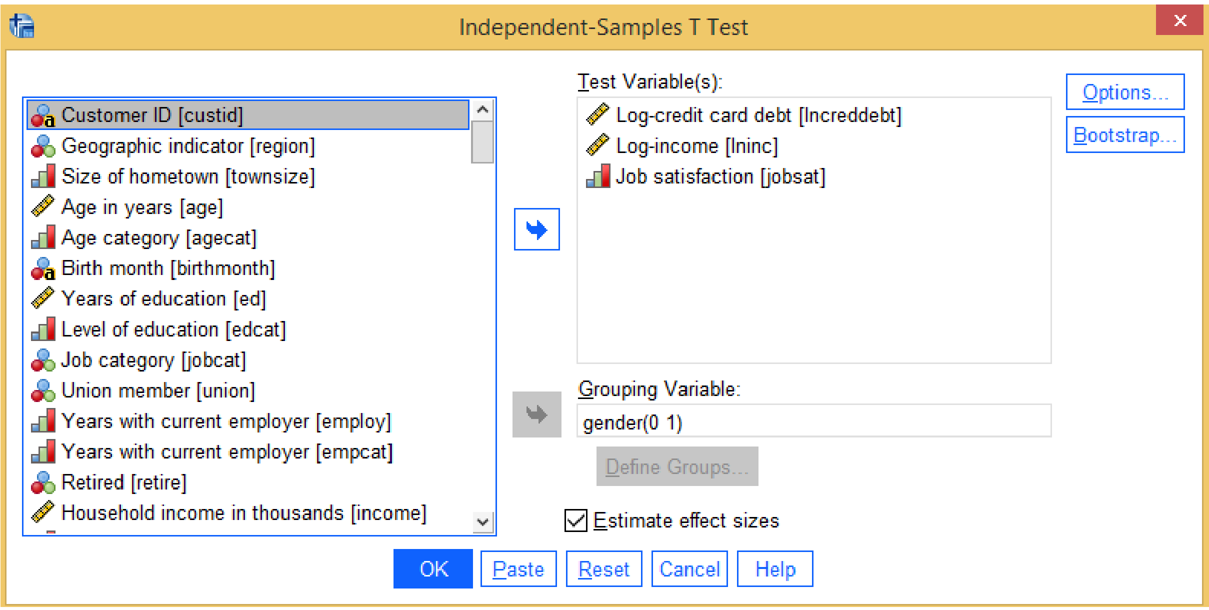


Figure 22. ***Variable Selection***

For this example we selected the variables that we are going to test:

Log-credit card debt

Log-income

Job satisfaction

We will examine these variables for Gender groups. So we selected gender from the left menu and put it to Grouping Variable section. Once you do this, you need to click on define groups to name the groups:

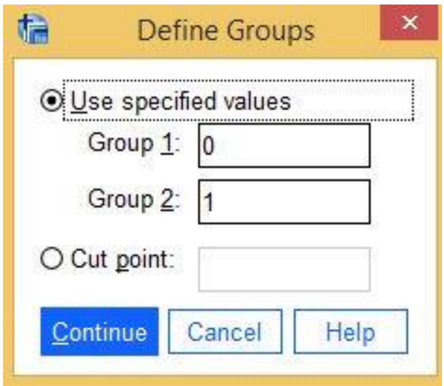


Figure 23. ***Defining Groups***

In this example, we named them as 0 (Male) and 1 (Female) because in the data each gender was stated as dummy variables. In order to prevent any mix ups, we named the data accordingly. Click on continue to proceed and switch to the main menu.

Optional: Before you continue to analysis from the main menu, you may click on Options button and re-arrange the confidence intervals, however the default confidence interval is 95%, so it is not necessary to change it. So, we skip this part.

On the main menu, click OK button for starting to analysis.

Table 26. ***Group Statistics***

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Log-credit card debt	0	2448	-,1259	1,28986	,02607
	1	2551	-,1235	1,26747	,02509
Log-income	0	2449	3,7029	,75631	,01528
	1	2551	3,7034	,74539	,01476
Job satisfaction	0	2449	2,95	1,379	,028
	1	2551	2,97	1,372	,027

The group statistics show the number of cross section units, mean, standard deviation and standard error mean of each variable for each gender group.

Table 27. ***Independent Samples Test***

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Log-credit card debt	Equal variances assumed	,083	,773	-,064	4997	,949	-,00233	,03617	-,07324	,06859
	Equal variances not assumed			-,064	4979,828	,949	-,00233	,03619	-,07327	,06861
Log-income	Equal variances assumed	,665	,415	-,027	4998	,979	-,00057	,02124	-,04221	,04107
	Equal variances not assumed			-,027	4982,737	,979	-,00057	,02125	-,04222	,04108
Job satisfaction	Equal variances assumed	1,248	,264	-,626	4998	,532	-,024	,039	-,101	,052
	Equal variances not assumed			-,626	4987,517	,532	-,024	,039	-,101	,052

Independent samples test shows our main analysis result. We mentioned about the equal variance assumption that should be hold for the analysis. SPSS allows us to check the results from the independent samples test. For that we need to check Levene's Test for Equality of Variances results. Sig. (p-value) should be higher than 0.05, so we can assume that the variances are equally distributed. For each variable, the results are higher than 0.05, so we can say that the variances are equally distributed.

For the second step, we need to check T-test for Equality of Means – Sig. (2-tailed) section to see whether there is a difference between two gender groups. In order for us to interpret that there is a statistically significant difference among groups, Sig. (p-value) should be lower than 0.05. As we can see, p-value of none of the variables are lower than 0.05. Therefore, we fail to reject the H0, so we can say that there is no statistically significant difference among gender groups in terms of credit card debt, income and job satisfaction.

If the variances were not equally distributed, we need to check the second line under the T-test for Equality of Means – Sig. (2-tailed) section. In this example both of the lines are equal.

h. Correlation & Regression Analysis (Practice)

Correlation and regression analyses can be used to examine the relationship between variables while T-test and ANOVA type analyses examine the differences among groups.

Correlation analysis shows the relationship of each variable with each other seperately. Coefficient of the relationship of variable can be either negative or positive. Unlike regression analysis, there may not be a causality between variables.

For this example we will use dataset from SPSS samples: customer_dbase.sav

Select customer_dbase.sav.

Click on Analyze section from the top menu.

Find Correlate section under Analyze. Then click on Bivariate Correlations button.

Once you clicked you will see the following menu:

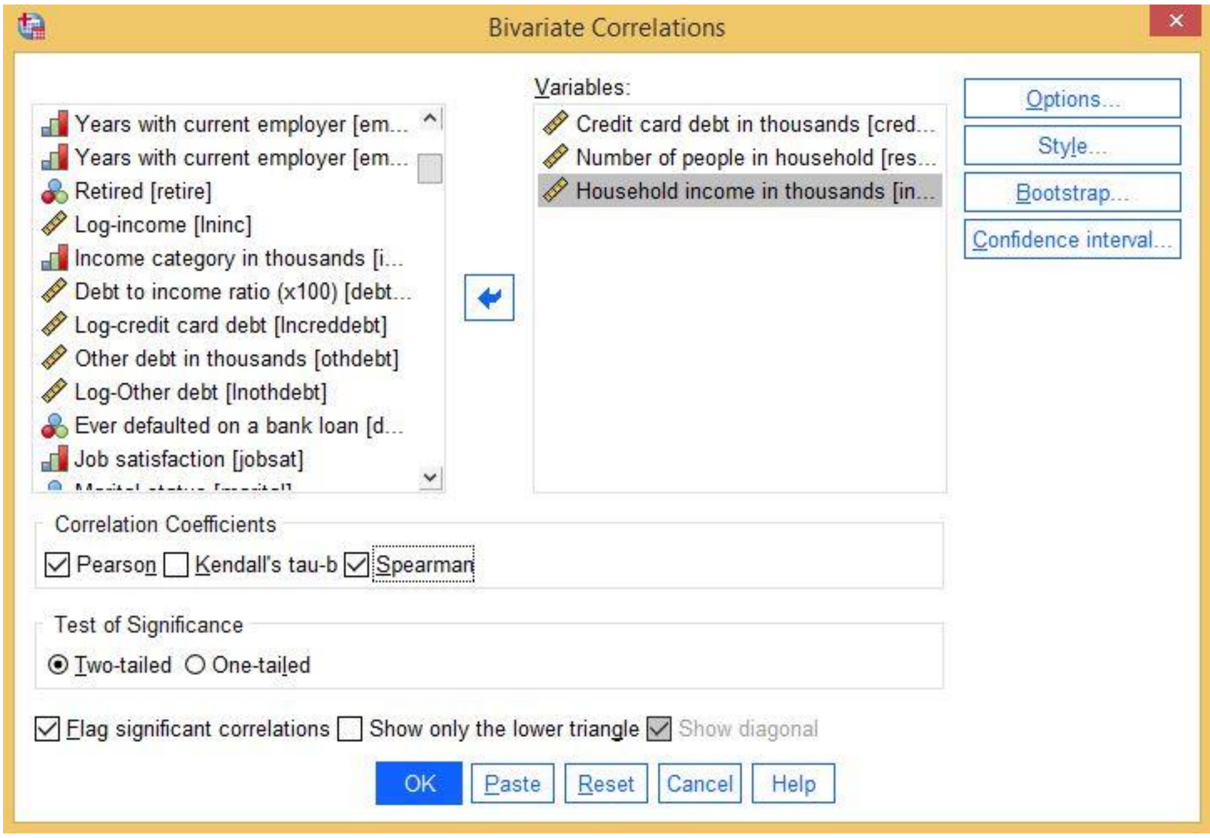


Figure 24. ***Variable Selection***

Normality assumption is important for the correlation analysis. So, if your variables are normally distributed you need to use pearson correlation coefficient, if not use spearman coefficient.

If you assume that there is only one way relationship between variables (i.e. you expect only a positive relationship between variables), you need to choose One-tailed test. If you are not sure or not foresee positive or negative relationship, choose Two-tailed test.

Once you finished, click OK to see the results. For this example, we selected both Pearson and Spearman Coefficients.

Table 28. ***Correlations***

		Credit card debt in thousands	Number of people in household	Household income in thousands
Credit card debt in thousands	Pearson Correlation	1	-,044**	,663**
	Sig. (2-tailed)		,002	,000
	N	5000	5000	5000
Number of people in household	Pearson Correlation	-,044**	1	-,068**
	Sig. (2-tailed)	,002		,000
	N	5000	5000	5000
Household income in thousands	Pearson Correlation	,663**	-,068**	1
	Sig. (2-tailed)	,000	,000	
	N	5000	5000	5000

**. Correlation is significant at the 0.01 level (2-tailed).

Table 29. ***Correlations***

			Credit card debt in thousands	Number of people in household	Household income in thousands
Spearman's rho	Credit card debt in thousands	Correlation Coefficient	1,000	-,042**	,589**
		Sig. (2-tailed)	.	,003	,000
		N	5000	5000	5000
	Number of people in household	Correlation Coefficient	-,042**	1,000	-,057**
		Sig. (2-tailed)	,003	.	,000
		N	5000	5000	5000
	Household income in thousands	Correlation Coefficient	,589**	-,057**	1,000
		Sig. (2-tailed)	,000	,000	.
		N	5000	5000	5000

** Correlation is significant at the 0.01 level (2-tailed).

** sign next to the coefficients show that the correlation is significant at 0.01 level.

If the sign was *, this would mean that the correlation is significant at 0.05 level.

In both coefficient tests, there are statistically significant relationship between each variable pairs.

Analysis results show that there is a negative correlation between Credit card debt and number of people in the household, and a positive correlation between Credit card debt and household income. We can intrepret that the household generate more income than they spent. That is the reason why there is such relationship.

Regression analysis can be used to examine the effect of the independent variable(s) on the dependent variable. A simple regression function can be illustrated as follows:

$$Yi = ß0 + ß1x + ε$$

Yi: Dependent variable

ß0: Constant / Intercept

ß1: Slope / Coefficient

x: Independent variable

ε: Error term

There can be more than one independent variables in the analysis. Effects of each variable on the dependent variable can be examined with their coefficients. The unobserved effects and variables will be represented by the error term.

For this example we will use dataset from SPSS samples: customer_dbase.sav

Select customer_dbase.sav.

Click on Analyze section from the top menu.

Find Regression section under Analyze. Then click on Linear button.

Once you clicked you will see the following menu:

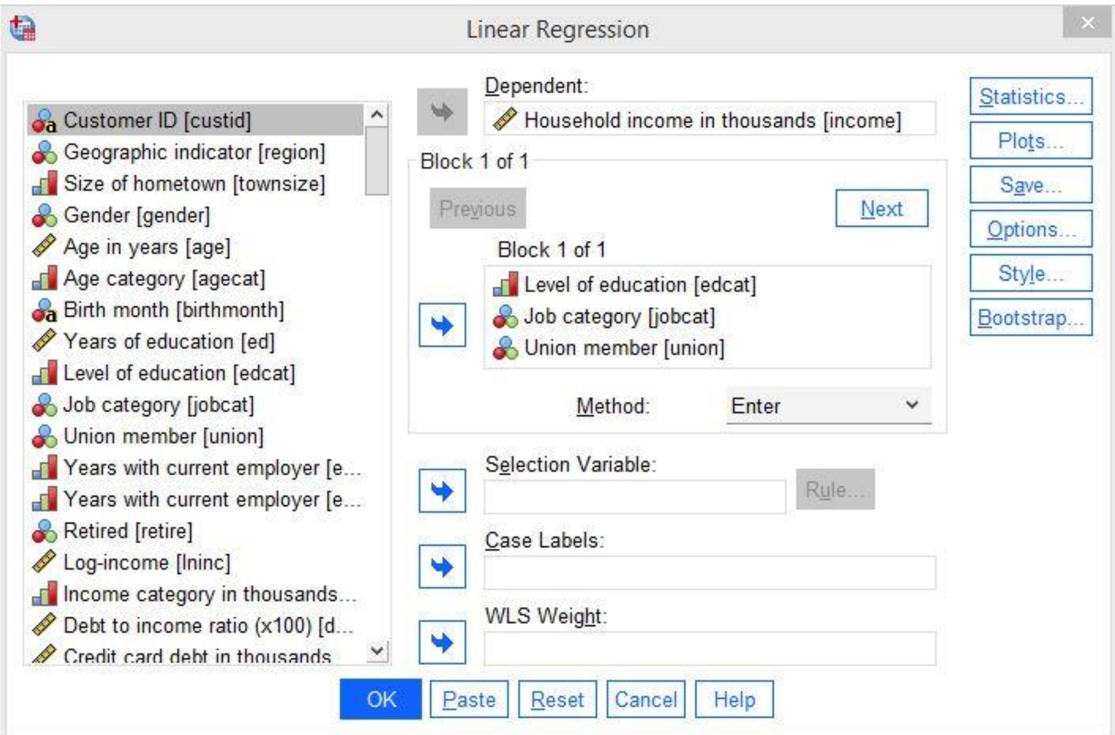


Figure 25: ***Variable Selection***

In this example, we will make a multiple regression analysis. We will examine the effects of level of education, job categories and union membership on Household income.

Before we start, I would like to remind you that, your variables should be normally distributed and have equal variance.

Once you selected your variables, click Statistics button on the right:

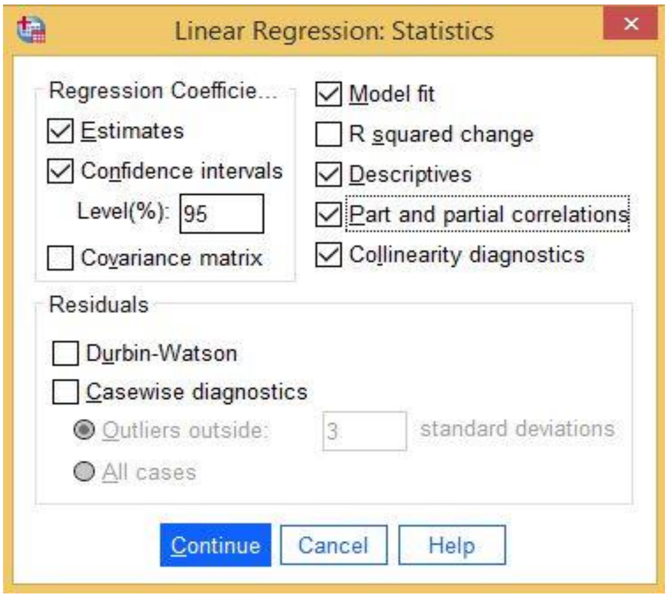


Figure 26: ***Statistics and Specifications***

Select Model fit, descriptives, part and partial correlations, colinearity diagnostics, confidence intervals (as 95%) and click on Continue.

On the main menu click OK to continue to analysis.

Table 30. Descriptive Statistics

	Mean	Std. Deviation	N
Household income in thousands	55,0406	55,54475	5000
Level of education	2,67	1,217	5000
Job category	2,76	1,741	5000
Union member	,15	,357	5000

Table 31. Correlations

		Household income in thousands	Level of education	Job category	Union member
Pearson Correlation	Household income in thousands	1,000	,176	,105	,013
	Level of education	,176	1,000	-,075	-,002
	Job category	,105	-,075	1,000	,082
	Union member	,013	-,002	,082	1,000
Sig. (1-tailed)	Household income in thousands	.	,000	,000	,186
	Level of education	,000	.	,000	,439
	Job category	,000	,000	.	,000
	Union member	,186	,439	,000	.
N	Household income in thousands	5000	5000	5000	5000

	Level of education	5000	5000	5000	5000
	Job category	5000	5000	5000	5000
	Union member	5000	5000	5000	5000

In the correlation matrix, it is important to not have a relationship above 0.70. This indicates a strong relationship among variables and yield spurious results. This would indicate multicolinearity problem. In this analysis, we see that there is not a strong relationship between variables. So, we can continue to analysis.

Table 32. Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Union member, Level of education, Job category[b]	.	Enter

a. Dependent Variable: Household income in thousands

b. All requested variables entered.

Table 33. Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,212[a]	,045	,044	54,29498

a. Predictors: (Constant), Union member, Level of education, Job category

Model summary shows the R values. Since we used a multiple regression model, we need to check Adjusted R Square. This value shows the power of the independent variables to explain the dependent variable. So, from level of education, job category and union membership, only 4.4% of Household income can be explained. This means that there are other contributors that we cannot currently observe and use in the model. If you have more variables, you need to use it in the regression model, otherwise analysis will be effected by unobserved variables.

Table 34. ***ANOVA*[a]**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	695077,819	3	231692,606	78,595	,000[b]
	Residual	14727932,940	4996	2947,945		
	Total	15423010,758	4999			

a. Dependent Variable: Household income in thousands

b. Predictors: (Constant), Union member, Level of education, Job category

When we checked the Sig. (p-value) of the ANOVA analysis, it can be seen that it is lower than 0.05. This means that at least one variable among independent variables has a statistically significant effect on the dependent variable. For more information about the effect we will examine next analysis.

Table 35. ***Coefficients***[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	95,0% Confidence Interval for B Upper Bound	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	21,950	2,302		9,537	,000	17,438	26,463					
	Level of education	8,440	,633	,185	13,335	,000	7,199	9,681	,176	,185	,184	,994	1,006
	Job category	3,787	,444	,119	8,535	,000	2,917	4,657	,105	,120	,118	,988	1,012
	Union member	,507	2,161	,003	,234	,815	-3,731	4,744	,013	,003	,003	,993	1,007

a. Dependent Variable: Household income in thousands

The first thing we need to check in this table is Sig. (p-value). It can be seen that level of education and job category have a significant effect on Household income, on the other hand being a union member has no statistically significant impact on it.

Unstandardized Coefficients show the effect of one unit increase on Household income. So, one level increase in the level of education and job category increase the Household income by 8,440 and 3,787 USD.

Standardized Coefficients show the effect of one unit increase in standard deviation on the standard deviation of Household income.

3. References

Ankarali, H., Cangur, S., & Ankarali, S. (2018). A new Seyhan's approach in case of heterogeneity of regression slopes in ANCOVA. *Interdisciplinary Sciences: Computational Life Sciences, 10,* 282-290.

Dattalo, P. (2013). *Analysis of multiple dependent variables.* Oxford University Press.

Gogtay, N. J., & Thatte, U. M. (2017). Principles of correlation analysis. *Journal of the Association of Physicians of India, 65*(3), 78-81.

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis*. John Wiley & Sons.

Huberty, C. J., & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In H. E. A. Tinsley and S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 183-208). Academic Press.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance* (No. 118). Sage.

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38*(1), 52.

Kim, T. K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology, 68*(6), 540.

Leech, N., Barrett, K., & Morgan, G. A. (2013). *SPSS for intermediate statistics: Use and interpretation.* Routledge.

Levene, H. (1960) Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics* (pp. 278-292). Stanford University Press.

Liu, Q., & Wang, L. (2021). t-Test and ANOVA for data with ceiling and/or floor effects. *Behavior Research Methods, 53*(1), 264-277.

Livingston, E. H. (2004). Who was student and why do we care so much about his t-test? 1. *Journal of Surgical Research, 118*(1), 58-65.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (Eds.). (1998). *Applied regression analysis: a research tool.* Springer.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21-33.

Rutherford, A. (2011). *ANOVA and ANCOVA: a GLM approach*. John Wiley & Sons.

Tian, C. H. E. N., Manfei, X. U., Justin, T. U., Hongyue, W. A. N. G., & Xiaohui, N. I. U. (2018). Relationship between Omnibus and Post-hoc Tests: An investigation of performance of the F test in ANOVA. *Shanghai Archives of Psychiatry, 30*(1).

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences, 106,* 234-240.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64*(1), 1-48.

Westfall, P. H., & Henning, K. S. (2013). *Understanding advanced statistical methods* (Vol. 543). Boca Raton, FL: CRC Press.

Digital Source: https://www.mathsisfun.com/data/standard-normal-distribution.html (Accessed: 01.06.2023)

SCALE DEVELOPMENT

Modules overview

Introduction:

Explanation of the significance of psychological scale development in research and assessment. Highlighting the relevance of competence-oriented methodologies in scale development. Outline of the training material's structure and objectives.

Module 1: Understanding Psychological Scales

Definition of psychological scales and their role in psychological research. Distinction between various types of psychological scales, including Likert scales, Thurstone scales, and others. Discussion of key characteristics of psychological scales, including reliability and validity.

Module 2: Identifying the Construct of Interest

Guidance on selecting and defining the psychological construct to be measured. Emphasis on the importance of a well-defined and clear construct. Examples of psychological constructs in various domains, such as personality traits, attitudes, and mental health indicators.

Module 3: Generating and Refining Scale Items

Explanation of the process of generating potential scale items related to the chosen construct. Guidance on refining and enhancing these items through expert reviews and pilot testing. Guidelines for item wording, formatting, and response options.

Module 4: Content Validity and Item Selection

Introduction to content validity and its role in ensuring that scale items adequately represent the construct. Explanation of the process for assessing content validity, including expert judgment and content validity ratio (CVR). Discussion of item selection and strategies to reduce redundancy, resulting in a concise and focused scale.

Module 5: Validity Assessment

Exploration of different types of validity, including content, criterion, and construct validity. Discussion of techniques and methods for establishing each type of validity. Illustration of validity assessment through examples and case studies.

Module 6: Reliability Assessment

Definition of reliability within the context of psychological scale development. Explanation of methods to measure reliability, such as Cronbach's alpha and test-retest reliability. Emphasis on the importance of internal consistency and scale stability.

Module 7: Data Collection and Analysis

Detailed guidance on planning and conducting data collection for scale validation. Introduction to exploratory and confirmatory factor analysis for assessing construct validity. Discussion of item analysis and statistical techniques for refining the scale based on data.

Module 8: Pilot Testing and Feedback Integration

Explanation of the role of pilot testing in the scale development process. Description of the process to collect feedback from pilot participants and integrate it into scale refinement. Emphasis on the iterative nature of scale development and the value of feedback loops.

Module 9: Practical Applications and Case Studies

Presentation of real-world applications of psychological scale development in research and assessment. Sharing of case studies and examples in various psychological domains. Discussion of ethical considerations in scale development and usage.

Module 10: Future Directions and Emerging Trends

Insights into the evolving field of psychological scale development. Discussion of emerging trends and technologies, such as computerized adaptive testing and online survey platforms. Encouragement for participants to stay updated with the latest advancements in the field.

Conclusion

Summary of key takeaways from the training material. Encouragement for participants to apply their newfound knowledge in research or professional practice. Suggested additional resources, references, and further reading materials.

Assessment

Inclusion of quizzes or self-assessment exercises after each module to reinforce learning. Suggestion of a final project where participants can develop and validate a psychological scale of their own.

Introduction

Explanation of the significance of psychological scale development in research and assessment.

Highlighting the relevance of competence-oriented methodologies in scale development.

Outline of the training material's structure and objectives.

Introduction: Significance of Psychological Scale Development

Psychological scale development, a fundamental cornerstone of empirical research and assessment, exerts profound and far-reaching implications across an expansive spectrum of academic and professional domains. These domains encompass disciplines as diverse as psychology, education, and the broader social sciences. The creation and validation of psychological scales, characterized by meticulous construction and rigorous examination, stand as indispensable tools that empower researchers and practitioners alike. These scales, meticulously designed, systematically and quantitatively measure multifaceted and intricate psychological constructs, which extend their reach to encompass diverse domains such as personality traits, mental health indicators, attitudes, cognitive abilities, and a myriad of others.

Psychological scale development is the scientific pursuit of precision and rigor in the exploration of intricate psychological phenomena. It constitutes a deliberate and systematic approach to construct instruments that provide researchers with the means to investigate, comprehend, and quantify the complexities of human behavior and cognition. Moreover, these well-constructed instruments not only serve as critical components of the research process but also open up a powerful avenue for the comparative analysis of research findings across different studies. This analytical capability, in turn, enriches our capacity to evaluate a wide array of interventions, educational programs, and real-world applications.

The essence of psychological scale development lies in the profound significance of precision and rigor. It is instrumental in the scientific pursuit of understanding and quantifying complex psychological constructs. A cornerstone in the domains of psychology, education, and social sciences, the development and validation of psychological scales are integral to empirical research and assessment.

Psychological scales, or psychometric instruments, serve as precision tools that provide a systematic means to measure intricate psychological constructs. These constructs encompass a vast spectrum, including personality traits, mental health indicators, attitudes, cognitive abilities, and numerous other dimensions of human psychology. Through meticulous construction and rigorous validation, psychological scales enable researchers and practitioners to dissect these complex constructs into quantifiable components. This deconstruction of complexity into measurable units is a hallmark of psychological scale development, and it has far-reaching implications.

The meticulous construction of psychological scales is underpinned by rigorous examination. This examination involves the verification of the instruments' reliability, validity, and their capacity to provide precise and accurate measurements. The reliability of a scale relates to the consistency and stability of its measurements, while its validity assesses whether it truly measures what it is intended to measure. Rigorous examination ensures that the scales can be relied upon as valid and dependable tools in empirical research and assessment.

Furthermore, these well-constructed instruments, with their reliability and validity firmly established, play a pivotal role in advancing research and assessment practices. They empower researchers to delve into intricate psychological phenomena with confidence, enabling the measurement and understanding of human behavior, attitudes, and cognitive abilities with a high degree of precision. This precision not only enriches the depth of knowledge in these domains but also enhances the ability to make informed decisions based on empirical evidence.

The significance of psychological scale development extends well beyond the confines of academic research. These meticulously crafted instruments have a broad range of applications in real-world scenarios. They serve as the bedrock upon which data-driven decision-making is built, facilitating critical judgments across numerous fields.

In the realm of education, psychological scales underpin the assessment of student performance, allowing educators to gauge not just what students know but also their capacity to apply knowledge and skills in practical contexts. Psychological scales are essential tools in evaluating educational programs, ensuring that learning outcomes are achieved, and students are adequately prepared for the challenges they will encounter.

In clinical psychology, psychological scales are vital in diagnosing and monitoring mental health conditions. They provide quantifiable indicators of psychological well-being and enable healthcare professionals to tailor interventions to individual needs.

Psychological scales are equally integral in organizational psychology and human resources. They assist in assessing the competency and aptitude of employees, guiding recruitment and training efforts, and supporting workforce development.

Moreover, in marketing research, psychological scales are used to measure consumer attitudes and preferences, providing valuable insights for product development and marketing strategies.

In conclusion, psychological scale development is an indispensable and enduring component of empirical research and assessment. The meticulous construction and rigorous validation of these scales offer a systematic means to quantitatively measure complex psychological constructs, extending their influence across diverse academic and professional domains. Through precision and rigor, psychological scales empower researchers and practitioners to investigate intricate psychological phenomena, compare research findings across studies, and make informed decisions in fields as diverse as education, clinical psychology, organizational psychology, and marketing research. The significance of psychological scale development resonates in its capacity to contribute to the advancement of scientific knowledge and applied practices across various domains.

In the words of Hays and Revicki (2016), psychometric instruments are instrumental in enabling researchers to quantitatively measure and analyze intricate psychological constructs, thus enhancing the scientific rigor of empirical investigations. This contribution is particularly pertinent in an era characterized by data-driven decision-making, where the systematic measurement and assessment of psychological phenomena are vital in guiding critical judgments. Psychological scales provide the means to assess a myriad of complex traits and behaviors, and they play a central role in guiding crucial decisions, including the diagnosis of psychological disorders and the evaluation of the effectiveness of educational programs (Boerma et al., 2014).

The importance of psychological scale development cannot be overstated. These instruments, often referred to as psychometric instruments, are heralded as the very bedrock upon which modern psychological science is built. They provide the means to assess specific psychological traits or behaviors in individuals, facilitating the classification of individuals into distinct categories or groups based on their psychological characteristics. Additionally, these instruments enable the quantification of the extent or intensity of these traits or behaviors, allowing for a nuanced and data-driven understanding of complex psychological phenomena (Cacioppo, Tassinary, & Berntson, 2016).

While the significance of psychological scale development is evident, it is also important to consider the evolving landscape of competence-oriented methodologies in scale development.

The Relevance of Competence-Oriented Methodologies in Scale Development

While the primary focus of this discussion has revolved around the significance of psychological scale development, it is crucial to briefly highlight the relevance of competence-oriented methodologies in the contemporary landscape. Competence-oriented methodologies represent a modern approach to measurement, emphasizing the assessment of an individual's proficiency, aptitude, or mastery in executing specific tasks and demonstrating specific knowledge and skills.

This paradigm aligns seamlessly with outcome-based education and competence-based curricula within the field of education. Competence-based assessments advocate for the alignment of assessments with envisioned learning outcomes, thereby shifting the focus from mere rote memorization to the practical application of acquired knowledge and skills in authentic, real-world settings (Mayer & Wittrock, 2015).

The application of competence-oriented methodologies in psychological scale development is a notable trend in the field. These methodologies call for rigorous and systematic processes of scale development, including content validation, expert judgment, and the utilization of advanced statistical techniques (Reis & Judd, 2000). Competence-oriented methodologies offer the potential to enhance the precision of scale construction and, by extension, to augment the validity and reliability of psychological scales. This alignment with contemporary educational and psychological practices, which prioritize the assessment of applied knowledge and skills, underscores the growing relevance of competence-oriented methodologies within the field of psychological scale development.

Training material's structure

The training material is structured into ten comprehensive modules, each designed to provide participants with a deep understanding of psychological scale development, encompassing its principles, applications, and emerging trends. These modules are carefully curated to ensure a comprehensive and structured learning experience. Let's explore each module in more detail to provide a comprehensive overview of the training program.

Module 1: Understanding Psychological Scales

In Module 1, participants are introduced to the foundational concepts of psychological scales. They learn the definition of psychological scales and their critical role in psychological research and assessment. Understanding the significance of psychological scales is the cornerstone of this module. Participants are introduced to the various types of psychological scales, including Likert scales, Thurstone scales, and others. Through practical examples, they gain insights into how different types of scales serve specific research and assessment needs. This module also delves into the fundamental characteristics of psychological scales, particularly reliability and validity. Participants learn how these attributes are pivotal in ensuring that the scales are dependable and accurate measurement tools.

Module 2: Identifying the Construct of Interest

Module 2 emphasizes the importance of selecting and defining the psychological construct of interest. This early stage in scale development is critical, as a well-defined and clear construct forms the basis for the entire process. Through examples across various domains, such as personality traits, attitudes, and mental health indicators, participants gain a deeper understanding of how to articulate and specify the construct they aim to measure. The module provides guidance on how to select a construct that aligns with the research or assessment objectives, ensuring that it is meaningful and relevant to the intended application.

Module 3: Generating and Refining Scale Items

This module takes participants through the process of generating and refining scale items related to the chosen construct. It explores the initial stages of scale development, where researchers brainstorm and create potential items. Participants learn about the importance of item wording, ensuring clarity and relevance to the construct. The module also addresses formatting and response options, underscoring the significance of consistency and precision in item presentation. Additionally, it provides guidance on how to enhance these items through expert reviews and pilot testing, ensuring that they effectively capture the nuances of the construct.

Module 4: Content Validity and Item Selection

Module 4 introduces the concept of content validity, which plays a vital role in ensuring that scale items adequately represent the construct. Participants learn about the process of assessing content validity, including techniques like expert judgment and content validity ratio (CVR). This module also delves into the crucial aspect of item selection. Participants discover strategies to reduce redundancy, resulting in a concise and focused scale that effectively measures the desired construct. The module highlights the iterative nature of scale development, emphasizing the importance of revisiting and revising items to enhance content validity.

Module 5: Validity Assessment

Module 5 is dedicated to the exploration of different types of validity, including content, criterion, and construct validity. Participants gain insights into how each type of validity serves as a quality indicator for psychological scales. The module offers an in-depth discussion of techniques and methods for establishing each type of validity, along with real-world examples and case studies that illustrate the process. By the end of this module, participants are well-versed in the crucial role of validity assessment in ensuring that the scale accurately measures the intended construct.

Module 6: Reliability Assessment

Reliability assessment is the focus of Module 6. Participants delve into the definition of reliability within the context of psychological scale development. They gain an understanding of methods used to measure reliability, including Cronbach's alpha and test-retest reliability. The module emphasizes the importance of internal consistency and scale stability, providing participants with the tools to ensure that their scales produce consistent and dependable results. Through practical exercises and examples, participants learn how to assess and enhance the reliability of their scales.

Module 7: Data Collection and Analysis

Module 7 provides detailed guidance on planning and conducting data collection for scale validation. Participants learn about the various data collection methods, such as surveys and questionnaires, and gain insights into best practices for data gathering. The module also introduces exploratory and confirmatory factor analysis as powerful techniques for assessing construct validity. Participants discover how to use statistical methods to analyze scale data, refine scale items, and ensure that their scale measures the intended construct effectively. This module equips participants with the skills needed to transform collected data into actionable insights.

Module 8: Pilot Testing and Feedback Integration

Pilot testing plays a critical role in Module 8. Participants explore the purpose and process of pilot testing in the scale development journey. They learn how to collect feedback from pilot participants and integrate it into scale refinement. The module emphasizes the iterative nature of scale development, where feedback loops lead to continuous improvement. Participants understand the value of pilot testing in uncovering potential issues and refining the scale to make it more accurate and user-friendly.

Module 9: Practical Applications and Case Studies

Module 9 shifts the focus to practical applications and case studies. Participants are presented with real-world examples of how psychological scale development is applied in research and assessment. Case studies across various psychological domains offer insights into how scales are utilized in diverse contexts. This module also addresses ethical considerations in scale development and usage, providing participants with a comprehensive view of the ethical responsibilities that come with the development and application of psychological scales.

Module 10: Future Directions and Emerging Trends

In the final module, participants gain insights into the evolving field of psychological scale development. They explore emerging trends and technologies, such as computerized adaptive testing and online survey platforms, that are shaping the future of scale development. The module encourages participants to stay updated with the latest advancements in the field and adapt to the changing landscape. It reinforces the idea that scale development is a dynamic field with ongoing developments and innovations.

Conclusion

The training program concludes with a summary of key takeaways from the entire material. Participants are encouraged to apply their newfound knowledge in research or professional practice. The module also suggests additional resources, references, and further reading materials for those who wish to deepen their understanding of psychological scale development.

Assessment

Throughout the training program, participants have opportunities for assessment. Quizzes or self-assessment exercises are provided after each module to reinforce learning and evaluate comprehension. Additionally, a final project option is presented, allowing participants to apply their knowledge by developing and validating a psychological scale of their own, applying the principles and techniques they've learned throughout the training.

This structured training material not only equips participants with a comprehensive understanding of psychological scale development but also empowers them to apply this knowledge effectively in their research and professional endeavors. The modular approach ensures a step-by-step and thorough exploration of the topic, making it accessible and practical for participants at all levels of expertise.

Module 1: Understanding Psychological Scales

Definition of psychological scales and their role in psychological research.

Distinction between various types of psychological scales, including Likert scales, Thurstone scales, and others.

Discussion of key characteristics of psychological scales, including reliability and validity.

Understanding Psychological Scales

Psychological scale development is a multifaceted and pivotal process that lies at the heart of various research endeavors and practical applications in psychology, education, and the social sciences. In this first module, we embark on a journey to comprehend the core concepts of psychological scales, their essential role in psychological research, and the differentiation between various types of scales. We will focus on exploring the definition of psychological scales and their vital function in psychological research.

Psychological scales, often referred to as psychometric instruments, are fundamental tools in the realm of psychological research. They serve as a structured approach to transform abstract psychological phenomena into concrete, quantifiable data. These phenomena can encompass a vast array of human experiences, emotions, and behaviors, including personality traits, attitudes, cognitive abilities, and mental health indicators. The primary purpose of psychological scales is to assign numerical values to these abstract constructs, facilitating systematic exploration, statistical analysis, and data-driven decision-making.

The significance of psychological scales becomes evident when we consider the intricate nature of psychological constructs. Human behavior, cognition, and emotion are multifaceted and often elusive to direct measurement. Without the aid of psychological scales, understanding, quantifying, and conducting empirical research on these phenomena would be a formidable task. Scales provide a structured framework to capture the nuances of these constructs in a manner that allows for rigorous analysis and comparison across individuals, groups, and contexts.

Psychological research seeks to delve into the complexities of human behavior, cognition, and emotion. It aims to answer fundamental questions about how individuals think, feel, and act, both individually and within the context of social interactions. At the heart of this research endeavor, psychological scales play a pivotal role, acting as a bridge between the abstract constructs under investigation and the empirical data that researchers collect, analyze, and utilize to draw conclusions.

The role of psychological scales in psychological research is multifaceted and indispensable:

- *Quantifying Psychological Constructs: Psychological scales provide a systematic and quantifiable means of measuring psychological constructs. For instance, a researcher examining self-esteem can employ a scale to assign a numerical value to an individual's self-esteem level, facilitating precise assessment and comparison.*
- *Measuring Change Over Time: Psychological scales enable researchers to track changes in psychological constructs over time. This feature is particularly valuable in longitudinal studies aimed at monitoring developments in attitudes, cognitive abilities, or well-being.*
- *Enabling Comparisons: Scales offer the ability to compare individuals, groups, or contexts with respect to a specific psychological construct. Researchers can assess whether one group exhibits significantly different levels of a psychological trait compared to another group or evaluate the impact of an intervention on a particular construct.*
- *Empowering Data-Driven Decision-Making: In an era marked by data-driven decision-making, psychological scales guide critical judgments. For example, in clinical psychology, these scales are indispensable for diagnosing mental health conditions. In the field of education, they inform decisions about the effectiveness of educational programs and interventions.*
- *Enhancing Scientific Rigor: The use of psychological scales enhances the scientific rigor of psychological research. It enables the replicability of findings as researchers in different settings can use the same scale to measure the same construct. This fosters cumulative knowledge and contributes to the robustness of psychological science.*

The significance of psychological scale development becomes particularly evident when considering the diverse fields in which they are applied. Whether in clinical psychology, educational psychology, marketing research, or organizational psychology, the creation and utilization of well-constructed scales play a pivotal role in advancing knowledge, improving decision-making, and enhancing our understanding of human behavior and cognition.

Psychological scale development is a complex process that demands careful consideration, precision, and a deep understanding of both the construct being measured and the principles of scale development. Over the course of this training, we will delve deeper into the intricacies of this process, addressing topics such as reliability, validity, and the distinct types of scales used in research.

In the subsequent sections of this module, we will explore the differentiation between various types of psychological scales, emphasizing the role each type plays in capturing psychological constructs. We will also delve into key characteristics of psychological scales, including reliability and validity, which are vital for ensuring the accuracy and trustworthiness of the measurements derived from these instruments.

Definition of Psychological Scales and Their Role in Psychological Research

The Essence of Psychological Scales

Psychological scales, also known as psychometric instruments, are fundamental tools in psychological research and assessment (DeVellis, 2016). They play a critical role in systematically and quantitatively measuring complex psychological constructs, which encompass a wide spectrum of human experiences, emotions, and behaviors (Streiner, Norman & Cairney, 2015). At their core, psychological scales serve as a means to transform abstract psychological phenomena into concrete, numerical data.

The primary function of psychological scales is to facilitate the measurement of aspects of human cognition, emotion, and behavior that are often intangible and challenging to quantify. Human psychology is marked by intricate and multifaceted constructs, and these constructs underpin the essence of psychological research. Whether the objective is to assess personality traits, mental health indicators, attitudes, or cognitive abilities, scales provide researchers with a structured framework to assign numerical values to these abstract concepts.

The transformation from abstract psychological phenomena to numerical data serves several essential purposes:

- *Systematic Exploration: Psychological scales enable the systematic exploration of psychological constructs. Researchers can precisely define what they intend to measure and apply this definition consistently.*
- *Statistical Analysis: The numerical nature of the data collected using psychological scales allows for rigorous statistical analysis. Researchers can employ various statistical techniques to derive meaningful insights from the data.*
- *Data-Driven Decision-Making: In an era characterized by data-driven decision-making, psychological scales provide the foundation for making informed judgments. These judgments can range from diagnosing psychological disorders to evaluating the effectiveness of interventions or educational programs (Hays & Revicki, 2016).*
- *Comparative Analysis: Scales facilitate comparative analysis across different studies, individuals, or groups. Researchers can determine whether one group exhibits significantly different levels of a psychological trait compared to another group or assess the impact of an intervention on a particular construct.*
- *Enhanced Replicability: The utilization of psychological scales enhances the replicability of findings. Researchers in different settings can employ the same scale to measure the same construct, thus contributing to the cumulative knowledge in the field.*

The significance of psychological scale development becomes especially evident when we consider the intricate and often elusive nature of psychological constructs. Human behavior, cognition, and emotion are multifaceted, and measuring them directly can be challenging. Psychological scales offer a structured approach to capture the nuances of these constructs in a manner that allows for rigorous analysis and comparison.

In essence, psychological scales serve as a bridge between the abstract constructs under investigation and the empirical data that researchers collect, analyze, and use to draw conclusions. They provide a means to assign numerical values to the abstract, making it possible to study, understand, and quantify intricate aspects of human psychology.

Psychological scale development is a multifaceted process that requires careful consideration and precision. Over the course of this training, we will delve deeper into the intricacies of this process, addressing topics such as reliability, validity, and the distinct types of scales used in research. We will explore how to create, validate, and use psychological scales effectively in various research and assessment contexts.

As we move forward in this module, we will further explore the differentiation between various types of psychological scales, emphasizing the unique roles each type plays in capturing psychological constructs. Additionally, we will delve into key characteristics of psychological scales, including reliability and validity, which are vital for ensuring the accuracy and trustworthiness of the measurements derived from these instruments.

The Role of Psychological Scales in Psychological Research

Psychological research is a multifaceted and dynamic field that seeks to unravel the mysteries of human behavior, cognition, and emotion. It is a discipline dedicated to understanding the intricacies of how individuals think, feel, and act, both as unique entities and within the context of social interactions. At the core of this research endeavor, psychological scales play a pivotal role, acting as a bridge between the abstract constructs under investigation and the empirical data that researchers collect, analyze, and utilize to draw meaningful conclusions.

Psychological scales are invaluable instruments in psychological research for several compelling reasons. First and foremost, they provide a systematic means of quantifying complex psychological constructs. These constructs encompass a wide array of human experiences, emotions, and behaviors, from personality traits to attitudes, cognitive abilities, and mental health indicators. For instance, a researcher examining self-esteem can employ a scale to assign a numerical value to an individual's self-esteem level, thus enabling a precise assessment and comparison across individuals or groups (Cronbach, 1951).

Furthermore, psychological scales offer researchers the ability to track changes in psychological constructs over time. This feature is particularly vital in longitudinal studies that seek to monitor the evolution of attitudes, cognitive abilities, or well-being over extended periods. Through the use of scales, researchers can assess not only the existence of change but also the magnitude and direction of change in these constructs, providing valuable insights into developmental processes and intervention outcomes (DeVellis, 2016).

In addition to enabling within-subject analyses over time, psychological scales facilitate comparisons across different individuals, groups, or contexts. Researchers can employ these scales to assess whether one group exhibits significantly different levels of a particular psychological trait in comparison to another group. This capability is especially valuable in experimental research, where interventions or treatments can be evaluated in terms of their impact on specific psychological constructs (Streiner & Norman, 2008).

In an era characterized by data-driven decision-making, psychological scales play an indispensable role in guiding critical judgments across a range of domains. In the realm of clinical psychology, for instance, these scales are essential for diagnosing and monitoring mental health conditions. They provide clinicians with a quantifiable and standardized means of assessing psychological well-being and functioning, thus informing the development of treatment plans and therapeutic interventions. Similarly, in educational psychology, psychological scales are instrumental in evaluating the effectiveness of educational programs, curricula, and instructional methods. By quantifying student outcomes and attitudes, educators can make data-informed decisions to enhance the learning experience (Hays & Revicki, 2016).

Moreover, the use of psychological scales enhances the scientific rigor of psychological research. It ensures the replicability of findings, as researchers in different settings can employ the same scale to measure the same construct. This not only fosters the cumulative growth of knowledge within the field but also contributes to the robustness and validity of psychological science. Rigor in measurement is a fundamental aspect of scientific inquiry, and psychological scales provide the means to achieve it, reinforcing the reliability and trustworthiness of research findings (DeVellis, 2016).

Psychological scale development is a complex and meticulous process that aims to create measurement tools capable of withstanding the scrutiny of scientific inquiry. This process is of paramount importance, as the quality and accuracy of the scales directly influence the reliability and validity of research outcomes. In the subsequent sections of this training, we will delve deeper into the intricacies of psychological scale development, addressing key aspects such as reliability, validity, and the various types of scales used in research. These topics are essential for ensuring that the measurements derived from psychological scales are not only accurate but also meaningful and robust, thus upholding the standards of scientific research.

Distinction Between Various Types of Psychological Scales

Likert Scales: Measuring Intensity and Agreement

Likert scales, named after their creator Rensis Likert, stand as one of the most commonly used and versatile types of psychological scales. These scales are designed to measure the intensity or strength of agreement or disagreement with a statement or a series of statements, allowing respondents to express their attitudes, opinions, and self-perceptions on a numerical scale. Likert scales offer a practical and straightforward means of data collection and have found widespread use in fields such as psychology, education, marketing, and more. In this section, we will explore the essence of Likert scales, their applications, and their significance in psychological research.

A classic Likert item typically presents respondents with a series of statements or questions, each accompanied by a numerical scale representing degrees of agreement or disagreement. Respondents are then asked to indicate their level of agreement or disagreement with each statement by selecting the appropriate point on the scale. The responses to these items are subsequently summed to create a score that reflects the respondent's position on the measured trait or construct.

For instance, consider the following Likert item: "On a scale from 1 to 5, please indicate how much you agree with the following statement: 'I am confident in my problem-solving abilities.' In this example, respondents can choose from a range of options, with 1 signifying "Strongly Disagree" and 5 indicating "Strongly Agree." By examining the collective responses to items like these, researchers can gain insights into individuals' perceptions and attitudes regarding specific topics or constructs.

The simplicity and flexibility of Likert scales are key reasons for their ubiquity in psychological research. They offer a practical and efficient way to capture subjective data, making them well-suited for assessing attitudes, opinions, and self-perceptions. Likert items can be readily adapted to various contexts and domains, making them a popular choice for researchers across diverse fields.

The applications of Likert scales extend far beyond their use in traditional research settings. They are commonly employed in fields such as psychology, education, marketing, and healthcare to measure a wide range of constructs. For example, in psychology, Likert scales are used to assess individuals' self-esteem, anxiety levels, or perceived social support. In education, they are utilized to evaluate teaching effectiveness or student satisfaction with educational programs. In marketing, Likert scales help gauge customer satisfaction, preferences, and brand loyalty. Their adaptability and ease of use make Likert scales a valuable tool in both research and practical applications (Likert, 1932).

It is essential to recognize that while Likert scales are a valuable resource, their design and implementation require careful consideration to ensure the validity and reliability of measurements. Researchers must pay attention to the wording and phrasing of items to avoid response bias, acquiescence bias, or other response artifacts that may influence the results (Krosnick & Presser, 2010). Moreover, the appropriate choice of response options (e.g., a 5-point scale, 7-point scale, or other variations) can impact the quality of the data collected. Proper analysis and interpretation of Likert scale data are equally crucial, involving statistical techniques like factor analysis to assess scale validity and reliability.

In summary, Likert scales, introduced by Rensis Likert, are a widely used type of psychological scale designed to measure the intensity or strength of agreement or disagreement with statements. They offer a practical and adaptable means of capturing attitudes, opinions, and self-perceptions. The simplicity of Likert items makes them a popular choice for researchers across various fields, including psychology, education, marketing, and healthcare. However, their design and application require careful attention to wording and response options to ensure the validity and reliability of measurements.

Thurstone Scales: Expertly Weighted Opinions

Thurstone scales, developed by Louis Leon Thurstone, represent a distinctive approach to psychological scale construction, distinct from the more straightforward Likert scales. Thurstone scales entail a more intricate process, incorporating a panel of experts who play a pivotal role in evaluating a set of items, typically statements, based on their relevance and representativeness of the construct under examination. The items that achieve consensus among the panel of judges are assigned higher weights, signifying their greater importance in capturing the essence of the construct. Conversely, items that fail to attain consensus receive lower weights. Respondents are then tasked with rating each of these items, and the final score is computed based on the weighted averages of their responses.

Thurstone scales, though less commonly used than Likert scales, offer a unique and valuable method of psychological measurement. They are particularly well-suited for situations where expert judgment is critical in defining the construct to be measured. This section delves into the intricacies of Thurstone scales, their distinctive features, and their applications in psychological research.

The process of constructing Thurstone scales is fundamentally different from Likert scales. It revolves around the involvement of a panel of experts who are well-versed in the domain or construct being measured. These experts play a central role in the scale development process by evaluating a pool of items to be included in the scale. The items typically consist of statements or questions that aim to capture various facets of the construct. The experts are tasked with assessing the relevance and representativeness of each item to the construct in question.

Items that achieve a consensus among the expert panel are deemed to be highly relevant and representative of the construct. These items are assigned higher weights, indicating their greater significance in capturing the essence of the trait or attribute under investigation. In contrast, items that do not receive consensus are assigned lower weights, as they are considered less central to the construct.

The next phase of the Thurstone scale construction process involves gathering responses from a sample of individuals who represent the target population. These respondents are asked to rate each of the items based on their personal opinions or experiences related to the construct. The rating is typically done on a numerical scale, such as a 5-point or 7-point scale, with higher values indicating stronger agreement with the statement.

The final score for each respondent is computed by taking a weighted average of their responses to the individual items. The weights assigned to the items are determined by the expert panel's judgment, reflecting the relative importance of each item in capturing the construct. The resulting score provides a quantitative representation of an individual's position on the trait or attribute under examination.

Thurstone scales are recognized for their ability to provide nuanced and context-specific measurements. This is particularly valuable when dealing with complex constructs or attributes that are difficult to assess using simple, one-dimensional Likert scales. By involving experts in the scale construction process, Thurstone scales ensure that the items selected are not only relevant but also reflect the multidimensional nature of the construct.

While Thurstone scales offer a robust approach to psychological measurement, they are less commonly used than Likert scales. This is primarily due to the more labor-intensive and resource-intensive nature of the scale construction process. Gathering a panel of experts and ensuring consensus among them can be time-consuming and costly. Additionally, the reliance on expert judgment may introduce subjectivity into the process, potentially affecting the objectivity of the measurements (Carmines & Zeller, 1979).

However, Thurstone scales shine in situations where expert consensus is paramount. They are particularly useful in cases where there is no universally accepted or standardized method for measuring a complex construct. In such instances, Thurstone scales provide a structured yet flexible approach to capturing the nuances of the attribute under study.

In summary, Thurstone scales, developed by Louis Leon Thurstone, offer a distinct and valuable approach to psychological measurement. These scales involve a panel of experts who play a central role in evaluating and weighting items based on their relevance and representativeness of the construct being measured. Thurstone scales provide a nuanced and context-specific means of measurement, making them especially useful in situations where expert judgment is essential for defining the construct.

Fuzzy rating Scales

Fuzzy rating scales represent a distinct and innovative approach to measurement in psychological research and assessment. Unlike traditional Likert or Thurstone scales, fuzzy rating scales offer respondents the flexibility to express their opinions or attitudes using a degree of membership to multiple categories or options, rather than a precise numeric value. In this section, we will explore the concept of fuzzy rating scales, their unique features, and their applications in psychological research, underpinned by relevant literature and citations.

Fuzzy rating scales, also referred to as fuzzy sets, derive their name from the principle of "fuzziness." This notion acknowledges that not all human experiences, opinions, or attitudes can be neatly categorized into discrete values or options. In the traditional Likert scale, for example, respondents are asked to choose from a finite set of options, such as "Strongly Disagree," "Disagree," "Neutral," "Agree," or "Strongly Agree," and assign a single numeric value to their response. Fuzzy rating scales, in contrast, recognize that human experiences and perceptions often exhibit degrees of ambiguity or overlap.

At the core of fuzzy rating scales is the concept of fuzzy logic, a mathematical framework that deals with imprecise or uncertain information. Fuzzy logic allows individuals to express their opinions or attitudes using degrees of membership in different categories. For instance, when asked about their level of agreement with a statement, respondents can assign partial membership to multiple categories simultaneously, indicating that their response may exhibit characteristics of both "Agree" and "Neutral."

The key advantage of fuzzy rating scales is their ability to capture the richness and complexity of human experiences and opinions. They provide a more nuanced and accurate representation of respondents' views by allowing them to express the degree to which they align with various response categories. This approach is particularly beneficial when dealing with abstract or multifaceted constructs that do not conform to a rigid, one-dimensional scale.

Fuzzy rating scales have found applications in a variety of psychological research domains. For example, in the field of linguistics and semantics, fuzzy logic has been used to model the vagueness and imprecision in natural language expressions. In educational assessment, fuzzy rating scales have been employed to evaluate students' competencies in a way that accounts for the gradual acquisition of skills rather than fixed, binary outcomes. In clinical psychology, these scales have been used to assess the intensity of symptoms in psychiatric disorders, acknowledging that mental health conditions often exhibit gradations in severity (Zadeh, 1973).

The use of fuzzy rating scales has been praised for its ability to capture the complexity and subtlety of human experiences. In a study by Dubois and Prade (1991), the authors highlighted the relevance of fuzzy logic in situations where human judgment involves making decisions under uncertainty or dealing with imprecise information. They argued that fuzzy logic enables decision-makers to incorporate both quantitative and qualitative assessments, resulting in more informed and contextually appropriate decisions.

In another study by Klir and Yuan (1996), the authors discussed the advantages of fuzzy logic in modeling complex systems, particularly when inputs are imprecise or ambiguous. They emphasized that fuzzy logic provides a formal framework for representing and processing vague information, allowing for a more realistic representation of complex phenomena in various domains, including psychology.

Despite the advantages of fuzzy rating scales, they are not without challenges. One key issue is the need for clear guidelines and training for respondents to effectively use these scales. Respondents may require a deeper understanding of how to assign degrees of membership, and researchers must provide guidance to ensure the meaningful interpretation of responses. Additionally, data analysis with fuzzy rating scales requires specific techniques, which may be less familiar to researchers accustomed to traditional measurement methods.

In conclusion, fuzzy rating scales offer a novel and flexible approach to psychological measurement, allowing respondents to express their opinions and attitudes using degrees of membership in multiple response categories. This approach recognizes the inherent ambiguity and complexity of human experiences, making it especially relevant for assessing abstract or multifaceted constructs. While fuzzy rating scales may require more guidance for respondents and specialized data analysis techniques, they have shown promise in fields where imprecise or uncertain information plays a significant role.

Other Types of Scales

In the realm of psychological measurement, a diverse array of scales is available to cater to specific research objectives, the nature of the construct being examined, and practical considerations. While Likert and Thurstone scales are among the most common, several other scale types offer unique advantages and are tailored to different research contexts. This section explores some of these alternative scale types, including Semantic Differential Scales, Visual Analog Scales (VAS), and Guttman Scales, delving into their applications and significance in psychological research, complemented by relevant literature citations.

Semantic Differential Scales: Semantic Differential Scales are a specialized form of psychological measurement that focus on assessing the connotations or emotional associations of a concept or item. Respondents are presented with pairs of bipolar adjectives, such as "happy" to "sad" or "friendly" to "unfriendly," and are asked to place a mark along a continuum to indicate their perception or evaluation of the concept. The choice of adjectives in these scales is variable, adapting to the specific construct under investigation.

Semantic Differential Scales offer a unique insight into the emotional or evaluative dimensions associated with a concept (Osgood, Suci, & Tannenbaum, 1957).

For example, in a study assessing public opinion on a political candidate, a Semantic Differential Scale might include pairs of adjectives such as "trustworthy" to "untrustworthy," "competent" to "incompetent," and "likable" to "unlikable." Respondents would then indicate where on the scale their perception of the candidate falls for each adjective pair. The resulting data could reveal the emotional and evaluative components of the public's view of the candidate.

Visual Analog Scales (VAS): Visual Analog Scales (VAS) represent another alternative to traditional Likert or Thurstone scales. They offer a continuous line or scale, often represented visually, where respondents are instructed to mark their position to indicate their response. The unique aspect of VAS is that they allow for finer granularity of measurement by presenting a continuous spectrum.

A common application of VAS is in the measurement of subjective experiences, such as pain intensity, mood, or satisfaction. In a clinical setting, patients might be asked to mark their level of pain on a VAS ranging from "no pain" to "worst imaginable pain." By placing a mark on the line, patients provide a quantitative measure of their pain intensity.

The advantages of VAS lie in their sensitivity to subtle differences, making them particularly valuable when dealing with constructs that involve gradations or nuances. This approach allows for more precise measurements, which can be especially relevant in clinical assessments or when assessing subjective experiences (Huskisson, 1974).

Guttman Scales: Guttman Scales, also known as cumulative scales, represent a unique approach to psychological measurement. They consist of a set of items or statements ordered hierarchically by difficulty or intensity. Respondents are required to endorse an item if they endorse all the preceding items in the scale. In essence, if a respondent agrees with a particular statement, it implies agreement with all previous statements in the scale, creating a cumulative measure of the construct.

Guttman Scales are particularly useful when measuring hierarchically organized constructs, where the endorsement of one item implies the endorsement of simpler or less intense items. These scales are often used in educational assessments, where learning is seen as a cumulative process. For example, in assessing mathematical knowledge, respondents might be presented with a series of mathematical problems, each building on the complexity of the previous one. By successfully solving a problem, they demonstrate their mastery of all preceding problems in the scale (Guttman, 1950).

The Choice of Scale Type: The selection of an appropriate scale type is a critical decision in the design of psychological research. Researchers must consider the nature of the construct, the specific research objectives, and practical considerations when choosing a scale. Each scale type has its unique strengths and limitations, making it essential to match the scale with the nuances of the construct.

For instance, Likert scales are well-suited for measuring straightforward attitudes or opinions, while Thurstone scales offer a more nuanced approach that can capture multidimensional constructs. Semantic Differential Scales are invaluable for exploring the emotional dimensions of concepts, VAS provide fine-grained measurements of subjective experiences, and Guttman Scales are ideal for hierarchically organized constructs.

It is also important to consider the ease of administration, respondent understanding, and data analysis requirements when selecting a scale. In some cases, researchers may opt for a combination of scales or use a scale that integrates elements from various types, depending on the research needs (Sarstedt, Ringle & Hair, 2021).

In summary, psychological research benefits from a diverse array of scale types beyond Likert and Thurstone scales. Semantic Differential Scales, Visual Analog Scales, and Guttman Scales offer unique advantages in measuring emotional associations, capturing nuanced experiences, and assessing hierarchically organized constructs, respectively. Researchers must carefully select the most suitable scale type based on the nature of the construct and the research objectives to ensure the accuracy and relevance of their measurements.

Module 2: Identifying the Construct of Interest

Guidance on selecting and defining the psychological construct to be measured.

Emphasis on the importance of a well-defined and clear construct.

Examples of psychological constructs in various domains, such as personality traits, attitudes, and mental health indicators.

Selecting and Defining the Psychological Construct of Interest

The realm of psychological research is built upon a solid foundation of well-defined constructs. These constructs, which encompass a diverse array of human behaviors, emotions, and cognitive processes, serve as the bedrock upon which researchers design studies, collect data, and draw conclusions. The process of selecting and defining these constructs is a fundamental and indispensable step that guides the entire research journey.

In psychological research, the journey begins with a fundamental and critical step: identifying and defining the psychological construct of interest. This process involves selecting the specific aspect of human behavior, cognition, or emotion that will be the focus of investigation. The construct serves as the cornerstone upon which the entire research endeavor is built, influencing the research questions, study design, data collection, and analysis. In this module, we delve into the intricacies of selecting and defining the psychological construct, emphasizing the significance of a well-defined and clear construct. We also explore examples of psychological constructs across various domains, from personality traits and attitudes to mental health indicators, drawing insights from published literature to guide our understanding.

The Significance of Selecting the Right Construct

At the heart of psychological research lies the notion of constructs—abstract concepts or ideas that researchers seek to understand and measure. Constructs represent aspects of human psychology that are not directly observable but can be inferred from observable behaviors, responses, and experiences. In essence, they serve as the bridge between the intricacies of the human mind and the empirical data collected in research.

Constructs can take on a multitude of forms, from personality traits and attitudes to cognitive abilities and mental health indicators. However, regardless of the specific construct under investigation, it is essential to approach construct selection and definition with precision and clarity.

Selecting the right psychological construct is paramount to the success of a research project. It influences the relevance of the study, the accuracy of the results, and the applicability of the findings in real-world contexts. The construct serves as the bridge that connects the abstract ideas within the researcher's mind to tangible, measurable phenomena in the external world. As such, it is essential to make informed decisions regarding the construct under investigation.

In a study by DeVellis (2003), the author highlights the importance of construct selection in scale development. He emphasizes that the chosen construct should align with the research question and objectives, ensuring that the scale measures what it is intended to measure. Misalignment between the construct and research goals can lead to irrelevant or misleading findings.

Why is it so vital to select and define constructs with care and precision? The reasons are manifold and extend to the very core of scientific inquiry:

- *Relevance: The construct should align with the research question, objectives, and the broader context of the study. An irrelevant or poorly defined construct can lead to misguided research efforts (Schmitt, 1996).*
- *Measurement Quality: A well-defined construct paves the way for the development of valid and reliable measurement tools. Clarity in construct definition ensures that researchers and respondents share a common understanding of what is being measured (Clark & Watson, 2015).*
- *Applicability: The findings of a study should be applicable to real-world situations or contribute to the advancement of knowledge. A well-defined construct increases the likelihood that research outcomes are practical and relevant (Smith & Glass, 1977).*

The significance of a well-defined and lucid construct cannot be overstated. It is the foundation upon which the entire research endeavor is built, influencing every aspect of the research process. For instance, in the study of anxiety, the construct's precise definition determines whether the research will focus on generalized anxiety, social anxiety, or a specific type of anxiety disorder.

The importance of well-defined constructs reverberates throughout the research process, and this begins with construct selection and definition. Precision in defining the construct is not a mere formality but a prerequisite for rigorous and meaningful research. Here's why:

1. Clarity and Consistency: A well-defined construct ensures that all researchers, both within a study and across different studies, share a common understanding of what is being measured. This clarity enables consistency in research design, data collection, and interpretation. Without a clear construct definition, the data collected, and conclusions drawn become muddled, diminishing the research's impact and credibility.

2. Measurement Validity and Reliability: Precisely defined constructs are the pillars upon which valid and reliable measurement tools are constructed. Measurement validity refers to the extent to which an instrument accurately measures what it is intended to measure. Measurement reliability is the instrument's consistency in producing the same results when applied repeatedly to the same phenomena. A well-defined construct serves as the benchmark for creating measurement tools that accurately and consistently capture the psychological phenomena of interest.

3. Research Relevance: Well-defined constructs ensure that research is relevant and applicable to the real world. Whether the goal is to contribute to scientific knowledge or inform practical applications, a precise construct is paramount. For example, in educational psychology, research findings on well-defined constructs like "self-efficacy" have direct implications for designing effective educational interventions.

4. Avoiding Ambiguity: Vague or ambiguous constructs can lead to confusion and misinterpretation of research findings. Without a clear definition, researchers and readers may have different interpretations of the study's implications, hindering the advancement of knowledge in the field.

Defining the Construct

Once the construct is selected, it is essential to define it clearly and precisely. A well-defined construct ensures that researchers, respondents, and other stakeholders share a common understanding of what is being measured. A lack of clarity in construct definition can lead to confusion, misinterpretation, and measurement error.

In their work on psychological assessment, Cohen and Swerdlik (2017) stress the importance of construct definition. They explain that a clear and concise definition provides a foundation for the development of valid and reliable measures. Researchers should articulate the construct in terms of its theoretical underpinnings, observable indicators, and relevance to the study.

Defining the construct of interest is a foundational step in psychological research, and it plays a pivotal role in shaping the trajectory and quality of the study. The precision and clarity of the construct's definition hold the key to a successful research endeavor, ensuring that all stakeholders, from researchers to respondents, share a common understanding of the phenomenon under investigation.

In psychological research, the construct definition serves as the North Star, guiding the way in which the research is designed, conducted, and interpreted. It impacts several critical aspects of the research process:

1. Measurement Development: A well-defined construct forms the basis for developing measurement tools. Without a precise definition, it would be challenging to create instruments that accurately capture the construct. Cohen and Swerdlik (2017) emphasize that construct definition is the starting point for the creation of valid and reliable measures.

2. Data Collection: Clear construct definition influences how data is collected. It determines what aspects of the construct are assessed and what indicators are relevant. For example, if the construct under investigation is "job satisfaction," a well-defined construct would specify the key elements, such as satisfaction with workload, work environment, and compensation.

3. Analysis and Interpretation: During data analysis, a well-defined construct allows for the meaningful interpretation of results. It ensures that the findings align with the construct's theoretical underpinnings, avoiding confusion and misinterpretation. For instance, in a study on "self-esteem," a clear definition prevents the misinterpretation of results as indicators of "self-confidence."

4. Comparability and Replicability: Clear construct definitions enable comparability and replicability of research. Researchers working on the same construct can use a common framework, making it easier to compare findings and replicate studies. This fosters cumulative knowledge in the field, which is essential for scientific progress.

Cohen and Swerdlik (2017) emphasize the pivotal role of construct definition in psychological assessment. They highlight the importance of articulating the construct in terms of its theoretical foundations, observable indicators, and relevance to the study. This comprehensive approach ensures that the construct's essence is captured accurately.

Theoretical Foundations: A well-defined construct should be situated within its theoretical context. This involves explaining the theoretical framework or model from which the construct emerges. For example, in studying "empathy," the definition might reference social psychology theories that underpin the construct.

Observable Indicators: The definition should specify the observable indicators or components of the construct. It provides a breakdown of what elements constitute the construct. In the case of "intelligence," the definition would detail components like problem-solving ability, memory, and learning capacity.

Relevance to the Study: The construct definition must connect the construct to the specific study at hand. It should explain why the construct is relevant and how it fits within the research context. If the study examines "stress in the workplace," the definition would justify why this construct is vital in the context of organizational psychology.

In psychological research, the precise definition of the construct of interest is not a mere formality but a fundamental and non-negotiable step. It serves as the cornerstone upon which the entire research endeavor is built, influencing the research questions, study design, data collection, and analysis. Clear construct definition ensures that research is rigorous, meaningful, and contributes to the body of knowledge in the field.

Examples of Psychological Constructs

In the realm of psychological research, the selection and definition of a psychological construct are central to the research process. To shed light on this crucial aspect, we'll delve into examples across various domains within psychology and the social sciences. These examples serve as case studies, highlighting the intricacies of selecting, defining, and operationalizing psychological constructs.

1. Personality Traits: The Complex Tapestry of Human Nature

Personality traits are fundamental psychological constructs that have intrigued researchers for decades. These enduring characteristics influence how individuals perceive, interact with, and adapt to the world around them. Let's consider the example of the "Big Five" personality traits: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience (John & Srivastava, 1999).

Selecting the Construct: In research focused on personality traits, the choice of the construct often begins with an overarching trait, such as extraversion.

Defining the Construct: Defining extraversion involves specifying the dimension, which may encompass attributes like sociability, assertiveness, and enthusiasm. The definition must also outline the observable behavioral indicators associated with varying levels of extraversion. For instance, a well-defined extraversion construct might encompass behavioral indicators like seeking social interactions, displaying talkative behavior, and showing enthusiasm in group settings.

Research Implications: The clear definition of personality traits is essential for conducting meaningful research, whether it's investigating the influence of personality on job performance, relationship satisfaction, or coping with stress. Precise definitions help in selecting or developing appropriate measurement tools, such as questionnaires or inventories, to assess personality traits accurately.

2. Attitudes: Evaluating the World Around Us

Attitudes are psychological constructs that reflect individuals' evaluations of people, objects, or ideas. They underpin our perceptions and guide our behaviors. Let's delve into the example of attitudes towards climate change, a topic of paramount importance in contemporary society (Ajzen & Fishbein, 2005).

Selecting the Construct: Research on attitudes towards climate change begins with the selection of the attitude construct.

Defining the Construct: The definition of this construct involves specifying several key aspects:

The Target: What is the focus of the attitude? In this case, it's climate change.

Valence: Is the attitude positive (supportive of action on climate change) or negative (skeptical or resistant to action)?

Strength or Intensity: How strongly do individuals hold this attitude? Is it a mild concern or a deeply ingrained belief?

Research Implications: The precise definition of attitudes towards climate change is vital for designing effective communication strategies, interventions, and policies. Clear definitions ensure that survey questions accurately capture the intended aspects of these attitudes. Without clarity, the measurement might not align with the construct's underlying dimensions.

3. Mental Health Indicators: Navigating the Depths of Well-Being

Mental health indicators are constructs of utmost significance in clinical and psychological research. Conditions like depression, anxiety, and well-being impact the lives of countless individuals. Defining these constructs is a complex and crucial task (American Psychiatric Association, 2013).

Selecting the Construct: In clinical research, the construct choice is often guided by the specific mental health condition under investigation, such as depression.

Defining the Construct: Defining depression involves specifying key elements:

Symptoms: What are the symptoms of depression? These may include persistent low mood, loss of interest in activities, changes in sleep and appetite, and feelings of hopelessness.

Duration: How long should these symptoms persist for a diagnosis?

Severity: What level of impairment or distress must individuals experience to meet the criteria for depression?

Research Implications: Precise definitions of mental health indicators are the foundation of accurate diagnosis and effective treatment. They allow clinicians and researchers to distinguish between clinical conditions and normal variations in emotional states. When definitions lack clarity, misdiagnosis and inappropriate interventions can occur.

To navigate the complexities of selecting and defining psychological constructs, researchers often turn to existing literature. Insights and definitions provided by experts in the field, as exemplified in the references above, guide the process. For example, the well-established criteria for diagnosing mental health conditions, such as depression, are outlined in diagnostic manuals like the DSM-5 (American Psychiatric Association, 2013).

The process of selecting and defining a psychological construct is a pivotal aspect of research design. It shapes the direction of the study, influences data collection and analysis, and determines the relevance of the research to the broader field. By examining these case studies across personality traits, attitudes, and mental health indicators, we gain a deeper understanding of the intricacies involved in this foundational step of psychological research.

Considerations in Construct Selection

In the vast landscape of psychological research, the selection and definition of a psychological construct are complex and critical tasks. These decisions shape the trajectory of research, influencing everything from study design to data analysis. This module explores essential considerations that should guide researchers when selecting and defining psychological constructs.

One of the primary considerations when selecting and defining a psychological construct is its theoretical foundation. In essence, the construct should have a solid footing in existing theories and empirical evidence. Researchers embark on a journey to understand and describe psychological phenomena. To ensure that this journey is meaningful, they must navigate it within the framework of established principles and models (Smith & Glass, 1977).

The Role of Theory: The importance of theory in construct selection cannot be overstated. A well-grounded theoretical framework provides a structured understanding of the construct and its potential relationships with other variables. It helps researchers ask relevant research questions and formulate hypotheses. For example, in studies of anxiety, researchers may ground their work in the well-established cognitive-behavioral theory of anxiety, which posits that cognitive distortions and behavioral avoidance play a central role.

Evidence-Based Selection: In practice, this means that the selection of a psychological construct is not arbitrary but is based on a thorough review of the existing literature. Researchers should delve into the wealth of prior research to ensure that their construct aligns with established principles. This alignment reinforces the scientific rigor of the study.

Another crucial factor in construct selection is relevance. Researchers should ask whether the chosen construct is relevant to the research question and objectives. In essence, they need to consider whether the construct contributes to the advancement of knowledge or addresses practical issues (Schmitt, 1996).

Addressing Research Questions: To make this determination, researchers must clearly articulate their research questions and objectives. The construct's relevance hinges on its capacity to shed light on these inquiries. For instance, in a study exploring job satisfaction, the construct's relevance is evident as it directly pertains to the research question about factors influencing employees' job contentment.

The Practical Dimension: Beyond theoretical significance, relevance can also encompass practical applicability. For example, in clinical psychology, the choice of a psychological construct like post-traumatic stress disorder (PTSD) is highly relevant due to its real-world implications for the diagnosis and treatment of trauma-related mental health conditions.

While the theoretical foundation and relevance are paramount, researchers should also consider the practical dimension of construct selection: measurement feasibility. It's essential to assess whether the construct can be effectively and reliably measured using available methods and tools (Clark & Watson, 2015).

Assessing Measurement Tools: Researchers must weigh the available measurement tools and methods to ensure they align with the chosen construct. For example, in the assessment of intelligence, researchers can utilize standardized tests like the Wechsler Adult Intelligence Scale (WAIS) to measure cognitive abilities effectively.

Reliability and Validity: Ensuring the reliability and validity of measurement instruments is a vital aspect of measurement feasibility. If researchers select a construct but lack the means to measure it reliably and validly, the scientific value of the research is compromised. For example, in studies of academic achievement, researchers need to ensure that the chosen tests accurately measure students' knowledge and skills.

The process of selecting and defining a psychological construct is a critical foundation for robust research. By anchoring the construct in established theory, ensuring its relevance to research questions, and considering measurement feasibility, researchers take the first steps on a path that can lead to valuable contributions to the field.

The examples from various domains, such as personality traits, attitudes, and mental health indicators, provide concrete illustrations of these considerations in action. As researchers navigate the complexities of construct selection, they must balance theoretical grounding, significance, and practicality to ensure their research is both scientifically sound and impactful.

In conclusion, the process of selecting and defining the psychological construct of interest is a foundational step in psychological research. Clear, precise definitions ensure that the construct aligns with the research goals, enabling the development of valid and reliable measures. Examples from various domains, such as personality traits, attitudes, and mental health indicators, illustrate the importance of construct definition. Researchers must also consider the theoretical foundation, relevance, and measurement feasibility when making construct selection decisions.

Module 3: Generating and Refining Scale Items

Explanation of the process of generating potential scale items related to the chosen construct.

Guidance on refining and enhancing these items through expert reviews and pilot testing.

Guidelines for item wording, formatting, and response options.

Basic Issues in Objective Scale Development

In the intricate journey of psychological scale development, Module 3 plays a pivotal role by focusing on the generation and refinement of scale items. These items serve as the building blocks of measurement, enabling the systematic quantification of complex psychological constructs. In this module, we dive into the process of creating potential scale items related to the chosen construct, providing guidance on refining and enhancing these items through expert reviews and pilot testing. Additionally, we explore the critical elements of item wording, formatting, and response options, which collectively shape the reliability and validity of the measurement scale.

The process of developing a psychological scale is a complex and multifaceted one that requires careful consideration of various factors to ensure its validity and reliability. Constructing a valid and reliable scale is crucial in psychological research, as the quality of data collected largely depends on the quality of the measurement instrument. This text explores the fundamental issues involved in objective scale development, with a focus on ensuring construct validity, which is essential to the success of any measurement instrument.

The first step in constructing a valid scale is defining the psychological construct of interest. The construct is the abstract concept that the scale aims to measure. This definition should be clear, precise, and grounded in existing psychological theories and empirical evidence. Without a well-defined construct, the scale's purpose becomes vague, and its ability to measure the intended psychological phenomena is compromised (Clark & Watson, 2015).

Once the construct is clearly defined, the next step is generating potential scale items. These items are statements or questions designed to elicit responses from respondents that reflect their position on the construct. Crafting clear and specific items is essential during this phase to ensure that each item directly relates to the construct, is free from double-barreled content, and maintains a balance between positively and negatively worded items (DeVellis, 2016).

Expert reviews are a critical component of item refinement. Experts in the field assess the generated items for their relevance, clarity, and representativeness of the construct. They provide valuable feedback, leading to revisions in item wording, format, or the removal of items deemed irrelevant or redundant. This iterative process helps enhance the content validity of the scale (Clark & Watson, 2015).

Pilot testing follows expert reviews and involves a small sample of individuals similar to the intended respondents of the final scale. This phase is essential for assessing the clarity of the items and understanding how respondents interpret and respond to them. Cognitive interviews conducted during pilot testing are particularly valuable, as they allow researchers to explore how participants think and feel while completing the scale. These interviews help identify potential sources of confusion and inform further item refinements (Willis, 2005).

Item wording, formatting, and response options are crucial elements in the scale development process. Item wording should be concise and specific, avoiding jargon or complex language to ensure accessibility to the target population. The formatting of scale items should be consistent and user-friendly, and response options should be clear and logically ordered to maintain the scale's reliability and validity (DeVellis, 2016).

A fundamental consideration in objective scale development is the theoretical foundation of the construct. A valid scale should be grounded in existing theories and supported by empirical evidence. Researchers need to thoroughly review the literature to ensure that their chosen construct aligns with established principles and models (Smith & Glass, 1977).

Relevance is another important factor in construct selection. The chosen construct should directly contribute to the advancement of knowledge or address practical issues. Irrelevant constructs can lead to data collection efforts that do not provide valuable insights or solutions. Therefore, the researcher must carefully assess the construct's relevance to the research question and objectives (Schmitt, 1996).

Measurement feasibility is a practical consideration in scale development. Researchers need to assess whether the chosen construct can be effectively and reliably measured using available methods and tools. If a construct is difficult to measure or requires specialized equipment, it may not be feasible for the study's scope or budget (Clark & Watson, 2015).

In conclusion, constructing an objective scale for psychological research is a meticulous and multifaceted process. Clear construct definition, careful item generation and refinement, and attention to item wording, formatting, and response options are crucial elements in the development of a valid and reliable scale. Additionally, researchers must consider the theoretical foundation, relevance, and measurement feasibility of the chosen construct to ensure the success of the measurement instrument.

Generating Potential Scale Items: The Creative Phase

The process of generating potential scale items is a pivotal and intellectually engaging phase in the development of a psychological scale. This creative process involves transforming abstract constructs of interest into tangible, measurable statements or questions that serve as the building blocks of the scale. This step is instrumental in bridging the gap between theoretical concepts and their operationalization (DeVellis, 2016).

Crafting clear and specific items is the initial step in the creative phase of generating potential scale items. These items should be constructed with clarity, specificity, and an absence of ambiguity. The goal is to ensure that each item directly pertains to the chosen construct and is designed to elicit responses that genuinely represent the respondent's position on the construct (DeVellis, 2016).

To illustrate, let's consider a psychological construct such as "self-esteem." This construct involves individuals' perceptions of their self-worth and self-efficacy. To create clear and specific items, potential statements might include, "I feel confident in my abilities," or "I often doubt myself." These statements address different aspects of self-esteem and capture the variations in how individuals perceive and experience this construct.

An essential consideration during item generation is the need to avoid double-barreled items. Double-barreled items are statements that encompass multiple concepts or ideas within a single item. These items can lead to confusion and compromise the validity of the scale. For instance, consider an item that combines self-esteem and body image, such as, "I feel good about myself and my appearance." This item confounds two distinct constructs, making it challenging for respondents to provide accurate responses, and potentially introducing measurement error (DeVellis, 2016).

Balancing positively and negatively worded items is another critical aspect of the item generation process. Achieving this balance is essential to mitigate response biases that can skew the results of the scale. Response biases can include the tendency to agree with all items, regardless of their content, or to consistently choose the opposite response. Balancing positively and negatively worded items helps to counteract these biases and ensure a more accurate reflection of the construct.

For example, in a scale measuring anxiety, it is important to include both positively worded items like "I feel calm and relaxed" and negatively worded items like "I often worry and feel nervous." By incorporating both types of items, the scale captures a broader range of responses related to the construct of anxiety. This approach ensures that individuals who experience different facets of anxiety can express their feelings and experiences more accurately through the scale (DeVellis, 2016).

In summary, the process of generating potential scale items is a dynamic and creative phase in the development of a psychological scale. Clear and specific items are crafted to ensure they directly relate to the chosen construct, while avoiding double-barreled items is essential to maintain item clarity and validity. Balancing positively and negatively worded items helps counteract response biases and provides a more accurate reflection of the construct being measured. These considerations collectively contribute to the development of a robust and reliable psychological scale.

Refining Through Expert Reviews: Ensuring Content Validity

The process of refining and enhancing potential scale items through expert reviews is a crucial step in the development of a psychological scale, with a primary focus on ensuring content validity. Content validity is vital to confirm that the items accurately and comprehensively represent the construct of interest (Clark & Watson, 2015). This phase engages experts to evaluate the items to assess their relevance, clarity, and conciseness.

Expert reviews play a central and irreplaceable role in the refinement process. These experts are individuals with subject matter knowledge who bring invaluable insights and judgment to the evaluation of scale items. Their expertise ensures that the items genuinely capture the essential aspects of the construct and that they are relevant, clear, and concise (Clark & Watson, 2015).

For example, in the development of a scale aimed at measuring teacher effectiveness, educational experts with experience in pedagogy and educational research are enlisted to assess the items. These experts evaluate the items to confirm that they effectively address critical elements of teaching quality, such as classroom management, instructional strategies, and student engagement. Their expertise is instrumental in identifying whether the scale items comprehensively represent the intended construct of teacher effectiveness.

The feedback provided by expert reviewers often leads to significant revisions in item wording and clarity. It may also result in the removal of items that are deemed irrelevant or redundant. This iterative process of refining and revising the scale items enhances the content validity of the scale (Clark & Watson, 2015).

In addition to clarifying item wording, experts may suggest alternative phrasing to ensure items are concise and unambiguous. This process contributes to the development of items that are easy to understand and unambiguous for respondents. It helps in preventing misinterpretation and response errors that could compromise the scale's validity.

Furthermore, expert reviews assist in identifying items that may not contribute to the measurement of the target construct. Redundant or irrelevant items are eliminated during this process. This not only streamlines the scale but also ensures that the remaining items more accurately and comprehensively reflect the construct of interest.

The involvement of expert reviewers is also beneficial in confirming the relevance of items to the construct within the specific context of the research. Depending on the research objectives and population, some items may need to be adapted or revised to ensure they are contextually appropriate.

The expert review process is characterized by a collaborative and iterative approach. Researchers work closely with the experts to address their feedback, make necessary revisions, and refine the items. This collaborative effort is essential for enhancing the content validity of the scale.

Additionally, expert reviews contribute to the overall credibility and trustworthiness of the scale. When it is known that experts in the field have reviewed and validated the scale items, it adds a layer of confidence in the scale's ability to measure the intended construct accurately and reliably.

In summary, the process of refining and enhancing potential scale items through expert reviews is essential for ensuring content validity. Expert reviewers play a pivotal role in evaluating the items, offering feedback, and identifying necessary revisions. This iterative process enhances the clarity, relevance, and conciseness of scale items, ultimately contributing to the development of a valid and reliable psychological scale.

Pilot Testing: Assessing Clarity and Response Patterns

Pilot testing is an integral step in the process of developing psychological scales, as it serves to assess the clarity of scale items and observe how respondents interpret and respond to them (Nunnally & Bernstein, 1994). This phase offers a valuable opportunity to ensure that the items are comprehensible to the target population and to identify and address potential issues in the scale's items and structure.

Pilot testing typically involves a small sample of individuals who are similar to the intended respondents of the final scale (Nunnally & Bernstein, 1994). This sample is selected to provide insights into the clarity, interpretability, and potential response biases associated with the items. It is essential that the pilot sample closely resembles the characteristics of the intended target population to ensure that the feedback is representative and meaningful.

Cognitive interviews are a key component of pilot testing (Willis, 2005). During these interviews, participants are asked to think aloud as they read and respond to the scale items. This method allows researchers to gain a profound understanding of how individuals interpret the items, identify potential sources of confusion, and refine item wording and formatting accordingly.

Cognitive interviews provide a window into the cognitive processes of the respondents. Researchers can observe how respondents approach each item, whether they find the wording clear or confusing, and if they have any difficulty in selecting a response option. These insights are invaluable in refining the items to ensure that they are as clear and unambiguous as possible.

Moreover, pilot testing often leads to item reduction. Items that are consistently misinterpreted, lead to response biases, or are deemed unclear by respondents may be removed from the scale (Nunnally & Bernstein, 1994). This is a critical step to ensure that the final scale is both reliable and valid. By eliminating problematic items, researchers improve the overall quality of the scale and enhance its effectiveness as a measurement tool.

The process of item reduction should be guided by empirical data gathered during the pilot testing phase. For example, items with high rates of non-response or those that exhibit little variability in responses may be candidates for removal. Additionally, items that respondents consistently find unclear, confusing, or irrelevant should be closely examined for potential revisions or removal from the scale.

The goal of pilot testing is to refine and enhance the scale's items to ensure they are easily understood, result in meaningful responses, and accurately capture the intended construct.

As a result of this phase, researchers obtain a more refined set of items that have undergone scrutiny in terms of clarity, interpretability, and response patterns.

In conclusion, pilot testing is a crucial phase in the development of psychological scales. It provides an opportunity to assess the clarity of scale items and observe how respondents interpret and respond to them. The use of cognitive interviews, as well as item reduction based on empirical data, ensures that the final scale is both reliable and valid. This iterative process enhances the quality of the scale, contributing to its effectiveness as a measurement tool in psychological research.

Guidelines for Item Wording, Formatting, and Response Options: Key Factors in Scale Development

In the development of psychological scales, the process extends beyond generating and refining scale items and involves critical considerations in item wording, formatting, and response options. These components significantly impact the reliability and validity of the measurement scale (DeVellis, 2016). To ensure that the scale effectively captures the intended construct, it is essential to pay meticulous attention to these factors.

Item wording is a pivotal element in the creation of scale items. Each item's wording should be concise, specific, and directly reflective of the construct's definition. Clarity in item wording is fundamental to ensure that respondents can easily comprehend and interpret the items (DeVellis, 2016). Avoiding jargon, overly complex language, or ambiguous phrasing is essential to make the items accessible to the target population. For instance, in a scale measuring job satisfaction, a well-constructed item could be worded as "I am content with my current job," rather than a more complex or convoluted statement. This level of clarity in item wording is crucial for obtaining meaningful and interpretable responses.

Formatting plays a substantial role in the usability of a psychological scale. It is imperative to maintain consistency and user-friendliness in the formatting of scale items. Utilizing a uniform response scale format across all items is recommended to ensure that respondents can easily understand and complete the scale (DeVellis, 2016). A common practice is to employ a 5-point Likert scale, which consists of options ranging from "Strongly Disagree" to "Strongly Agree." This standardized format simplifies the response process, allowing respondents to provide their feedback systematically. Consistent formatting not only eases the respondents' experience but also enhances the scale's reliability and interpretability.

The choice of response options is another critical aspect to consider in the development of scale items (DeVellis, 2016). Response options are the set of choices available to respondents for each item. They should be clear, logically ordered, and aligned with the specific construct and research objectives. Common response formats include Likert scales, numerical rating scales, and visual analog scales. These response options serve as the anchors for the respondent to express their agreement, disagreement, intensity, or other relevant dimensions of the construct. The choice of response options should be made thoughtfully, considering the specific characteristics of the construct and the objectives of the research. For instance, in a scale measuring self-efficacy, response options could range from "Not at all confident" to "Extremely confident," allowing respondents to convey the degree of their self-efficacy beliefs.

In summary, the development of scale items necessitates meticulous attention to item wording, formatting, and response options. Item wording should emphasize clarity, specificity, and alignment with the construct's definition. Consistency in formatting enhances user-friendliness and aids in maintaining uniformity throughout the scale. Lastly, the choice of response options should be well-considered, ensuring they are clear and logically ordered to capture the nuances of the construct under investigation. These aspects collectively contribute to the reliability, validity, and interpretability of the psychological scale, allowing for meaningful data collection and analysis in psychological research.

In conclusion, Module 3 delves into the critical phase of generating and refining scale items, which are the fundamental elements of psychological measurement. The process involves creativity in crafting clear and specific items, refinement through expert reviews to ensure content validity, and pilot testing to assess clarity and response patterns. Additionally, careful attention to item wording, formatting, and response options is essential to enhance the reliability and validity of the measurement scale.

Module 4: Content Validity and Item Selection

Introduction to content validity and its role in ensuring that scale items adequately represent the construct.

Explanation of the process for assessing content validity, including expert judgment and content validity ratio (CVR).

Discussion of item selection and strategies to reduce redundancy, resulting in a concise and focused scale.

Content validity is an intrinsic element of psychological research and scale development that warrants comprehensive attention and careful consideration. As researchers, the ability to construct measurement tools that accurately and comprehensively represent the construct under investigation is a fundamental pursuit. In Module 4, we embark on a journey to delve into the core concepts of content validity and its crucial role in the development of effective and reliable measurement scales.

Understanding Content Validity

Content validity stands as a central and indispensable concept in the realm of psychological scale development. But what exactly does content validity signify? In essence, content validity is a measure of the extent to which the items contained within a psychological scale genuinely, comprehensively, and accurately represent the specific construct targeted for measurement (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). In simpler terms, content validity ensures that the items within the scale are not only relevant but also closely aligned with the psychological trait, behavior, or attribute being examined.

The core essence of content validity is akin to capturing the true essence of the construct within the measurement tool. It's akin to ensuring that no vital aspect of the construct goes unexplored, and that the items within the scale faithfully mirror the intended phenomenon. Devoid of robust content validity, a scale may fall short of fulfilling its intended purpose, potentially generating results that do not accurately or comprehensively represent the construct under investigation (Clark & Watson, 2015).

The concept of content validity is a linchpin in scale development for a multitude of reasons. Most notably, it directly impacts the quality and reliability of research and assessment. The robustness of content validity is pivotal in determining the credibility and interpretability of the data derived from the scale. It underscores the fundamental importance of placing content validity at the forefront of the scale development process. Doing so ensures that the data produced are not only meaningful but also accurate, aligning harmoniously with the overarching objectives of the research.

As researchers, understanding the significance of content validity equips us with the knowledge and tools to develop measurement scales that genuinely encapsulate the constructs we aim to study. By embracing content validity as a cornerstone of our research endeavors, we enhance the precision and depth of our measurements, resulting in a more profound understanding of the phenomena under investigation. In summary, content validity is not just a technicality; it is the bedrock upon which reliable and impactful psychological scales are built.

The Process of Assessing Content Validity

Ensuring content validity is far from a matter of guesswork or subjectivity; rather, it is a systematic and methodical process. This process encompasses a series of meticulously designed steps aimed at rigorously evaluating whether the scale's items genuinely and comprehensively represent the targeted construct. Two fundamental components within the assessment of content validity are expert judgment and the Content Validity Ratio (CVR). Both of these elements work in concert to refine the scale and eliminate items that do not effectively capture the essence of the construct (Lawshe, 1975).

The process of assessing content validity is multifaceted, encompassing several critical steps that are essential in the creation of a reliable and valid measurement tool. These steps include item generation, expert reviews, and content validity ratio calculations. Let's delve deeper into each of these steps, highlighting the use of expert judgment and the Content Validity Ratio (CVR) as pivotal tools in this process.

The initial step in content validity assessment is the generation of potential scale items. This phase involves crafting a series of statements or questions that are conceptually related to the construct under investigation. The items must be framed in a way that is clear, specific, and unambiguous to ensure that they accurately capture the essence of the construct. This creative process requires a deep understanding of the construct and a careful choice of wording to prevent ambiguity or confusion. Crafting items that effectively measure the intended psychological trait is fundamental in establishing content validity.

Once potential scale items are generated, the subsequent step involves expert reviews. Expert reviews are an essential component in the refinement of scale items. Researchers enlist the expertise of individuals who possess subject matter knowledge related to the construct being measured. These experts meticulously evaluate each item to determine whether they accurately represent the construct, are clear and relevant, and exhibit concise wording. This expert judgment provides valuable insights into the suitability of items for inclusion in the final scale. Feedback from experts often results in revisions to item wording, the clarification of ambiguous statements, or the elimination of items that are considered irrelevant or redundant. It is an iterative process aimed at enhancing the content validity of the scale.

Moreover, when expert judgment is applied to the assessment of content validity, it bolsters the overall quality and effectiveness of the scale. Expert reviewers assess items with a discerning eye, ensuring that each item aligns with the construct's definition and relevance to the study. They consider the clarity of the items, their conciseness, and the extent to which they accurately reflect the intended psychological trait. This comprehensive evaluation by experts helps in identifying and eliminating items that do not meet the stringent criteria for content validity, thereby enhancing the scale's robustness.

In parallel with expert reviews, the Content Validity Ratio (CVR) plays a vital role in content validity assessment. The Content Validity Ratio (CVR) is a statistical index that quantifies the extent of agreement among experts regarding the relevance of each item within the scale (Lawshe, 1975). It helps to objectively identify items that have a low level of content validity, as determined by the expert panel. The CVR process involves experts rating each item as "essential," "useful but not essential," or "not necessary" for measuring the construct. The scores are then calculated to derive a CVR value for each item. Items that receive a low CVR score are generally considered for removal from the scale since they do not achieve the required level of consensus among experts regarding their relevance to the construct.

The careful interplay between expert judgment and the CVR ensures that the scale items are thoroughly evaluated, and only those that genuinely represent the construct are retained. This iterative process, combining expert reviews and CVR calculations, ultimately contributes to the content validity of the scale.

Thus, the pursuit of content validity in scale development is a systematic journey that encompasses several essential steps. The generation of potential scale items requires a deep understanding of the construct and careful crafting of clear and unambiguous statements. Expert reviews, which involve subject matter experts, provide valuable feedback to refine the items, enhance their clarity, and eliminate irrelevant or redundant items. Additionally, the incorporation of the Content Validity Ratio (CVR) brings objectivity to the content validity assessment, allowing researchers to objectively gauge the consensus among experts regarding the relevance of each item. The interplay between expert judgment and CVR is pivotal in the creation of a valid and reliable measurement scale. Ultimately, content validity is not a singular step but an ongoing process of refinement, ensuring that the scale accurately and comprehensively captures the essence of the targeted construct.

Leveraging Expert Judgment

In the pursuit of content validity, the integral role of expert judgment cannot be overstated. It serves as a foundational and indispensable component in the rigorous assessment of content validity. The assessment of content validity draws upon experts with specialized knowledge in the subject matter or field related to the construct under investigation. These experts play a pivotal role in evaluating the items within the scale, employing their extensive expertise to critically assess whether these items accurately and comprehensively represent the intended construct (Clark & Watson, 2015).

The experts enlisted to evaluate the scale's items are often individuals who have devoted their careers to the study, practice, or application of the specific domain of the construct. Their deep understanding of the nuances and intricacies of the subject matter empowers them to offer informed and precise judgment regarding the alignment of scale items with the intended construct. These experts may include seasoned researchers, experienced practitioners, or professionals with specialized knowledge, depending on the context of the research.

The primary function of these experts is to apply their professional judgment to determine the extent to which each item genuinely mirrors the construct in question. For instance, consider a scenario where researchers are developing a scale to measure the construct of "teacher effectiveness." In this case, educational experts with extensive experience and expertise in teaching, educational psychology, or pedagogical research would be engaged to evaluate the scale items. These experts are uniquely positioned to assess the items, ensuring that they accurately capture the critical elements that define teacher effectiveness.

The collaboration with experts is characterized by an iterative and thorough process. These experts engage in the rigorous evaluation of each item, taking into account several important factors to assess content validity. Among the key considerations are the relevance of the items to the construct, the clarity of the item wording, and the overall representativeness of the items concerning the construct (Clark & Watson, 2015).

- *Relevance: Experts scrutinize each item to determine its relevance to the construct. They assess whether the item genuinely taps into the essential components of the construct under investigation. If an item fails to address a pivotal aspect of the construct, experts may recommend revisions or removal from the scale.*
- *Clarity: The clarity of item wording is of paramount importance. Experts assess the phrasing of items to ensure that they are clear, concise, and readily comprehensible to the intended respondents. Ambiguity, vagueness, or overly complex language is carefully examined, and suggestions for improvements are provided.*
- *Representativeness: Experts evaluate the overall representativeness of the items concerning the intended construct. They consider whether the items cover a broad spectrum of relevant aspects of the construct, avoiding undue emphasis on one aspect to the detriment of others.*

The input of experts is invaluable, as it ensures that the scale's items are not only closely aligned with the construct but also intelligible and unambiguous. Furthermore, the collaboration with experts fosters a dynamic and iterative process, enabling continual refinement of the scale. Experts often provide feedback on the items, suggesting revisions or clarifications when necessary. Their input guides researchers in the fine-tuning of the scale's items to enhance their clarity, relevance, and overall content validity.

In essence, the utilization of expert judgment in content validity assessment is fundamental in the development of a reliable and valid measurement tool. It leverages the expertise of individuals deeply immersed in the subject matter, ensuring that the items effectively capture the essence of the construct. This collaborative and iterative approach enriches the scale development process, contributing to the content validity of the scale. The result is a measurement tool that comprehensively and accurately represents the targeted construct, ultimately producing meaningful and trustworthy research outcomes.

Content Validity Ratio (CVR)

The Content Validity Ratio (CVR) is a quantitative tool used to assess the relevance and importance of scale items based on expert judgment (Lawshe, 1975). It helps researchers determine whether an item should be retained or eliminated from the scale. To calculate the CVR, experts evaluate each item and determine its essentiality for measuring the construct. Items that receive a favorable assessment from the experts, indicating their significance for the construct, are retained, while items that do not reach a predefined level of agreement are considered for removal. The CVR is calculated using the following formula:

$$CVR = (n_e - N/2)/(N/2)$$

Where:

- n_e *represents the number of experts indicating an item as essential.*
- N *represents the total number of experts.*

The CVR is a quantitative measure that serves as a valuable indicator of item essentiality. It quantifies the degree of consensus among experts regarding the relevance and necessity of each item within the scale. The CVR value can range from -1 to 1, and the interpretation of this value provides insights into the consensus among experts.

A positive CVR value carries significant implications. It signifies that there is a consensus among the experts that the item is essential for capturing the construct under investigation. In other words, the positive CVR value indicates that the item aligns with the construct and is seen as relevant and necessary by the experts. This consensus is a strong indicator of the item's content validity and its ability to comprehensively and accurately represent the construct.

Conversely, a negative CVR value raises concerns about the item's alignment with the intended construct. It suggests a lack of agreement among the experts, indicating that they do not universally consider the item to be essential for capturing the construct. A negative CVR value should prompt careful consideration of the item's wording, relevance, and overall contribution to the scale. Such items may warrant revision or potential removal to enhance the scale's content validity.

The utilization of the CVR in content validity assessment provides several critical advantages. First and foremost, it introduces an objective and systematic element to the evaluation process. The quantitative nature of the CVR allows researchers to standardize the assessment of item essentiality, minimizing subjectivity in the content validity evaluation. This objectivity is crucial for producing reliable and valid measurement scales.

The CVR also reinforces the significance of item relevance and consistency in content validity assessment. It emphasizes that the evaluation of item essentiality should not solely rely on expert judgment but should also incorporate a quantitative measure. The combination of expert judgment and the CVR contributes to a comprehensive and rigorous content validity assessment.

In summary, the Content Validity Ratio (CVR) is a vital tool in content validity assessment, providing a quantitative measure of item essentiality. A positive CVR value indicates a consensus among experts regarding the relevance and necessity of the item for capturing the construct, reinforcing its content validity. In contrast, a negative CVR value highlights the lack of consensus among experts and suggests a need for further examination and potential revision. By incorporating the CVR into content validity assessment, researchers enhance the objectivity and comprehensiveness of the evaluation, ultimately contributing to the development of reliable and valid measurement scales.

Item Selection and Reduction

One the process of scale development within psychological research is characterized by multiple phases, each integral to the creation of a valid and reliable measurement tool. One pivotal stage in this journey is the selection and reduction of scale items, a process deeply influenced by the assessment of content validity. This phase is crucial for refining the scale, enhancing its focus, and ultimately producing a measurement tool that accurately captures the intended construct.

Content validity, as previously discussed, plays a central role in content validity assessment. It is the degree to which scale items genuinely, comprehensively, and accurately represent the specific construct targeted for measurement. In essence, content validity ensures that the items within the scale are not only relevant but also closely aligned with the psychological trait, behavior, or attribute being assessed. The objective of content validity is to capture the true essence of the construct, leaving no critical aspect unexamined (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014).

The concept of content validity can be likened to capturing the true spirit of the construct within the measurement tool. It's about making sure that no critical aspect of the construct is left unexamined, and that the items within the scale are a faithful reflection of the targeted phenomenon. Without robust content validity, a scale might fail to serve its intended purpose, producing results that do not faithfully represent the construct under investigation (Clark & Watson, 2015).

The journey of content validity assessment offers valuable guidance on which items to retain within the scale and which to consider for removal. Content validity assessment involves expert judgment, an essential component that allows experts in the field related to the construct to evaluate the items within the scale. Their expertise is harnessed to critically assess whether the items accurately and comprehensively represent the construct. The process includes evaluating factors such as relevance, clarity, and overall representativeness of the items (Clark & Watson, 2015).

This phase of the scale development process is an iterative one, characterized by expert feedback and revisions to the scale items. Experts often provide feedback on the items, suggesting revisions or clarifications as needed. This iterative process helps to enhance the content validity of the scale, ensuring that the items are not only reflective of the construct but also intelligible and unambiguous.

Expert judgment, in this context, is more than a subjective evaluation. It's the culmination of years of knowledge and experience in the field, making it a reliable and objective tool in content validity assessment. The collaboration of experts ensures that the items effectively tap into the essential elements that define the construct, providing a strong foundation for the scale's development and effectiveness (Clark & Watson, 2015).

The assessment of content validity, often conducted through expert reviews, yields essential information for item selection. Items that receive favorable Content Validity Ratio (CVR) scores are deemed to be essential for accurately capturing the construct under investigation. A positive CVR value signifies consensus among experts regarding the relevance and necessity of the item. These items are retained in the scale, as they are considered vital for representing the construct comprehensively and accurately (Lawshe, 1975).

However, content validity assessment also serves another crucial purpose. It highlights potential redundancy within the scale. Redundant items can lead to respondent confusion and compromise the scale's efficiency. To mitigate this issue, items that consistently overlap or duplicate the measurement of the same facet of the construct are considered for removal. This process of item reduction streamlines the scale, resulting in a more concise and focused set of items, free from redundancy (Clark & Watson, 2015).

The significance of content validity in scale development has been widely recognized within the psychological and educational research communities. Various studies and publications have addressed the importance of content validity in the context of scale development.

For example, a study by Haladyna, Downing, and Rodriguez (2002) provides a comprehensive exploration of content validity as part of test development. The authors discuss the various elements of content validity, emphasizing the crucial role of expert judgment in evaluating items. Their work underscores the significance of aligning test items with the targeted construct, which is central to content validity.

In the realm of psychological research, O'Neill, Goffin, and Tett, (2009) explores content validity as an essential component of personality assessment. The study emphasizes the importance of expert judgment and theoretical alignment in creating personality measures with strong content validity. It highlights the need to develop measurement tools that accurately reflect the richness and complexity of personality constructs.

Content validity is not merely a technical aspect of scale development; it is the bedrock upon which the effectiveness and accuracy of psychological scales rest. Through a methodical process that engages expert judgment and the Content Validity Ratio (CVR), researchers ensure that their scale items genuinely and comprehensively represent the construct. The outcome of this rigorous assessment is a focused and reliable measurement tool that generates results aligned with the construct under investigation.

The importance of content validity resonates through psychological research, with a myriad of studies emphasizing its pivotal role. As researchers, by prioritizing content validity, we elevate the quality and impact of our research, ensuring that our measurement scales are robust, meaningful, and aligned with the constructs we aim to explore. In the ever-evolving landscape of psychological research, content validity remains a steadfast beacon, guiding us toward the creation of precise, reliable, and valuable measurement tools.

Module 5: Validity Assessment

Exploration of different types of validity, including content, criterion, and construct validity.

Discussion of techniques and methods for establishing each type of validity.

Illustration of validity assessment through examples and case studies.

Module 5 delves into the critical topic of validity assessment in the realm of psychological research and scale development. Understanding and establishing validity is paramount to ensure that a measurement scale accurately measures what it intends to measure. In this module, we explore various types of validity, including content, criterion, and construct validity, each playing a distinct role in the assessment of a scale's effectiveness.

Content validity, which we've previously discussed, focuses on the alignment between the items within a scale and the construct under investigation. In contrast, criterion validity assesses the scale's ability to predict or correlate with external criteria, providing evidence of the scale's effectiveness in real-world applications. Construct validity, on the other hand, is concerned with the theoretical underpinnings of the construct and how well the scale measures the intended psychological trait.

To establish each type of validity, researchers employ specific techniques and methods. These may involve comparing a new scale to an established one, conducting predictive studies, or using statistical analyses to assess relationships and patterns. Real-world examples and case studies serve as practical illustrations of how these techniques are applied in research, making the complex concept of validity more tangible and accessible for scale developers and researchers.

In Module 5, we embark on a journey to unravel the intricate layers of validity assessment, equipping us with the knowledge and tools to ensure that our measurement scales are not only reliable but also valid in capturing the constructs we aim to explore.

Content, Criterion, and Construct Validity

In the field of psychological measurement, the concept of validity stands as a cornerstone, ensuring that a scale accurately measures the psychological constructs it purports to assess. It is the backbone of measurement, reflecting the degree to which an instrument measures what it intends to measure. Without a strong foundation of validity, the data derived from psychological measurements may lack the credibility and relevance needed to inform research and practice. To address this fundamental issue, psychologists have developed different types of validity that serve distinct purposes and provide nuanced insights into the quality of measurement tools. This comprehensive review delves into content, criterion, and construct validity, unraveling their definitions, historical developments, contemporary perspectives, and practical applications. We will explore their significance in psychological research and the methods employed to assess them.

Content validity is the extent to which the items within a psychological scale genuinely, comprehensively, and accurately represent the specific construct targeted for measurement (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). In simpler terms, it ensures that the items within the scale are not only relevant but also closely aligned with the psychological trait, behavior, or attribute being examined.

The concept of content validity can be traced back to the development of educational and psychological tests in the early 20th century. Its importance was recognized by educational psychologists like E. L. Thorndike, who emphasized the need for test items to represent the content they aimed to measure. Content validity has since evolved and found applications in diverse fields of psychology, ranging from clinical and counseling psychology to educational and industrial-organizational psychology.

Content validity remains a critical component of psychological scale development. With advancements in measurement theory and technology, the methods for assessing content validity have become more sophisticated and systematic. Researchers are continually refining their approaches to ensure that scales comprehensively capture the constructs of interest.

Content validity is assessed through techniques such as expert reviews, item generation, and the Content Validity Ratio (CVR). Expert reviews involve the evaluation of scale items by individuals with subject matter knowledge, who assess factors like relevance, clarity, and representativeness. The CVR provides a quantitative measure of expert consensus on item relevance.

Content validity is vital when developing measurement scales that cover multifaceted constructs, such as self-esteem, job satisfaction, or personality traits. To ensure the comprehensiveness and accuracy of the scale, researchers engage with experts in the field, refine items, and calculate CVR scores. The iterative process guarantees that items effectively tap into the essential elements of the construct, enhancing the quality and validity of the measurement tool.

Criterion validity assesses the extent to which a psychological scale can predict or correlate with external criteria or outcomes. It involves comparing the scores obtained on a new scale with those from a well-established criterion, which serves as a benchmark for the construct being measured.

The development of criterion validity is rooted in the early days of psychometrics, when pioneers like Charles Spearman laid the groundwork for statistical approaches to validity. Criterion validity has seen significant use in educational testing, clinical psychology, and industrial-organizational psychology.

Criterion validity continues to be a crucial type of validity, especially in applied fields. Advances in statistical techniques and the development of more sophisticated criterion measures have refined the methods for establishing criterion validity.

Researchers assess criterion validity through methods like concurrent and predictive validity. Concurrent validity examines the relationship between the new scale and an established criterion measure at the same time. Predictive validity assesses the scale's ability to predict future criteria or outcomes.

Criterion validity is instrumental when researchers aim to determine the real-world applicability of a measurement tool. For example, in the field of clinical psychology, a depression scale may be validated by assessing its ability to predict clinical diagnoses made by mental health professionals. This type of validity helps establish the practical relevance and usefulness of a measurement tool.

Construct validity is concerned with the theoretical underpinnings of a psychological construct and how well a scale measures the intended psychological trait or characteristic. It assesses the degree to which a measurement tool captures the complexity and multifaceted nature of a construct.

The concept of construct validity gained prominence with the advent of psychological tests. In the mid-20th century, pioneers like Lee J. Cronbach provided insights into the multifaceted nature of constructs and the need for rigorous validation. Since then, construct validity has been a central focus in the field of psychological measurement.

Construct validity remains a vibrant area of research, with contemporary psychologists exploring its nuances in the context of multifaceted constructs and modern measurement methods. Researchers are developing and refining advanced statistical techniques to assess construct validity.

Researchers employ techniques like factor analysis, convergent and discriminant validity, and multitrait-multimethod (MTMM) analysis to assess construct validity. Factor analysis helps identify underlying dimensions of a construct, while convergent and discriminant validity examines the scale's relationships with related and unrelated constructs. MTMM analysis provides insights into the relationships between multiple traits and methods of assessment.

Construct validity is essential when researchers investigate complex constructs like intelligence, personality, or attitudes. For example, a new intelligence test must demonstrate that it measures the multifaceted nature of intelligence by showing convergent validity with other established intelligence tests and discriminant validity from unrelated constructs. Establishing construct validity ensures that the measurement tool captures the richness and complexity of the targeted psychological trait.

Examples of Validity Assessment

Content Validity Example

Let's consider the development of a scale to measure "career satisfaction" in a specific industry. Initially, items are generated, and experts, including experienced professionals and academics in the field, assess the items. After feedback and revisions, a Content Validity Ratio (CVR) analysis is conducted. Items that achieve a high CVR score are retained, while those with lower scores are modified or excluded. This iterative process ensures that the scale comprehensively represents the facets of career satisfaction relevant to that industry.

Criterion Validity Example

Imagine a scenario in clinical psychology where a newly developed depression scale is assessed for criterion validity. Researchers administer the scale to a sample of individuals seeking mental health treatment. Concurrent validity is examined by comparing the scale scores to clinical diagnoses made by experienced psychologists. High concordance between the scale scores and the diagnoses indicates strong concurrent validity, supporting the scale's ability to accurately measure depression.

Construct Validity Example

In the realm of educational assessment, researchers develop a test to measure students' problem-solving skills. Construct validity is established by conducting factor analysis to identify underlying dimensions within the construct of problem-solving. Additionally, convergent and discriminant validity analyses explore the relationships between the problem-solving test and other measures of related and unrelated constructs. The findings provide evidence of the test's ability to comprehensively capture the construct of problem-solving.

In the diverse landscape of psychological research, the exploration of validity types is integral to the development of reliable and meaningful measurement tools. Content validity ensures that a scale covers the relevant facets of a construct, criterion validity demonstrates its applicability to real-world criteria, and construct validity assures it captures the multifaceted nature of a psychological trait. Researchers employ various methods to assess these validity types, such as expert reviews, criterion comparisons, and advanced statistical techniques.

This section has illuminated the historical foundations, contemporary perspectives, and practical applications of content, criterion, and construct validity. It has underscored the importance of these validity types in different areas of psychology, from clinical and educational to industrial-organizational and personality assessment. Moreover, examples have illustrated the role of validity assessment in the development of measurement tools.

In conclusion, the pursuit of validity in psychological measurement is a dynamic and evolving journey. Researchers must carefully navigate the landscape of content, criterion, and construct validity, employing a range of techniques and methods to ensure their measurement tools are accurate, meaningful, and applicable. By embracing the nuances of each validity type and their historical evolution, psychologists can continue to refine their practices and create measurement tools that stand up to rigorous scrutiny. The ongoing advancements in the field of psychological measurement underscore the centrality of validity and its unwavering importance in the pursuit of scientific knowledge.

Discussion of techniques and methods for establishing each type of validity

The establishment of different types of validity—content, criterion, and construct—demands specific techniques and methods tailored to the unique characteristics of each validation process. The robust validation of psychological measurement tools hinges upon the careful selection and implementation of these techniques. In this comprehensive exploration, we delve into these methods for each type of validity, providing a detailed understanding of their application.

Content Validity

Content validity, the first pillar of scale validation, ensures that the items within a scale genuinely and comprehensively represent the construct of interest (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Achieving content validity necessitates a diligent evaluation of the scale items to verify their relevance and clarity. Several techniques contribute to this process:

Expert Judgment: As discussed earlier, expert judgment is a pivotal element in content validity. Experts assess the scale's items, considering factors such as relevance, clarity, and overall representativeness (Clark & Watson, 2015). The collaboration of experts ensures that the items closely align with the construct. An example of expert judgment is found in a study by Haladyna, Downing, and Rodriguez (2002), which evaluates items for educational assessments.

Content Validity Ratio (CVR): CVR offers a quantitative approach to content validity. It quantifies the extent of agreement among experts regarding the relevance of each item within the scale (Lawshe, 1975). By assigning experts to rate items as "essential," "useful but not essential," or "not necessary" for measuring the construct, CVR provides objective insight into item relevance. This method helps to ensure that scale items align with the intended construct, as exemplified in nursing research when validating a pain assessment scale (Polit & Beck, 2006).

Content Validity Index (CVI): The CVI focuses on assessing item-level content validity through expert ratings. It is particularly useful when evaluating content validity in health-related fields (Polit et al., 2007). Researchers often use CVI to validate scales such as a patient satisfaction survey in healthcare contexts (Lawshe, 1975).

Content validity is a critical aspect of scale development, as it ensures that the scale captures the essential elements of the construct comprehensively and accurately. By implementing these techniques, researchers can enhance content validity and create reliable measurement tools.

Criterion Validity

Criterion validity assesses the extent to which a scale correlates with or predicts an external criterion. There are two primary techniques for establishing criterion validity:

Concurrent Validation: In concurrent validation, the scale in question is administered simultaneously with a criterion measure that represents the same construct. Researchers then assess the correlation between the two sets of scores (Anastasi & Urbina, 1997). For instance, when validating a new scale for measuring depression, it might be administered alongside a well-established depression inventory. The study can then examine the correlation between the two sets of scores to assess the concurrent validity of the new scale (Beck et al., 1996).

Predictive Validation: Predictive validation, on the other hand, aims to determine whether the scores from the scale can predict future criteria. In the context of employment settings, this often involves assessing the ability of a job applicant's test scores to predict their future job performance. For example, a study may investigate whether scores on a pre-employment aptitude test can predict the subsequent job performance of candidates (Murphy & Davidshofer, 2005).

Construct Validity

Construct validity, the third type of validity, pertains to the underlying theoretical structure of the scale and its ability to assess the theoretical construct of interest. Numerous techniques contribute to the establishment of construct validity:

Factor Analysis: Factor analysis is a common technique used to assess the underlying structure of a scale. It helps uncover the latent constructs that drive item responses. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are often employed to examine the relationships among observed variables (items) and their underlying latent constructs (factors) (Brown, 2006).

Convergent and Discriminant Validity Analysis: Convergent validity demonstrates that constructs that are theoretically expected to be related are, in fact, related. Researchers evaluate the correlations between the construct being measured and other constructs that should theoretically be correlated (Campbell & Fiske, 1959). Discriminant validity, on the other hand, verifies that constructs that should not be related theoretically exhibit low correlations (Fornell & Larcker, 1981). A study by Netemeyer, Bearden, and Sharma (2003) exemplifies the use of these techniques in assessing the construct validity of a consumer satisfaction measure.

Multitrait-Multimethod Matrix Examination: This technique aids in distinguishing the impact of different traits and methods on scale scores (Campbell & Fiske, 1959). Researchers employ this method to examine the relationships among multiple traits (constructs) and the different methods used to measure them. It ensures that the scale genuinely assesses the construct of interest rather than other related but distinct constructs.

As an example, a study by La Greca and Lopez (1998) utilized factor analysis to validate a scale measuring social anxiety in adolescents. The researchers identified and confirmed the underlying factor structure of the scale, ensuring its construct validity in assessing social anxiety. This demonstrates how factor analysis can be instrumental in the validation of psychological scales.

In sum, establishing validity in psychological measurement tools is a multi-faceted process. Content validity relies on expert judgment and quantitative measures like CVR and CVI to confirm item relevance and alignment with the construct. Criterion validity involves concurrent and predictive validation methods, while construct validity employs factor analysis and assessments of convergent and discriminant validity. These methods ensure that psychological measurement tools accurately capture the constructs they are designed to assess, contributing to the overall reliability and validity of psychological research.

Convergent Validity and the Nomological Network: A Comprehensive Examination

One fundamental aspect of validity, known as convergent validity, plays a pivotal role in this process. Convergent validity assesses the extent to which a particular measurement is correlated with other measures that it theoretically should be related to, based on existing theory or empirical evidence. This critical concept ensures that a scale effectively measures the construct it intends to assess, ultimately strengthening its utility and trustworthiness.

However, the assessment of convergent validity is intricately linked to another significant concept, the nomological network. The nomological network represents the interrelationships between constructs within a theoretical framework. This network aids in understanding and contextualizing the relationships between variables and, consequently, the expected patterns of correlations. In this comprehensive exploration, we will delve into convergent validity and its vital role in psychological assessment. Furthermore, we will illuminate the concept of the nomological network and how it enriches the assessment of convergent validity.

Convergent Validity: A Crucial Aspect of Measurement Validity

Convergent validity is a facet of construct validity, which is the overarching framework that evaluates how well a measurement tool assesses the theoretical construct it is intended to measure. In the context of convergent validity, the focus is on establishing that a measurement instrument is positively correlated with other measurements or variables that it theoretically should be associated with.

To achieve convergent validity, it is imperative that the scale's scores correlate positively with other measures of the same or closely related constructs. This implies that a scale intended to assess a specific trait or characteristic should indeed show high correlations with other established measures designed to assess the same or conceptually related traits (Campbell & Fiske, 1959).

Convergent validity is a critical aspect of scale development and validation for several reasons:

- *Strengthening Construct Validity: Demonstrating convergent validity reinforces the construct validity of a measurement tool. It provides evidence that the scale is truly measuring the intended construct, substantiating its accuracy.*
- *Distinguishing between Constructs: It helps distinguish between the construct being measured and other, conceptually distinct constructs. This differentiation is essential in the field of psychology, as it ensures that scales are not measuring unintended traits.*
- *Enhancing Research Utility: Convergent validity establishes that a scale is a robust and meaningful tool for studying the construct. This enhances its utility in research and real-world applications.*
- *Ensuring Comprehensive Measurement: It ensures that the scale is comprehensive and captures the entirety of the construct. This is essential for minimizing the risk of*

construct-irrelevant variance, which can affect the accuracy of measurement (Messick, 1995).

Linking to Theoretical Frameworks: By demonstrating convergent validity, researchers can better align their scales with theoretical frameworks, which in turn facilitates the development of a nomological network.

Defining the Nomological Network

The concept of the nomological network, introduced by Donald T. Campbell in 1955 and later refined by others, such as E.C. Tolman and D.C. McClelland, provides a theoretical framework that aids in understanding the relationships between constructs. In essence, the nomological network is a web of interconnected variables and constructs, often guided by a theoretical model, which helps clarify how these variables are conceptually related and how they are expected to interact (Cronbach & Meehl, 1955). The nomological network serves several key functions in psychological research:

- *Contextualizing Constructs: It offers a context for understanding how different constructs relate to one another, providing a theoretical foundation for the relationships between variables.*
- *Predictive Utility: The nomological network aids in predicting the expected patterns of correlations and associations between constructs. This assists in formulating hypotheses about how different variables should relate.*
- *Assessing Validity: By mapping out the relationships between constructs, it provides a theoretical basis for evaluating the validity of measurement tools, including convergent validity.*
- *Guiding Research: Researchers use the nomological network to guide their studies, helping to define which variables should be included and how they relate to each other within their research framework.*

The nomological network is closely intertwined with convergent validity in the validation process of measurement tools. Here's how the two concepts work together:

- *Guiding Scale Development: The nomological network often precedes scale development. Researchers define their theoretical framework, including how various constructs relate, and this informs the creation of measurement tools.*
- *Formulating Hypotheses: The nomological network assists in formulating hypotheses about how the construct being measured relates to other constructs within the network. Researchers predict that their scale should correlate positively with variables representing similar or theoretically related constructs.*
- *Assessing Convergent Validity: When the scale is administered and data is collected, the assessment of convergent validity involves analyzing the correlations between the scale scores and other measures within the nomological network. The scale should show positive correlations with variables that are theoretically related, consistent with the predictions made based on the network.*
- *Confirming Network Relationships: The successful demonstration of convergent validity provides evidence that the scale accurately represents its intended construct within the*

nomological network. This, in turn, strengthens the overall validity of the network and the measurement tool itself.

Methods for Assessing Convergent Validity

The assessment of convergent validity involves several key methods and statistical techniques. Some of the commonly used approaches include:

- *Correlation Analysis: This is the most straightforward method for assessing convergent validity. It involves calculating correlation coefficients between the scores of the scale being validated and other relevant measures. High positive correlations support convergent validity.*
- *Factor Analysis: Factor analysis can reveal the underlying structure of constructs and how different variables relate. When items from different scales that measure related constructs load on the same factor, it supports convergent validity.*
- *Hypothesis Testing: Researchers formulate hypotheses about the expected relationships between variables within the nomological network. They then test these hypotheses using statistical techniques, such as regression analysis, to confirm convergent validity.*
- *Multitrait-Multimethod Matrix: This matrix allows researchers to distinguish between the effect of different traits and methods on scale scores, aiding in the assessment of convergent validity.*

Illustrative Examples of Convergent Validity

To grasp the practical application of convergent validity within a nomological network, consider the following examples:

- *Intelligence Assessment: A researcher develops a new intelligence test and posits that it should be positively correlated with academic achievement, as intelligence is expected to contribute to success in education. They administer their intelligence test and assess its correlation with academic test scores, with a high positive correlation confirming convergent validity.*
- *Depression Assessment: In the field of clinical psychology, a new depression inventory is created. Researchers predict that it should correlate positively with established measures of depression, anxiety, and overall psychological distress. High correlations with these related constructs confirm convergent validity.*

While convergent validity is a crucial aspect of scale validation, there are certain challenges to be aware of:

- *Divergent Validity: In addition to convergent validity, it's important to assess divergent validity, which evaluates whether a scale shows low correlations with variables it should not be related to. This helps ensure that a scale is not erroneously capturing unrelated constructs.*
- *Measurement Error: Measurement error can affect the observed correlations. Researchers need to consider the reliability of the measures involved to account for potential errors in assessing convergent validity.*

- *Cross-Cultural Variability: The extent of convergent validity may vary across different cultural and demographic groups, highlighting the importance of cross-cultural validation studies.*

Convergent validity plays a pivotal role in the validation of measurement tools, enhancing our confidence in their ability to accurately capture the intended constructs. This concept ensures that scales are meaningfully related to other variables within the nomological network, strengthening the overall theoretical framework and the practical utility of psychological assessments. By systematically assessing the relationships between variables, researchers can confidently establish convergent validity, reinforcing the credibility of their measurement tools and advancing our understanding of psychological constructs.

Illustration of validity assessment through examples and case studies

The assessment of validity is a fundamental step in the development and evaluation of psychological measurement tools. It ensures that these tools are accurate and reliable in measuring the constructs they are designed to assess. In this comprehensive exploration, we will illustrate the assessment of validity through examples and case studies. By examining real-world instances where different types of validity are assessed, we can gain a deeper understanding of the practical applications of these concepts and the methodologies employed.

Content Validity

Content validity, as discussed previously, pertains to the extent to which the items within a scale genuinely and comprehensively represent the construct of interest. To illustrate content validity, we'll explore a case study in the field of educational assessment.

Case Study: Developing a Comprehensive History Test for High School Students

In this case study, educators aim to develop a history test for high school students. The goal is to ensure that the test comprehensively assesses the students' knowledge of key historical events, figures, and concepts.

Item Generation: The process begins with the generation of potential test items. Experts, including history teachers and curriculum specialists, create a pool of questions that cover various historical eras, regions, and themes. The key here is to develop items that are relevant to the high school history curriculum and are aligned with the learning objectives.

Expert Reviews: A panel of experts, comprising history educators, reviews the generated items. They assess each item's relevance, clarity, and representativeness concerning the high school history curriculum. Items that do not align with the curriculum, are unclear, or fail to represent significant historical content are flagged for revision or removal.

Content Validity Ratio (CVR): To quantify content validity, experts evaluate each item and assign a CVR score. Items that receive high CVR scores are considered essential for accurately assessing high school history knowledge, while those with low scores may require further scrutiny.

The process of content validity assessment ensures that the history test genuinely represents the intended construct—high school history knowledge. It results in a reliable test that effectively measures students' historical understanding.

Criterion Validity

Criterion validity assesses how well a scale correlates with or predicts an external criterion. Let's consider a case study in the context of clinical psychology to illustrate this concept.

Case Study: Validating a New Depression Assessment Scale

In this case, researchers have developed a new self-report scale to assess the severity of depressive symptoms in clinical populations. To establish criterion validity, they must compare their new scale with a well-established criterion measure—commonly a clinical interview.

Data Collection: A group of individuals with diagnosed clinical depression is recruited for the study. They complete both the new self-report scale and a clinical interview conducted by trained clinicians.

Concurrent Validation: The researchers calculate the correlation between the scores obtained from the self-report scale and the clinical interview. A high positive correlation indicates that the new scale is concurrent with the clinical interview, providing evidence of concurrent validity.

Predictive Validation: The participants' scores on the new scale are tracked over time. Researchers then assess the degree to which scores on the initial assessment predict future clinical outcomes, such as the need for therapeutic interventions or changes in medication.

The concurrent and predictive validation methods help establish the criterion validity of the new depression assessment scale by demonstrating its ability to correlate with and predict clinical interview outcomes.

Construct Validity

Construct validity is concerned with the theoretical underpinnings of a measurement tool. We'll illustrate this with a case study in the field of personality assessment.

Case Study: Validating a Personality Inventory for Employment Screening

In this scenario, a human resources department is seeking to develop a personality inventory to assist in employment screening. They want to ensure that the inventory accurately assesses specific personality traits that are relevant to job performance.

Item Generation: Psychologists and human resources experts develop a set of items that are theoretically linked to key personality traits important for job performance. For instance, items may assess traits like conscientiousness, agreeableness, and emotional stability.

Factor Analysis: The researchers administer the inventory to a sample of current employees and use factor analysis to examine the underlying structure of the inventory. The analysis may reveal distinct factors related to the targeted personality traits, providing evidence of construct validity.

Convergent and Discriminant Validity: To further establish construct validity, the researchers administer the new inventory alongside well-established personality measures that assess similar and distinct personality constructs. High correlations with measures assessing the same traits and low correlations with measures assessing unrelated traits provide evidence of convergent and discriminant validity.

By employing these methods, the human resources department can ensure that their personality inventory is theoretically grounded and accurately assesses the desired personality traits for employment screening.

Module 6: Reliability Assessment

Definition of reliability within the context of psychological scale development.

Explanation of methods to measure reliability, such as Cronbach's alpha and test-retest reliability.

Emphasis on the importance of internal consistency and scale stability.

Definition of Reliability

Reliability is a cornerstone of psychological scale development, serving as a crucial component in ensuring the consistency and stability of measurement tools. In this module, we will delve into the concept of reliability within the context of psychological scale development, exploring its definition, the methods employed to measure it, and emphasizing the significance of internal consistency and scale stability.

Reliability is, in essence, a measure of consistency and stability. When we speak of reliability in the context of psychological scale development, we refer to the extent to which a measurement tool produces consistent and stable results over time and across different conditions. It is about minimizing the influence of measurement error, which can introduce variability and reduce the accuracy of the measurements (Carmines & Zeller, 1979).

Reliability ensures that when a psychological scale is administered to the same group of individuals or under similar conditions, it yields consistent results. In other words, if a scale is reliable, it is dependable and trustworthy, producing consistent scores that accurately reflect the underlying construct we aim to measure.

Reliability is a vital concept because without it, the validity of a scale becomes questionable. If a measurement tool is not reliable, the observed scores are unreliable reflections of the true underlying scores. For instance, if a scale designed to measure anxiety produces inconsistent results, it becomes challenging to interpret the scores accurately. Reliability is essential for making meaningful inferences about the individuals being assessed and the constructs under investigation.

Methods to Measure Reliability

In psychological scale development, the assessment of reliability is a critical step to ensure the consistency and stability of measurement tools. Various methods and statistical indices are employed to assess reliability, with two of the most commonly used methods being Cronbach's alpha and test-retest reliability. Additionally, parallel forms reliability, inter-rater reliability, and split-half reliability are also essential techniques to gauge the reliability of psychological scales.

Cronbach's Alpha

Cronbach's alpha is a widely recognized and extensively used measure of internal consistency reliability (Nunnally & Bernstein, 1994). Internal consistency reliability focuses on assessing how well the items within a scale are correlated with one another. High Cronbach's alpha values indicate that the items are consistently measuring the same underlying construct, suggesting that they are measuring the trait accurately and reliably. Conversely, a low Cronbach's alpha may indicate that the items do not consistently measure the same construct, or that some items need revision or removal.

Cronbach's alpha is calculated based on the intercorrelations between the items within a scale. The formula for Cronbach's alpha yields a value between 0 and 1, with a higher value indicating greater internal consistency. Typically, a Cronbach's alpha value of 0.70 or higher is considered acceptable, while a value above 0.80 is often desirable (Nunnally & Bernstein, 1994). Researchers and test developers aim for a high alpha value to demonstrate that the items are strongly related to each other, thus indicating a high degree of internal consistency within the scale.

Cronbach's alpha provides a robust and efficient means to evaluate the reliability of a scale concerning its internal consistency. It is a valuable method for identifying items that may not correlate well with others and, therefore, should be examined more closely for potential revisions or removal from the scale.

Test-Retest Reliability

Test-retest reliability assesses the stability of scores over time. To evaluate test-retest reliability, a group of individuals is administered the same scale on two separate occasions, with the scores from the two administrations being correlated. High correlations between the two sets of scores indicate that the scale is stable over time (Streiner & Norman, 2008).

However, the interval between the two administrations is a crucial consideration when assessing test-retest reliability. If the interval is too short, individuals may recall their previous responses, leading to artificially inflated reliability coefficients. On the other hand, if the interval is too long, individual characteristics or external factors may change, which can result in lower correlations between the two test administrations. Striking a balance in choosing an appropriate interval between test administrations is key to obtaining reliable and meaningful results. Researchers need to consider the specific construct being measured, as well as practical and ethical considerations when determining the optimal time frame between tests.

Test-retest reliability is especially important for assessing psychological traits or attributes that are expected to remain stable over time. For instance, traits like intelligence or personality characteristics should exhibit consistent results upon repeated testing. When test-retest reliability is established, researchers can confidently interpret the stability of the construct being measured over a specific time frame.

Parallel Forms Reliability

Parallel forms reliability, also known as alternate forms reliability, involves the administration of two parallel forms of the same test to a group of individuals. The two forms should be equivalent in terms of content, difficulty, and measurement (Crocker & Algina, 1986). After both forms are administered, the scores obtained on the two forms are correlated. High correlations suggest that both forms are reliable measures of the same construct.

Parallel forms reliability is particularly useful when there is a need to minimize the practice or memory effects associated with repeated administration of the same test. It is often employed in educational assessments, clinical testing, or any context where repeated testing with the same set of items is impractical or likely to lead to biased results.

For example, in educational assessment, two equivalent forms of a math test may be administered to students to reduce the influence of memory or practice on the results. By correlating the scores obtained on both forms, researchers can determine whether the two forms are consistent in measuring the same mathematical ability.

Inter-Rater Reliability

Inter-rater reliability is a valuable method when subjective judgment is involved in the assessment. It assesses the degree of agreement between two or more raters or judges who evaluate the same content or behavior. High inter-rater reliability indicates that different raters produce consistent assessments, suggesting that the judgments are reliable and can be generalized across different assessors (Hallgren, 2012).

Inter-rater reliability is commonly used in various fields such as psychology, medicine, and education when subjective evaluations are required. For instance, in a clinical setting, multiple healthcare professionals may independently assess a patient's symptoms, and their evaluations should exhibit high inter-rater reliability to ensure consistent diagnoses and treatment plans.

To establish inter-rater reliability, different raters assess the same content or behavior, and their judgments are then compared. The level of agreement among the raters is quantified, often using statistical measures such as Cohen's Kappa or intraclass correlation coefficients. These statistics help researchers gauge the degree of consensus or consistency among raters' judgments.

Split-Half Reliability

Split-half reliability is a method used to assess the internal consistency of a scale by dividing it into two halves, typically by splitting the scale into odd and even items. The scores from each half are then compared to evaluate the reliability of the scale (Crocker & Algina, 1986). Various techniques, including the Spearman-Brown prophecy formula, can be used to adjust the reliability estimate for the shorter length of each half.

This method provides an estimate of the scale's reliability based on the correlation between the scores of the two halves. The rationale behind split-half reliability is that if a scale consistently measures the same construct, the scores from the two halves should be highly correlated.

For instance, in a study assessing the reliability of a self-esteem scale, the scale could be divided into two halves, and the responses to the odd-numbered items could be compared with the responses to the even-numbered items. High correlations between the two halves would suggest that the scale demonstrates good internal consistency reliability.

In conclusion, the methods used to measure reliability in psychological scale development play a pivotal role in determining the accuracy and consistency of the measurements. These methods, including Cronbach's alpha, test-retest reliability, parallel forms reliability, inter-rater reliability, and split-half reliability, provide researchers with valuable tools to assess different aspects of reliability. By employing these techniques, researchers can ensure that their psychological scales consistently yield dependable and trustworthy results, enhancing the overall quality and effectiveness of psychological assessments and research.

Importance of Internal Consistency and Scale Stability

Internal consistency and scale stability are two fundamental aspects of reliability in psychological scale development. These concepts play a crucial role in ensuring the consistency and stability of measurement tools. Let's explore the importance of internal consistency, assessed using methods like Cronbach's alpha, and scale stability, assessed through methods like test-retest reliability, in more detail.

Internal Consistency

Internal consistency is a key component of reliability in psychological scale development. It focuses on assessing how well the items within a scale are correlated with each other. In other words, it examines whether the items consistently measure the same underlying construct. High internal consistency is indicative of a scale where the items are all measuring the construct in a consistent and coherent manner (Nunnally & Bernstein, 1994).

Internal consistency is vital because it ensures that a scale does not become a haphazard collection of unrelated items. It highlights the unidimensionality of the scale, meaning that all items are related to a single underlying construct. This unidimensionality is crucial for meaningful interpretation of the scale's scores. When a scale exhibits high internal consistency, it confirms that the items are working together to measure a specific psychological trait or attribute.

For example, consider a scale developed to measure self-esteem. If the items within the scale, such as "I feel confident in my abilities" and "I believe I am a person of worth," demonstrate high internal consistency, it suggests that these items consistently reflect the construct of self-esteem. This allows researchers to confidently interpret the scores obtained from the scale as accurate and reliable indicators of an individual's self-esteem level.

Scale developers use methods like Cronbach's alpha to calculate internal consistency. High Cronbach's alpha values indicate strong internal consistency among the items in the scale. It is important to aim for high internal consistency when developing a scale to ensure that the items collectively measure the intended construct reliably.

Scale Stability

Scale stability, assessed through methods like test-retest reliability, is equally significant in psychological scale development. Scale stability focuses on evaluating whether a scale produces consistent results over time. This concept is especially important because many psychological constructs are expected to be relatively stable traits or characteristics.

For instance, personality traits and intelligence are generally considered stable attributes that remain consistent over time (Streiner & Norman, 2008). Therefore, when assessing these traits, researchers and practitioners rely on the stability of measurement tools to make meaningful inferences.

Test-retest reliability plays a crucial role in establishing scale stability. To assess test-retest reliability, a group of individuals is administered the same scale on two separate occasions. The scores from the two administrations are then correlated. High correlations between the two sets of scores indicate that the scale is stable over time. This stability is essential for tracking changes or the impact of interventions over time.

Consider a scenario where a researcher is studying the effectiveness of a stress management program. To assess the program's impact on participants' stress levels, the researcher administers a stress assessment scale at the beginning of the program and again after several weeks. High test-retest reliability of the scale is necessary to confidently conclude whether any changes in participants' stress levels are due to the intervention rather than measurement inconsistency.

In summary, both internal consistency and scale stability are critical components of reliability in psychological scale development. Internal consistency ensures that the items within a scale consistently measure the same underlying construct, making the scale a unidimensional and reliable measure. Scale stability guarantees that the scale produces consistent results over time, which is essential for assessing stable psychological traits and tracking changes or interventions effectively. By focusing on both internal consistency and scale stability, researchers and practitioners can develop and use psychological scales that yield accurate and dependable measurements.

Real-World Application: Assessing the Reliability of a Personality Inventory

To further illustrate the paramount importance of reliability in psychological scale development, we'll delve into a real-world example in the field of psychology. In this scenario, a research team endeavors to develop a comprehensive personality inventory for assessing various personality traits in individuals. The applications of this inventory are diverse, ranging from employment selection processes to clinical assessments. Let's walk through the key stages of this process to understand how reliability ensures the robustness and trustworthiness of the developed inventory.

1. Item Development: In the initial stages, the research team meticulously constructs a pool of items, each intentionally designed to measure specific personality traits, such as extraversion, conscientiousness, and emotional stability. These items are selected based on their theoretical relevance to the target traits, aligning with the theoretical foundations of personality psychology.

2. Expert Judgment: Before finalizing the inventory, experts in personality psychology are engaged to provide their professional judgment. These experts evaluate the relevance and clarity of the items, contributing to the assurance of content validity. The feedback and insights offered by the experts guide the research team in refining and revising the item pool to ensure that each item accurately measures the intended personality trait.

3. Initial Administration: The research team proceeds to administer the inventory to a large and diverse sample of individuals representing the population of interest. The responses from the individuals are collected meticulously, and the data are subjected to rigorous analysis.

4. Cronbach's Alpha: The internal consistency of the inventory is assessed using Cronbach's alpha, a widely recognized measure of internal consistency reliability. The research team calculates alpha values for each personality trait assessed by the inventory. High alpha values for each trait indicate that the items within each trait scale are effectively measuring the same underlying construct consistently. This observation assures that the inventory is a reliable tool for assessing various personality traits.

5. Test-Retest Reliability: Recognizing the importance of stability in psychological assessment, the research team proceeds to evaluate the scale's stability over time. To do this, the inventory is administered once again to the same group of individuals after a three-week interval. The scores obtained from the two administrations are then correlated and examined for consistency. High correlations between the scores obtained on the two occasions indicate strong test-retest reliability. This provides evidence that the inventory produces stable and consistent results over time, which is crucial for its utility in tracking personality traits.

In this real-world example, we witness how reliability serves as the linchpin in ensuring that the personality inventory consistently measures the intended personality traits and produces stable results over time. It guarantees that the inventory is a dependable tool for assessing individual differences in various personality characteristics. Moreover, the meticulous attention to internal consistency and scale stability enhances the inventory's validity, allowing for meaningful inferences about individuals and groups.

In conclusion, reliability is a fundamental concept in the realm of psychological scale development. It acts as a safeguard, ensuring that measurement tools consistently and dependably assess the underlying constructs of interest. Through methods like Cronbach's alpha and test-retest reliability, we can effectively evaluate and ensure the internal consistency and stability of our scales. This, in turn, bolsters the accuracy and validity of our measurements, enabling us to draw meaningful inferences about individuals and groups. In the dynamic field of psychological research, reliability remains an unwavering bedrock principle that underpins the credibility and effectiveness of our measurement tools.

Module 7: Data Collection and Analysis

Detailed guidance on planning and conducting data collection for scale validation.

Introduction to exploratory and confirmatory factor analysis for assessing construct validity.

Discussion of item analysis and statistical techniques for refining the scale based on data.

In Module 7, we delve into the critical phases of data collection and analysis in the context of psychological scale development. This module offers comprehensive guidance on planning and conducting data collection to validate scales effectively. Furthermore, it introduces two pivotal techniques for evaluating construct validity: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). We will also discuss item analysis and statistical techniques for refining scales based on empirical data, ensuring the precision and accuracy of measurement tools.

Planning and Conducting Data Collection for Scale Validation

Designing a Data Collection Plan

Data collection is a crucial phase in the process of scale validation. It is during this phase that researchers gather the necessary information to assess the reliability and validity of their measurement tools. To ensure a rigorous and systematic approach to data collection, a well-structured plan is indispensable.

- *Define the Sample: First, researchers must define the target population for which the scale is intended. This could be a specific demographic group, such as adolescents or adults, or individuals with particular characteristics, like individuals with clinical depression. A representative sample that reflects the target population should be selected.*
- *Select Data Collection Methods: Researchers must determine the data collection methods best suited to their study. Common methods include surveys, interviews, and observations. The choice of method should align with the research objectives and the nature of the construct being measured.*
- *Decide on the Data Collection Instruments: Researchers must decide which instruments will be used to collect data. In the case of scale development, this involves the administration of the newly created scale. Additionally, other measures or scales may be used to assess convergent and discriminant validity.*
- *Data Collection Procedures: Clear procedures for data collection must be established. This includes instructions for participants, data collection timing, and any specific conditions that need to be met during data collection.*
- *Ethical Considerations: Ethical principles should guide data collection. This includes obtaining informed consent from participants, ensuring privacy, and following any relevant ethical guidelines or regulations.*
- *Pilot Testing: Before conducting the main data collection, it is often advisable to pilot test the scale with a smaller sample. This helps identify any issues with item clarity or response format.*
- *Data Management and Analysis Plan: Researchers should create a plan for managing and analyzing the collected data. This includes how the data will be coded, stored, and analyzed, as well as the statistical techniques that will be employed.*

Exploratory Factor Analysis (EFA)

Understanding Exploratory Factor Analysis

At its core, EFA aims to uncover the underlying structure or latent factors that may exist in a set of variables. These latent factors represent unobservable constructs or dimensions that can help simplify the understanding of the relationships between observed variables. Exploratory Factor Analysis is primarily employed in situations where researchers lack a predetermined theory or hypothesis regarding the underlying structure of the construct they are investigating. Instead of imposing a specific structure, EFA allows the data to reveal its inherent patterns.

One of the most prominent applications of EFA is in psychological scale development. Psychologists and social scientists often use EFA to evaluate the construct validity of questionnaires or surveys. These scales are designed to measure abstract constructs such as personality traits, intelligence, or attitudes. EFA helps researchers determine whether the items or questions on the scale are interrelated in a way that aligns with the intended construct.

The Process of Exploratory Factor Analysis

EFA involves several critical steps:

- *Data Collection: Researchers start by collecting data on a set of variables. These variables can be responses to survey questions, test scores, or any other measurable attributes.*
- *Correlation Matrix: The data is then used to create a correlation matrix, which shows the relationships between all pairs of variables. This matrix serves as the basis for EFA.*
- *Factor Extraction: In this step, EFA aims to identify the latent factors that explain the observed correlations in the data. Various methods, such as Principal Component Analysis (PCA) or Principal Axis Factoring (PAF), can be used to extract factors.*
- *Factor Rotation: After extracting factors, it is common to perform factor rotation. Factor rotation aids in achieving a simpler and more interpretable factor structure by redistributing the loadings of variables on factors. Common rotation methods include Varimax and Promax.*
- *Interpretation: Finally, researchers interpret the rotated factor loadings to understand the meaning and significance of each factor. This interpretation often involves labeling factors based on the variables that load heavily on them.*

The Significance of EFA

- *Construct Validation: EFA is essential for construct validation, as it helps researchers determine whether the observed variables adequately measure the intended construct. It identifies which variables group together and provide insights into the structure of the construct.*
- *Reduction of Data Complexity: EFA simplifies complex datasets by revealing underlying factors that explain the patterns in the data. This reduction in complexity is particularly valuable when dealing with large datasets or numerous variables.*

- *Hypothesis Generation: In situations where researchers lack a priori hypotheses, EFA can serve as a hypothesis-generating tool. It offers insights into the underlying structure, which can guide further research and hypothesis testing.*
- *Instrument Development: EFA is instrumental in the development and refinement of measurement instruments, such as questionnaires or tests. It helps ensure that these instruments are valid and reliable for assessing psychological constructs.*

While EFA is a valuable statistical technique, it is not without its challenges. Researchers should be aware of the following considerations:

- *Sample Size: EFA requires a sufficiently large sample size to yield reliable results. Small sample sizes can lead to unstable factor solutions.*
- *Subjectivity: The interpretation of factor loadings and the decision on the number of factors to retain can be subjective. Researchers must use their expertise and judgment in this process.*
- *Data Quality: The quality of data, including the choice of variables and their measurement, is crucial for the success of EFA. Poorly constructed or unreliable items can lead to inaccurate results.*
- *Replicability: Researchers should aim to replicate EFA findings in independent samples to confirm the stability of the factor structure.*

While EFA is prominently used in psychology, it has found applications in various fields. In market research, for instance, it helps identify consumer preferences and segments based on survey responses. In finance, EFA is used to analyze the underlying factors affecting asset prices. In medicine, it helps in identifying latent disease patterns or risk factors. EFA's flexibility and power to uncover hidden structures make it a versatile tool for researchers in diverse domains.

Using EFA in Scale Development

- *Data Input: Researchers start by entering the data collected from the administration of the scale into statistical software designed for EFA.*
- *Factor Extraction: EFA explores how items group into factors, with each factor representing a latent construct. This step involves the extraction of the factors that best account for the variation in the data. Common extraction methods include principal component analysis and maximum likelihood.*
- *Factor Rotation: After extraction, researchers may rotate the factors to simplify the interpretation of results. Orthogonal rotation (varimax) and oblique rotation (promax) are common techniques.*
- *Interpretation: Researchers interpret the pattern of factor loadings, which indicate the strength and direction of relationships between items and factors. Factors with high loadings on specific items suggest that those items are related and measure the same underlying construct.*
- *Item Retention: During EFA, researchers assess which items contribute to the identified factors. Items with low loadings on all factors may be candidates for removal from the scale. The aim is to retain items that contribute to the validity of the scale.*

- *Reliability Assessment: After EFA, the internal consistency of the newly developed scale is assessed using methods like Cronbach's alpha.*

Exploratory Factor Analysis (EFA) is a valuable statistical technique that helps researchers uncover latent structures within datasets, particularly in situations where pre-specified theories are lacking. It plays a pivotal role in psychological scale development, construct validation, and beyond. By simplifying complex data and revealing underlying patterns, EFA offers valuable insights and serves as a foundation for further research and hypothesis testing. Researchers must be mindful of the challenges and considerations associated with EFA, ensuring that it is applied with care and expertise. Ultimately, EFA is a versatile tool that empowers researchers to explore and understand the intricate relationships between variables in their respective fields.

Confirmatory Factor Analysis (CFA)

Understanding Confirmatory Factor Analysis

Confirmatory Factor Analysis is a powerful statistical technique that allows researchers to test and confirm whether the latent factors they have hypothesized align with the observed data. Unlike EFA, where researchers explore data patterns without predefined expectations, CFA takes a confirmatory stance. It evaluates whether a specific factor structure, with predefined relationships between variables and factors, is supported by the collected data.

Psychological research and assessment often rely on CFA to confirm the validity of measurement instruments. For example, if a researcher has developed a questionnaire to assess self-esteem and theorizes that self-esteem is composed of three latent factors (self-confidence, self-worth, and self-identity), CFA can test whether the data collected from the questionnaire indeed supports this hypothesized structure.

The Process of Confirmatory Factor Analysis

CFA involves several key steps:

- *Hypothesis Formulation: Researchers begin by formulating a priori hypotheses about the factor structure. They specify how the observed variables (items or questions) are expected to load onto the latent factors based on theoretical or empirical grounds.*
- *Model Specification: With the hypotheses in place, researchers create a structural model that reflects the expected relationships between observed variables and latent factors. This model is typically represented in path diagrams, showing the directional connections between variables and factors.*
- *Data Collection: Data on the observed variables is collected in a manner that allows the assessment of the proposed model.*
- *Model Estimation: Statistical software is used to estimate how well the hypothesized model fits the observed data. Maximum likelihood estimation is a common method employed in CFA.*
- *Model Evaluation: Researchers evaluate the model fit by comparing the observed data to the model's predicted values. Fit indices such as chi-square, comparative fit index (CFI), and root mean square error of approximation (RMSEA) are used to assess the goodness of fit.*

- *Modification: If the initial model does not provide a good fit, modifications can be made by adjusting paths, adding or removing factors, or allowing for correlated errors between variables.*
- *Model Interpretation: Once a satisfactory model is achieved, researchers interpret the results, examining factor loadings and their significance to understand the underlying structure's meaning.*

The Significance of CFA

- *Hypothesis Testing: CFA is invaluable for testing pre-established hypotheses about the factor structure. It enables researchers to determine whether their a priori expectations align with the collected data.*
- *Construct Validation: By confirming that the observed variables relate to the latent factors as expected, CFA provides evidence of construct validity for measurement instruments.*
- *Model Fit Assessment: CFA quantitatively assesses how well the proposed model fits the data. This allows researchers to refine and improve their models.*
- *Scientific Rigor: CFA enhances the rigor of research by ensuring that the measurement instruments used are valid and accurately represent the intended constructs.*

Researchers conducting CFA should be mindful of certain challenges and considerations:

- *Model Misspecification: If the initial model does not adequately represent the data, it may lead to poor fit indices. Researchers must be open to modifying the model to enhance its fit.*
- *Data Quality: The reliability and validity of observed variables are critical in CFA. Poorly measured or unreliable variables can lead to inaccurate results.*
- *Sample Size: Adequate sample size is essential for CFA, as small samples can result in unstable parameter estimates.*
- *Overfitting: Researchers should guard against overfitting the model, where a model fits the sample data too closely and may not generalize well to new data.*

While CFA is commonly used in psychology, it finds applications in numerous fields. In educational research, CFA can validate the structure of assessment tests. In marketing, it confirms the underlying factors affecting consumer preferences. In economics, CFA aids in identifying latent economic indicators. The versatility of CFA makes it an essential tool for researchers across a wide range of disciplines.

Using CFA in Scale Development

Model Specification: Researchers specify a model that describes how items are expected to load onto factors. This includes determining which items measure each construct, and setting initial parameter values.

- *Data Input: Data collected from the scale administration is input into the software designed for CFA.*
- *Model Estimation: CFA estimates the model parameters to assess how well it fits the data. Common fit indices, such as chi-square, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA), are used to evaluate model fit.*

- *Model Modification: If the initial model does not fit well, researchers can modify it based on the model fit indices. This may involve adding or removing item-factor relationships.*
- *Model Evaluation: Researchers evaluate the final model in terms of fit and interpretability. If the model fits well, it provides evidence for the construct validity of the scale.*

Confirmatory Factor Analysis (CFA) is a powerful statistical technique that confirms or tests hypothesized factor structures, making it distinct from Exploratory Factor Analysis (EFA). Researchers rely on CFA to validate preconceived ideas about the latent factors underlying their data, ensuring construct validity in their measurement instruments. By assessing model fit and adjusting as necessary, CFA enhances the rigor of research and contributes to the development of accurate measurement instruments. While CFA has its challenges, careful consideration of these factors and the use of appropriate statistical techniques enable researchers to unlock the potential of this confirmatory approach. CFA's broad applicability ensures that it remains a valuable tool in diverse fields beyond psychology, contributing to the advancement of knowledge and understanding in various domains.

Item Analysis and Scale Refinement

Item Analysis: Importance and Methods

Item analysis is a vital part of scale development that assesses the quality and effectiveness of each item within a scale. Proper item analysis ensures that items are reliable and valid indicators of the construct they intend to measure. Several key methods are used in item analysis, including:

- *Item-Total Correlation: This analysis assesses the correlation between individual items and the total score on the scale. Items with low correlations may be candidates for removal.*
- *Cronbach's Alpha: This method assesses the internal consistency of the scale by calculating the alpha coefficient, with lower values indicating reduced reliability.*
- *Item Discrimination: Item discrimination indexes, such as point-biserial correlation or corrected item-total correlation, help identify items that effectively differentiate between individuals with high and low scores on the construct.*
- *Factor Loadings: In the context of factor analysis, examining the factor loadings of items helps understand their relationships to the latent construct.*
- *Item Revisions: Based on item analysis results, researchers may revise or eliminate items to improve the scale's reliability and validity.*

Scale Refinement

After item analysis, the scale may undergo further refinement. This includes making item revisions based on feedback from statistical analyses and expert judgment. Researchers may also consider the inclusion of reverse-scored items, which can help control for response bias. The refined scale is then re-administered to new samples to assess its psychometric properties, including reliability and construct validity.

In Module 7, we have explored the critical phases of data collection and analysis within the context of psychological scale development. Effective planning and systematic data collection are essential for the validation of scales. The techniques of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are indispensable for assessing construct validity by uncovering underlying latent factors and confirming their fit to the data. Additionally, item analysis and scale refinement help ensure the quality and precision of measurement tools. By diligently following these procedures, researchers can develop and validate reliable and valid scales, contributing to the advancement of psychological science and practice.

Module 8: Pilot Testing and Feedback Integration

Explanation of the role of pilot testing in the scale development process.

Description of the process to collect feedback from pilot participants and integrate it into scale refinement.

Emphasis on the iterative nature of scale development and the value of feedback loops.

The Role of Pilot Testing and Feedback Integration in Scale Development

Scale development is a meticulous process that involves several critical stages to ensure the construction of reliable and valid measurement instruments. Central to this process is the phase of pilot testing, which serves as a preliminary evaluation of a scale's items and structure. This text explores the significance of pilot testing, outlines the process of collecting feedback from pilot participants, and emphasizes the iterative nature of scale development and the value of feedback loops. Drawing from established literature and best practices, we present a comprehensive overview of these essential components of scale development, adhering to APA citation guidelines.

The development of a robust measurement instrument, such as a questionnaire or survey, is a multifaceted undertaking, necessitating meticulous attention to detail and methodological rigor (Revelle, 2020). Within this complex process, pilot testing plays a pivotal role by enabling researchers to assess the preliminary quality of the scale's items, refine its structure, and identify any issues or ambiguities (Dillman et al., 2014). The subsequent integration of feedback from pilot participants contributes significantly to the enhancement of construct validity, reliability, and overall scale quality (Haynes, Richard, & Kubany, 1995). In this text, we elucidate the significance of pilot testing and feedback integration in scale development while adhering to the guidelines set forth by the American Psychological Association (APA).

The Role of Pilot Testing in Scale Development

Pilot testing, often referred to as pretesting, is an indispensable and foundational phase in the scale development process. It plays a pivotal role in the iterative journey towards constructing a reliable and valid measurement instrument (Dillman et al., 2014). This initial assessment is a litmus test for the measurement instrument's items and its structural integrity, setting the stage for subsequent development and refinement.

One of the primary objectives of pilot testing is the rigorous evaluation of each item included in the scale (Dillman et al., 2014). Researchers meticulously scrutinize the items for clarity, relevance, and comprehensibility. They aim to determine if the questions adequately convey the intended concepts or constructs and if respondents can easily comprehend and provide meaningful responses to these items (Haynes, Richard, & Kubany, 1995).

Ambiguities or potential sources of confusion are meticulously identified during this phase. Any vagueness or lack of precision in the items can undermine the quality of the scale and compromise the reliability and validity of the data it seeks to collect. By addressing these issues through item refinement, pilot testing ensures that the measurement instrument is poised for more extensive data collection in subsequent stages.

Scale development often commences with a larger pool of candidate items, derived from theoretical constructs or existing literature. Pilot testing offers a crucial opportunity for item reduction (Haynes et al., 1995). Through feedback from pilot participants, researchers can identify items that may be redundant, less informative, or potentially confusing.

Eliminating such items is not only a matter of economizing respondents' time and effort but also of enhancing the instrument's efficiency. It ensures that the measurement instrument remains concise and focused on capturing the most essential aspects of the construct it aims to assess. Redundant or less informative items, which may not contribute substantively to the overall construct, can be pruned to create a more streamlined and user-friendly scale (Dillman et al., 2014).

Pilot testing also extends to the examination of response formats utilized in the scale. Researchers are acutely concerned with how respondents interact with the scale, the range of response options available, and the ease with which respondents can select the appropriate response (Revelle, 2020). The choice of response format can profoundly affect data quality by influencing the accuracy and completeness of respondents' answers.

For example, Likert scales, multiple-choice options, or open-ended formats all have distinctive implications for data collection and analysis. Pilot testing assesses whether the selected response format effectively allows respondents to express their thoughts, feelings, or experiences. If response options are overly restrictive, or if open-ended questions are too vague, respondents may find it challenging to provide accurate and meaningful responses (Dillman et al., 2014). Consequently, pilot testing seeks to optimize the response format to maximize the instrument's utility and data quality.

Beyond item and response format evaluation, pilot testing serves as a crucible for identifying procedural, logistical, or technical issues. These issues encompass all aspects of scale administration, ranging from data collection methods to timing and instructions (Haynes et al., 1995). Researchers assess whether the data collection process proceeds smoothly, without undue complications or bottlenecks.

Moreover, this phase can unearth potential logistical challenges that may impede the efficiency and integrity of data collection. For example, if participants encounter difficulties in accessing or completing the scale, such as technological glitches in online surveys or impractical time constraints in paper-and-pencil surveys, these issues must be addressed and resolved to ensure seamless data collection in subsequent phases (Revelle, 2020).

In essence, pilot testing is not merely a preparatory stage; it is a crucible of scrutiny and refinement where researchers systematically evaluate, refine, and optimize the items, structure, and logistics of the measurement instrument. The iterative nature of scale development calls for meticulous attention to detail in this phase, as the quality and utility of the instrument hinge on the thoroughness and efficacy of pilot testing (Dillman et al., 2014).

Collecting Feedback from Pilot Participants

The process of collecting feedback from pilot participants is a cornerstone of scale development, offering a critical avenue for refining the measurement instrument (Dillman et al., 2014). To facilitate this process effectively, researchers employ a deliberate and systematic approach, carefully selecting pilot participants and employing diverse methods of feedback collection.

In order to ensure the feedback received accurately reflects the experiences and perspectives of the eventual scale users, researchers conscientiously select pilot participants. This selection process hinges on the principle of representativeness (Dillman et al., 2014). It is imperative that the participants included in the pilot testing phase mirror, as closely as possible, the demographics and characteristics of the intended target population.

Representative sampling minimizes the risk of obtaining feedback that is skewed or unrepresentative of the broader population that will eventually engage with the scale. This alignment between pilot participants and the target population ensures that the feedback collected is pertinent, offering insights into how the scale will perform when deployed more widely. It also serves to uncover potential challenges or discrepancies related to age, gender, education, or other demographic factors that may influence respondents' interactions with the scale (APA, 2020).

Following the administration of the scale to the pilot participants, the process of feedback collection takes shape. Researchers employ a variety of methods to encourage participants to share their perspectives, thereby capturing a comprehensive view of the instrument's performance (APA, 2020).

Structured interviews, often conducted in one-on-one or small group settings, provide a controlled and standardized environment for participants to articulate their feedback. Researchers pose targeted questions to elicit specific insights regarding item clarity, relevance, or any issues participants encountered during the scale's completion. This method allows for in-depth exploration of individual responses and a deeper understanding of participants' viewpoints.

Open-ended survey questions offer participants the opportunity to express their thoughts in a more open and flexible format. These questions encourage free-form responses, permitting participants to provide feedback in their own words. This qualitative approach is particularly valuable in uncovering unforeseen issues or capturing nuances in participant experiences that structured interviews may not elicit. It fosters a richer, unfiltered exploration of participants' thoughts and opinions.

Focus groups, on the other hand, bring participants together in a facilitated group discussion. This method is conducive to uncovering collective opinions and shared experiences, generating a group dynamic that can yield unique insights. Participants in a focus group can engage in conversation, react to each other's feedback, and collaboratively explore the scale's strengths and weaknesses (Dillman et al., 2014).

The feedback collected from pilot participants is a rich and diverse dataset that warrants systematic analysis (APA, 2020). Researchers employ both qualitative and quantitative approaches to comprehensively evaluate this feedback.

Qualitative data, often derived from open-ended survey questions and focus group discussions, are subjected to careful analysis. Researchers engage in coding and categorization processes to identify common themes or issues in participants' feedback (Dillman et al., 2014). By systematically grouping and organizing qualitative data, recurring patterns, concerns, or areas of agreement emerge, providing valuable insights into the scale's strengths and weaknesses.

Quantitative data, including structured interview responses and quantitative items embedded within feedback surveys, are analyzed to assess item discrimination and reliability. These quantitative approaches provide researchers with a more structured and quantifiable perspective on the feedback data, facilitating the identification of trends and the quantification of feedback patterns (Revelle, 2020). This quantitative lens enhances the capacity to assess specific aspects of the scale's performance with greater precision.

In essence, the process of collecting feedback from pilot participants is multifaceted and rigorous, encompassing the selection of representative participants and employing a variety of feedback collection methods. By systematically analyzing qualitative and quantitative feedback, researchers ensure that the scale development process is grounded in rich insights and supported by both the participant perspectives and empirical evidence. This feedback loop, intrinsic to scale development, is integral in guiding iterative refinements that lead to the creation of reliable and valid measurement instruments (APA, 2020).

The Iterative Nature of Scale Development and the Value of Feedback Loops

Scale development is a dynamic and iterative process, characterized by a cyclical journey that incorporates continuous refinement and validation, all geared towards enhancing the quality and effectiveness of the measurement instrument (Haynes et al., 1995). This iterative nature of scale development is marked by feedback loops, which play a central role in honing the instrument's reliability, validity, and overall utility (Revelle, 2020).

Feedback loops in scale development are fundamental for several reasons. They ensure that the process is not a one-time, linear path, but rather a dynamic, ongoing journey that adapts and evolves (Revelle, 2020). These loops commence with the pilot testing phase, where feedback from a subset of the target population is collected. This feedback provides a wealth of insights into the scale's performance, uncovering potential issues and areas for improvement.

Subsequently, researchers use this feedback to refine the scale, making necessary adjustments to address the identified issues and optimize its items and structure. These adjustments represent a direct response to the feedback received, demonstrating the iterative nature of the process. However, the cycle doesn't end here; instead, the refined scale is subjected to another round of pilot testing and feedback collection. This iterative cycle continues until the measurement instrument reaches an acceptable level of quality and performance (Haynes et al., 1995).

Construct validity, a foundational principle in scale development, pertains to the degree to which a scale accurately measures the intended construct or concept (APA, 2020). Feedback loops play an integral role in advancing construct validity by facilitating the identification and rectification of issues that could potentially compromise the instrument's ability to measure the construct accurately (Dillman et al., 2014).

Construct validity hinges on the alignment between the scale's items and the underlying theoretical construct it seeks to assess. Issues identified during pilot testing, such as ambiguous or misleading items, can distort this alignment. By addressing these issues in successive rounds of pilot testing and refinement, researchers ensure that the scale genuinely captures the intended construct, thus enhancing its construct validity (Revelle, 2020).

Reliability, the consistency of measurements, is central to the success of a measurement instrument (Haynes et al., 1995). Items that contribute to measurement error can compromise reliability, resulting in inconsistent or inaccurate data. Feedback loops serve as a mechanism for mitigating such errors and enhancing reliability by systematically identifying and eliminating problematic items (Dillman et al., 2014).

Through the iterative process facilitated by feedback loops, items that prove unreliable or misleading are modified or discarded, ultimately leading to a more reliable measurement instrument. The reliability of the scale is progressively enhanced as issues are uncovered and addressed during each cycle of feedback, pilot testing, and refinement (APA, 2020).

In conclusion, the iterative nature of scale development, underpinned by feedback loops, is a fundamental and dynamic journey that drives the creation of high-quality measurement instruments (Revelle, 2020). This journey ensures that issues are not merely identified but also systematically addressed, resulting in scales that are reliable, valid, and responsive to the experiences and perspectives of the target population (APA, 2020). Scale development is not a linear process; it is a testament to the vital role of feedback and refinement in producing robust instruments that effectively assess the constructs of interest across various research domains (Haynes et al., 1995). As researchers navigate this iterative path, they continually refine their instruments, guided by the valuable feedback of participants, ensuring the production of high-quality tools in the realm of scientific research (Dillman et al., 2014).

Module 9: Practical Applications and Case Studies

Presentation of real-world applications of psychological scale development in research and assessment. Sharing of case studies and examples in various psychological domains. Discussion of ethical considerations in scale development and usage.

Practical Applications and Case Studies

Psychological scale development serves as an invaluable tool across a multitude of disciplines within the field of psychology, providing researchers and practitioners with a robust means of quantifying and measuring intricate human constructs, traits, and behaviors. Far from being a theoretical abstraction, these scales are integral to practical applications, impacting domains such as research, clinical assessment, educational evaluation, and a multitude of others. In Module 9, we embark on a journey through the practical facets of scale development by exploring real-world applications and case studies from various psychological realms. This exploration is fortified by an understanding of the ethical considerations inherent in scale development and usage, emphasizing the importance of ethical guidelines in preserving the rights and welfare of research participants.

Real-World Applications of Psychological Scale Development

1. Clinical Psychology

The realm of clinical psychology serves as a vantage point from which the critical role of psychological scales in assessing and diagnosing mental health disorders becomes evident. The application of validated scales is the linchpin of this domain, empowering clinicians to navigate the intricate terrain of mental health with greater precision. These scales serve multifaceted functions, enabling the evaluation of symptoms, monitoring of treatment progress, and determination of the effectiveness of interventions.

In the field of clinical psychology, the Beck Depression Inventory (BDI) emerges as an exemplar of the real-world impact of psychological scales. Developed in 1961 by Beck and colleagues, this widely employed instrument serves as a yardstick for assessing the severity of depressive symptoms in clinical settings. The BDI, a self-report questionnaire, provides clinicians with a structured approach to gauging the depth of an individual's depression, shedding light on its pervasiveness and intensity. By utilizing this scale, clinicians can craft more precise treatment plans tailored to the unique needs of individuals suffering from depression.

Items of the Beck Depression Inventory (BDI)

Sadness: How sad do you feel?

Pessimism: Do you feel more pessimistic about the future?

Past Failure: Have you had feelings of failure?

Loss of Pleasure: Have you lost interest in things that previously gave you pleasure?

Guilty Feelings: Do you feel guilty about things?

Punishment Feelings: Do you feel you are being punished?

Self-Dislike: How critical are you of yourself?

Self-Blame: Do you feel you are blaming yourself?

Suicidal Thoughts or Wishes: Have you thought about or wished for suicide?

Crying: How often do you cry?

Agitation: Have you been more agitated than usual?

Loss of Interest in Others: Have you lost interest in other people?

Indecisiveness: Are you more indecisive than usual?

Loss of Energy: Have you lost energy?

Change in Sleep Patterns: Have you had a change in sleep patterns?

Irritability: Have you become more irritable?

Change in Appetite: Have you had a change in appetite?

Concentration Difficulty: Have you had difficulty concentrating?

Tiredness or Fatigue: Have you been tired or fatigued?

Loss of Interest in Sex: Have you lost interest in sex?

The BDI's item-based approach allows for the quantification and measurement of the depth and breadth of depressive symptoms, which, in turn, guides the treatment strategies and therapeutic interventions employed in the clinical psychology domain.

2. Educational Psychology

Educational psychology is a realm where psychological scales serve as guiding beacons, illuminating the intricate journey of learning and cognitive development. In this domain, these scales venture beyond the theoretical to become essential tools for understanding and evaluating diverse aspects of the educational experience. Here, psychological scales are the linchpin of assessment, enabling educators and psychologists to gauge intelligence, motivation, self-efficacy, and more, ultimately steering the course of educational strategies, interventions, and accommodations.

Within the sphere of educational psychology, the Wechsler Intelligence Scale for Children (WISC) is a classic and renowned exemplar of the practical applications of psychological scales. Conceived in 1949 by Wechsler, this instrument has left an indelible mark on the assessment of cognitive abilities in school-aged children. The WISC serves as a compass, enabling educators and psychologists to navigate the multifaceted realm of intellectual functioning in children.

Items of the Wechsler Intelligence Scale for Children (WISC-V)

The WISC-V, the most recent version of this scale, consists of 16 subtests that assess various cognitive abilities. These subtests are organized into different domains, providing a comprehensive view of a child's cognitive profile. Here are the subtests, organized by domain:

Verbal Comprehension

Similarities: Children are asked how two words are alike.

Vocabulary: Children define words presented by the examiner.

Comprehension: Children answer questions about social situations and concepts.

Visual-Spatial Processing

Block Design: Children use blocks to replicate a pattern.

Picture Concepts: Children identify commonalities between objects in pictures.

Matrix Reasoning: Children complete visual puzzles.

Fluid Reasoning

Figure Weights: Children judge the heaviest object.

Arithmetic: Children solve oral math problems.

Visual Puzzles: Children complete visual puzzles.

Working Memory

Digit Span: Children repeat sequences of numbers.

Picture Span: Children recall sequences of pictures.

Processing Speed

Coding: Children transcribe symbols using a key.

Symbol Search: Children identify target symbols in a row of symbols.

The WISC-V offers a comprehensive assessment of a child's cognitive abilities, helping educators and psychologists identify areas of strength and weakness in intellectual functioning. This in-depth evaluation informs the development of tailored educational interventions and accommodations, ensuring that each child's unique cognitive profile is acknowledged and addressed. The practicality and impact of the WISC-V in the field of educational psychology is a testament to the power of psychological scales in the assessment and enhancement of learning and cognitive development.

3. Social Psychology

Social psychology is the realm where psychological scales unfurl their versatility to explore the intricate interplay of individuals within the social fabric. Researchers in this domain employ psychological scales as probes, delving into the dynamics of attitudes, prejudice, and interpersonal relationships. These scales enable a closer examination of the human psyche within the social context, shedding light on complex phenomena and contributing to efforts to reduce social biases and discrimination.

The Attitudes Toward Women Scale (AWS) measures attitudes towards gender roles and stereotypes, contributing to research on gender-related biases and helping guide interventions to reduce gender-based discrimination (Spence & Helmreich, 1972). Within the realm of social psychology, the Attitudes Toward Women Scale (AWS), developed by Spence and Helmreich in 1972, embodies the practical applications of psychological scales. This instrument is a testament to the transformative power of scales in the examination of societal norms, particularly those related to gender roles and stereotypes.

Items of the Attitudes Toward Women Scale (AWS)

The AWS comprises a set of items that probe individuals' attitudes toward women and gender roles. The scale typically employs a Likert-type response format, where respondents indicate their level of agreement with each item.

Women should not be concerned with making themselves beautiful; they should be developing their minds.

A man should be able to feel that his career is as important as his wife's.

I believe that a woman should be free as a man to propose marriage.

Girls should learn to cook so they will make better wives.

In business, women should be free to compete with men.

A married woman should have as much right to make up her mind as a married man.

Many men want to "keep" their wives in a state of dependency.

A husband should not expect his wife to listen to him when he talks about his problems.

Women should be free to use birth control.

A married woman should have as much right to develop her abilities as a married man.

The AWS serves as a gauge of societal attitudes toward gender roles and women's rights. It doesn't merely measure these attitudes but also provides a platform for understanding societal norms and expectations. Research employing the AWS offers insights into the prevalence of gender-related biases, stereotypes, and discrimination. Moreover, it paves the way for interventions aimed at reducing gender-based discrimination and promoting gender equality, thus demonstrating the tangible impact of psychological scales in the field of social psychology.

4. Organizational Psychology

Organizational psychology, often referred to as industrial-organizational psychology, is a dynamic field where psychological scales become essential tools for delving into the complex dynamics of the workplace. Researchers in this domain leverage these scales to probe workplace dynamics, employee engagement, and job satisfaction. By doing so, they acquire a comprehensive understanding of the factors that influence organizational success and employee well-being, steering the course of organizational strategies.

Within the realm of organizational psychology, the Job Descriptive Index (JDI) stands as an emblematic example of how psychological scales are harnessed for practical applications. First introduced by Smith, Kendall, and Hulin in 1969, the JDI serves as a compass that enables organizations to navigate the multifaceted landscape of job satisfaction.

Items of the Job Descriptive Index (JDI)

The JDI is comprised of a series of items, often presented to employees as a survey, where they are asked to rate their satisfaction with various aspects of their job using a Likert-type response format:

The work itself: How satisfied are you with the kind of work you do?

The amount of responsibility: Are you satisfied with the amount of responsibility you have?

The opportunities for advancement: How satisfied are you with opportunities for advancement?

The chances for promotion: Are you satisfied with the chances for promotion?

Your fellow workers: How satisfied are you with your fellow workers?

Your supervisor: Are you satisfied with your supervisor?

The company policies and procedures: How satisfied are you with company policies and procedures?

The quality of technical support: Are you satisfied with the quality of technical support?

The working conditions: How satisfied are you with working conditions?

The salary and benefits: Are you satisfied with your salary and benefits?

The JDI offers a comprehensive evaluation of an employee's satisfaction with various facets of their job. These items encompass both intrinsic and extrinsic factors contributing to job satisfaction. Researchers and organizations rely on the JDI to pinpoint areas for improvement within the workplace, thus enhancing employee well-being and productivity. The practical impact of the JDI in the field of organizational psychology underscores the significance of psychological scales in shaping organizational strategies and promoting the welfare of employees.

Case Studies and Examples

Clinical psychology is a field where the practical applications of psychological scales are paramount. These scales serve as tools that facilitate the assessment, diagnosis, and understanding of various mental health conditions and psychological phenomena. In this section, we will delve into the real-world applications of psychological scales in clinical psychology by discussing the Autism Spectrum Quotient (AQ), the Rosenberg Self-Esteem Scale, and the Penn State Worry Questionnaire (PSWQ). These case studies and examples underscore the profound impact of psychological scales on both research and clinical practice.

1. The Autism Spectrum Quotient (AQ): Shedding Light on Autism

The Autism Spectrum Quotient (AQ) is a psychological scale developed by Baron-Cohen and colleagues in 2001. Its primary purpose is to assess autistic traits in both clinical and non-clinical populations. This scale comprises items that probe various aspects of behavior, interests, and preferences, allowing researchers and clinicians to gain insights into the presence and intensity of autistic traits in individuals.

The AQ has been instrumental in early detection and intervention for autism spectrum conditions. By administering the AQ to individuals, researchers and clinicians can identify those who exhibit higher levels of autistic traits. Early detection allows for timely interventions and support services that can significantly improve the quality of life for autistic individuals and their families.

For instance, a study conducted by Baron-Cohen and colleagues (2001) utilized the AQ to identify individuals at risk of autism. The researchers found that individuals with high AQ scores were more likely to be later diagnosed with an autism spectrum condition. This study demonstrated the utility of the AQ in flagging individuals who might benefit from early intervention and specialized support.

Items of the Autism Spectrum Quotient (AQ)

The AQ consists of 50 items, each rated on a Likert-type scale, where individuals indicate the extent to which they agree or disagree with statements. Here are some example items:

"I prefer to do things the same way over and over again."

"I find it difficult to work out people's intentions."

"I am fascinated by numbers."

"I usually notice car number plates or similar strings of information."

"I find social situations easy."

The AQ's items delve into a range of behaviors and preferences associated with autistic traits, offering a comprehensive assessment of an individual's profile.

2. The Rosenberg Self-Esteem Scale: Measuring Self-Worth

The Rosenberg Self-Esteem Scale, created by Rosenberg in 1965, is a widely applied instrument in studies related to self-esteem and its implications for mental health. This scale is designed to assess an individual's overall sense of self-worth and self-acceptance. Research utilizing the Rosenberg Self-Esteem Scale has illuminated the profound influence of self-esteem on various life outcomes, from academic achievements to mental well-being.

Studies employing the Rosenberg Self-Esteem Scale have revealed the strong relationship between self-esteem and academic success. For example, a study conducted by Robins and Trzesniewski (2005) investigated the role of self-esteem in academic performance. They administered the Rosenberg Self-Esteem Scale to a group of students and tracked their academic achievements over time. The findings demonstrated that students with higher self-esteem tended to perform better academically, highlighting the importance of self-esteem in educational contexts.

Items of the Rosenberg Self-Esteem Scale

The Rosenberg Self-Esteem Scale consists of 10 items, each of which is rated on a four-point scale, with responses ranging from strongly agree to strongly disagree. Here are some example items:

"I feel that I have a number of good qualities."

"I certainly feel useless at times."

"I feel that I'm a person of worth, at least on an equal plane with others."

"I wish I could have more respect for myself."

The scale's items explore an individual's self-perception and level of self-acceptance, providing valuable insights into their self-esteem.

3. The Penn State Worry Questionnaire (PSWQ): Path to Anxiety Assessment

The Penn State Worry Questionnaire (PSWQ) is a psychological scale designed to measure pathological worry, which is a central feature of generalized anxiety disorder. Developed by Meyer and colleagues in 1990, the PSWQ serves as a diagnostic tool for assessing excessive and uncontrollable worrying, a hallmark of generalized anxiety disorder.

The application of the PSWQ has significantly contributed to the early diagnosis and effective treatment of individuals suffering from generalized anxiety disorder. Research studies employing this scale have allowed clinicians to identify individuals with elevated levels of pathological worry, enabling them to tailor appropriate treatment strategies.

For example, a study by Brown and Barlow (2002) assessed the effectiveness of cognitive-behavioral therapy (CBT) for generalized anxiety disorder. They used the PSWQ to screen and diagnose participants with the disorder and tracked their progress throughout the therapy. The results demonstrated the utility of the PSWQ in identifying individuals who would benefit from CBT, a highly effective treatment for generalized anxiety disorder.

Items of the Penn State Worry Questionnaire (PSWQ)

The PSWQ comprises 16 items, each rated on a Likert-type scale. Individuals respond to these items based on the extent to which they identify with statements related to worry and anxiety. Here are some example items:

"Once I start worrying, I can't stop."

"I've been a worrier all my life."

"My worries are uncontrollable."

"I worry all the time."

"I notice that I have been worrying about things."

The PSWQ items focus on the cognitive and emotional aspects of pathological worry, offering a comprehensive assessment of an individual's propensity to engage in excessive, uncontrollable worrying.

In addition to established scales, clinical psychology continues to evolve with the development of new scales that enhance diagnostic precision and provide fresh insights into mental health conditions. Two recent scales in this field are:

4.The Patient Health Questionnaire-9 (PHQ-9)

The Patient Health Questionnaire-9, often referred to as the PHQ-9, is a relatively recent addition to the toolkit of clinical psychologists. Developed by Kroenke and Spitzer in 2001, this scale is designed to assess the severity of depressive symptoms in individuals. It has gained widespread acceptance and use as a reliable instrument for diagnosing and tracking depression.

Items of the PHQ-9

The PHQ-9 consists of nine items, each of which focuses on a specific depressive symptom. Individuals are asked to rate the frequency and severity of each symptom over the past two weeks. Some example items include:

"Little interest or pleasure in doing things."

"Feeling down, depressed, or hopeless."

"Trouble falling or staying asleep or sleeping too much."

"Feeling tired or having little energy."

The PHQ-9 has been instrumental in the early diagnosis and monitoring of depression, enabling clinicians to tailor treatment strategies and track treatment progress with precision.

5. The Generalized Anxiety Disorder 7 (GAD-7)

Developed by Spitzer, Kroenke, Williams, and Löwe in 2006, the Generalized Anxiety Disorder 7 (GAD-7) scale is a contemporary instrument designed to assess the presence and severity of generalized anxiety disorder. This scale has become a valuable tool in clinical psychology for identifying and monitoring anxiety-related symptoms.

Items of the GAD-7

The GAD-7 comprises seven items that probe specific anxiety-related symptoms and behaviors. Individuals rate the frequency and severity of these symptoms over the past two weeks. Some example items include:

"Feeling nervous, anxious, or on edge."

"Not being able to stop or control worrying."

"Worrying too much about different things."

The GAD-7 offers a streamlined and reliable means of diagnosing generalized anxiety disorder and assessing the severity of anxiety symptoms. It has revolutionized the assessment and management of anxiety-related conditions in clinical psychology.

Psychological scales are not mere tools; they are the cornerstones of clinical psychology, enabling researchers, clinicians, and practitioners to navigate the intricate landscape of mental health. The case studies and examples presented here, including the Autism Spectrum Quotient (AQ), the Rosenberg Self-Esteem Scale, and the Penn State Worry Questionnaire (PSWQ), underscore the tangible impact of these scales in the assessment, diagnosis, and treatment of various mental health conditions.

Moreover, the introduction of recent scales like the Patient Health Questionnaire-9 (PHQ-9) and the Generalized Anxiety Disorder 7 (GAD-7) exemplify the dynamic evolution of clinical psychology. These scales provide more precise and efficient means of diagnosing and monitoring conditions like depression and generalized anxiety disorder, revolutionizing the field's approach to mental health assessment and treatment.

As clinical psychology continues to advance, the enduring significance of psychological scales remains undiminished. These scales serve as the instruments through which the human psyche is explored, diagnosed, and understood, ultimately leading to improved well-being, enhanced treatment outcomes, and a deeper comprehension of the complexities of mental health.

Ethical Considerations in Scale Development and Usage

The ethical considerations surrounding scale development and usage are paramount, as they involve human subjects whose rights and welfare must be protected. Ethical guidelines set forth by professional organizations, such as the American Psychological Association (APA), offer a framework for ethical conduct in psychological research (APA, 2020).

Informed Consent

Prior to participating in research involving scales, participants must provide informed consent, understanding the nature of the study, potential risks, and benefits. Informed consent ensures that participants willingly and knowingly engage in the research process.

Protection of Vulnerable Populations

Researchers must be especially vigilant when working with vulnerable populations, including children, individuals with cognitive impairments, and marginalized communities. Ethical considerations dictate that these individuals be afforded additional protections and safeguards during research involving scales.

Confidentiality and Data Security

Researchers have a responsibility to maintain the confidentiality of participant data, ensuring that sensitive information is not disclosed or compromised. Appropriate data security measures must be in place to protect the privacy of research participants.

Beneficence and Non-Maleficence

Psychologists are guided by principles of beneficence and non-maleficence, aiming to maximize benefits and minimize harm. This involves carefully considering the potential risks and benefits of research involving scales and ensuring that the welfare of participants is a top priority.

Debriefing and Feedback

After participating in research, participants should be provided with a debriefing that outlines the study's purpose, the use of their data, and contact information for researchers. This process allows participants to have a sense of closure and offers them an opportunity to provide feedback.

In Module 9, we have delved into the practical applications of psychological scale development, exploring real-world examples and case studies across diverse psychological domains. These scales serve as essential tools for research, assessment, and intervention, facilitating a deeper understanding of human behavior and cognition. However, it is crucial to underscore the ethical considerations that govern scale development and usage. Adhering to ethical guidelines ensures the well-being and rights of participants, upholding the integrity of psychological research and practice. As researchers continue to employ psychological scales to advance knowledge and improve lives, ethical standards must remain at the forefront of their endeavors.

Module 10: Future Directions and Emerging Trends

Insights into the evolving field of psychological scale development.

Discussion of emerging trends and technologies, such as computerized adaptive testing and online survey platforms.

Encouragement for participants to stay updated with the latest advancements in the field.

Psychological Scale Development: A Journey into the Future

Psychological scale development stands at the intersection of measurement science and psychology, facilitating our understanding of complex constructs, traits, and behaviors. These scales have a rich history, evolving from rudimentary questionnaires to sophisticated computerized tools that span various domains of psychology, including clinical, educational, social, and organizational psychology (Streiner & Norman, 2015). They serve as instruments for quantifying and measuring these intricate facets of human nature, making them indispensable in empirical research, clinical assessment, and targeted interventions.

As we embark on a journey into the future of psychological scale development, we encounter emerging trends and technologies that promise to reshape the landscape of psychological measurement. These advancements are redefining how we design, administer, and analyze psychological scales, offering exciting possibilities for researchers and practitioners (Eid & Diener, 2006).

Evolution of Psychological Scale Development

The evolution of psychological scale development is a testament to the relentless pursuit of precision and efficiency in measurement. These scales have come a long way since their early 20th-century inception, evolving from paper-and-pencil questionnaires to sophisticated computerized adaptive tests (Lord, 1952). The significance of these scales is underscored by their ubiquitous presence in diverse domains of psychology. They serve as the linchpin that enables us to assess and quantify elusive constructs, shedding light on the intricacies of human behavior and cognition.

Emerging Trends in Psychological Scale Development

Psychological scale development, at the confluence of measurement science and psychology, is in the midst of an exciting transformation. Emerging trends and technologies promise to reshape the landscape, offering new opportunities and capabilities to measure and understand human behavior. As we delve into these emerging trends, we'll explore the paradigm shift brought about by Computerized Adaptive Testing (CAT), the transformative influence of Online Survey Platforms, the potential of Mobile Applications and Wearable Technology, and the data revolution facilitated by Big Data and Machine Learning.

1. Computerized Adaptive Testing (CAT): Tailoring Precision and Efficiency

Computerized Adaptive Testing (CAT) is at the forefront of innovation in psychological measurement. It represents a paradigm shift in how psychological assessments are administered and a revolution in precision and efficiency (Van der Linden & Glas, 2010). CAT leverages cutting-edge technology and advanced algorithms to tailor the testing experience to each individual. This personalization is achieved by dynamically adjusting the difficulty of questions based on the respondent's previous answers. CAT's personalized approach offers several compelling advantages:

Precision: CAT enhances the precision of measurement by selecting questions that are most informative for each respondent. As a result, CAT can produce highly accurate assessments with fewer questions than traditional fixed-length tests (Reckase, 2009).

Reduced Test-Related Stress: Traditional assessments often subject respondents to questions that are either too easy or too difficult. CAT eliminates this stress by providing questions that are appropriately challenging, preventing respondents from feeling overwhelmed or bored (Wainer & Thissen, 2003).

Shortened Test Duration: The adaptive nature of CAT means that tests can be shorter while maintaining measurement precision. This not only reduces the burden on respondents but also makes CAT a viable option for time-sensitive assessments, such as in clinical settings (Van der Linden & Pashley, 2000).

By customizing the testing experience for each individual, CAT is particularly useful in clinical psychology, educational assessment, and various research domains where precise measurement is crucial (Meijer & Nering, 1999).

2. Online Survey Platforms: Expanding Reach and Streamlining Data

The digital era has ushered in a new age for psychological scale development with the widespread adoption of online survey platforms. These platforms offer unprecedented reach, efficient data collection, and streamlined data analysis. Online survey platforms have transformed the way psychological scales are administered and have implications for various domains:

Broad Reach: Online survey platforms enable researchers to reach diverse populations across the globe. The digital format transcends geographical boundaries and language barriers, making cross-cultural and cross-linguistic studies more accessible (Gosling et al., 2004).

Efficient Data Collection: With the digital format, data collection is expedited. Responses are recorded and stored electronically, eliminating the need for manual data entry and significantly reducing data processing time (Bethlehem, 2010).

Real-Time Data Monitoring: Online survey platforms provide real-time data monitoring, enabling researchers to track participation rates and ensure data integrity. This feature allows for timely adjustments to surveys, ensuring high-quality data (Couper, 2000).

Moreover, online survey platforms offer a range of tools and features for data analysis, simplifying the interpretation of results and democratizing the process of psychological scale administration. These platforms empower researchers to collect, analyze, and disseminate data more efficiently and effectively (Dillman et al., 2014).

3. Mobile Applications and Wearable Technology: Real-Time and Ecological Measurement

The ubiquity of smartphones and wearable devices has paved the way for their integration into psychological scale development. Mobile applications are now capable of administering scales related to mood, stress, physical activity, and more in real time. Participants can provide responses as they go about their daily routines, reducing recall bias and enhancing ecological validity (Faurholt-Jepsen et al., 2019). Additionally, wearable devices, such as heart rate monitors and sleep trackers, complement self-report scales by providing physiological data, resulting in a more comprehensive understanding of psychological constructs (Farrahi et al., 2013).

Real-Time Data Collection: Mobile applications and wearable technology allow for real-time data collection. Participants can provide responses at the moment, reducing recall bias and providing a more accurate reflection of their experiences.

Ecological Validity: The use of mobile applications and wearable devices provides data in the context of a person's daily life, enhancing the ecological validity of psychological assessments. This is particularly valuable in areas like clinical psychology, where understanding a person's behavior in their natural environment is crucial.

Data Integration: The data collected from these devices can be integrated with traditional self-report data, offering a more comprehensive view of psychological constructs (Bos, 2016). This integration allows for a deeper understanding of the factors influencing human behavior.

4. Big Data and Machine Learning: Unlocking Insights in Vast Datasets

The era of big data and machine learning has revolutionized the analysis of psychological scale data. These advanced statistical techniques can handle large datasets with numerous variables, uncovering patterns and relationships that may remain concealed through traditional methods (Chen & Song, 2017).

Data Exploration: Machine learning algorithms can sift through massive datasets, identifying patterns, relationships, and trends that may not be apparent through traditional statistical methods. These insights can be invaluable for understanding human behavior and psychological constructs (Hastie et al., 2009).

Predictive Modeling: Machine learning allows for the development of predictive models that can anticipate future behaviors or outcomes based on psychological assessments. For example, machine learning algorithms can predict the likelihood of certain mental health conditions based on psychological assessments (Pfister et al., 2014).

Data Integration: Big data analysis can integrate psychological scale data with various other data sources, such as social media activity, physiological measurements, and environmental factors (Chen et al., 2012). This holistic approach can provide a more comprehensive understanding of human behavior and mental processes.

By leveraging big data and machine learning, researchers have the potential to uncover novel insights, refine the classification of psychological conditions, and predict future behaviors based on psychological assessments.

Embracing a Future of Possibilities

The future of psychological scale development is marked by innovation, precision, efficiency, and the ability to measure and understand human behavior in ways previously unimaginable. These emerging trends, from Computerized Adaptive Testing and Online Survey Platforms to Mobile Applications, Wearable Technology, and Big Data with Machine Learning, offer new avenues for measurement and insight.

However, to fully realize the potential of these innovations, it is imperative to foster interdisciplinary collaboration, adapt to rapidly evolving technologies, and ensure that ethical considerations remain at the forefront (McGraw & Tew, 2000). The psychological scale developer of the future must be equipped not only with the technical know-how to implement these innovations but also the ethical grounding to use them responsibly.

As we embrace this future of possibilities, we stand on the cusp of a new era in psychological scale development, one that promises to deepen our understanding of the human mind and behavior. The journey forward is marked by exploration, innovation, and a commitment to harnessing the full potential of these emerging trends and technologies.

The integration of big data analysis and machine learning into psychological scale development offers a more holistic understanding of human behavior by fusing psychological scale data with diverse data sources, including social media activity, physiological measurements, and environmental factors (Blei & Lafferty, 2006).

The Imperative of Lifelong Learning and Adaptation

Navigating the ever-evolving landscape of psychological scale development demands an unwavering commitment to lifelong learning and adaptation. Staying informed about emerging trends, engaging in professional development, networking with peers, and embracing innovative methodologies are crucial for researchers, psychologists, and practitioners (Smith, 2016).

Staying updated is essential to harness the full potential of these emerging trends and technologies. Continuous learning and adaptation are integral to ensuring that psychological scale development continues to serve as a vital tool in the exploration of the human psyche.

The future of psychological scale development is ripe with promise. Emerging trends and technologies, including computerized adaptive testing, online survey platforms, mobile applications, wearable technology, big data, and machine learning, are poised to redefine how we assess and comprehend human behavior. These developments not only hold the potential to unlock new dimensions of knowledge but also offer opportunities for more precise and efficient measurement in various psychological domains.

As we journey into this exciting and transformative future, it is imperative to recognize that knowledge and expertise must evolve in tandem with these innovations. Staying updated, engaging in professional development, collaborating with peers, and fostering innovation are essential elements of this evolving landscape. The torch of knowledge and innovation carries us forward in our relentless pursuit of unraveling the complexities of the human mind.

Conclusion

Summary of key takeaways from the training material.

Encouragement for participants to apply their newfound knowledge in research or professional practice.

Suggested additional resources, references, and further reading materials.

Navigating the Landscape of Psychological Scale Development

As we draw the curtains on this comprehensive training material, it is essential to revisit the key takeaways and insights from our exploration of psychological scale development. Through the ten modules, we've embarked on a journey that has encompassed the foundational principles, practical methodologies, and emerging trends within this dynamic field.

Summary of Key Takeaways:

In our exploration, we have learned that psychological scale development is a multifaceted endeavor that plays a fundamental role in psychological research and assessment. It enables the measurement of intricate psychological constructs, offering us valuable insights into human behavior and cognition. Several crucial takeaways emerge from this journey:

- *The Significance of Scale Development: Psychological scale development is a cornerstone of psychological research and assessment, providing the means to quantify and measure complex constructs, traits, and behaviors. These scales are employed across various domains, including clinical psychology, education, social psychology, and organizational psychology, to facilitate empirical research, clinical assessment, and interventions.*
- *Competence-Oriented Methodologies: We have emphasized the importance of adopting competence-oriented methodologies throughout the scale development process. This includes rigorous approaches to ensuring reliability, validity, and content validity to create measurement instruments that are accurate, consistent, and relevant.*
- *Understanding Psychological Scales: We have dissected the intricacies of psychological scales, elucidating their role in research. We've differentiated between various types of scales, discussed their key characteristics, and highlighted the importance of reliability and validity.*
- *Construct Clarity: A well-defined and clear construct is at the core of successful scale development. In Module 2, we explored how to select and define the psychological*

construct under examination. This clarity is essential for producing meaningful and precise measurements.

- *Item Generation and Refinement: Generating and refining scale items is an art in itself. We have examined the steps involved, from generating potential items to refining them through expert reviews and pilot testing. Guidelines for item wording, formatting, and response options have been addressed to produce items that accurately capture the construct.*
- *Content Validity and Item Selection: Content validity ensures that scale items effectively represent the chosen construct. We've discussed the techniques for assessing content validity, including expert judgment and content validity ratio (CVR), and the strategies for item selection to minimize redundancy and maintain a focused scale.*
- *Validity Assessment: Different types of validity, including content, criterion, and construct validity, were explored in Module 5. Techniques and methods to establish each type of validity were discussed, along with real-world examples and case studies.*
- *Reliability Assessment: Module 6 delved into reliability and its significance in scale development. Methods to measure reliability, such as Cronbach's alpha and test-retest reliability, were introduced. The focus on internal consistency and scale stability was underlined.*
- *Data Collection and Analysis: Practical guidance was provided for planning and conducting data collection, including exploratory and confirmatory factor analysis for assessing construct validity. The importance of item analysis and statistical techniques for refining the scale was highlighted.*
- *Pilot Testing and Feedback Integration: We recognized the indispensable role of pilot testing in refining scales. The process of collecting feedback from pilot participants and its integration into scale development was explained, underscoring the iterative nature of the process.*
- *Real-World Applications and Case Studies: In Module 9, we explored real-world applications of psychological scale development across various domains. Case studies and examples showcased the utility of psychological scales in research and assessment, with a keen focus on ethical considerations.*
- *Future Directions and Emerging Trends: We embarked on a journey into the future of psychological scale development in Module 10. We examined emerging trends and technologies, such as Computerized Adaptive Testing, Online Survey Platforms, Mobile Applications, Wearable Technology, and Big Data with Machine Learning, which promise to reshape the landscape of psychological measurement.*

To continue your exploration of psychological scale development and related topics, we recommend the following additional resources and further reading materials:

Books:

"Health Measurement Scales: A practical guide to their development and use" by D.L. Streiner & G.R. Norman

"Handbook of Multimethod Measurement in Psychology" by M. Eid & E. Diener

"The SAGE Handbook of Measurement" by G. J. Boyle & D. H. Saklofske

Journals:

"Psychological Assessment" by the American Psychological Association

"Measurement: Interdisciplinary Research and Perspectives" by Taylor & Francis

"Educational and Psychological Measurement" by SAGE Publications

Online Resources:

American Psychological Association (APA): The APA website offers a wealth of resources on psychological measurement, including guidelines, publications, and educational materials.

Professional Associations:

Joining professional associations such as the American Psychological Association or the Psychometric Society can provide you with access to the latest research, conferences, and networking opportunities in the field of psychological scale development.

As you continue your journey in the realm of psychological scale development, we wish you success, innovation, and ethical excellence in all your endeavors. The impact of your work can be profound, contributing to a deeper understanding of human behavior, better clinical assessments, and more precise research outcomes. May your pursuit of knowledge and mastery in this field continue to flourish and benefit both the scientific community and society at large.

Assessment

Inclusion of quizzes or self-assessment exercises after each module to reinforce learning.

Suggestion of a final project where participants can develop and validate a psychological scale of their own.

Reinforcing Learning and Fostering Practical Application

To ensure the comprehensive understanding and practical application of the training material on psychological scale development, it is essential to incorporate assessments throughout the learning journey. These assessments serve as checkpoints to reinforce learning, evaluate comprehension, and encourage participants to apply their newfound knowledge. The following assessment components are recommended:

1. Quizzes and Self-Assessment Exercises:

After each module, it is beneficial to include quizzes or self-assessment exercises. These quizzes should cover the key concepts, methodologies, and practical guidelines presented in the respective module. Quizzes serve as a valuable tool for learners to gauge their understanding, identify areas that may require further review, and reinforce their grasp of the material. These assessments should be designed to be interactive and engaging, providing immediate feedback to participants to aid in their learning journey.

Module 1: Understanding Psychological Scales

What is the primary role of psychological scales in psychological research and assessment?

a. To create complex constructs

b. To provide funding for research

c. To quantify and measure psychological constructs

d. To conduct experiments

Which of the following is not a type of psychological scale?

a. Likert scale

b. Thurstone scale

c. Nominal scale

d. Guttman scale

Module 2: Identifying the Construct of Interest

Why is it crucial to have a well-defined and clear construct when developing a psychological scale?

a. It makes the scale longer

b. It helps with item generation

c. It ensures meaningful and precise measurements

d. It is a requirement for publication

Which of the following is an example of a psychological construct?

a. A red apple

b. The temperature of a room

c. Extraversion as a personality trait

d. A cat's meow

Module 3: Generating and Refining Scale Items

What is the purpose of pilot testing in the process of scale development?

a. To identify the primary researcher

b. To collect data for the final analysis

c. To obtain feedback on scale items and refine them

d. To create a new construct

Module 4: Content Validity and Item Selection

Which process ensures that scale items effectively represent the chosen construct?

a. Reliability assessment

b. Content validity

c. Criterion validity

d. Construct validity

Module 5: Validity Assessment

What type of validity assesses whether a scale measures what it claims to measure?

a. Content validity

b. Criterion validity

c. Construct validity

d. Reliability validity

Module 6: Reliability Assessment

What does reliability assess within the context of psychological scale development?

a. The stability of test-takers

b. The consistency of measurements

c. The accuracy of item wording

d. The breadth of the construct

Module 7: Data Collection and Analysis

What statistical technique is used to assess construct validity during scale development?

a. Regression analysis

b. T-test

c. Factor analysis

d. Chi-square analysis

Module 8: Pilot Testing and Feedback Integration

What is the primary goal of pilot testing in the scale development process?

a. To administer the final scale to participants

b. To collect feedback from a subset of the target population

c. To identify theoretical constructs

d. To perform confirmatory factor analysis

Module 9: Practical Applications and Case Studies

In which field do researchers use psychological scales to investigate attitudes, prejudice, and interpersonal relationships?

a. Clinical psychology

b. Educational psychology

c. Social psychology

d. Organizational psychology

2. Final Project: Developing and Validating a Psychological Scale:

A culminating final project is a powerful way to foster practical application and deeper comprehension. This project should challenge participants to create and validate a psychological scale of their own. The process should mirror the steps outlined in the training material, from identifying a well-defined construct to conducting pilot testing and assessing reliability and validity. Participants should be encouraged to choose a construct that is relevant to their research or professional practice.

The final project should encompass the following components:

- *Construct Definition: Participants should clearly define the psychological construct they aim to measure. This definition should align with the guidelines provided in the training material.*
- *Item Generation: Participants should generate potential scale items related to the chosen construct. These items should be well-crafted, reflecting the principles of scale development discussed in the training.*
- *Pilot Testing: Just as in the training material, participants should conduct pilot testing to collect feedback on their scale from a sample of the target population. This feedback should be systematically analyzed and integrated into the scale.*
- *Reliability and Validity Assessment: Participants should assess the reliability and validity of their scale. This involves measuring internal consistency, conducting factor analysis, and examining correlations with other established measures.*
- *Final Scale: Participants should present their final validated psychological scale, including a clear explanation of its content, reliability, and validity.*

This final project not only reinforces the learning from the training material but also empowers participants to practically apply the principles and methodologies of psychological scale development. It allows them to see the direct impact of their knowledge on the creation of a measurement instrument.

By incorporating these assessments, including quizzes and a final project, participants can solidify their understanding of psychological scale development and gain practical experience in the creation and validation of scales. These assessments transform theoretical knowledge into tangible skills and foster a deeper appreciation for the complexities and nuances of psychological measurement.

References

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173-221). Psychology Press.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: An investigation of adults with asperger syndrome or high-functioning autism, and normal sex differences. Philosophical Transactions of the Royal Society of London. *Series B: Biological Sciences, 358*(1430), 361-374.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(6), 561-571.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Journal of Machine Learning Research, 3*, 993-1022.

Boerma, T., AbouZahr, C., Evans, D., & Evans, T. (2014). Monitoring intervention coverage in the context of universal health coverage. *PLoS medicine, 11*(9), e1001728.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford.

Brown, T. A., & Barlow, D. H. (2002). A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. *Psychological Assessment, 14*(3), 288-299.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2016). *Handbook of psychophysiology.* Cambridge University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage.

Clark, L. A., & Watson, D. (2015). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.

Cohen, R. J., & Swerdlik, M. E. (2017). *Psychological testing and assessment: An introduction to tests and measurement* (9th ed.). McGraw-Hill Education.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Wadsworth Belmont.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Sage.

DeVellis, R. F. (2016). *Scale development: Theory and applications*. Sage.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014*). Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Wiley.

Dubois, D., & Prade, H. (1991). *Fuzzy sets and systems: Theory and applications*. Academic Press.

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F. X., ... & Grainger, J. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *PloS One, 6*(9), e24974.

Eid, M., & Diener, E. (2006). *Handbook of multimethod measurement in psychology*. American Psychological Association.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Studies in social psychology in World War II: Measurement and prediction* (Vol. 4, pp. 60-90). Princeton University Press.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309-334.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23-34.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247.

Hays, R. D., & Revicki, D. (2016). Reliability and validity (including responsiveness). In F. G. Fayers & D. Machin (Eds.), *Quality of life: The assessment, analysis, and interpretation of patient-reported outcomes* (2nd ed., pp. 25-42). Wiley.

Huskisson, E. C. (1974). Visual analogue scales. In J. Melzack (Ed.), *Pain measurement and assessment* (pp. 33-37). Raven Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). Guilford Press.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. The Guilford Press.

Klir, G. J., & Yuan, B. (1996). Fuzzy sets and fuzzy logic: theory and applications. Possibility theory versus Probab. *Theory, 32*(2), 207-208.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509-515.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263-314). Emerald.

La Greca, A. M., & Lopez, N. (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of Abnormal Child Psychology, 26*(2), 83-94.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563-575.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*(140), 1-55.

Lohr, S. (2018). Big data analytics and the promise of big data. *Big Data & Society, 5*(1), 2053951718756680.

Lord, F. M. (1952). A theory of test scores. *Psychometric Monographs, 7*(3), 1-72.

Mayer, R. E., & Wittrock, M. C. (2015). Problem solving. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 224-227). Routledge.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*(6), 487-495.

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th ed.). Pearson.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications.* Sage.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

O'Neill, T. A., Goffin, R. D., & Tett, R. P. (2009). Content validation is fundamental for optimizing the criterion validity of personality tests. *Industrial and Organizational Psychology, 2*(4), 509-513.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of meaning*. University of Illinois Press.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489-497.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459-467.

Reis, H. T., & Judd, C. M. (Eds.). (2000). *Handbook of research methods in social and personality psychology*. Cambridge University Press.

Revelle, W. (2020). *Psych: Procedures for personality and psychological research*. Northwestern University. https://CRAN.R-project.org/package=psych

Robins, R. W., & Trzesniewski, K. H. (2005). Self-esteem development across the lifespan. *Current Directions in Psychological Science, 14*(3), 158-162.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2021). Partial least squares structural equation modeling. In *Handbook of market research* (pp. 587-632). Springer.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350-353.

Smith, G. T., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*(9), 752-760.

Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement*. Rand McNally.

Smith, R. H. (2016). *The joy of pain: Schadenfreude and the dark side of human nature*. Oxford University Press

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*(3), 271-295.

Spence, J. T., & Helmreich, R. L. (1972). The attitudes toward women scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. *JSAS Catalog of Selected Documents in Psychology, 2*, 66-67.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097.

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use.* Oxford University Press.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. Oxford University Press.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Thurstone, L. L. (1928). Attitudes can be measured. *The American Journal of Sociology, 33*(4), 529-554.

Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Lawrence Erlbaum.

Wechsler, D. (1949). *The Wechsler intelligence scale for children*. The Psychological Corporation.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design.* Sage.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics,* (1), 28-44.

STRUCTURAL EQUATION MODELING (SEM) WITH AMOS

1. Introduction

The statistical approach, Structural Equation Modeling (SEM), investigates the relationships between independent and dependent variables (Ullman, 2006). Numerous software programs, such as AMOS, EQS, LISREL, and Mplus (Streiner, 2006), can be utilized for Structural Equation Modeling, which is often employed, particularly in the social sciences (Jin et al., 2021). This chapter explains the basic concepts of Structural Equation Modeling and application steps of Structural Equation Modeling in AMOS software to guide researchers.

1.1. Key Concepts of Structural Equation Modeling

In this section, the basic concepts of Structural Equation Modeling are explained.

1.1.1. Observed and Latent Variables

Characteristics like attitudes, IQ, personality traits, and socioeconomic status that are not directly observable are referred to as latent variables in the social and behavioral sciences (Yuan & Bentler, 2007). Ellipses are used as a visual representation of latent variables in AMOS analysis. It is necessary to establish measurable behaviors that are assumed to reflect this latent variable because latent variables can not be directly assessed. Thus, observed variables are used to gather information on latent variables. Observed variables, also known as indicator variables, are visually represented by rectangles or squares (Schreiber et al., 2006). Scales used in research to measure a certain emotion, thought or behavior are examples of observed variables. For example, Mindfulness in Marriage Scale (Erus & Deniz, 2018) is a latent variable, and sub-dimensions of the scale formed by the scale items are observed variables that make the latent variable observed. Scale items are brought together to make the concept of mindfulness in marriage observed. Figure 1 shows an example of observed and latent variables.

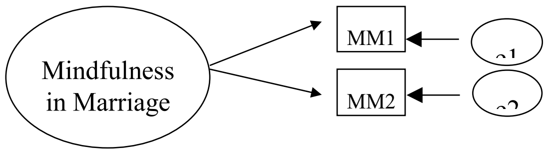


Figure 1. *Observed and Latent Variables*

As seen in Figure 1, MM1 and MM2 sub-dimensions are observed variables and Mindfulness in Marriage is a latent variable. MM1 and MM2 are taken as indicators of Mindfulness in Marriage. Mindfulness in Marriage Scale is unidimensional, but in order to create Structural Equation Modeling, the scale was divided into two sub-dimensions by "parceling method". If there are no sub-dimensions of a scale, dimensions should be created by parceling method for Structural Equation Modeling. In order to analyze latent variables, the parceling method involves building "item plots" based on the totals of the responses to various items (Russell et al., 1998). An aggregate indication called a parcel is made up of the mean of two or more items, behaviors or answers (Little et al., 2002). There are several techniques for parceling, including exploratory factor analysis (for more information about parceling; see Matsunaga, 2008). Random parceling is one of the easiest ways to constructing parcels. The goal of random parceling is to assign each item to a parcel group at random, without change. There can be two, three, or four parcels constructed, depending on the number of items that need to be assigned (Little et al., 2002). Alternatively, you can take the sum of the even numbered items of the scale as one parcel and the sum of the odd numbered items as the other parcel. It should be noted, however, that a latent variable must have at least two observed variables. However, e1 and e2 are error terms. The influence of measurement error on the observed variables is shown by the one-way arrows linking the error terms to the variables.

1.1.2. Exogenous and Endogenous Variables

Independent (predictor) variables are referred to as exogenous in Structural Equation Modeling, whereas dependent (predicted) variables are called endogenous (Bodoff & Ho, 2016). Figure 2 shows an example of exogenous and endogenous variables in the model.

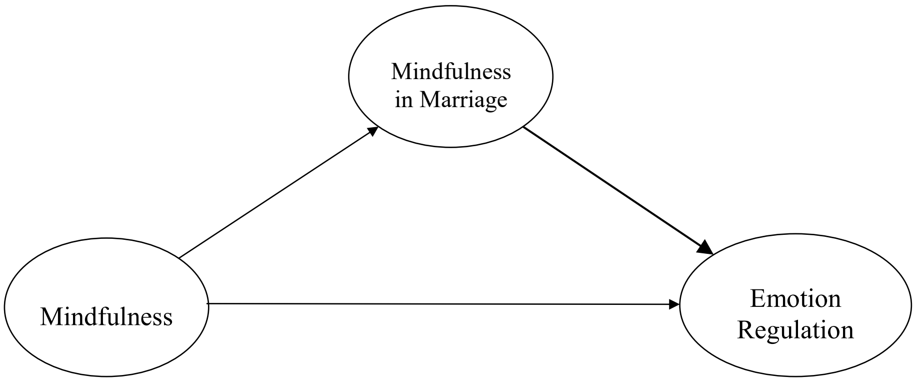


Figure 2. *Exogenous and Endogenous Variables*

As seen in Figure 2, the independent variable is "Mindfulness". This variable is also an exogenous and predictor variable. "Emotion Regulation" is a dependent, endogenous and predicted variable. "Mindfulness in Marriage" is also an endogenous and predicted variable.

1.1.3. Mediator and Moderator Variables

For a variable to be a mediator variable, it must meet some requirements. These requirements are as follows (Baron & Kenny, 1986):

- *Changes in the independent variable significantly explain changes in the hypothesized mediator variable,*
- *Changes in the mediator variable significantly explain the cause of changes in the dependent variable,*
- *A significant relationship between a dependent and independent variable is no longer significant or the strength of the relationship has decreased.*

In the model given in Figure 2, "Mindfulness in Marriage" is the mediator variable.

A moderator variable influences the strength and/or direction of the association between an independent or predictor variable and a dependent or predicted variable. Examples of moderator variables include gender, race, and class, whereas quantitative variables include education level (Baron & Kenny, 1986). The model for the moderator variable is given as an example in Figure 3.

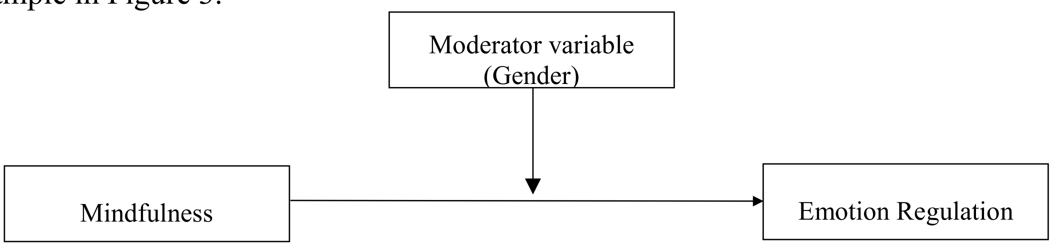


Figure 3. *Moderator Variable Model*

In the moderator variable model in Figure 3, "Mindfulness" is the independent variable, "Emotion Regulation" is the dependent variable and "Gender" is the moderator variable. The primary goal of moderator analysis is to determine how the variable chosen as a moderator influences the strength of the link between the dependent and independent variables. In other words, depending on the gender, the correlation between mindfulness and emotion regulation may be stronger or weaker.

1.1.4. Confirmatory Factor Analysis Models

As opposed to being a paradigm for establishing theories, confirmatory factor analysis tests theories. Before to the analysis, a hypothesis must be established for confirmatory factor analysis. Which variables are associated to which factors and which factors are related to each other are determined by this hypothesis and, by extension, by the model (Stapleton, 1997). Figure 4 presents the confirmatory factor analysis model consisting of two factors. The confirmatory factor analysis model given in Figure 4 aims to confirm the hypothesis that the Mindfulness in Parenting Questionnaire (Aslan Gördesli et al., 2018; McCaffrey et al., 2017) consists of two sub-scale, namely "Parental Self Efficacy" and "Being in the Moment with the Child".

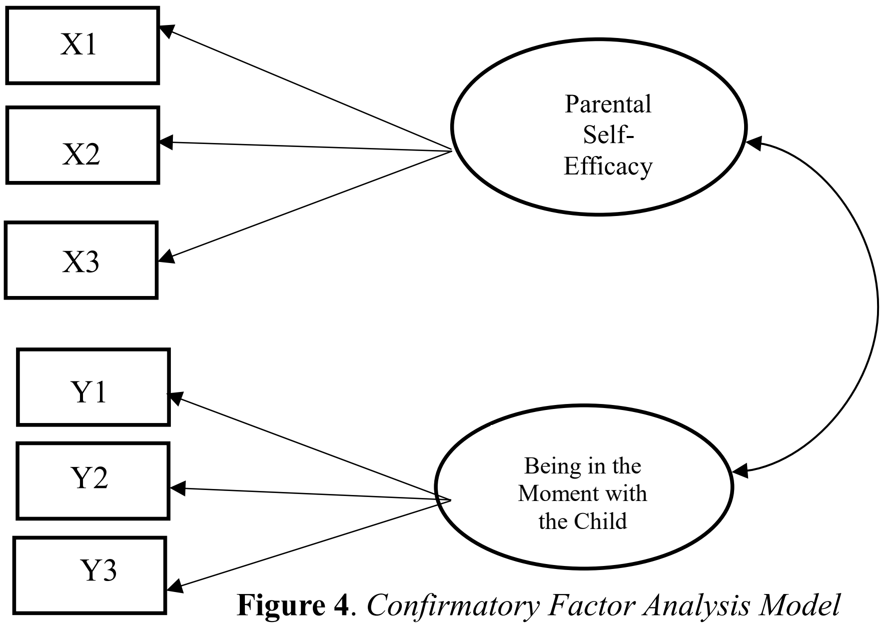


Figure 4. *Confirmatory Factor Analysis Model*

The figure shows the relationships between the "Parental Self-Efficacy" and "Being in the Moment with the Child" sub-scales and the scale items that constitute these sub-scales. In this model, two sub-scales were shown to be correlated with each other.

1.1.5. Structural Equation Models

Modeling approaches that include measurement errors, multiple concept measurements, and multi-equation models are called structural equation models (Bollen & Noble, 2011). Figure 5 presents the structural equation model consisting of 3 latent variables.

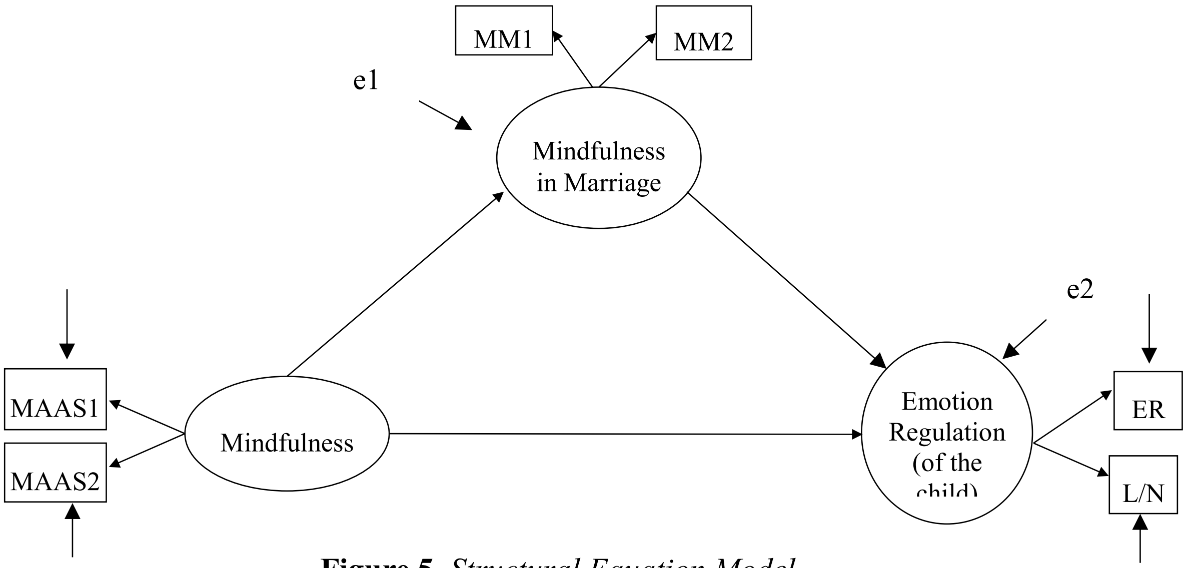


Figure 5. *Structural Equation Model*

One of the hypotheses for the structural equation model presented in Figure 5 is "Mindfulness in marriage has a mediator role in the relationship between parents' mindfulness and emotion regulation of their children." Based on this hypothesis, parents' mindfulness predicts mindfulness in marriage and mindfulness in marriage predicts emotion regulation of their children. Thus, parents' mindfulness predicts emotion regulation of their children through mindfulness in marriage. The figure clearly shows the observed and latent variables. For example, "Emotion Regulation" consists of two observed variables, ER and L/N. In other words, Emotion Regulation Checklist (Kapçı et al., 2009; Shields & Cicchetti, 1997) consists of two sub-dimensions.

The error terms are shown by the one-way arrows in the figure that point to the observed variables. The error terms, e1 and e2, are shown above the latent and dependent variables, mindfulness in marriage and emotion regulation. The error terms represent the effect of the error in the estimation of the latent variable. Each path in the model shows the hypothesis being tested.

Until this section, general information about Structural Equation Modeling has been presented. In the next section, it will be explained how to conduct Structural Equation Modeling with AMOS software.

2. Steps of Structural Equation Modeling (SEM) with AMOS

One of the most widely used programs in Structural Equation Modeling analysis is AMOS. AMOS is a shortened form of "Analysis of Moment Structures". Because of its extensive menu of plotting tools, which are thoughtfully developed with SEM guidelines in mind, and the convenience and velocity with which a publishable quality route diagram can be created, AMOS is the program of choice for most researchers when conducting their studies (Byrne, 2001). Therefore, it is one of the programs that researchers should know how to use. This section explains the steps required to conduct Structural Equation Modeling (SEM) through AMOS with examples. The example models used to explain the steps should be considered independently. In the sample models, “Mindfulness” was used as the independent variable, “Emotion Regulation” as the dependent variable, and “Mindfulness in Marriage” and “Mindfulness in Parenting” as both dependent and mediator variables.

First, the AMOS Graphics main window is opened with Start → Programs → AMOS 26.0 → AMOS Graphics commands. The screenshot of the AMOS Graphics main window is shown below.

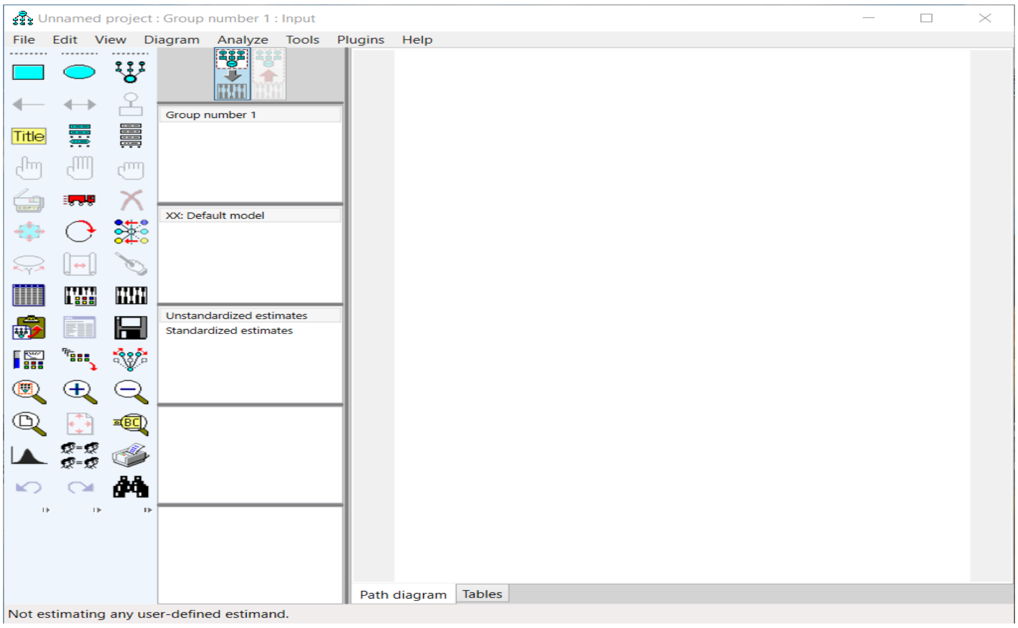
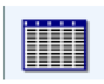


Figure 6. *AMOS Graphics Main Window*

As can be seen in Figure 6, the most commonly used icons are located on the far left of the AMOS Graphics main window. The middle section provides information about the modeling process and on the right side, there is a workspace for drawing the structural model. The functions of the icons on the left side of the screen are given where necessary when describing the steps.

2.1. Opening the Data File to be Analyzed

The raw data file to be analyzed is opened by selecting the location where the file is saved in the window that opens with the Data Files → File Name option in the File menu in the AMOS Graphics main window. AMOS software can open SPSS, Excel, MS Access, Lotus and Dbase data files. As an alternative to the File menu, the same can be done by clicking the  icon on the toolbar. After selecting the data file, the name of the data file and the number of data can be displayed in the “Data Files” window. Below is a screenshot showing the name of the data file and the number of data.

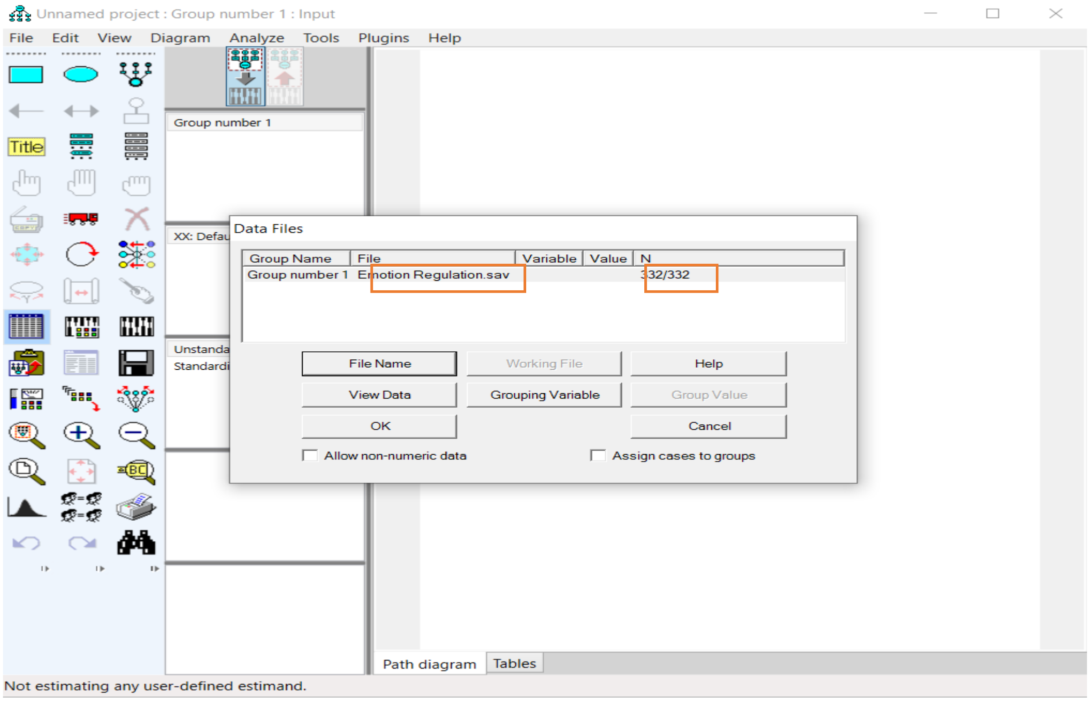
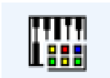
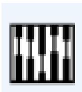


Figure 7. *Data Files*

As seen in Figure 7, the data file named "Emotion Regulation" consists of 332 data.

2.2. Determination of Normality Assumption

A normality test is required before the equations are defined. After selecting the data file to be analyzed on the AMOS Graphics screen, the observed variables to be included in the analysis are selected from the "Variables in Dataset" window opened with the "list variables in data set"  option and transferred to the working screen on the right side by dragging and dropping with the mouse. Descriptive statistics can be done by clicking on the relevant icon  or through the "Analysis Properties" option accessed via the menu. The values of the minimum and maximum as well as the univariate kurtosis and skewness, are included in these descriptive statistics. Mahalanobis distance values are provided by outliers, which identify the data that is farthest distant from the center (Byrne, 2001). In the "Output" tab, select the "Test for normality and outliers" option. After the selection is made, the window is closed. Click on the "Calculate Estimates" or icon  from the "Analyze" menu. To view the results, click on "Text output" in the "View" menu or the  icon. In the "AMOS Output" window, click on "Assessment of normality" and the results are displayed. Normality assumption is determined by "Mardia's coefficient" in AMOS software program. In the computation, the formula $p*(p+2)$ is applied, where p is the total number of observed variables (Raykov & Marcoulides, 2008). Data can be deemed multivariate normal if Mardia's coefficient is lower than the value derived from the formula

(Khine, 2013). For instance, in a model with two observed variables, the Mardia's coefficient should be less than eight (Raykov & Marcoulides, 2008). If the normality assumption can not be met, identifying outliers and excluding them from the data set may be an alternative way to ensure the assumption.

2.3. Creating the Structural Model

After the normality assumption is met, the structural model is constructed according to the hypotheses of the research. The latent variable and its observed variables are drawn on the workspace on the right by clicking on the  icon from the left side menu. The observed variables are selected from the window opened by clicking on the  icon and transferred to the screen by dragging and dropping with the mouse. Latent variables are represented by ellipses while observed variables are represented by rectangles. The error values associated with the observed variables are also shown as ellipses and all observed variables have error values. The figure below shows the latent variable, observed variables and the error values of these observed variables in the AMOS Graphics main window.

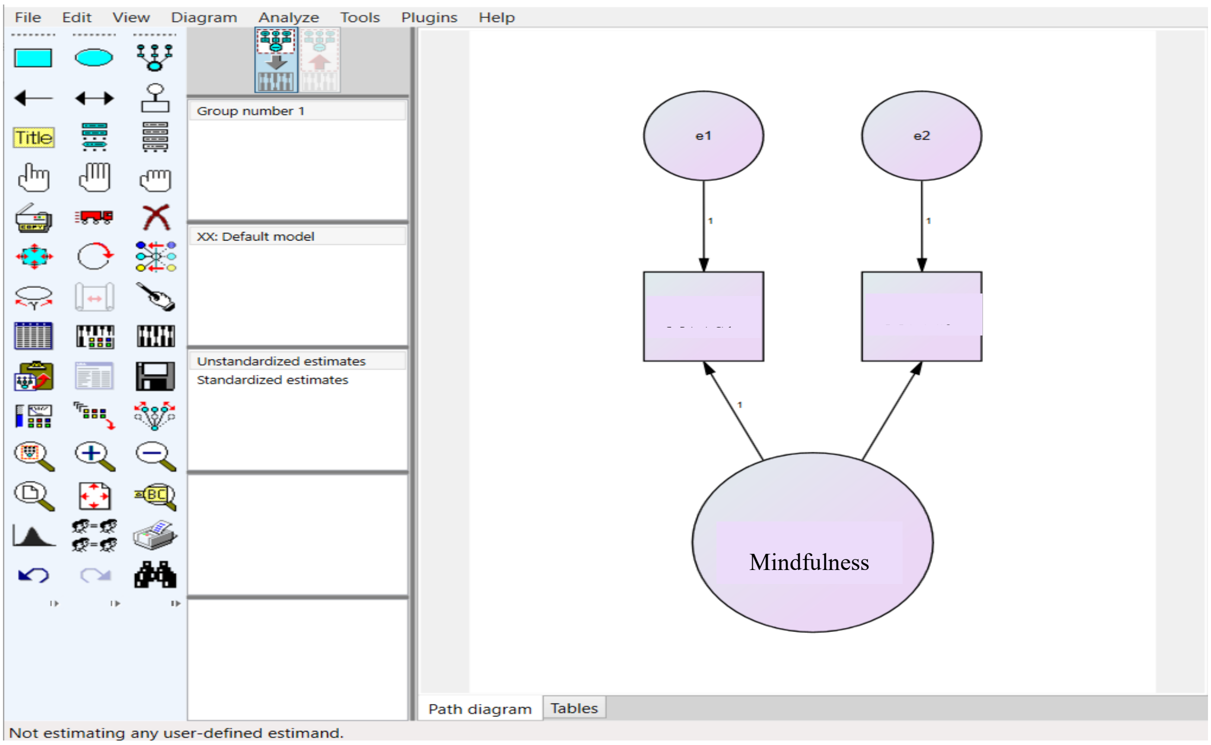
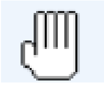
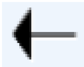
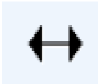


Figure 8. *AMOS Graphics Image of Latent Variable and Observed Variables*

In Figure 8, "Mindfulness" is the latent variable, "MAAS1" and "MAAS2" are observed variables, "e1" and "e2" are error values. According to this figure, the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003; Özyeşil et al., 2011), which measures mindfulness, consists of two sub-dimensions determined through the parceling method:

MAAS1 and MAAS2. According to the number of latent variables used in the model, the diagram is selected with the  icon and copied with the  icon.

2.4. Testing the Measurement Model

An evaluation of how well the observed variables represent the latent variable they are supposed to measure is made possible by the measurement model (Gallagher et al., 2008). To draw up the measurement model, the latent variable and the observed variables are connected by a one-way arrow . In addition, covariance  is drawn between the latent variables. An example measurement model is given below (The measurement model in the example has two latent variables and is created for example purposes only. However, all latent variables in your model will be included in the measurement model and covariance will be drawn between all latent variables).

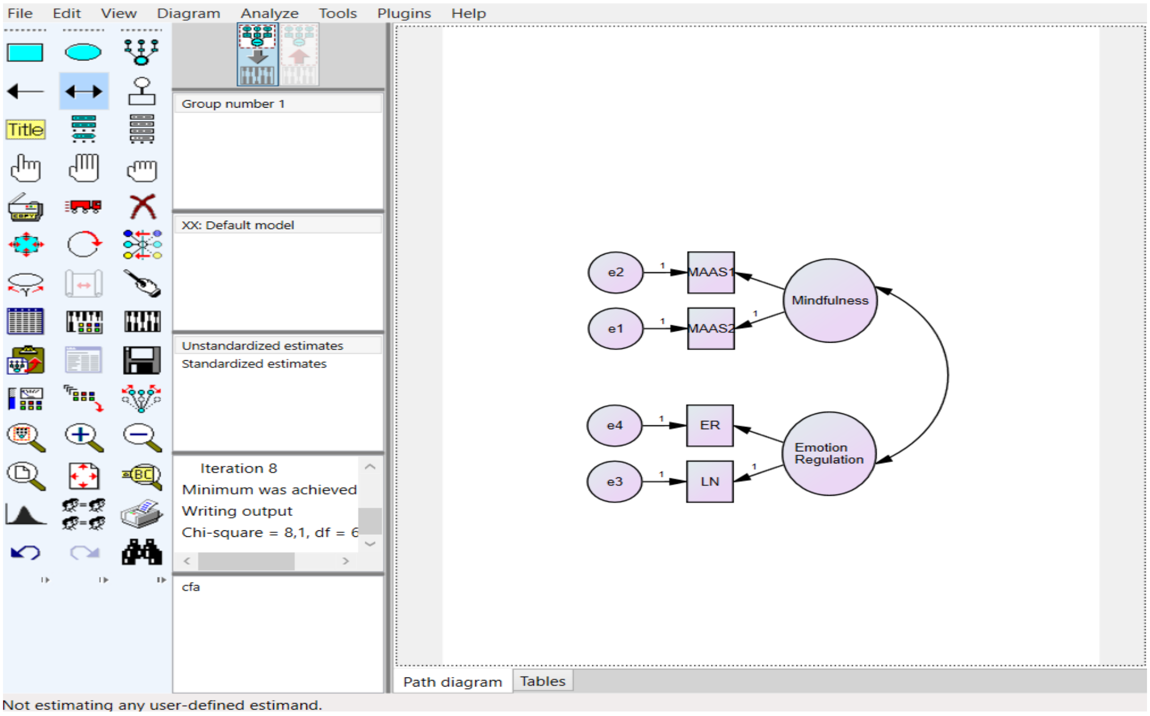


Figure 9. *Measurement Model*

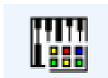
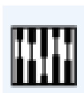
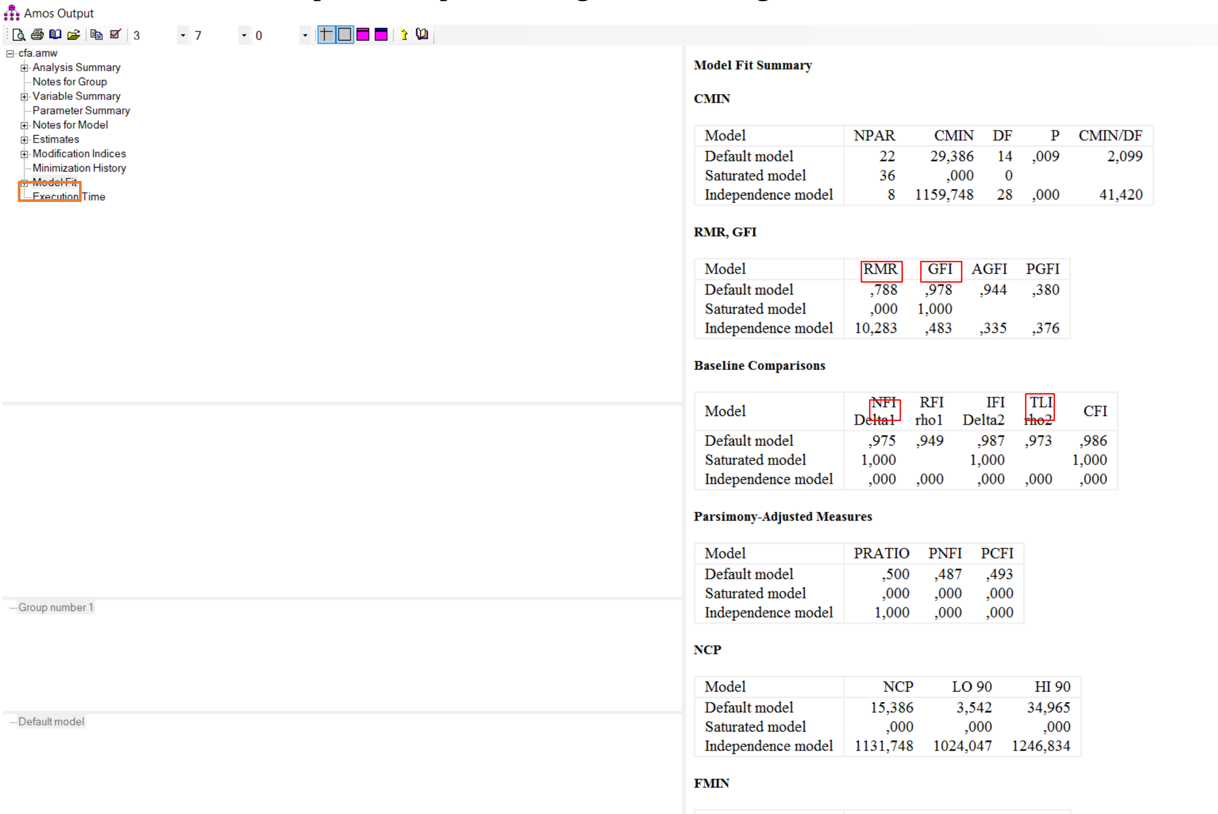
Figure 9 shows the measurement model consisting of the independent variable "Mindfulness", the dependent variable "Emotion Regulation" and the sub-dimensions of these variables. After creating the measurement model, select the  icon from the menu on the left side of the screen. In this screen, "modification indices", "standardized estimates" and "direct and indirect effects" boxes should be checked. After closing the analysis properties window, click on the "Calculate Estimates"  icon and perform the calculation. Select the “View Text”  icon on the toolbar to open the output file. In the measurement model, factor loadings should be examined first. In the “Output” window, select “Estimates” from the list on the left. With statistical significance, standardized factor loadings must be higher than .50 and ideally above .70 (Hair et al., 2019). After evaluating the factor loadings, the model’s overall fit should be examined through the goodness of fit indexes given under the heading “Model Fit”. The cutoff criteria of fit indexes recommended by Schermelleh-Engel et al. (2003) are given in Table 1.

Table 1. *Cutoff Criteria of Fit Indexes*

Fit Indexes	Good Fit	Acceptable Fit
χ^2	$.05 < p \leq 1.00$	$.01 < p \leq .05$
χ^2 /df	$0 \leq \chi^2/df \leq 2$	$2 < \lambda 2 /df \leq 3$
RMSEA	$0 \leq$ RMSEA $\leq .05$	$.05 <$ RMSEA $\leq .08$
SRMR	SRMR $< .05$	$.05 \leq$ SRMR $< .10$
CFI	$.97 \leq$ CFI ≤ 1.00	$95 \leq$ CFI $< .97$
NFI	$.95 \leq$ NFI ≤ 1.00	$.90 \leq$ NFI $< .95$
AGFI	. $90 \leq$ AGFI ≤ 1.00	$85 \leq$ AGFI $< .90$
GFI	$.95 \leq$ GFI ≤ 1.00	$.90 \leq$ GFI $< .95$
TLI	$.97 \leq$ TLI ≤ 1.00	$95 \leq$ TLI $< .97$

RMSEA= Root Mean Square Error of Approximation, SRMR= Standardized Root Mean Square Residual, CFI = Comparative Fit Index, NFI = Normed Fit Index, AGFI = Adjusted Goodness-of-Fit- Index, GFI = Goodness-of-Fit Index, TLI= Tucker-Lewis Index

Below is a sample output image with goodness of fit indexes.



Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	22	29,386	14	,009	2,099
Saturated model	36	,000	0		
Independence model	8	1159,748	28	,000	41,420

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	,788	,978	,944	,380
Saturated model	,000	1,000		
Independence model	10,283	,483	,335	,376

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,975	,949	,987	,973	,986
Saturated model	1,000		1,000		1,000
Independence model	,000	,000	,000	,000	,000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	,500	,487	,493
Saturated model	,000	,000	,000
Independence model	1,000	,000	,000

NCP

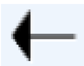
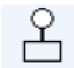
Model	NCP	LO 90	HI 90
Default model	15,386	3,542	34,965
Saturated model	,000	,000	,000
Independence model	1131,748	1024,047	1246,834

FMIN

Figure 10. *Goodness of Fit Indexes Output Image*

When compared with the cutoff criteria of fit indexes in Table 1, GFI=.978, AGFI=.944, NFI=.975 and TLI=.973 are evaluated among the "good fit" criteria. The goodness of fit indexes of the measurement model are expected to be within the range of good fit or acceptable fit criteria. The structural model should be tested after evaluating the goodness of fit indexes of the measurement model.

2.5. Testing the Structural Model

After drawing the diagram, the endogenous and exogenous variables should be connected to each other with a one-way arrow . The errors of the endogenous variables are included in the model by clicking on the variable with the "Add a unique variable to an existing variable"  icon. Plugins → Name unobserved variables to name these variables (if the naming is not done, a warning will appear during the analysis phase). A sample structural model is given below.

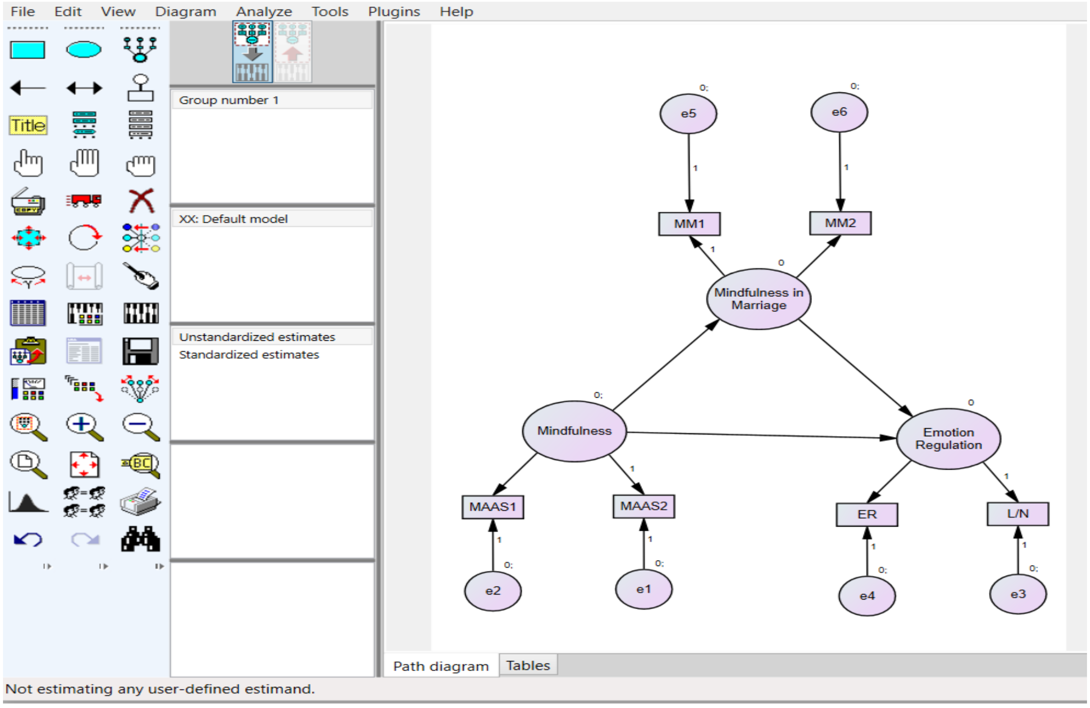


Figure 11. *AMOS Graphics Image of the Structural Equation Model*

One of the hypotheses of the structural equation model presented in Figure 11 is "The link between parents' mindfulness and emotion regulation of their children is mediated by mindfulness in marriage." Based on this hypothesis, parent's mindfulness predicts mindfulness in marriage and mindfulness in marriage predicts emotion regulation of their children. Thus, it can be said that parent's mindfulness predicts emotion regulation of their children through mindfulness in marriage. If you try to analyze the structural model in the figure, you will see a warning about not including error terms. Therefore, adding an error term to the endogenous (dependent) variables is important.

Once the structural model is completed, the stages of testing the measurement model described in the previous section are followed. First, whether the parameter estimates are statistically significant should be examined. In the "Output" window, select "Estimates" from the list on the left. Direct, indirect and total effects can be seen on the screen that opens after selection. Significant path coefficients are important for the overall fit of the model (Baron & Kenny, 1986). The regression coefficient should be checked for the significance of path coefficients. To examine whether these values are significant or not, the *p* value of each of them should be examined. If this value is less than .05, it can be said that the path coefficient is significant. Below is a sample output image showing the regression coefficients and standardized regression coefficients.

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
Mindfulness_in_Marriage	<---	Mindfulness	,218	,029	7,462	***	par_4
EmotionRegulation	<---	Mindfulness_in_Marriage	-,331	,073	-4,541	***	par_5
EmotionRegulation	<---	Mindfulness	-,095	,030	-3,156	,002	par_6
	<---	Mindfulness	1,000				
	<---	Mindfulness	,824	,068	12,109	***	par_1
	<---	Mindfulness_in_Marriage	1,000				
	<---	Mindfulness_in_Marriage	,855	,080	10,674	***	par_2
	<---	EmotionRegulation	1,000				
	<---	EmotionRegulation	1,486	,248	5,997	***	par_3

Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
Mindfulness_in_Marriage	<---	Mindfulness	,469
EmotionRegulation	<---	Mindfulness_in_Marriage	-,454
EmotionRegulation	<---	Mindfulness	-,280
	<---	Mindfulness	,935
	<---	Mindfulness	,850
	<---	Mindfulness_in_Marriage	,915
	<---	Mindfulness_in_Marriage	,781
	<---	EmotionRegulation	,586
	<---	EmotionRegulation	,634

Figure 12. *Standardized Regression Coefficients*

In Figure 12, firstly, whether the path coefficients are significant or not in the "Regression Weights" table should be examined. The "Standardized Regression Weights" table should be examined for path coefficients. While it is desirable for the model that all paths are significant, the non-significance of the path between the dependent variable and the independent variable is not seen as a problem since it reflects the strength of the mediator variable in the relationship. After evaluating whether the path coefficients are significant or not, the goodness of fit indexes given under the heading "Model Fit" in the Output file are examined and evaluated according to the cutoff criteria of fit indexes. Models with the best fit to the data are indicated by goodness of fit indexes (Smith & McMillan, 2001). The goodness of fit indexes are evaluated within the frame of the cutoff criteria of fit indexes recommended by Schermelleh-Engel et al. (2003) and presented in Table 1. As a result of these assessments, the structural model is accepted, rejected or modified.

2.6. Modification

After examining the fit indices, model modifications can be made for a better fitting model. Modification indices suggest that these changes be made. However, care should be taken when making modifications and each modification to the model should be based on a theoretical foundation. As a result, the researcher should explain why the change was made and whether it is theoretically significant for the model (Schreiber et al., 2006).

Given that the single degree of freedom chi-square table value is 3.84 (Whittaker, 2012), 4 is used as the modification indices cutoff value. If there are no modification indices as a result of the analysis, there is no value exceeding the cutoff value. Below is an output image with the suggested modification indices after model testing.

Modification Indices (Group number 1 - Default model)

Covariances: (Group number 1 - Default model)

	M.I.	Par Change
e1 <--> e5	4,123	-1,068

Variances: (Group number 1 - Default model)

	M.I.	Par Change

Regression Weights: (Group number 1 - Default model)

	M.I.	Par Change

Means: (Group number 1 - Default model)

	M.I.	Par Change

Intercepts: (Group number 1 - Default model)

	M.I.	Par Change

Figure 13. *Modification Indices*

In the image above, the modification indices between error values e1 and e5 is calculated as 4.123. Thus, it is recommended to draw the covariance ↔ between these two error values and retest the model. Once the modifications have been theoretically evaluated (the modifications should be theoretically grounded) and made, the model should be tested again. If good or acceptable fit indices are obtained after the model is retested, the model is accepted; otherwise, the model should be modified again if possible. The modification process is continued until a new modification can not be made and the model is accepted or rejected according to the fit indices obtained due to these modifications.

2.7. Determination of Model Validity

The last step of Structural Equation Modeling is to determine the model validity. One of the most common methods used to determine the validity of the structural model is the "Bootstrapping" method. The bootstrapping process consists of resampling the original data (Scharkow, 2017). Using this strategy, we may ascertain the model's validity in larger samples. For bootstrapping, the "Analysis Properties" menu item opens and the "Indirect, direct, and total effects" section is selected. In the "Bootstrap" tab, "Perform Bootstrap", "Bias-Corrected Confidence Intervals" and "Bootstrap ML" options are selected. Type a number between 500 and 1000 for the "Number of bootstrap samples" option (Cheung & Lau, 2008). This number represents the number of resampling of your model. Write 95 for the "BC Confidence Level" option. Thus, the analysis is performed with a 95% confidence interval. After everything has been done so far, click on the "Calculate Estimates" button and get the results. A sample "Analysis Properties" menu screenshot is given below.

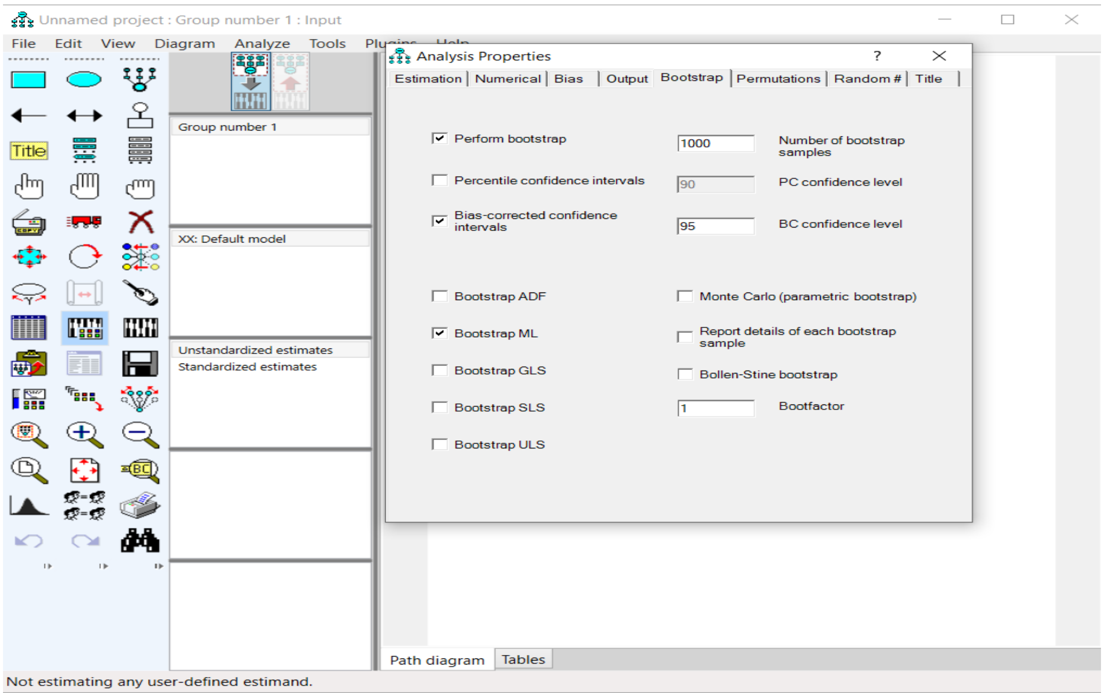
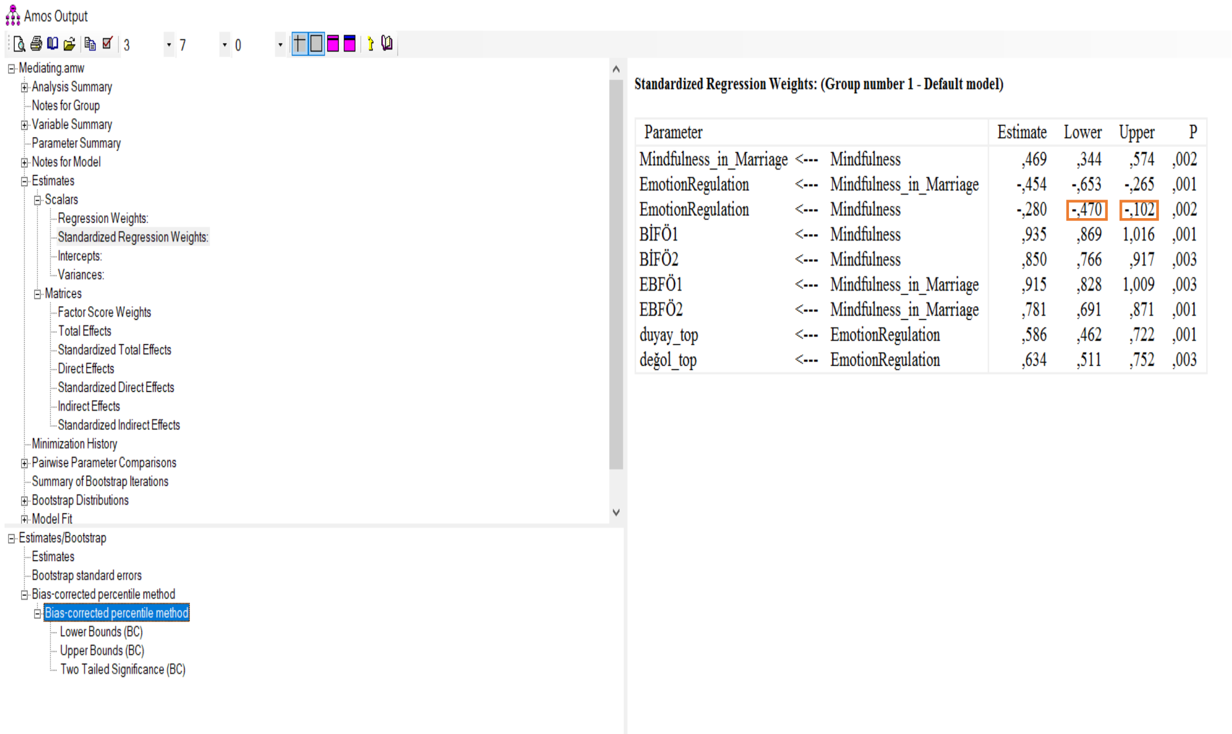


Figure 14. *Analysis Properties Menu View for Bootstrapping*

The "Standardized Indirect Effect" tab in Output shows the independent variable's indirect effect on the dependent variable. Whether this effect is significant is determined by looking at the lower and upper bounds of the confidence interval. For this, the number between the independent variable and the dependent variable should be checked in the "Lower Bounds" and "Upper Bounds" tables under the "Standardized Indirect Effect" tab. These numbers represent the lower bound and upper bound of the confidence interval. Hayes (2022) stated that an indirect effect is significant if there is no zero between the lower and upper bounds of the confidence interval (for example, the lower bound is -.470 and the upper bound is -.102). The screenshot showing the lower bound and upper bound in the table is given below.



Parameter			Estimate	Lower	Upper	P
Mindfulness_in_Marriage	<---	Mindfulness	,469	,344	,574	,002
EmotionRegulation	<---	Mindfulness_in_Marriage	-,454	-,653	-,265	,001
EmotionRegulation	<---	Mindfulness	-,280	-,470	-,102	,002
BİFÖ1	<---	Mindfulness	,935	,869	1,016	,001
BİFÖ2	<---	Mindfulness	,850	,766	,917	,003
EBFÖ1	<---	Mindfulness_in_Marriage	,915	,828	1,009	,003
EBFÖ2	<---	Mindfulness_in_Marriage	,781	,691	,871	,001
duyay_top	<---	EmotionRegulation	,586	,462	,722	,001
değol_top	<---	EmotionRegulation	,634	,511	,752	,003

Figure 15. *Lower Bound and Upper Bound*

In Figure 15, since our independent variable is "Mindfulness" and our dependent variable is "Emotion Regulation", when we look at the lower bound (-.470) and upper bound (-.102) values between these two variables, we see that there is no zero between them. Thus, as a result of bootstrapping, the indirect effect in the structural model is significant at a 95% confidence interval. In the end, all analyses conducted within the scope of Structural Equation Modeling should be reported.

3. *Summary*

AMOS (Analysis of Moment Structures) is one of the most frequently used and practical software in Structural Equation Modeling. Therefore, researchers need to know how to conduct Structural Equation Modeling analysis through AMOS. The first step of Structural Equation Modeling in AMOS is to open the data file to be analyzed. After opening the data file, the first thing to do is decide the normality assumption. Normality is assumed using Mardia's coefficient and the formula $p*(p+2)$. In this formula, p is the number of observed variables. To meet the normality assumption, the value derived from the formula should be higher than Mardia's coefficient. After the normality assumption is met, the structural model is created according to the research hypotheses. Observed and latent variables in the model are drawn on the main window with the help of the icons on the AMOS toolbar. After all the observed and latent variables are drawn on the main window, we should proceed testing the measurement model. In testing the measurement model, covariance between latent variables should be drawn. Once the measurement model is drawn, the analysis should proceed. First, factor loadings should be examined. Standardized factor loadings must be higher than .50 and ideally above .70 with statistical significance. After examining the factor loadings, the goodness of fit indexes of the model should be evaluated according to the criteria of Schermelleh et al. (2003). The goodness of fit indexes of the measurement model are expected to be within the range of good fit or acceptable fit criteria. Testing of the structural model requires the following testing of the measurement model. The same paths are followed for testing the measurement model. An essential point in the structural model is the addition of error terms to the dependent variables. Another critical point is that not all paths have to be meaningful in the structural model. It should be kept in mind that the insignificance of the path between the independent and dependent variables may be due to the strength of the mediator variable in the relationship. It is desired that the path between the independent variable and the dependent variable becomes insignificant when the mediator variable is included in the model. After that, the goodness of fit indexes of the structural model are examined and modifications are made if necessary. The important thing here is that all modifications should have a theoretical basis. If the modification indices are higher than 4, it is necessary to draw covariance between the two variables or error terms shown in the table. The last step of the structural equation model is determining the model validity. The bootstrapping method is used to determine the model validity. The important point here is to make the correct markings in the "Analysis Properties" menu. In the output that opens, standardized indirect effect of the independent variable on the dependent variable is examined. If there is no 0 between the lower bound and upper bound values of the standardized indirect effect, it is concluded that the mediation effect is significant. Lastly, proper reporting of all analyses is required.

References

Aslan Gördesli, M., Arslan, R., Çekici, F., Aydın Sünbül, Z., & Malkoç, A. (2018). The psyhometric properties of Mindfulness in Parenting Questionnaire (MIPQ) in Turkish sample. *European Journal of Education Studies, 5*(5), 175–188. https://doi.org/10.5281/zenodo.1477467

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bodoff, D., & Ho, S. Y. (2016). Partial least squares Structural Equation Modeling approach for analyzing a model with a binary indicator as an endogenous variable. *Communications of the Association for Information Systems, 38*(23), 400-419. https://doi.org/10.17705/1CAIS.03823

Bollen, K. A., & Noble, M. D. (2011). Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences*, *108*(3), 15639-15646. https://doi.org/10.1073/pnas.1010661108

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822–848. https://doi.org/10.1037/0022-3514.84.4.822

Byrne, B. M. (2001). Structural Equation Modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing*, *1*(1), 55-86. https://doi.org/10.1207/S15327574IJT0101_4

Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, *11*(2), 296-325.

Erus, S. M., & Deniz, M. E. (2018). Evlilikte bilinçli farkındalık ölçeğinin geliştirmesi: Geçerlik ve güvenirlik çalışması [Development of mindfulness in marriage scale (MMS): Validity and reliability study]. *The Journal of Happiness & Well-Being*, *6*(2), 96–113.

Gallagher, D., Ting, L., & Palmer, A. (2008). A journey into the unknown; taking the fear out of Structural Equation Modeling with AMOS for the first-time user. *The Marketing Review*, *8*(3), 255-275.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2019). *Multivariate data analysis.* Cengage.

Hayes, A.F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression based approach.* The Guilford Press.

Jin, S., Noh, M., Yang-Wallentin, F., & Lee, Y. (2021). Robust nonlinear Structural Equation Modeling with interaction between exogenous and endogenous latent variables. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 547-556. https://doi.org/10.1080/10705511.2020.1857255

Kapçı, E. G., Uslu, R., Akgün, E., & Acer, D. (2009). İlköğretim çağı çocuklarında duygu ayarlama: Bir ölçek uyarlama çalışması ve duygu ayarlamayla ilişkili etmenlerin belirlenmesi [Emotion regulation in elementary school children: A scale adaptation study and determination of factors related to emotion regulation]. *Turk J Child Adolesc Ment Health*, *16*(1), 13–20.

Khine, M.S. (2013). *Application of Structural Equation Modeling in educational research and practice.* Sense Publishers.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Matsunaga, M. (2008). Item parceling in Structural Equation Modeling: A primer. *Communication Methods and Measures*, *2*(4), 260-293. https://doi.org/10.1080/19312450802458935

McCaffrey, S., Reitman, D., & Black, R. (2017). Mindfulness in Parenting Questionnaire (MIPQ): Development and validation of a measure of mindful parenting. *Mindfulness, 8*, 232–246. https://doi.org/10.1007/s12671-016-0596-7

Özyeşil, Z., Arslan, C., Kesici, Ş., & Deniz, M. E. (2011). Adaptation of the mindful attention awareness scale into Turkish. *Education and Science, 36*(160), 224–235. http://eb.ted.org.tr/index.php/EB/article/view/697

Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Routledge. http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/ResourcesWkshp/2008_RaykovMarcoulides_Ch3.pdf

Russell, D. W., Kahn, J. H., Spoth, R., & Altmaier, E. M. (1998). Analyzing data from experimental studies: A latent variable Structural Equation Modeling approach. *Journal of Counseling Psychology, 45*(1), 18–29. https://doi.org/10.1037/0022-0167.45.1.18

Scharkow M. (2017). Bootstrapping. In: J. Matthes, C.S. Davis, & R.F. Potter (Eds), *The international encyclopedia of communication research methods* (pp. 1-5). John Wiley & Sons. http://doi.org/10.1002/9781118901731.iecrm0017

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of Structural Equation Models: Tests of significance and descriptive goodness- of-fit measures. *Methods of Psychological Research, 8*(2), 23-74. https://psycnet.apa.org/record/2003-08119-003

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338.

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale [Abstract]. *Developmental Psychology, 33*(6), 906–916. https://doi.org/10.1037/0012-1649.33.6.906

Smith, T. D., & McMillan, B. F. (2001). *A primer of model fit indices in Structural Equation Modeling.* Paper presented at the Annual Meeting of the Southwest Educational Research Association, New Orleans, LA.

Stapleton, C. D. (1997, January). *Basic concepts and procedures of confirmatory factor analysis.* Paper Presented at the Annual Meeting of the Southwest Educational Research Association, Austin.

Streiner, D. L. (2006). Building a better model: an introduction to Structural Equation Modelling. *The Canadian Journal of Psychiatry, 51*(5), 317-324. https://doi.org/10.1177/070674370605100507

Ullman, J.B. (2006). Structural Equation Modeling: Reviewing the basics and moving forward. *Journal of Personality Assessment, 87*(1), 35-50. https://doi.org/10.1207/s15327752jpa8701_03

Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *The Journal of Experimental Education*, *80*(1), 26-44. https://doi.org/10.1080/00220973.2010.531299

Yuan, K.-H., & Bentler, P.M. (2007). Structural equation modeling. In C.R. Rao & S. Sinharay (Eds.), *Handbook of statistics 26: Psychometrics* (pp. 297–358). https://doi.org/10.1016/S0169-7161(06)26010-3

QUALITATIVE RESEARCH DESIGNS

1. INTRODUCTION

1.2. Scientific Research Process

A research process represents a specific way of accumulating knowledge, which is supposed to be the guarantee of truth, based on logical principles and criteria, instruments and actions. However, society cannot be covered in its entirety by experiential research. The human spirit, for example, will never be fully explored.

Speaking about scientific research, there is a rule that one should never try to explore each segment of a phenomenon separately. Instead, all the segments should be explored together. By exploring one segment, we also explore the other ones, and then make connections between them. There is a feedback loop between the parts and the whole, and researchers move from the former to the latter, and backwards. This is actually the essence of research, and this is how the connections between phenomena are made, and understanding achieved.

If we explore reality only as a whole, we might not understand it properly. Each scientist can extract from the whole what is important to him/her, but only when one starts dealing with what reality is made of, with its parts, the horizons of reality will broaden, and one will manage to get out of a single frame. One's opinion about something can change significantly once one becomes familiar with its parts, and especially when a majority of them become familiar. This is like a circle, which can hardly be closed. '*Research resembles a man trying in vain to quench the thirst by drinking seawater*' (Pečujlić, 1982, p. 44).

Even if we manage to collect all the facts, we will not get to know reality in its entirety. The knowledge which we discover gradually is only a tiny fraction of the light which illuminates the darkness of ignorance. Society as a whole is very well regulated. It has a logical structure, where everything is interwoven. Getting to know reality is like getting to know a person. At first glance, and based on the first impression, only a general opinion can be made. And later, after we spend some time with the person, we become familiar with their traits, personality, their reactions in specific situations, and our opinion changes over time until we manage to become fully familiar with the person's character and form our final opinion. But even then, we cannot be completely sure we are right. What is on the outer side will never perfectly match what is inside. When exploring reality, certain facts and parts of a whole are taken into consideration. When making a selection, it is important to select those parts that will help us reveal as much as possible. This does not imply that the right selection will always be made. Some facts may help us reveal the true character of reality much better than the ones we have already selected. It all depends on how researchers position themselves, on their own understanding of what they have discovered.

> *'The dialectic of society as a whole, its understanding, is always our ultimate goal, the ultimate object of our research. But unfortunately, or maybe fortunately, there is no direct way to it. It can be reached only by exploring a more specific object, which represents our immediate and specific research assignment'* (Pečujlič, 1982, p. 44).

The theoretical frame is wide, and definitely nothing can be checked precisely, everything can be as it seems, and does not have to be so. What is needed is time and experience. Life is the best test of any theory that may be developed.

As mentioned above, in order to gain the true knowledge of reality, facts have to be collected and analysed, and the more the facts the better, because it is the facts that illuminate the path, helping us discover reality. It is very important to take all the facts that will help to reveal reality into consideration, and reality itself can help to understand the facts. A fact reveals certain parts, but it itself becomes revealed with the help of other parts. *'Facts are the codes of reality, but they are decoded by means of a whole to which they belong'* (Pečujlić, 1982, p. 45). A collection of facts will be understood once we discover their role and place in reality.

When carrying out some research, there is an order to be followed and therefore there are some stages that any research has to pass through. The first stage implies defining the object, and it can be theoretical and practical. Theoretical defining is done using mental terms. Practical defining implies certain indicators, which have to be tested and examined. Once the object is defined, a hypothesis, i.e. an assumption, is formulated. A hypothesis has a guiding role throughout the research process – it connects all research stages. At the next stage, data are collected, and classified. The classification is followed by the next stage, and that is the scientific explanation. *'Scientific explanation generally narrows down to determining the types of correlation, functional, and causal relationships'* (Pečujlić, 1982, p. 45). The final research stage implies testing the scientific explanation.

When carrying out some research, a multivariate analysis is used to check if a phenomenon has been actually caused by what is thought to have caused it, or by something completely different. A researcher comes across a large number of indicators, and there is always a dilemma as to which one to select, and how the selection should be made. In his doctrine of interchangeable indices, Lazarsfeld (1966, p. 190) claims that it does not matter which indicator is taken into consideration, but science would not agree with that because each indicator is different and will not necessarily provide representative data about the essence of something. For example, the degree of power or the attitude towards the means of production is a more important indicator of the class structure than of the degree of prestige.

The positivist research is much simpler, but boring to a certain extent, because it requires no thinking as there is a pattern where everything is predefined and cannot be changed. The analysis of each indicator which provides a true image of a society is what is missing. Research cannot be performed so superficially, without finding out how it affects the society, classes. The percentage of the unemployed can be taken as a general data. But the data does not mean anything if we do not know how it affects the population, how they manage to earn a living.

It is important to determine all the relationships between phenomena. The cause is a phenomenon which is sufficient to produce another phenomenon. Defining the relationships between phenomena is just a step forward to obtaining knowledge. Each phenomenon discovered has to be thoroughly analysed because there are more phenomena hiding behind it. In order to reach to the core, one has to discover everything surrounding it. Even though each phenomenon is analysed separately, it does not mean that each phenomenon is distinct. It means that they are all related to each other, and only if viewed together, can lead to knowledge acquisition. It is worth noting that a phenomenon can reflect a society but can also be the cause of its changes.

1.3. Research Paradigms

Research paradigms are the theoretical perspectives that shape the way research is formulated and implemented (Mackenzie & Knipe, 2006). A paradigm represents a worldview, and various philosophical assumptions associated with that point of view. Each paradigm comprises four elements:

1. **Epistemology** – (from the Greek 'epistēmē' meaning 'knowledge') – describes how we get to know something (the truth, reality), how we know what we know, i.e. it focuses on the nature of human knowledge and comprehension and therefore influences the way of uncovering knowledge in the social context that we investigate. According to Slavin (1984), there are four sources of knowledge or epistemological bases of research. They are as follows:

- *intuitive knowledge: beliefs, faith, and intuition;*
- *authoritative knowledge: data gathered from leaders in organisations, from books;*
- *logical knowledge: reason is emphasized as the surest path to knowing the truth;*
- *empirical knowledge: knowledge is best derived from sense experiences, and objective facts.*

2. **Ontology** – (from the Greek compound comprising 'ὄντος', meaning 'being' or 'that which is', and '-λογία', meaning 'logical discourse') – philosophical assumptions about the nature of reality or essence of the social phenomenon being researched, crucial to understanding how the researcher makes meaning of the collected data (Scotland, 2012).

3. **Methodology** – (from the Greek 'μέθοδος', meaning 'pursuit of knowledge', and 'λόγος' meaning 'science') – research designs, methods, approaches and procedures used in the research that is well planned to find out something (Keeves, 1997). It includes data collection, participants, instruments used, as well as data analyses, and focuses on how the researcher gets to know the world or gain knowledge about part of it (Moreno, 1947).

4. **Axiology** – (from the Greek 'axios', meaning 'worthy', and 'logos', meaning 'science') – or the Theory of Value: ethical issues, i.e. defining, evaluating, and understanding the concepts of right and wrong behaviour related to the research.

In social sciences, different paradigms can be adopted, such as positivism, interpretivism/constructivism, subjectivism, pragmatism, critical realism.

- Positivism

Positivism is based on the belief that there is only one and single reality, which can be explored scientifically using deductive methodology, and usually quantitative research methods such as experiment or survey. Therefore, it focuses on identifying or testing causal relationships in an objective manner.

- Interpretivism/Constructivism

4. *Interpretivism/Constructivism is based on the belief that there is no single reality, and reality is socially constructed and reconstructed and therefore cannot be objectively observed from outside. The only way to better understand it is by experiencing it, so it must be observed through the direct experience of people, and therefore knowledge is subjective. Its aim is to explore, not just explain the underlying meanings of reality (Crotty, 1998). So, it uses inductive and usually qualitative methods, such as interviews, and observation, aimed at generating theory.*

- Realism

5. *Realism is based on the belief that reality is independent from the human mind (Saunders et. al., 2012). Direct realism perceives the world through personal, human senses, whereas critical realism argues that senses can be deceptive, and aims at uncovering, understanding and explaining the mechanisms underlying a phenomenon from multiple perspectives. It uses both qualitative and quantitative approach and therefore a combination of qualitative and quantitative methods (Sayer, 2000).*

- Pragmatism

Pragmatism focuses on the belief that there are many different ways of interpreting the world, that knowledge can never be truly representative of reality, and that the combination of different methodological approaches provides a better understanding of the phenomena being researched. It is based on 'what works best' in finding answers to the questions under investigation and therefore uses all available approaches to understand the problem (Morgan, 2007). Its emphasis is on experience, and instead of focusing on finding the truth, it focuses on what is useful to believe (Dewey, 1941). It is action-based, and advocates human rights and individual freedom. It perceives truth as constantly changing according to practical necessities of the present (Creswell, 2009).

1.4. Research Methods

A method is a way or manner of discovering the truth. A scientific method is often equalised with the essence of science. The scientific methodology defines logical, technical, organisational, and strategic rules by means of which certain knowledge is obtained. Moreover, the methodology of science provides instructions about what is to be done and how in order to make scientific discoveries. On the other hand, methods can be approached from two points of view, which make a distinction between the essential and technical aspects of the very research:

- *General method as a scientific research strategy and*
- *Research method as a research tactic, i.e. technique.*

A method is also the way in which the knowledge about the object of study belonging to a specific scientific field is obtained. The object of study in a specific scientific field is determined by the method. The human body can be taken as an example. Physics studies it from the aspect of the motion of the physical body, biology from the aspect of the functions of the living organism, psychology through conscious behaviour, sociology determines how such behaviour and other people's behaviour intertwine to form a social phenomenon. However, even when we define the method, we do not know what it comprises, which results in different interpretations of the concept of the scientific method. Basically, the purpose of a method is to ensure obtaining a more profound knowledge on the object of study within a certain scientific field based on the existing knowledge (Pečujlić, 1982, p. 175).

Broadly speaking, a method comprises three elements:

- *the process of obtaining knowledge,*
- *the knowledge on the very object of study,*
- *the means by which the characteristics of the object of study are discovered, which is the goal of scientific research (Pečujlić, 1982, p. 176).*

The research process implies all the activities necessary to obtain the knowledge about an object. The existing knowledge is usually insufficient. It is later into the research process that the existing knowledge is complemented. The means are something specific (actions and tools) used to study an object. The concept of a method is very often related only to the general rules of the scientific research process, whereas the means are referred to as research techniques (Lukić, 1989 p. 48). A method, as a research practice of a scientific field, always depends on its content and therefore cannot be determined only by general, logical principles and technical means used. Different scientific fields use the same technical means in their research, but not the same methods because of the different object of study, which requires a different approach to reality. So, the specifics of methods of a certain scientfic field can be determined only if the main objectives of its research are taken into account. The contemporary methodological practicism has completely neglected the content-based definition of a sociological method, tending to replace the term *sociological method* itself with the term *social research* (Milić, 1996, p. 233).

The progress in sociological methodology mostly involves the improvement of the existing and discovery of new data collection methods, which has improved the exactness of Sociology, and developed it from the scientific perspective so that it can compete with natural sciences. Such sudden development of methods has some disadvantages, too. Most sociologists have started shying away from the scientific explanation of facts exactly because it cannot be reached by the exact methods used for the collection of data. This leaves us without any valuable scientific conclusion. As for the contemporary data collection methods, team work on data collection is important because social phenomena are becoming more complex, and it is impossible for one scientist to notice and collect everything properly. Such work requires a precisely determined, thorough and strict work plan, especially when a wider social entity is explored with regard to all the aspects of its social life. Some examples of this include the study of a settlement, class, stratum (Pečujlić, 1982). Data collection methods, or observation in a broader sense, are divided into six basic types:

- *observation in the narrow sense – direct or indirect observation of a phenomenon;*
- *examination – direct observation;*
- *measuring – precise quantitative observation;*
- *statistics – observation of mass phenomena, which can be counted and measured by means of mathematics;*
- *experiment – observation of phenomena induced by the observer and altered for the purpose of the research;*

- *comparison – comparative observation of several phenomena (Pečujlić, 1982, p. 493).*

As for the methods of knowledge acquisition, we can distinguish between:
- *basic general and*
- *basic specific methods.*

General methods are more or less used in all scientific fields. They include the general and specific dialectical method, modelling method, statistical method, and axiomatic method. Besides general methods, there is a range of specific fundamental methods such as analysis and synthesis, classification and generalisation, induction and deduction, as well as basic scientific procedures and cognitive processes such as concept defining, hypothesis setting, proving attitudes, checking the obtained scientific knowledge, etc. (Šešić, 1979, p. 7).

1.5. Research Designs

Research designs are plans and procedures for carrying out research that span the decisions from broad assumptions to detailed methods of data collection and analysis. The overall decision relates to which design should be used to study a topic. The selection of a research design is also based on the nature of the research problem or issue being addressed, the researchers' personal experiences, and the audiences for the study (Cresswell, 2009, p. 22).

There are three types of designs: qualitative, quantitative, and mixed methods designs. The three approaches are not as discrete as they first appear. Qualitative and quantitative approaches should not be viewed as polar opposites or dichotomies; instead, they represent different ends on a continuum (Newman & Benz, 1998). A study tends to be more qualitative than quantitative or vice versa. Mixed methods research resides in the middle of this continuum because it incorporates the elements of both qualitative and quantitative approaches.

Quantitative research is a means of testing objective theories by examining the relationships between variables. These variables, in turn, can be objectively measured, typically using some instruments, so as to obtain numerical data, which can be analysed using statistical procedures (Dörnyei, 2007, p. 24). The final written report has a set structure consisting of introduction, literature and theory, methods, results, and discussion (Creswell, 2008). Those who engage in this form of inquiry have assumptions about testing theories deductively, building in protections against bias, controlling for alternative explanations, and being able to generalize and replicate the findings. Quantitative methods include: statistical method, surveys, experiments, measurement scales, etc.

Qualitative research is a means of exploring and understanding the meaning individuals or groups ascribe to a social or human problem. The process of research involves emerging questions and procedures, data typically collected in the participant's setting, data analysis inductively building from particulars to general themes, and the researcher making interpretations of the meaning of the data. The final written report has a flexible structure. Those who engage in this form of inquiry support a way of looking at research that honors an inductive style, a focus on individual meaning, and the importance of rendering the complexity of a situation (Creswell, 2007). Some examples of qualitative methods are: observation, interview, focus group, case study, etc.

Mixed methods research is an approach to inquiry that combines or associates both qualitative and quantitative forms. It involves philosophical assumptions, the use of qualitative and quantitative approaches, and the mix of both approaches in a study. Thus, it is more than the simple collection and analysis of both kinds of data. It also involves the use of both approaches in tandem so that the overall strength of a study is greater than either qualitative or quantitative research (Creswell et al., 2007). Nowadays, these methods are gaining importance.

2. QUALITATIVE RESEARCH

2.1. Definition and Distinctive Characteristics

> *'Qualitative research begins with assumptions, a worldview, the possible use of a theoretical lens, and the study of research problems inquiring into the meaning individuals or groups ascribe to a social or human problem. To study this problem, qualitative researchers use an emerging qualitative approach to inquiry, the collection of data in a natural setting sensitive to the people and places under study, and data analysis that is inductive and establishes patterns or themes. The final written report or presentation includes voices of participants, the reflexivity of the researcher, and a complex description and interpretation of the problem, and it extends to literature or signals a call for action*' (Creswell, 2007: 37).

The above-given definition covers all the main characteristics of qualitative research. They include as follows:

- ***Natural setting*** *– data are collected in the field or at the site where participants experience the issue or problem being studied by talking to the people directly, and watching them behave and act within their context.*
- ***Researcher as the key instrument*** *– the researchers are the ones who actually gather the information without using or relying on questionnaires or instruments developed by other researchers. They collect data themselves through examining documents, observing participants' behaviour, and interviewing participants.*
- ***Multiple sources of data*** *– multiple forms of data are gathered by means of interviews, observation and documents instead of relying on a single data source. Then the researchers review all of the data, and make sense of them by organising them into categories or themes that cut across all of the data sources.*
- ***Inductive data analysis*** *– qualitative researchers use the bottom-up approach to build their patterns, categories, and themes, i.e. they organise the data into increasingly more abstract units of information. This requires going back and forth between the themes and the database until a comprehensive set of themes is established. It may also involve interactive collaboration with participants, who are given a chance to shape the themes or abstractions that emerge from the process.*

- ***Participants' meaning*** *– throughout the research process, the researchers keep a focus on discovering the views that the participants hold about the problem or issue, not the meaning that the researchers bring to the research or find in the literature.*
- ***Emergent design*** *– the qualitative research process is emergent, which means that the initial plan for research cannot be tightly prescribed, and that all the phases of the process may change or shift once the researcher enters the field, and starts collecting the data (the questions may change, the form of data collection may shift, the individuals and sites under study may be modified) with an aim to learn about the problem from participants.*
- ***Theoretical lens*** *– the theoretical lens, such as the concept of culture, gendered, racial or class differences, is often used to view the research.*
- ***Interpretive inquiry*** *– the researchers make an interpretation of what they see, hear, and understand, so these interpretations cannot be separated from the researchers' own background, history, context, and prior understanding. Once the research report is issued, the readers, as well as the participants, interpret it, offering the new interpretation to the study, and thus multiple views of the problem emerge.*
- ***Holistic account*** *– a complex picture of the problem under study is developed by identifying the complex interactions of factors in any situation, i.e. by reporting multiple perspectives, identifying the many factors involved in a situation, and generally sketching the larger picture that emerges.*

Qualitative research is appropriate when there is a need to study a group or population, hear the silenced voices, obtain a complex, detailed understanding of an issue, or context/setting in which the participants in a study address a problem, which cannot be done without talking directly to people, going to their homes or places of work, and allowing them to tell the stories unaffected by what we expect to find or what we have read in the literature. The qualitative research empowers the individuals to share their stories, have their voices heard, as well as to collaborate with the researcher throughout the data analysis and interpretation phases of the research. Qualitative research is often used as a follow-up of quantitative research, providing the explanation of why people reacted as they did, of the context in which they responded, and their deeper thoughts that governed their responses. Qualitative research helps to capture the interactions between people, including their individual differences, which cannot be accomplished by quantitative measures, which level all individuals to a statistical mean.

Qualitative research is time-consuming regarding both data collection and data analysis. The researcher spends many hours in the field, collecting data, trying to gain access, and establish rapport. The data analysis implies sorting through large amounts of data, and reducing them to a few themes or categories, which is followed by writing long reports, showing multiple perspectives, and incorporating quotes to support these perspectives (Creswell, 2007, p. 41).

2.2. Qualitative Theory Use

Qualitative researchers use theory in their studies in several ways. First, like in quantitative research, it is used as a broad explanation for behaviours and attitudes, and it may be complete with variables, constructs, and hypotheses. For example, ethnographers employ cultural themes or 'aspects of culture' (Wolcott, 1999, p. 113) to study in their qualitative projects, such as social control, language, stability and change, or social organization systems, such as kinship or families. Themes in this context provide a ready-made series of hypotheses from literature to be tested. Although researchers might not refer to them as theories, they provide broad explanations that anthropologists use to study the culture-sharing behavior and attitudes of people.

Secondly, researchers increasingly use a theoretical lens or perspective in qualitative research, which provides an overall orienting lens for the study of issues of gender, class, and race (or other issues of marginalized groups). This lens becomes an advocacy perspective that shapes the types of questions asked, informs how data are collected and analysed, and provides a call for action or change.

Qualitative research of the 1980s underwent a transformation to broaden its scope of inquiry to include these theoretical lenses. They guide the researchers as to what issues are important to examine (e.g. marginalization, empowerment), and the people that need to be studied (e.g. women, homeless, minority groups). They also indicate how the researcher positions himself or herself in the qualitative study (e.g. up front or biased from personal, cultural, and historical contexts), and how the final written accounts need to be written (e.g. without further marginalizing individuals, by collaborating with participants). In critical ethnography studies, researchers begin with a theory that informs their studies. This causal theory might be one of emancipation or repression (Thomas, 1993). Creswell (2007) provides a list of some of these qualitative theoretical perspectives available to the researcher, which include as follows:

- *Feminist perspectives – they view as problematic women's diverse situations, and the institutions that frame those situations. Research topics may include policy issues related to ensuring social justice for women in specific contexts or raising the awareness about oppressive situations for women (Olesen, 2000).*
- *Racialized discourses – they raise important questions particularly about people and communities of color (Ladson-Billings, 2000).*
- *Critical theory perspectives – they are concerned with empowering human beings to transcend the constraints placed on them by race, class, and gender (Fay, 1987).*
- *Queer theory – it focuses on individuals calling themselves lesbians, gays, bisexuals, or transgendered people. The research using this approach does not objectify individuals, it is concerned with cultural and political means, and conveys the voices and experiences of individuals who have been suppressed (Gamson, 2000).*
- *Disability inquiry – it addresses the meaning of inclusion in schools, and encompasses administrators, teachers, and parents who have children with disabilities (Mertens, 1998).*

Rossman and Rallis (1998) capture the sense of theory as critical and postmodern perspectives in qualitative inquiry. At the end of 20th century, traditional social science came under increasing scrutiny and attack as those espousing critical and postmodern perspectives challenged objectivist assumptions and traditional norms for the conduct of research. Central to this attack are four interrelated notions:

- *research fundamentally involves issues of power;*
- *the research report is not transparent, but rather authored by a raced, gendered, classed, and politically oriented individual;*
- *race, class, and gender are crucial for understanding experience; and*
- *historic, traditional research has silenced members of oppressed and marginalized groups (Rossman and Rallis, 1998, p. 66).*

Thirdly, distinct from this theoretical orientation are qualitative studies in which theory (or some other broad explanation) becomes the end point. It is an inductive process of building from data to broad themes, to a generalized model or theory (Punch, 2005). The researcher begins by gathering detailed information from participants, and then transforms this information into categories or themes. These themes are developed into broad patterns, theories, or generalizations, which are then compared with personal experiences or with the literature that exists on the topic. The development of themes and categories into patterns, theories, or generalizations suggests varied end points for qualitative studies. For example, in case study research, Stake (1995, p. 86) refers to an assertion as a propositional generalization – the researcher's summary of interpretations and claims, to which the researcher's own personal experiences are added, is called 'naturalistic generalizations'. As another example, grounded theory provides different end points. Inquirers hope to discover a theory that is grounded in information collected from participants (Strauss & Corbin, 1998). Lincoln and Guba (1985) refer to 'pattern theories' as explanations that develop during naturalistic or qualitative research. Rather than the deductive form found in quantitative studies, these pattern theories or generalizations represent interconnected thoughts or parts linked to a whole.

Finally, some qualitative studies do not employ any explicit theory. However, the case can be made that no qualitative study begins from pure observation, and that prior conceptual structure composed of theory and method provides the starting point for all observations (Schwandt, 1993). Still, one sees qualitative studies that contain no explicit theoretical orientation, such as in phenomenology, in which inquirers attempt to build the essence of experience from participants (Riemen, 1986). In these studies, the inquirer constructs a rich, detailed description of a central phenomenon.

2.3. Qualitative Research Questions

According to Creswell (2009), in a qualitative study, inquirers state research questions, not objectives. These research questions assume two forms: the central question, and associated subquestions. The central question is a broad question that asks for an exploration of the central phenomenon or concept in a study. The inquirer poses this question, consistent with the emerging methodology of qualitative research, as a general issue so as not to limit the inquiry. To arrive at this question, one should ask: 'What is the broadest question that I can ask in the study?' Beginning researchers trained in quantitative research might struggle with this approach because they are accustomed to the reverse approach: identifying specific, narrow questions or hypotheses based on a few variables. In qualitative research, the intent is to explore the complex set of factors surrounding the central phenomenon, and present the varied perspectives or meanings that participants hold. Creswell (2009) also provides the guidelines for writing broad, qualitative research questions:

- *Ask one or two central questions followed by no more than five to seven subquestions. Several subquestions follow each general central question; the subquestions narrow the focus of the study, but leave open the questioning. The subquestions, in turn, can become specific questions used during interviews (or in observing or when looking at documents). In developing an interview protocol or guide, the researcher might ask an icebreaker question at the beginning, for example, followed by five or so subquestions in the study. The interview would then end with an additional wrap-up or summary question, or ask: 'Who should I turn to, to learn more about this topic?' (Asmussen & Creswell, 1995).*
- *Relate the central question to the specific qualitative strategy of inquiry. For example, the specificity of the questions in ethnography at this stage of the design differs from that in other qualitative strategies. In ethnographic research, Spradley (1980) advanced a taxonomy of ethnographic questions that included a mini-tour of the culture-sharing group, their experiences, use of native language, contrasts with other cultural groups, and questions to verify the accuracy of the data. In critical ethnography, the research questions may build on a body of the existing literature. These questions become working guidelines rather than truths to be proven (Thomas, 1993, p. 35). Alternatively, in phenomenology, the questions might be broadly stated without specific reference to the existing literature or a typology of questions. Moustakas (1994) talks about asking what the participants experienced, and about the contexts or situations in which they experienced it. In grounded theory, the questions may be directed towards generating a theory of some process. In a qualitative case study, the questions may address a description of the case, and the themes that emerge from studying it.*
- *Begin the research questions with the words what or how to convey an open and emerging design. The word why often implies that the researcher is trying to explain why something occurs, and this suggests a cause-and-effect type of thinking associated with quantitative research instead of the more open and emerging stance of qualitative research.*
- *Focus on a single phenomenon or concept. As a study develops over time, factors will emerge that may influence this single phenomenon, but one should begin a study with a single focus to explore in great detail.*
- *Use exploratory verbs that convey the language of emerging design:*
  - *discover (e.g. grounded theory);*
  - *seek to understand (e.g. ethnography);*
  - *explore a process (e.g. case study);*
  - *describe the experiences (e.g. phenomenology);*
  - *report the stories (e.g. narrative research).*
- *Use these more exploratory verbs that are non directional rather than directional words that suggest quantitative research, such as 'affect', 'influence', 'impact', 'determine', 'cause', and 'relate'.*

- *Expect the research questions to evolve and change during the study in a manner consistent with the assumptions of an emerging design. In qualitative studies, the questions are often under continual review and reformulation (as in a grounded theory study). This approach may be problematic for individuals accustomed to quantitative designs, in which the research questions remain fixed throughout the study.*
- *Use open-ended questions without reference to the literature or theory unless otherwise indicated by a qualitative strategy of inquiry.*
- *If closed-ended questions, which are considered quantitative, are necessary (e.g. rating something/the satisfaction with something on a given scale), a text box should be added that asks for extra comments on why a specific rating was chosen, thus providing qualitative insights alongside their respective quantitative research question responses.*
- *Specify the participants and the research site for the study, if the information has not been given yet.*
- *Ensure that research questions are ethical and free from bias (it is always good to have another person check for unconscious bias).*
- *Consider the language used, and make sure it is clear and easy to understand. Therefore, jargon, acronyms, and overly technical language should be avoided.*

2.4. Types of Qualitative Research Questions

Marshall and Rossman (1989) identified four qualitative research question types, each with its own typical research strategy and methods:

- ***Exploratory questions:*** *these questions are used when relatively little is known about the research topic. The researchers usually interview participants, hold focus groups, or a case study to delve deeper into the phenomena.*
- ***Explanatory questions****: the research topic is approached with the aim to understand the causes lying behind phenomena. Multiple interconnected factors that have influenced a particular group or area are analysed*
- ***Descriptive questions:*** *these questions aim to document and record what is happening, and in order to answer them, researchers might interact directly with participants using surveys and interviews or observational and etnography studies that collect data on how participants interract with their wider environment.*
- ***Predictive questions:*** *these questions start from the phenomena of interest and investigate their future ramifications. They may involve looking back as well as forward. The researchers use content analysis, questionnaires, and studies of non-verbal communication.*

3. QUALITATIVE RESEARCH DESIGNS

A research design is 'a set of guidelines and instructions to be followed in addressing the research problem' (Creswell et al., 2007, p. 238). The definition implies that a research problem or research question informs the choice of design, whereas a research design serves as the plan that the researcher will use in conducting the research and addressing the problem in such a way that will ensure the maximum validity of the findings. These questions are open-ended, calling for views supplied by participants in a study (Creswell et al., 2007, p. 238). The research design influences the choice of research methods, i.e. the strategies to be used to implement the plan.

Some of the most common qualitative research designs include: narrative research, grounded theory, phenomenology, participatory action research (PAR), ethnomethodology, and historical study.

3.1. Narrative Research

Narrative research is a specific type of qualitative designs in which 'narrative is understood as a spoken or written text giving an account of an event/action or series of events/actions, chronologically connected' (Czarniawska, 2004, p. 17). It implies studying one or two individuals, collecting data through their stories in which they report individual experiences, and then chronologically ordering the meaning of the experiences (Creswell et al., 2007, p. 240). Narratives may have a guiding theoretical lens or perspective (e.g. a feminist lens).

According to Creswell (2007), various forms of narrative study can be found in the narrative research practice, such as:

- *biography – the researcher writes about/records the experiences of another person's life;*
- *autobiography – the individuals who are the subjects of the study write about/record the experiences of their lives;*
- *life histories – portrayals of an individual's whole life;*
- *oral histories or personal-experience stories – a study of someone's personal reflections of events, and their causes and effects, collected from one or several individuals (Plummer, 1983).*

The narrative study procedure implies:

- *determining if the research problem or question best fits the narrative study;*
- *selecting one or two individuals of interest to the study and having them tell their stories (field texts), thus collecting raw data. The field texts may include a record of their stories in a journal or diary. The researcher may also observe them, and record field notes, or collect letters sent by them, gather documents about them, stories about them from family members, photographs, etc.;*
- *collecting information about the context of these stories (the participants jobs, homes, culture, historical contexts, etc.);*
- *the researcher analyses the participant's stories searching for key elements, and orders them into a chronological presentation. It is this chronology with an emphasis on the sequence that represents the distinctive feature of narrative research. The final story might include the elements typically found in novels, such as time, place, plot, and scene;*
- *collaborating with participants by actively involving them in the research and negotiating the meaning of stories with them to add to the validity of the analysis (Creswell & Miller, 2000).*

These narrative research procedures and characteristics make this research design challenging to use – an extensive amount of information needs to be collected, the context of the individual's life needs to be clearly understood, active collaboration with the participant is mandatory, a keen eye is needed to identify the particular stories that capture the individual's experiences in the gathered source material, and the researcher has to be reflective about their own personal and political background, which determines how they present the participant's stories (Creswell, 2007, p. 57).

SELF-ASSESSMENT

1. Are the statements given below true or false?

- *Random samples are used in narrative research.*
- *Narrative is a spoken text.*
- *Questionnaires are the best data collection instruments in qualitative research.*
- *The researcher's prior knowledge and experience cannot affect the data analysis.*
- *Chronology is unimportant in narrative research.*

Additional source of information:

https://www.youtube.com/watch?v=zKTH5_k50vM&t=37

3.2. Grounded Theory

Grounded theory is a qualitative research design in which the researcher generates a general explanation or a theory of a process, action or interaction based on the views of a large number of participants, i.e. theory is grounded in data from the field, not off the shelf (Creswell, 2007, p. 62). Despite the diversity of the collected data, the grounded theory approach presumes it is possible to discover fundamental patterns, or basic social processes, in social life.

The grounded theory procedure includes as follows:

- *deciding if the grounded theory is best suited to study the research problem (there is no theory available to explain the problem, or there are available theories but developed on samples and populations other than those of interest to the researcher);*
- *data collection, typically in one-on-one interviews, from a larger number of individuals who have directly experienced an action, interaction or process. Other forms of data collection may be used such as observations, documents, and audiovisual materials;*
- *data analysis aimed at forming categories of information by segmenting information, or grouping statements into broad ideas (open coding), while trying to identify the core idea. The researcher carefully reads and re-reads the data, breaking it down into discrete incidents or ideas. Each of these incidents is then given a code – a word or short phrase that represents the essence of that piece of data. Open coding is also where constant comparison begins. As each piece of data is coded, it is compared to other data coded in the same way, and during this stage, categories start to emerge.*

- *axial coding – the researcher constantly compares data within a category, as well as compares categories to each other. The researcher usually returns to the participants to ask more detailed questions in order to develop or saturate the model.*
- *selective coding – the researcher has a clear idea of the main categories and how they relate to each other. The researcher connects the categories, discussing the relationships between these categories and the core category. The core category represents the main theme or process that the theory explains.*
- *developing a theoretical model, with which the study may end or which may be tested later for its empirical verification with quantitative data to determine if it can be generalized to a sample and population.*

Both inductive and deductive approach to theory development are used in the grounded theory design because concepts are grounded in data, and hypotheses are tested as they arise from the research (Field & Morse, 1985, p. 23), but theory generation is more important than theory testing. Purposeful sampling is used, i.e. the researcher looks for certain subjects who will be able to shed new light on the phenomenon being studied. Diversity rather than similarity is sought in the participants. Data collection takes place in natural settings, and primarily involves participant observation and interviews. Data collection and analysis occur simultaneously, and new data are constantly compared with the already gathered data through a process called *constant comparison.*

Theoretical saturation is a critical concept in grounded theory. It refers to the point at which no new insights or concepts can be found in the data, indicating that the categories are well-developed, and that further data collection is unnecessary.

A grounded theory study is challenging in that it requires the researcher to set aside as much as possible theoretical ideas or notions so that that an analytic, substantive theory can emerge. It is difficult to determine when the saturation of the categories is reached or when the theory is sufficiently detailed. One way of ensuring this is discriminant sampling, i.e. gathering additional information from individuals similar to those people initially interviewed to determine if the theory holds true for these additional participants is one way of ensuring this.

SELF-ASSESSMENT

1. ***Which of the following statements are correct?***

   1 *The purpose of the research is to test an existing theory.*
   2 *Social phenomena are understood from the perspective of those experiencing it.*
   3 *Selective coding implies detailed, line-by-line data examination.*
   4 *Data collection is an iterative process and continues throughout the research process.*
   5 *Selective coding ends when theoretical saturation is reached.*
   6 *The researcher plays a critical role in grounded theory.*

3.3. Phenomenological Research

As in the grounded theory, the views of a number of participants are collected, but instead of theorizing from these views, phenomenologists describe what all the participants have in common as they experience a phenomenon, with an aim to reduce their experiences with a phenomenon to a description of the universal essence. So, the aim is not an explanation or analysis, but a description of the essence of people's experiences as they are lived every day, which requires the researcher's bracketing or setting aside their own experiences as much as possible to take a fresh perspective towards the phenomenon studied (Creswell, 2007, p. 59). The procedure involves the following steps:

- *identifying the phenomenon;*
- *the researchers identify their own experiences with the phenomenon, as well as what they expect to discover, and then deliberately put aside these ideas, thus bracketing their own views with an aim to be as objective as possible, and see the experience from the eyes of the person who has lived the experience;*
- *selecting the participants who have experienced the phenomenon. It is recommended that from 5 to 25 individuals be interviewed;*
- *collecting information, most often through interviews or multiple interviews, but the participants may also write about their experiences. According to Moustakas (1994), there are two broad, general questions that have to be asked in order to perform phenomenological research: (1) What have you experienced in terms of the phenomenon? (2) What contexts or situations have typically influenced or affected your experiences of the phenomenon? Of course, they may be followed by other open-ended questions;*
- *data analysis – highlighting significant statements, sentences or quotes that provide an understanding of the overall experience;*
- *collapsing these statements into broader themes, then going back through the transcripts to look at the themes more closely;*
- *describing the essence of the experience, i.e. the common experiences of the individuals studied (Creswell et al., 2007, p. 255).*

In order to carry out a phenomenological study, the researcher needs to have at least some understanding of the broader philosophical assumptions. The participants need to be carefully selected to make sure that they have experienced the phenomenon in question. Bracketing personal experiences may be difficult for the researcher to do (Creswell, 2007, p. 62).

3.4. Participatory Action Research

> *'Action research brings together action and reflection, as well as theory and practice, in participation with others, in the pursuit of practical solutions to issues of pressing concern' (Bradbury, 2015, p. 1).*

Unlike other qualitative designs, the main aim of PAR is to produce social change and improve the quality of life in opressed and exploited communities (Stringer, 1999). Its unique feature is that the researcher and members of the community collaborate at all levels of the research process to help find a suitable solution for the social problem that significantly affects the community (Creswell et al., 2007, p. 255).

PAR is a social process in which the researcher deliberately explores the relationship between the individual and other people in order to find out how individual relationships are formed and reformed through social interaction. 'Participatory' means that people are involved in examining their understandings, skills, and values, 'action' relates to the aim of improving the lives of individuals by studying the problems they face.

There is no clear procedure for conducting this type of research. The request for the research might originate from the members of the community who ask the researchers to help them or from the researcher who is an active member of the community, and they establish a mutually respectful and collaborative relationship. Then they identify and define community problems, develop research questions to be explored, and discuss the resources needed to address the problems. Then the plan of research is developed, and data are collected. Community members are included in data analysis, and recommendations for policy changes are made, i.e. an action plan is designed, the implementation of which is afterwards monitored (Creswell et al., 2007, p. 258).

SELF-ASSESSMENT

TASK 1. Watch the recording and identify seven activities the PAR usually includes:

https://www.youtube.com/watch?v=8ISl7JKQuxw

3.5. Ethnomethodology

Ethnomethodology is a new research design. As its name implies, it refers to the study of a nation and characteristics of its communities. It is a bit related to ethnograhy, which provides the description of the collective values of a community, though it also describes other important elements of a social group, such as its tools, food, culture, and way of life. Ethnomethodology places an emphasis on the attitudes of a society, and the way in which they are expressed linguistically. Unlike sociology, which does not have a data collection system, ethnomethodology endeavours to develop one.

Ethnomethodology uses facts obtained from the speech of an ordinary person, a man from the people. It is aimed at defining the attitudes of individuals towards their society, formed on the basis of their environment, its influences, and interpersonal communication, i.e. at understanding how people make sense of their own lived realities and practices. People have a perception of something, be it what they simply see or what they accept or do not accept through various media, and they use it to form an opinion. The purpose of the ethnomethodological design is to collect the facts about how, i.e. in which social conditions and situations, individuals gain emotional perceptions of reality (Ristić, 2016, p. 228).

The object of the ethnomethodological research is defined by people, because people decide what is important to them by expressing their emotions. The aim is to express one's feelings and perceptions regardless of those who disagree with them. Every person needs to have a personal perception of society, and if they find the collective perception unacceptable, the attitudes can change. People create society and therefore people can change it.

Ethnomethodology is a new scientific design, and based on the type of perceptions, can be divided into limited and unlimited ethnomethodology, as well as ethnomethodology of direct and indirect influence. Limited ethnomethodology deals with a specific group of people and their attitudes. This method is used to analyse the relationships within a group, and how people are perceived. Unlimited ethnomethodology goes beyong a specific group to other social wholes. It shows what the members of a group think of their group, and whether they accept other groups. Direct influence ethnomethodology is characterised by personal and autonomous perceptions of the members of a society or group. The aim is to identify similar opinions, and group them. On the other hand, indirect influence ethnomethodology is characterised by external influences. The members' opinions are corrected under the external pressure in order to conform to the majority group.

Data is usually collected through the extended observation of the behaviour, language and the interaction among members of a culture-sharing group, and interviews with the most knowledgeable members of a community, who are called key informants. Researchers have to bracket themselves. The data collection and analysis occur simultaneously. According to Creswell (2007), the procedure of conducting such research includes the following steps:

- *determining if etnomethodology is the most appropriate design to use to study the research problem (it is appropriate when it is necessary to describe how a cultural group functions, and to explore the beliefs, language, behaviours, and issues such as power, resistance and dominance);*
- *identifying or locating a culture-sharing group to study – the one that has been together for an extended period of time so that their shared language, patterns of behaviour, and attitudes have merged into a discernable pattern, or a group marginalised by society;*
- *selecting cultural themes or issues to study about the group (such as learning, socialisation, cognition, domination, inequality, etc.);*
- *fieldwork/gathering information where the group works and lives, while respecting the daily lives of individuals on the site. Observations, interviews, tests, surveys, audiovisual methods can be used for data collection;*
- *data analysis – description of themes that emerge from the group, and an overall analysis of how the group functions and lives;*
- *providing a holistic cultural portrait of the group that incorporates the views of the participants, as well as the views of the researcher, so that the reader learns about the culture-sharing group from both the participants' and researcher's interpretation. It might advocate for the needs of the group or suggest changes to address these needs. Theatre productions, plays or poems can be some of the end products of ethnomethodology.*

In order to engage in ethnomethodology, the researcher needs to have a grounding in cultural anthropology and the meaning of a socio-cultural system. Data collection is extensive and time-consuming. There is a possibility that the researcher will go native, and be unable to complete the study. The researcher needs to acknowledge his or her impact on the people and places being studied (Creswell, 2007, p. 72).

3.6. Historical Study

Historical studies imply the identification, location, evaluation, and synthesis from the past, with an aim not to only discover the events of the past, but to relate these past events to the present, and to the future. According to Leininger (1985, p. 109), 'Without a past, there is no meaning to the present, nor can we develop a sense of ourselves as individuals and as members of groups'.

The process of conducting a historical study includes some typical steps such as problem identification, literature review, data collection and analysis. The data are usually found in documents, artefacts and relics, but are also obtained through oral reports. The data sources may be found in libraries, archives, or in personal collections.

The sources of historical data are divided into primary and secondary sources, the former providing first-hand information or direct evidence, and the latter providing second-hand information. Primary sources include: oral histories, written records, diaries, eyewitnesses, photographs, and physical evidence. Secondary sources often use the primary sources to analyse the topic.

The collected data undergo two types of evaluation – external and internal criticism. External criticism is concerned with the authenticity of the data (validity), whereas internal ctiticism examines the accuracy of the data (reliability), and it follows the extrenal criticism. Internal criticism is more difficult to conduct because when evaluating the material in a document and determining if the material is accurate, motives and possible biases of the author must be taken into consideration.

3.7. Case Study

A case study is a common framework for conducting qualitative research (Stake, 1995). A case-study design is defined as systematic binding of a whole based on its unique character, regardless of whether it is a phenomenon, process, relationship, individual, group of people or even a whole society. It focuses on an issue with the case selected to provide insight into the issue, and this is what distinguishes it from a narrative study, especially when an individual is selected as the case – the focus is not on the individual and their stories, but on the issue, with the individual case selected to help understand the issue (Creswell, 2007, p. 245). According to Yin (2003), case studies ensure understanding of complex social phenomena by allowing investigators to retain the holistic and meaningful characteristics of real-life events, because a detailed description of the case and the setting of the case within contextual conditions are provided, whereas the presentation need not be chronological. It is the study of an issue explored through one or more cases within a bounded system, and provides an in-depth understanding of the case, covering contextual conditions, and relying on multiple data sources (Creswell, 2007, p. 73).

A case study is the analytical research used to study a specific real situation or an imagined scenario. It belongs to the group of relatively young research designs. However, almost all natural sciences can be said to have 'studied cases'. The design appeared quite early in historical sciences, where certain events or societies were treated as cases, whereas in other social sciences, it appeared much later. Psychology and Psychiatry were the last to adopt it when studying certain diseases or their specific manifestations.

Anything that can be bounded or fenced in a unique whole with all its real characteristics can be the object of a case study. It can be an individual, family, settlement, work organization, etc. The case as a bounded whole is studied over time, using multiple sources of information (questionnaires, observations, interviews, documents, reports) in order to discover its characteristics, and learn more about the unfamiliar or poorly understood situation (Leedy & Ormrod, 2005). So, case studies are typically regarded as qualitative research because a single case cannot be representative of the population, though they often include quantitative data collection instruments such as questionnaires (Dörnyei, 2007, p. 152). According to Dörnyei, a case study is not a specific design, but a method of collecting and organising data so as to maximise our understanding of the unitary character of the social being or object studied.

The complexity of the case-study design is evident in the theoretical and operational defining of the relevant properties and relationships of the studied case, and it is particularly complex to preserve the wholeness of the case and meaningful characteristics of actual relationships. If the case is a process or relationship, then it is extremely difficult to define its beginning and ending point, its relevant context, its spacial and temporal demarcation. So, it is here that assumptions, as an important element of the research plan, play an important role. They serve to connect the initial questions and research objectives with relevant data and results obtained later.

A single-case study is used when an individual, phenomenon, process, etc. is the object of a study, and as such, the best representative of the population. A single-case study is convenient in the following situations:

- *when an extreme or unique character of the studied phenomenon is to be determined;*
- *when a phenomenon which has never been researched before is to be studied, and the case study will result in some discovery, even if it only describes the phenomenon;*
- *when carrying out preliminary research as the foundation for further research – this is useful when some unknown or insufficiently investigated or very complex phenomena are to be studied, and the aim is to use the knowledge obtained to develop the research plan, or make the existing plan more precise.*

Sometimes research has an immediate experience as its starting point, and then empirical regularities are discovered, whereas the conceptual framework results from the very process of the research in the form of assumptions. In such a process, the investigator's previous experience and creativity are very important because theory cannot be developed only from observation, nor can generalizations be made using induction only. Theory development is a creative process, which exceeds simple recording.

A single-case strategy can develop into a multiple-case method in the situations where the case relates to something general, and needs to be explored in all its varieties (Pečujlić & Milić, 1995, p. 130). A multiple-case design is used when a research study examines more than one case, following the repeating logic – new cases are studied under the theoretically precisely defined conditions, as is done in multiple experiments. Plans for both types of the case-study method are developed within the same methodological framework. However, using the multiple-case approach requires more serious standardization of the process, more serious selection of key dimensions and relevant data, but fewer details than when a single-case method is used. A multiple-case study is more difficult to perform, more time-consuming, and usually requires more than one researcher. It serves as a strategy for the strengthening of epistemological foundations of the approach (Ševkušić, 2008, p. 242).

According to some authors, there are four distinguishing properties of a case-study method:

- *the data must be diverse, and have the demographic, financial, political, cultural and historical character;*
- *the data must be comprehensive, and reveal the most important characteristics of the case studied;*
- *the type of the case is defined based on precise criteria;*
- *it includes the temporal dimension of the data, which is important for the case development (Pečujlić & Milić, 1995, p. 130).*

The use of multiple data sources enhances the cognitive value of the result, and increases the reliability of the whole study (Yin, 2003). That is why the draft research itself should predict the sources that can be considered relevant with regard to the research objectives (biographies, diaries, personal history, data collected through observation, informal interviews). Thus, the data are supplemented and converged, and gradually certain patterns are identified, theoretical assumptions are supported by concrete content, new hypotheses relating to the central topic are formulated. It is this convergence of data from multiple sources that adds strength to the findings, and facilitates better understanding of the case (Baxter & Jack, 2008, p. 554).

In case of conflicting data from multiple sources, additional evidence is sought. The more evidence sources there are the more measures of a single phenomenon are obtained. The data sources used in a case study can be divided:

- *according to the way they are organised, into formal and informal;*
- *according to the reasons for their creation, into those created for the purpose of the research, and those created independently of the research, but used for the research purposes.*

Formal sources are those developed by some official institutions and organisations, the most important being those developed by the state authorities at all levels. Informal sources are created by people, and they include different types of personal documents, records, letters, notes, diaries. This type of sources comprises anything that relates to the life of individuals within a whole which represents the case and can be used for research purposes.

The sources created for the purpose of the research are ordered and scientifically oriented; they can be of an institutional origin but reshaped for the purpose of the research. The sources created independently are the same as formal and informal sources.

When conducting a case study, the most important thing is to have a good research plan, which helps to define the case thoroughly, determine its essence, categorise the data collected by observation, define the type and scope of the materials to be used, and thus the protocol is designed. There is no doubt that the success of a case-study method depends on the research plan or protocol (Pečujlić & Milić, 1995, p. 130).

The closeness that a case study establishes with real-life situations, and the wealth of details, are important for researchers for two reasons: first, it is important to get as much a nuanced image of the social reality and human behaviour as possible, because they only exist as such; then, by studying individual cases, the researchers improves their learning process, and develop their research skills.

A great distance from the object of study, the so-called highly desired objectivity, together with a lack of feedback from the research participants, can easily lead to 'ritual' academic studies, with vague effects and benefits, which cannot be tested. As a methodological approach, the case study can be an efficient means against such tendencies.

Types of Case Study

Case studies differ in respect to what is studied, and can be divided into:

- *those relating to formal institutions (official) – this process is based on regulated relationships, where a whole which represents the case is always part of wider society. So, within such society, case boundaries are clearly defined, and it can be objective;*
- *those relating to informal organisations (unofficial) – they are usually social groups whose boundaries are not clear-cut, and therefore attention should be paid to setting the criteria as to how to determine what constitutes the case.*

Based on the way in which the case is generated, case studies can be:

- *spontaneous – self-induced cases, interesting for research because of their social significance (these are usually different types of incidents);*
- *intentionally induced – the case is scientifically designed so that one can, in compliance with their needs, study an important problem.*

In terms of the intent of the analysis, there are:

- *a single or multiple-case instrumental case studies – the researcher focuses on an issue, and then selects one bounded case or several cases to illustrate this issue;*
- *an intrinsic case study – the focus is on the case itself because the case presents an unusual or unique situation (Creswell, 2007, p. 74).*

According to Creswell (2007, p. 76), a case study procedure includes:

- *defining the case within a bounded system;*
- *data collection using multiple information sources to ensure in-depth understanding;*
- *data analysis (a holistic analysis of the entire case or an embedded analysis of a specific aspect of the case) resulting in a detailed case description;*
- *focusing on a few key issues (themes) not for the purpose of generalisation, but to understand the complexity of the case;*

- *a broad interpretation of findings and reporting of the lessons learned from the case.*

The research methodology literature also deals with different ways of reporting the findings of a case study. The chronological and problem-based approaches are most often mentioned, as well as their combination. In addition to them, the standard linear-analytical approach is mentioned, too, which starts with the theoretical defining of the problem, after which methods and data collection and analysis processes are described, and then conclusions are drawn, and their implications explained (Ševkušić, 2008, p. 254).

Advantages and Disadvantages of Case Study

The advantages of the case study research are closely related to its characteristics. There is no other design that can so comprehensively study the whole experience of an individual, phenomenon, organization, etc. There is no other way in which one could collect more comprehensive data about the case, and examine it in its entirety. Another advantage is the in-depth analysis of the case, which is possible due to the comprehensive materials, as well as due to the possibility of discovering the essence, which consists of a series of relationshps established within the whole case.

The design is also important for the examination of the case dynamics, and its development, which reflects the nature of social phenomena, and its history of changes. A case study is extremely easy to combine with other designs, because the wholeness of the procedure itself, as well as diverse facts collected, imply the use of other research methods (Pečujlić & Milić, 1995, p. 132).

A case study also has certain disadvantages. As there are no criteria for defining a case, and the criteria is defined for each specific case, the subjectivity of the design is a disadvantage, i.e. it is up to the researcher to decide what will be explored as a case. Moreover, this method is considered unreliable to a certain extent because it is impossible to generate a unique image based on a single case. Also, the fact that it is time-consuming is a disadvantage. Sometimes it takes several years, and at the end the case is not the same as it was at the beginning, when the research started. Subject dropout may occur during the study due to various reasons. Multi-year research requires huge investments that are necessary at all stages, which may be the reason why researchers avoid using this design (Pečujlić & Milić, 1995, p. 133).

3.8. Key Differences between Some Qualitative Research Designs

Narrative study vs phenomenology: Whereas a narrative study reports the life of a single individual, a phenomenological study describes the meaning for several individuals or their lived experiences of a concept or a phenomenon. Phenomenologists focus on describing what all participants have in common as they experience a phenomenon.

Phenomenology vs grounded theory: Phenomenology describes the experience of a number of individuals, whereas the intent of a grounded theory study is to move beyond the description, and to generate or discover a theory.

Grounded theory vs ethnomethodology: Although a grounded theory researcher develops a theory by examinating many individuals who share the same process, action, or interaction, the study participants are not likely to be located in the same place or interacting on so frequent a basis that they develop shared patterns of behaviour, beliefs, and language. These shared patterns are at the focus of ethnomethodology, and entire cultural group (not only 20 or so individuals) are needed for it.

Case study vs ethnomethodology: An entire culture-sharing group in ethnomethodology may be considered a case, but the intent in ethnomethodology is to determine how the culture works rather than to understand an issue or a problem using the case as its specific illustration.

4. QUALITATIVE RESEARCH METHODS

4.1. Observation

Qualitative observation is the observation in which the researcher takes field notes on the behavior and activities of individuals at the research site, in an unstructured or semistructured way (using some prior questions that the researcher wants to find the answers to). Qualitative observers may engage in a variety of roles from a non-participant to a complete participant.

Observation in its narrowest sense is the oldest and most natural method of collecting facts, significantly improved nowadays. Simple observation is unsystematic, unplanned, random, imprecise, and instead of obtaining accurate data, the observer gets only insufficiently arranged impressions. The beginnings of science in ancient Greece were based on observation. From biology to history, and from Aristotle to Herodotus, the eyes played an important role in registering the phenomena and events of the natural and social world. This is even more evident in modern sciences, such as physics, astronomy, chemistry, etc., whose development and differentiation actually began with observations directed at various elements of the subject reality (Pečujlić, 1982, p. 103). As a specific method, observation spread to social sciences, although it was developing in them as well, primarily in history and psychology. Therefore, it is used as one of the basic data collection methods in all sciences, from economics to sociology.

Unlike the common-sense, vulgar observation, with which it shares some common features and origins, scientific observation is defined as a planned and systematic collection of facts through the direct sensory perception of social phenomena. The plan or system has a number of features. First of all, it implies orchestrated observation, which further implies the creation of all the conditions related to a certain system, and order in conducting the observation, which ensures overcoming voluntarism and arbitrariness of the procedure. Then it implies the precision, which refers to the fulfillment of all the norms that determine the accuracy of the measurement of sensory perceptions. Finally, there is the objectivity of observation, which is also achieved by means of norms that guide the observer's personal wishes and aspirations. The aim is to record details in as objective a manner as possible, by avoiding interpretations and inferences, and by setting aside one's own preconceptions (Angrosino, 2007, p. 42). The specific nature of observation therefore lies in the procedures developed to control it so that it can yield fruitful scientific results (Pečujlić, 1982, p. 103).

4.1.1. Observation Types

A lot of different types and sub-types of observation have developed, and now it has the most complex structure of all data collection procedures.

- ***Cross-sectional Studies***

A cross-sectional study is a type of research that captures information from a population at a single point in time (Figure 1). The information can be obtained from a group of participants with certain characteristics known as variables (e.g. age, gender, origin, education, religion, geographic location, etc). The variables used in a particular study are determined by the research to be carried out, and by the goals of the study. Researchers can use cross-sectional studies to focus on one independent variable, and see how it affects one or more dependent variables. These studies enable researchers to examine several variables at the same time. Cross-sectional studies do not involve the manipulation of variables. They are descriptive studies.

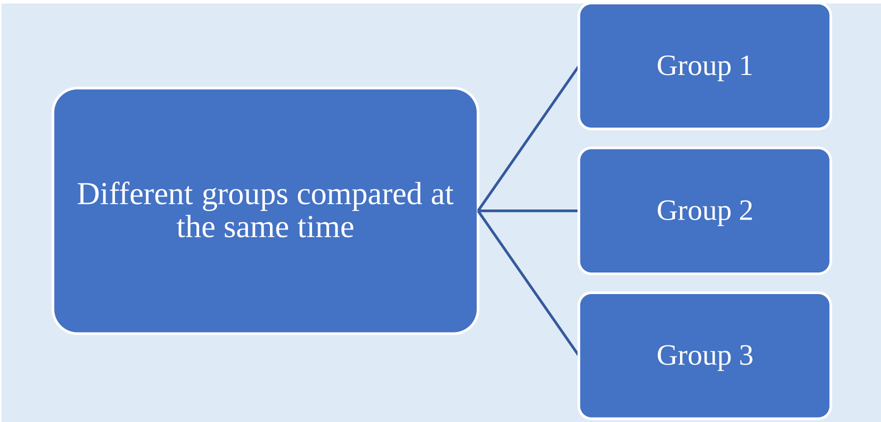


Figure 1. *Cross-Sectional Studies*

Some examples of cross-sectional studies are given below:

- ***Descriptive study****: conducting market research by collecting data on consumer behaviour within the given period of time with an aim to predict the future trends of consumer behaviour or introduce new sales strategies or launch new products and services. In such a study, the researcher does not analyse the causes, but focuses on the description of the current situation on the market.*
- ***Analytical study****: a study examining the lack of calcium in the population of women above 50 years of age. Whereas the focus of the descriptive study would be to determine the level of calcium in women above 50 years of age, the analytical study would focus on examining the factors such as hormonal status, physical activity, nutrition, socio-economic status, education level, etc., in order to explain why some women belonging to such an age group are more likely to develop calcium deficiency than other women, and therefore have a greater probability of developing other illnesses such as osteoporosis.*

- Case-control Studies

Case-control studies are a type of observational research often used in the fields such as medical research or environmental health. A case-control study is an experimental design that compares two or more groups of participants. One of these groups is the ‘case’ group, and other groups are ’control’ groups. It is important to note that the case group is selected because it is already characterised by the attribute of interest for the study, and the purpose of the control group is to find out if the case group systematically exhibits that attribute more than the control group does. Therefore, by analysing the data collected from the control group, the research hypothesis is either confirmed or rejected. Commonly, tables are used to present data (Table 1).

Table 1. *Case-Control Study Data*

	Cases	Controls
Exposed	a	b
Unexposed	c	d

The risk ratio (or rate ratio) is calculated by using the exposure odds ratio (OR):

- *The odds of exposure among cases: OR = a/c,*
- *The odds of exposure among controls: OR = b/d,*

e.g. OR = (axd)/(cxb).

If OR = 1, then the odds of risk is the same for the exposed and non-exposed.

If OR > 1, then the exposure increases the odds of risk.

If OR < 1, then the exposure reduces the odds of risk.

Example: Case-control study in medicine

The aim of the research was to examine the connection between the contamination of potable water and potental stomack diseases (e.g. gastritis). The case-study group comprised the people diagnosed with a stomach disease, whereas the control group comprised the people without the disease. During the research, the data on the exposure of the participants to the contaminated potable water were collected, with the focus on the variables such as the source of the water, and duration of the exposure, for each group. The aim was to compare the results in order to determine whether there was a correlation between the water contamination and the risk of developing a stomach disease.

Practice tasks:

Researchers are conducting a case-control study of breast cancer using incident cases. The study includes 100 cases and 100 controls. 70% of the cases report smoking. Among the controls, 50% report smoking.

a) Prepare a 2x2 table with the given data.

b) Calculate the exposure odds ratio.

c) Interpret the exposure odds ratio in a sentence.

Answer:

a)

	Cases	Controls
Exposed	70 (a)	50 (b)
Unexposed	30 (c)	50 (d)

b) Exposure odds ratio = (a/c)/(b/d) = (a*d)/(c*b) = (70*50)/(50*30)=2.33

c) An odds ratio of 2.33 means that the odds of smokers being a case are 2.33 times the odds of non-smokers being a case.

- Cohort Studies

Cohort studies are a type of analytic studies used to test the hypotheses about cause-and-effect relationships. The term 'cohort' is defined as a group of people, usually comprising 100 or more people who share a common characteristic or experience within a specified period of time (e.g. age, occupation, exposure to a medicine or vaccine, citizenship, etc.). A cohort is divided into 2 parts – those exposed to the supposed risk factor, and those who are not (e.g. smokers and non-smokers). Then the cohort is observed over a certain period of time in order to find out the incidence of death or disease development (which are supposed to be caused by the exposure to the present risk factor), and at the end of the observation period, the comparison between the incidence of death and illness in the exposed and unexposed participants is performed.

Cohort studies are an important method of medical research, convenient for the identifiction of the causes of a disease, because groups of people are observed before they develop a disease. It means that researchers can examine whether there is a cause-and-effect relationship between the participants' selected way of life and their health (Figure 2).

The ongoing Millenium cohort study is following the lives of 19,000 babies born in Great Britain between 2000 and 2001 in order to demonstrate how circumstances in the very first stages of life can influence later health and development. In addition to collecting the data about the health of these babies and their parents, the study is examining the behaviour and cognitive development of the children, as well as a number of other social factors.

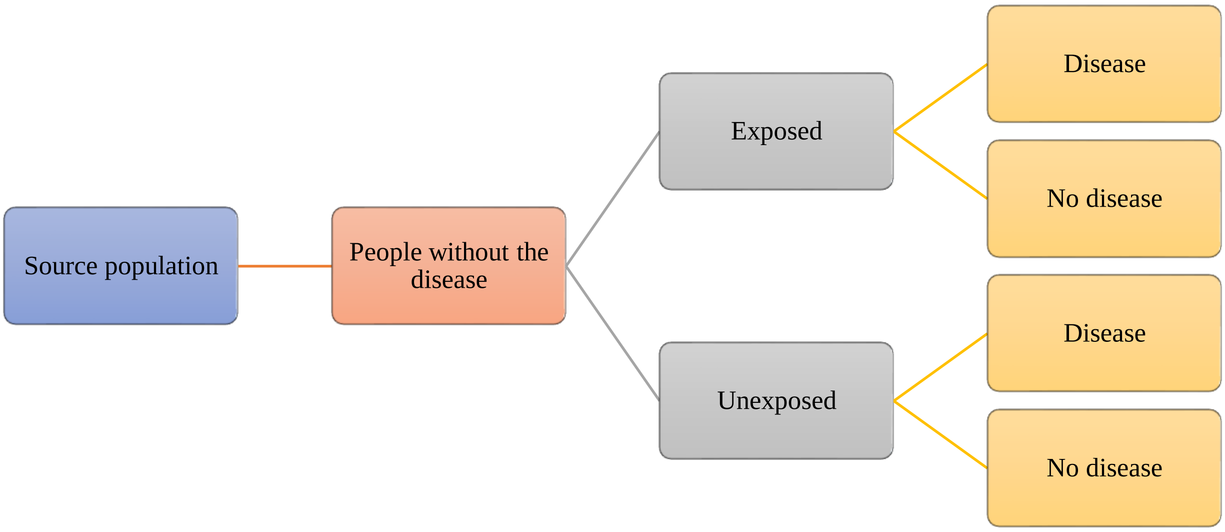


Figure 2. *Example of Cohort Study Design*

The design of a cohort study is shown in the following table:

Table 2. *Cohort Study Design*

Cohort	Diseased	Non-Diseased	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

- *Incidence rates (IR):*

6. *Among the exposed: IR = a/(a+b)*
7. *Among the unexposed: IR = c/(c+d)*

An IR is the indicator of the occurence of an illness in the exposed and unexposed cohort members over the total period of time.

- *Relative risk (RR) – the ratio between the risk of disease in exposed individuals and that in the unexposed:*
  8. *[a/(a+b)]/[c/(c+d)] or a(c+d)/c(a+b)*

The relative risk shows the incidence of an event in the exposed over that in the unexposed individuals.

- *Attributable risk (AR)=[a/(a+b)-c/(c+d)]/[a/(a+b)]*100 is the difference in the incidence rates in the exposed and unexposed individuals over the incidence in the exposed.*

The attributable risk shows the probability of disease prevention provided there is an effective measure of eliminating the exposure.

- Naturalistic Observation

Naturalistic observation is a method of observing how people or animals behave in their natural settings. It is a kind of field research, which means that researchers collect data outside the laboratory or clinical setting. Researchers tend to immerse with the environment as much as possible so as not to affect the participants' natural behaviour in any way.

Based on the position of the observer, observation can be participant and non-participant. The participant observation is the observation when the researchers immerse themselves in the observed phenomenon they are collecting data about. The observer has to integrate with a social group, and take part in their actions. This type of observation is used for two reasons. The first reason is the fact that a lot of groups, such as religious sects, elites, secret associations, are closed and inaccessible to observation. So, in order to obtain the necessary data, the observer has to become part of the group, and 'lose' their identity, i.e. hide their role as the observer. This leads to a lot of ethical and technical problems, which almost raise doubts about the possibility of conducting covert observation.

The second reason why participant observation is used is that it helps to overcome the difference between the observed object and dominant subject, thus introducing the dialectic action into the research. Participant observation revolutionises the relationship between the observer and the observed so that all the subjects become one phenomenon. Therefore, it is not surprising that the action research has accepted this type of observation as important for achieving not only a specific scientific, but also a wider social objective – the change of tradional research. The abovementioned ethical problems are eliminated in this case because the observer's performance is public. However, the methodical and technical difficulties, such as objectivity and immersion with the group persist, but if the traditional research method is neglected, then these problems disappear too. There is also non-participant observation. The non-participant observation implies observing the participants without actively participating. So, the observer enters a community or social system involved, but stay separate from the activities being observed.

According to the number of the observed cases, the distinction between single-case and multiple-case observation can be made. Single-case observation is the observation of a single case or a single type of a phenomenon. The systematic collection of data on a single case has been termed 'Case Study'. The aim of this procedure is to explain an individual reaction within a collective situation and therefore it is also called the development of personal documentation. It has developed into a separate research method. Closely related to single-case observation is clinical observation, which has developed in medicine and psychiatry. It is aimed at establishing a direct contact between the observer and the patient on the one hand, and obtaining thorough information about what affects an illness, on the other hand, i.e. which complex individual and social factors cause a certain mental condition of the patient. On the other hand, in the multiple-case observation, several cases or groups of the same or different phenomena are the object of observation. It implies registering facts using numerous examples owing to which the structure of phenomena, and even their development are explained. This data collection method is similar to creating social registers on the one hand, and to the mass registration of population – a census, on the other hand

Based on its duration, observation can be snapshot and longitudinal. Snapshot observation is used at a specific moment, which is usually crucial for the development of a process and recorded in a similar manner as a photo is taken. The aim of the snapshot is to spot the main factor that determines further development of a series of actions caused by it. All dynamic activities are observed in this way, because it allows for discerning the important, and eliminating the unimportant characteristics of a phenomenon. Longitudinal observation is used in different continuous sequences of a process with an aim to spot the changes over a longer period of time (Figure 3). Social processes last longer, and take place over a longer period of time and therefore such an observation procedure is important for discovering the dynamics of all events, and the axis they evolve about.

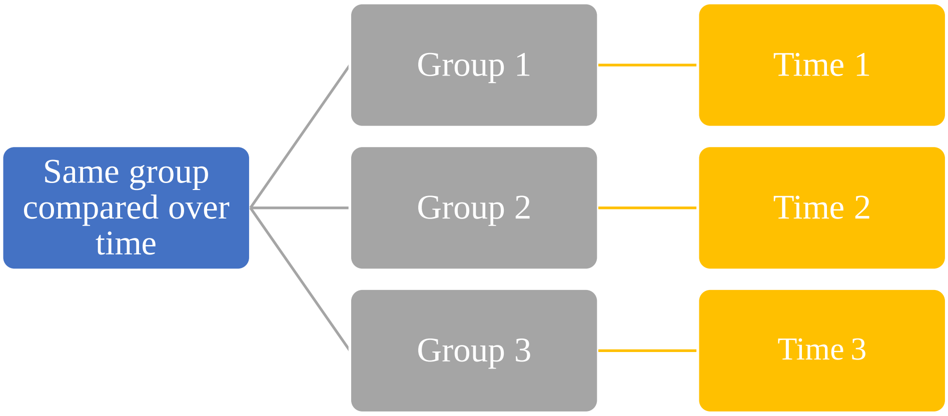


Figure 3. *Longitudinal Studies*

- Structured Observation

Structured or methodical observation is the observation performed according to a pre-defined plan, and the plan is important because it ensures the insight into those facts that are within the observer's scope of interest. The observation remains 'elastic', but deeper insight into the key data or knowledge thoroughly describing a phenomenon is obtained.

All the above-mentioned observation types and subtypes have different applications depending on the nature of the specific research. It is this nature that almost entirely directs the observer as to which procedure to select. This, of course, depends on the researcher's intuition and imagination, as well. Therefore, it is very important to be familiar with the specific types of observation, as well as with its scopes and limitations.

Scopes and Limitations of Observation

Due to the objectivity of the five senses of the researcher, observation is an extremely objective research method. It seems to require little or no interaction between the researcher and those being studied (Angrosino, 2007, p. 37). This procedure does not disturb the natural, continuous flow of events at all or it does so to the smallest extent. The natural dynamics of the procedural unfolding of an event, and all the important elements of the event, are successfully identified by means of observation. Observation can also be directed to some crucial moments, which can be isolated, and thoroughly studied. But more than anything else, observation, and especially its participant forms, directly establishes communication between people, and opens up the closed groups of people, bringing some changes to them that did not exist before (Pečujlić, 1982, p. 105).

In spite of being a very fruitful data collection procedure, observation has some significant limitations, which relate both to the research method, and the observer as a central figure in the procedure. Observation is a long-lasting and complex process. Even when thoroughly planned, it depends on the so-called dispersion of senses. Namely, due to the long-lasting process of observation and its complex organization within a place, senses cannot detect all the facts. So, the unconscious process of suppressing some facts with others, or suppressing old facts with new ones, inevitably takes place.

The success of keeping structured and organised field notes depends on a thoroughly prepared plan for note taking, but even then, if there is not enough time, it is difficult to record all the data. So, this is usually done after the observation, but then it loses the elements of real observation and becomes recollection, which implies that some data processing has been done in the mind that interprets the facts, thus revealing the deceptiveness of human senses. A note taking plan can help to solve the problem but cannot completely eliminate it.

The observer as the central figure in the observation process is limited by the deceptiveness of human senses, but also by his or her own skills, because skills differ with different people. When it comes to observation, they differ so much that the same two people will not perceive the same fact in the same way. Most people are in fact persuaded to see what they want to see. A series of the observer's mental characteristics such as speed, precision, mental ability, recollection, mood, together with a series of social characteristics such as sociability and talkativeness, form the so-called 'personal equation' of the observer, which affects the observation, and can distort the originality of facts. Bias is even more dangerous than this latent danger accompanying observation. Very often the observer distorts the accuracy of facts not only unconsciously, but quite consciously as well. The traditional methodology considers this not only a serious limitation, but also the breach of the scientific rigour of the procedure. In the dialectic action research model, this is not a problem, and the bias itself, which breaks through the established ideological veil, is the precondition for reaching objectivity and truthfulness. Bias serves as a social engagement that provides a path to good observation (Pečujlić & Milić, 1995, p. 107).

4.2. Interview

Interviewing is a planned process of eliciting verbal responses from the person we are talking to. In the research process, it is performed in order to examine individuals and groups of people, and obtain new knowledge. Since it elicits a phenomenon, it is considered a type of an experiment. Interviewing is mostly used in professional orientation. When we are certain or at least suppose that the easiest way of obtaining certain knowledge is through personal contact, we should use interview rather than survey or some other technique. When used by extremely experienced researchers, it can be very convenient, and yields successful results.

There are numerous reasons and needs for using an interview, and they may depend on specific categories of interviewees. We usually opt for an interview when the objective of the interview is clearly stated in the project. We need to get familiar with the setting, groups or interviewees before we begin an interview. The participants should voluntarily answer the questions. The main answers or impressions should be recorded in the form of minutes, and finally, the results should be sorted out, and answers and impressions arranged so as to have a scientific value (Krulj, 2007, p. 64).

Types of Interviews

An interview as a research procedure can be classified according to several different criteria. Based on the content and the way they are carried out, interviews can be divided into:

- ***structured*** *– the questions are pre-defined, asked in a pre-prepared order. It is similar to a questionnaire. It is usually analytical, and easy to implement. Such tightly controlled interviews ensure that the interviewee focuses on the target topic area, and that the interview covers a well-defined domain, which makes the answers comparable across different respondents. On the other hand, there is little room for variation or spontaneity in the responses because the interviewer is to record the responses according to a coding scheme. There is also very little flexibility in the way questions are asked because by adopting a standardized format, it is hoped that nothing will be left to chance. This interview type is appropriate when the researcher is aware of what he/she does not know, and can frame questions that will yield the needed answers (Dörnyei, 2007, p. 135).*
- ***unstructured*** *– it is more like a discussion, i.e. the answers are spontaneous, but it is more difficult to implement. It allows maximum flexibility to follow the interviewee in unpredictable directions, with only minimal interference from the research agenda. The intention is to create the atmosphere in which the interviewee may reveal more than he/she would in formal contexts, with the interviewer assuming the listening role. No detailed interview guide is prepared in advance, although the researcher usually thinks of a few (1–6) opening questions to elicit the interviewee's story. During the interview, the researcher may ask an occasional question for clarification, and may give some reinforcement feedback to keep the interview moving, but interruptions are kept to a minimum. So, a positive rapport needs to be established with the interviewee. This kind of interview is most acceptable when a study focuses on the deeper meaning of a particular phenomenon, or when some personal historical account of how a particular phenomenon has developed is required (Dörnyei, 2007, p. 136).*
- ***semi-structured*** *– there is a set of pre-prepared open-ended questions and topics that need to be covered during the conversation, but the interviewer may follow the lines of enquiry in the conversation that stray from the guide if they find it appropriate, and elaborate on the issues raised in an exploratory manner. It is suitable when the researcher has a good enough overview of the phenomenon or domain in question, and is able to develop broad questions about the topic in advance, but does not want to use the ready-made response categories that would limit the depth and breadth of the respondent's story. Usually the interviewer will ask the same questions to all the participants, but not necessarily in the same order or wording, and would supplement the main questions with the various probes (Dörnyei, 2007, p. 136).*

Based on the participants, i.e. interviewees, we can distinguish between:

- *direct interview (the conversation with the interviewees) and*
- *indirect interview (the conversation with the interviewee's family, friends, etc.).*

Depending on the number of interviewees, we can distinguish between:

- *individual interview (the conversation with one person) and*
- *group interview (the conversation with several people) (Mužić, 1977, p. 250).*

Preparation for Interview

As for the preparation for an interview, it is very important to define the place where the interview will be held. This is so because the setting itself may significantly affect the atmosphere during the interview. It is important to select such a place that will help to build the trust between the interviewer and interviewees.

The next step involves defining the time, and making sure that the people who are to be interviewed appear. It would be good to get informed about the interviewees in advance, though it may lead to some prejudice, which can affect the results.

The interviewees should not be prepared for the interview in advance, though there are different opinions about this. Sometimes the interviewees participate in such activities for the first time, and due to enormous pressure, may provide less accurate information. In such cases, some prior preparation of participants is justified.

As for the content of the interview, the interviewer has to determine clearly and precisely which information is needed and which is not, and explain the concepts thoroughly according to it. A good interview guide requires careful planning followed by some piloting in order to ensure that the questions elicit sufficiently rich data (Dörnyei, 2007, p. 137). The guide serves as the main research instrument – it ensures that the domain is properly covered, and that nothing important is left out by accident; it suggests appropriate question wording, and offers a list of probe questions to be used if needed, as well as a template for the opening statement, and it lists some comments to bear in mind (Dörnyei, 2007, p. 137).

According to Dörnyei (2007, pp. 137–138), a variety of questions can be included in an interview, but they only provide a framework whereas the real meaning is usually uncovered through explanatory and unstructured responses that deviate from the interview guide. The first few questions are particularly important, not so much from the content point of view, but rather because they set the tone, and create the initial rapport. If they make the interviewees feel competent, they will relax and be encouraged to open up. These questions are often quite personal. As for the content of the questions, there are 6 main types that may be asked on any possible topic, and they focus on: (a) experience and bahaviour, (b) opinions and values, (c) feelings, (d) knowledge, (e) sensory information, and (f) background or demographic information. The emergent nature of qualitative interview data can be enhanced by applying various probes, using what the interviewee has said as a starting point to go further and increase the richness and depth of the responses. They may include detail-oriented and clarification questions. The final closing question permits the interviewee to have the final say. As simple a question as: *Is there anything else you would like to add?* has proven to have the power to elicit extremely rich data (Dörnyei, 2007, p. 138).

Procedure

Creating a relaxed atmosphere and establishing rapport: it has already been mentioned that a favourable atmosphere is one of the preconditions for the successful implementation of interviews. In addition to the atmosphere, attention should be paid to the interviewees' attitudes towards the interview. Moreover, when it comes to some sensitive topics, anonimity should be promised, and the promise should be kept. It is necessary to explain that there are no good and bad answers, only sincere and insincere ones, and that the interview is not an exam. If the interviewee is in a hurry, we should show them that we respect their time. If the interviewer and interviewees have not met before, the former is obliged to introduce himself/herself, to be cordial, and start the conversation in a professional manner.

All the above-mentioned factors are extremely important for creating a relaxed atmosphere and establishing the first contact. However, the success of an interview depends very much on the interviewer's personality, as well as on their gender, age, social status, but also on how familiar they are with the subject matter, how interested in the conversation with the interviewees they are, as well as on their approach, tone of voice, clothes, etc. (Mužić, 1977, p. 255).

Carrying out the interview: the way in which an interview is carried out depends on the type of the interview and the topic of the conversation. It is important to keep the pace, not to stray from the topic, not to enter into a discussion with the interviewee, except when it is necessary to provoke them to express their opnion more accurately. The interviewers should not express their opinion because it may affect the interviewees' answers, thus affecting the results.

As for the topics, they can be more or less complex. With less complex topics, there are rarely problems in conversation. The only problem may be extremely short answers or answers not relating to the question. With complex topics, the interviewer has to be skillful and experienced, showing in certain situations that they are well familiar with the topic, that they have knowledge on the subject matter, and understand it (Mužić, 1977, p. 256).

A good qualitative interview flows naturally, with the various parts connecting seamlessly, and is rich in detail. The interviewer is there primarily to listen, and let the interviewee dictate the pace without being rushed or interrupted, trying to be as neutral as possible. The questions should be asked in a straightforward, clear and non-threatening way, without any cues that would lead the interviewee in a particular direction, and the interviewer should give the message that he/she enjoys the interview. The end of the interview can be signalized by summarizing or recapping the main points, which will also allow the interviewee to correct anything that may have been misunderstood, and to make additional points.

Interview Protocol

The protocol or minutes represents the result of interviewing. It is a document which contains the information obtained during an interview. The general information on the interviewer, interviewees, time and place of the interview are provided in the header of the protocol, whereas the title of the project within the framework of which interviewing has been performed is provided in the right-hand corner.

The content of an interview depends on the type of the interview. If the interview is structured, the interviewees answer the questions and the interviewer notes the answers, which means that the interviewer has the minutes in front of themselves, whereas in case of an unstructured interview, the interviewer usually does not have a protocol in front of themselves, and thus establishes a good rapport with the interviewees. The responses are recorded afterwards, preferably immediately after the interview in order to prevent forgetting. The interview can also be recorded, which is extremely useful because then the responses cannot be forgotten. Even though by doing audio recording some information such as non-verbal cues are inevitably lost, it is more often used than video recording because it is less obtrusive and much less difficult to perform, whereas there is no difference in the complexity of data amalyses (Dörnyei, 2007, p. 139).

The protocol contains the objectives of the interview, plan with the predefined duration of the interview, instructions about the interview tactics, etc. There should be some space for notes on the protocol itself, where the interviewer records their impressions about the interviewees, whether they showed signs of fatigue, had a repulsive attitude, etc. (Mužić, 1977, pp. 252–259).

Online Interview

In spite of being the most frequently used data collection method of qualitative research, traditional face-to-face interviews are accompanied by certain problems such as geographical distance and restricted physical mobility of participants, as well as financial issues, etc. Over the last few decades, technological changes brought forth by the development of the Internet have led to the development of the online interview in qualitative research, thus overcoming some of the limitations relating to the traditional interview method.

An online interview (also called a digital interview or virtual interview) is an interview conducted remotely using technology and appropriate software platforms. An e-interview research framework is presented in Figure 4:

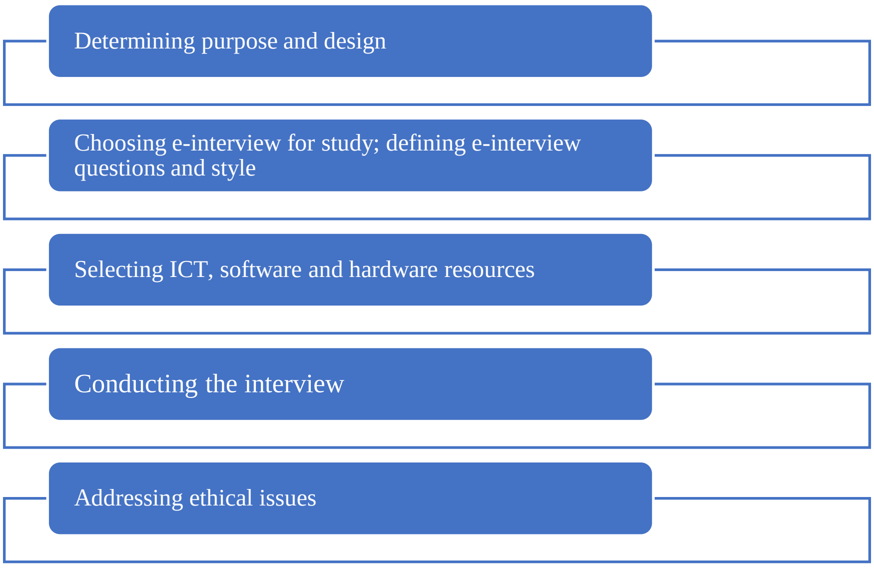


Figure 4. *E-Interview Research Framework*

Both focus group and one-on-one interviews can be conducted in the synchronous mode (real-time) and asynchronous mode (non-real-time). E-mail, Google questionnaires, online bulletin boards and discussion groups are the most commonly used technologies in asynchronous online interviews. To access the respective software platform, the applicants usually receive a URL link to the meeting, or meeting ID, or a username and password. Applicants must be familiar with the software chosen to conduct an interview. Applicants complete the questionnaire when they have time, and according to their ability. In this case, the focus is on the applicant's key skills in their natural environment, when they are free of pressure and restrictions.

The synchronous mode is appropriate for applications such as Skype, Zoom, Google Hangouts, Microsoft Teams, video conferencing, and so on. These technologies allow for various options and combinations of one-to-one, one-to-many, and many-to-many communication. The advantages of using these applications are that they provide a free communication service, improve teamwork and interaction among participants, allow for flexible content sharing – without temporal or geographic limitations – and allow for audio or video interviews using a microphone and web camera.

Chat-based recruitment is on the rise because new generations of young people are entering the labor market (Millennials and Generation Z employees). So, employers use tools such as WeChat, WhatsApp and Facebook Messenger to perform pre-screening and early-stage interviews in the hiring process. Vervoe, VidCruiter, SparkHire, interviewstream, Outmatch, AllyO, Mya, and many other platforms offer a set of tools designed to support employers in the hiring process. Despite the abovementioned advantages, it should be noted that the interviewees must be digitally literate, have an access to the high speed Internet, and have some experience in on-line communication in order to take part in such interviews (Janghorban et al., 2014). Moreover, inconsistent Wi-Fi or Internet connections, poor camera/image quality, or sound issues are some of the potential problems that can occur when conducting remote online interviews.

Table 3. *Interview Advantages and Disadvantages*

Interview advantages	Interview disadvantages
- *personal contact* - *adaptable to individuals* - *oriented towards an individual* - *possibility of communicating with interviewees*	- *neglecting the basic research factors* - *simple data collection* - *inefficiency/time-consuming* - *requires good communication skills* - *difficult to record* - *changeable interview propositions*

Source: Krulj (2007, p. 66).

4.3.Focus Groups

Group interviews are administered to groups of people (usually 6–12 members) selected particularly for the purpose of the ongoing research. The same interviews are administered to different people and therefore different attitudes and responses to a specific topic are obtained. During the group brainstorming, the participants think together, inspire and challenge each other, and react to the emerging issues and points (Dörnyei, 2007, p. 144). Later on, similarities and differences between the obtained responses are analysed.

The process of the focus group research includes:

- ***The selection of the topic*** *– the topic has to be important to participants, and they are supposed to have some knowledge about it. However, some topics may seem safe to the interviewer, but are very sensitive to participants. According to Farquhar and Das (1999), all research topics have the potential to be sensitive because the sensitivity of a topic is not fixed but socially constructed;*

9. ***Tthe selection of participants*** *– the participants' characteristics should be in line with the interview objectives and the topic discussed. Focus group participants are not selected by means of systematic random sampling. Given that the success of a group depends on the dynamics between group members, the group composition should be carefully considered by the researcher. The interaction between participants is a key feature of the focus group method, and there has to be sufficient diversity to encourage discussion. However, too heterogeneous groups may result in a conflict. So, researchers have to be familiar with the differences, aware of the potential problems, and have the strategies to deal with them, and even then, the researcher will not be able to anticipate or control the direction of the group discussion (Bloor et al. 2001, p. 20);*

- ***Organisation*** *– all the relevant participants for a specific topic should be organised at the same time, and at the same venue;*
- ***Implementation*** *– the group moderator should be experienced and competent for such a type of interview. The objective of the focus group research is not to elicit the group's answers to pre-determined questions, but to stimulate discussion, and through subsequent analysis, understand the meanings and norms that underlie the group answers. Hence, instead of being asked to answer a question, the group may be required to perform a specific task, such as a ranking exercise, or description of a photograph (Bloor et al., 2001, p. 43). For the successful implementation of the focus group research, the role of the researcher is very important. Control is necessary, but the researcher should facilitate the group discussion, not control it, as the group interaction may be distorted by too much external control. At the same time, the facilitator must avoid the domination of the group by individual members, and also seek to encourage contributions from the more timorous ones (Bloor et al., 2001, p. 49). Silences may be quite embarrassing for facilitators, but there are certain things they can do to break them, such as commenting on non-verbal clues (Bloor et al., 2001, p. 52).*
- ***Recording*** *– most often, audio-video recordings are made;*
- ***Analysis*** *– a quality analysis of the collected material, during which the moderator's presence is mandatory. According to Bloor et al. (2001, p. 59), the transcription of the recording of the focus group discussion is mandatory for academic research, as simple listening to the recording or the memory of the moderator may lead to the loss of much of the richness of the data and will risk a selective and superficial analysis. On the other hand, the data that focus group discussions produce are chaotic because people*

often talk at once, sentences remain unfinished, people misinterpret other's comments, their arguments develop as they discuss the topic, etc. Therefore, all the recorded speech should be transcribed, which means all the speakers, if more than one person is talking, not just the dominant voice, the unfinished or interrupted speech, very brief extracts of speech, even laughter, and body language, etc. Moreover, the speaker should be identified (Bloor et al., 2001, p. 72). In order to make this possible, Bloor et al. (2002) suggest that, as the first group exercise, people be asked to say their name, and tell a few sentences about themselves, which can serve as a reference point or a basis for the identification during the transcription. Moreover, the quantity of data is overwhelming, whereas the analysis must be systematic and rigorous, reflecting the views of all cases, not only those that fit the researcher's own agenda (Bloor et al., 2001, p. 62);
- ***Reporting** – the research findings are properly transformed into reports.*

The analysis of the data collected by means of interviews comprises the following elements:

- *generating codes;*
- *designing analytic concepts;*
- *identifying patterns within the concepts;*
- *producing an explanation;*
- *moving forwards-backwards – constantly moving from the beginning to the end of the text and vice versa.*

By coding, data are selected, connected, and theories are selected. From the process of coding, classes of phenomena and a clearly defined concept are obtained. Open coding is the initial procedure, which brings the elementary order into a huge amount of information. Focused coding is the process of removing less productive and less important codes, and focusing on a smaller number of key selected codes. A code is good if it contains a clear label, defined topic, description which explains when exactly something related to the topic appeared, as well as positive and negative examples in order to avoid confusion.

When documenting behaviour, focus groups are less suitable than individual interviews because there is a tendency for atypical behaviour to be unreported or under-reported in group settings, especially when groups are to reach a consensus on a given topic (Bloor et al., 2001, p. 8). The advantages of group interviews involve social orientation, flexibility, validity, clarity, efficiency, practicality, whereas the disadvantages include a lack of control, difficult data analyses, heterogeneity of groups, complex organization.

4.3. Biographical Research Method

The method of personal documents or biographical method is a concept related to a group of loosely connected, differently termed research tools, ranging from the narrative interview, life histories, life stories, oral history, (auto)biography, biographical interpretive method, story telling to ethnography (Poleti-Ćosić, 2019, p. 29).

The biographical method is one of the qualitative research methods used in sociological research. Its use is the subject of a lot of disputes. Even though it has gained the status of an independent method, a lot of theorists do not consider it a method sufficient on its own.

It is a new data collection method, grounded on two assumptions. According to the first one, the man as an individual creates social phenomena and therefore we have to get to know the man's psychology because his mental condition cannot be separated from social events. According to the second assumption, personal (biographical) documents contain extremely important information, collected on the basis of sincere experiences of a phenomenon and therefore can provide a very good explanation of both an individual's life and social trends (Pečujlić & Milić, 1995, p. 137). The development of the biographical model is ascribed to Chicago School and the study of V. Thomas and F. Znaniecki 'The Polish Countryman in Europe and America', based on the use of personal documents (Poleti-Ćosić, 2019, p. 30).

Personal documents, which the method was named after, can be divided into two groups. The first group implies the documents that provide information about a person, and can be obtained in official institutions – formal documents. These are various archives following the happenings in the life of an individual – the court, tax administration, police archives, etc. The second group includes documents designed by individuals themselves, their records, in which they describe their roles in social events. There are different types of such documents:

- *letters or correspondence with other people who had an important role in the life of a person, and took part in recording a social event,*
- *diaries, which deal with phenomena and events in a systematic and detailed manner,*
- *autobiographies, which describe the whole life of an individual, and are very important because of the accuracy of events, and the order in which they took place,*
- *biographies, which are an individual's recordings about other people,*
- *memoires, similar to autobiographies, which describe specific events important for an individual,*
- *brief notes and messages related to a single event, which we are trying to preserve from oblivion (Pečujlić & Milić, 1995, p. 138).*

The method of personal documents represents the planned classification of materials, and identification of categories related to the personal and social life of an individual. The classification of materials is done in the following manner:

- *the main register, which contains all the collected documentation,*
- *the analytical register, which represents the revised main register,*
- *the diary register, which represents the scientist's own recording made while examining the respondents' personal documents.*

In addition to these registers, auxiliary registers such as financial reports and bibliographies of papers related to the research topic can be used. The data arranged in this way dictates certain steps of the analysis. The first step of the data analysis implies their chronological ordering. The use of specific sources of information which are already chronologically ordered makes this step easier. Then the researcher identifies the relationships between specifc events and respondent's experiences in order to critically examine the credibility of the story. The next step is dealing with individual documents, and mapping characteristic types of behavior, social relations, etc. This step of the analysis involves the use of the sequential analysis. The categories depend on the problems being researched. However, there is a general category resulting from the desired objectives of the method, and it is to determine the relationships between personal and social situations of the individual who is the object of the research.

Types of Biographical Method

Even though it represents a whole, the method of personal documents can be divided into several types. The most important classification is the one performed according to the types of the sources of material or the classification into the personal history method, related to formal documents, and the life story method, based on the respondents' records.

In addition to this classification, the classification according to the types of events, different in scope and significance, is also extremely important. It is the classification into the method of personal documents about the life of the respondent, and method of personal documents about specific events.

Furthermore, there is the classification into the method of personal documents intended to be used for scientific purposes, and the method of personal documents independent of them.

The difference between the method of formal personal documents and other similar methods is that they were written by other people, and the material was processed in the mind of the person who recorded them, which makes its reliability questionable. On the other hand, there are subjective records, made by the respondent. In spite of being subjective, they can be considered as a true research method (Pečujlić & Milić, 1995, p. 139).

As for the method which relates to a majority of events which left some traces on the respondent's life, it can be said to have a more universal meaning because, in addition to the important events in the life of the respondent, this method also reveals the way in which the respondent developed as a person, every change that took place during that period of time, changes in attitudes, etc.

The importance of the method which relates to fewer events is not in the amount of events, but in their effects on the respondents themselves on the one hand, and in some specific event itself on the other hand.

The biographical method which was developed for scientific purposes is intended to ensure achieving the goal, i.e. examine the relationship between an individual and social phenomena. Therefore, it could be said that this method yields the best results. However, even though an individual is the typical object of biographical research, the validity of this methodological approach has been confirmed in the studies based on the collective units of analysis as well (Poleti-Ćosić, 2019, p. 34).

The role of the researcher using the biographical method is opposite to the role of the researcher performing traditional qualitative research: at the beginning of the research, the researcher collects experiential data, reads or listens to the life stories of people, while doing their best to remain invisible and unnoticeable so as not to affect the data authenticity and quality anyhow, and then carefully reads the collected materials, and reconstructs the real life accordingly.

In addition to bringing some changes into the role of the researcher in a reseach process, the biographical method also brings some changes into the observation of social reality. Traditional research seeks to derive a cause-and-effect chain going from the so-called independent variables to the dependent ones, whereas the biographical method seeks to reconstruct the chronological chain of events and study the logic of their occurrence (Pečujlić & Milić, 1995, p. 139).

Example: When hiring a new worker with a company, the biographical method can be used alongside the questionnaire administration or interviewing. During the narrative interview, the initial introductory question is asked to elecit an account of events and experiences from the interviewee's own life (the so-called 'main narrative'), which should not be interrupted by asking additional qustions, but stimulated by non-verbal gestures and expressions of interest. This allows the interviewee to constract the narrative as they wish. During the second part of the interview, i.e. during the 'interrogation period', the researcher initiates the broadening of the topic by asking narrative questions or asking for further details about the aspects of life described in the narrative. Talking to a person about certain elements from their biography, in a relaxed atmosphere, is a good way of getting to know the person, and their positive and negative traits better, thus contributing to the prediction of the prospective future business cooperation.

Advantages and Disadvantages of Biographical Method

The main advantage of the biographical method is primarily the fact that it takes into account the subjective dimension of the event or phenomenon being examined. This subjective dimension is presented directly by the subject, thus ensuring that the data is honest. However, many authors believe that an excessive emphasis on the subjective dimension of a phenomenon or event poses a risk to the objectivity and credibility of the obtained research results. This problem of the biographical method is solved by applying the principle of complementarity. The principle implies the use of different data sources which complement each other. In the case of the biographical method, these are institutional data, which are considered the objective data that the personal documents miss.

Another advantage of the biographical method is the temporal dimension that it brings to the research process. The information sources were mostly created immediately after the event being examined took place.

Unlike other methods, the method of biographical documents is the only one directed to the role of an individual in social processes. Furthermore, it explains how events influence the development of the respondent's personality, and how the respondent reacts both to their own problems and the problems of the social environment (Pečujlić & Milić, 1995, p. 140).

Finally, one of the advantages of this model relates to the very sources of experiential materials, which are much easier to control than other sources, which require the researcher's presence in the data collection process.

However, the biographical data can be said to have more disadvantages than advantages primarily due to their subjectivity. The first problem or disadvantage is the inaccessibility of data. A small amount of personal documents is easily accessible to researchers, forcing them to announce the calls for submitting biographies, which leads to another problem – the problem of authenticity. The same problem appears with the sources designed for the purpose of the research, i.e. ordered sources, collected orally (narrative biographies). The problem with this type of sources is actually the degree in which the researcher influenced the development of the source. The solution to such a problem can be found in as precise a definition of the topic and research objectives as possible, and then in defining the broadest possible experiential framework of the research itself. The process of creating documents is very slow and long-lasting, and in addition to time, requires huge financial resources (Pečujlić & Milić, 1995, p. 140). Another problem that might be encountered when using the personal documents method is the selection of participants/respondents. Finally, there is the problem of the representativeness of the results, which can be encountered when using the personal documents method, and it relates both to sampling, and the analysis and presentation of the research results.

4.4. Qualitative Survey

Qualitative surveys use open-ended questions eliciting long, written answers with an aim to reveal opinions, experiences, narratives or accounts. They are often a useful precursor to interviews or focus groups as they help identify themes or issues that should be further explored in the research. The main aim of a qualitative survey is not to establish frequencies, means or other parametres, but to determine the diversity of some topic of interest within a given population – it is a study of diversity, not distribution, in a population (Jansen, 2010).

Qualitative surveys may be inductive (open) or deductive (pre-structured). In inductive surveys, relevant objects/topics, dimensions and categories are identified through the interpretation of raw data, whereas in deductive ones, the diversity to be studied is defined beforehand, and the aim of a descriptive analysis is only to see which of the pre-defined characteristics exist empirically in the population under study. Though qualitative researchers equate qualitative research with induction, Jansen (2010) underlines that pre-structured surveys focused on the diversity analysis as opposed to the numerical distribution are considered qualitative research, too.

As the aim of a qualitative survey is to study the diversity of a phenomenon within the target population, the diversity sample should be purposefully selected, with an aim to cover all existing varieties of the phenomenon (saturation).

The data collection is generally conducted by questioning people, whereas the researcher wants to find out about the relationships between the characteristics of units, i.e. the patterns of categories, in order to explain the diversity in the subject of study. Three main methods of administering qualitative surveys are as follows:

- *face-to-face surveys – the researcher asks the participant one or more open-ended questions, and observes the participant's reactions, which enables them to ask follow-up questions to elicit a more detailed response. These surveys are audio recorded and transcribed.*
- *telephone surveys – the researcher asks the participant one or more questions, but cannot see their facial or behavioural responses, thus missing potential cues for asking follow-up questions.*
- *online surveys – open-ended questions are presented to participants in writing via email, often alongside quantitative research questions on the same topic. Some contextual information or key definitions may be provided to frame the participants' perception of the questions. Diverse responses are usually obtained (from brief to detailed, vague).*

The advantage of a survey over an interview relates to the fact that a lot of people can be questioned at the same time; it can be sent by e-mail, thus collecting data from a wider region. It is also cost-effective. However, the reliability of responses depends on the repondents' sincerety, whereas questions can be misleading/leading, eliciting answers that are not core to the research subject. Questions must be as neutral as possible in order not to reveal the researcher's expectations.

In order to ensure the success of a survey, it is important to describe its purpose politely at the top of the questionnaire. The fact that the survey is anonymous should be emphasised, as well as the fact that the responses will be used only for the research purposes, and that the aim of the questionnaire is not to test the respondents' knowledge and therefore there cannot be correct or incorrect answers. It is important to state that the respondents should feel free to express their opinions or attitudes.

5. References

Angrosino, M. (2011). *Doing etnographic and observational research*. Sage.

Asmussen, K. J., & Creswell, J. W. (1995). Campus response to student Gunman. *Journal of Higher Education*, *66*(5), 575–596.

Baxter, P, & Jack, S. (2008) Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, *13*, 544–559.

Bloor, M., Frankland, J., Thomas, M., Robson, K. (2001). *Focus groups in social research: Introducing qualitative methods*. Sage.

Bradbury, H. (2015). Introduction: How to situate an define action research. In H. Bradbury (Ed.), *The Sage handbook of action research* (3rd ed.). Sage.

Connolly, M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher, 19,* 2–14.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Sage.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (3rd ed.).* Pearson.

Creswell, J. W. (2009). *Research design qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*, 124–130.

Creswell, J. W., Hanson, W. E., Clark Plano, V. L. (2007). Qualitative research designs. *The Counselling Psychologist*, *35*(2), 236–264.

Crotty, M. (1998). The foundations of social research. (1st Ed.) Sage.

Czarniawska, B. (2004). *Narratives in social science research*. Sage.

Dewey, J. (1941) Propositions, warranted assertibility and truth. *The Journal of Philosophy.*

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Farquhar, C., & Das, R. (1999). Are focus groups suitable for 'sensitive' topics? In R. S. Barbour & J. Kitzinger (Eds.), *Developing focus group research: Politics, theory and practice* (pp. 47–63). Sage.

Fay, B. (1987). *Critical social science: Liberation and Its Limits.* Cornell U. P.

Field, P. A., & Morse, J. M. (1985). *Nursing research: The application of qualitative approaches.* Aspen.

Gamson, J. (2000). Sexualities, queer theory, and qualitative research. *Handbook of Qualitative Research*, *2*, 347–365.

Janghorban, R., Roudsari, R. L., & Taghipour, A. (2014). Skype interviewing: The new generation of online synchronous interview in qualitative research. *International Journal of Qualitative Studies on Health and Well-being*, *9*(1), 24152.

Jansen, H. (2010). The logic of qualitative survey research and its position in the field of social research methods. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research, 11*(2).

Keeves, J. P. (1997). *Educational research methodology and measurement*. Cambridge University Press.

Krulj, R. S. (2007). *Uvod u metodologiju pedagoških istraživanja sa statistikom.* Učiteljski fakultet, Centar za naučno-istraživački rad.

Ladson-Billings, G. (2000). Fighting for our lives. *Journal of Teacher Education, 51*(3), 206–214.

Lazarsfeld, S. (1966). The courage for imperfection. *American Journal of Individual Psychology*, *22*(2).

Leedy, P. D., & Ormrod, J. E. (2005). *Practical research: Planning and design.* Upper Saddle River.

Leininger, M. M. (1985). Ethnography and ethnonursing: Models and modes of qualitative data analysis. In M. M. Leininger (Ed*.)*, *Qualitative research methods in nursing* (pp. 33–72). Grune& Stratton.

Lincoln, Y. & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Lukić, R. (1989). *Metodologija društvenih nauka*. Savremena administracija.

Mackenzie, N. & Knipe, S. (2006). Research dilemmas: paradigms, methods and methodology. *Issues In Educational Research*, 1–15.

Marshall, C., & Rossman, G. B. (1989). *Designing qualitative research*. Sage.

Mertens, D. M. (1998). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches.* Sage.

Mertens, D. M. (2003). Mixed methods and the politics of human research: The transformative-emancipatory perspective. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 135-164). Sage.

Milić, V. (1996). *Sociološki metod.* Zavod za udžbenike i nastavna sredstva.

Moreno, J. (1947). Contribution of sociometry to research methodology in sociology. *American Sociological Review, 12*(6): 287–292.

Morgan, D. L. (2007). Focus groups as qualitative research. (2nd Ed.). Sage.

Moustakas, C. E. (1994). *Phenomenological research methods.* Sage.

Mužić, V. (1977). *Metodologija pedagoškog istraživanja*. Zavod za udžbenike.

Newman, I. & Benz, C. (1998*). Qualitative-quantitative research methodology: Exploring the interactive continuum.* SIU Press

Olesen, V. L. (2000). Feminist qualitative research and grounded theory: Complexities, criticisms, and opportunitites. *The SAGE Handbook of Grounded Theory.* Sage.

Pečujlić, M. (1982). *Metodologija društvenih nauka*. Savremena administracija.

Pečujlić, M., & Milić, V. (1995). *Metodologija društvenih nauka.* Grafika.

Plummer, K. (1983). *Documents of life: An introduction to the problems and literature of a humanistic method.* Allen & Unwin.

Poleti-Ćosić, D. (2019). Biographical method and its use in the field of migration studies. *Sociologija*, *62*(1), 24–41.

Punch, K. (2005) *Introduction to social research: Quantitative and qualitative approaches* (2nd ed.). Sage.

Riemen, D. J. (1986). Non-caring and caring in the clinical setting: Patients' descriptions. *Topics in Clinical Nursing*, *8*(2), 30–36.

Ristić, Ž. (2016). *Objedinjavanje kvantitativnih i kvalitativnih istraživanja*. Evropski centar za mir i razvoj (ECPD), Univerzitet za mir Ujedinjenih nacija.

Rossman, G. B., & Rallis, S. F. (1998). *Learning in the field: An introduction to qualitative research*. Sage.

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students*. Pearson.

Sayer, A. (2000). Realism and social science. (1st ed.). Sage.

Schwandt, T. A. (2003). *Back to the rough ground! Beyond theory to practice in evaluation*. Sage.

Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English Language Teaching*, *5*(9), 9–16.

Slavin, R. E. (1984). *Research methods in education: A practical guide*. Prentice-Hall.

Spradley, J. (1980). *Participant observation*. Holt, Rinehart and Winston.

Stake, R. (1995). *The art of case study research*. Sage.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage.

Stringer, E. (1999). *Action research* (2nd ed.). Sage.

Šešić, M. (1979). *Opšta metodologija*. Naučna knjiga.

Ševkušić, S. (2008). Kvalitativna studija slučaja u pedagoškim istraživanjima: Saznajne mogućnosti i ograničenja. *Zbornik Instituta za pedagoška istraživanja*, *40*(2), Beograd, 2008.

Thomas, A. S. (1993). *Doing critical ethnography*. Sage.

Wolcott, H. F. (1990). Making a study "More ethnographic". *Journal of Contemporary Ethnography*, *19*(1): 44–72.

Yin, R. K. (2003). *Case study research design and methods* (3rd ed.). Sage.

QUALITATIVE DATA ANALYSES

1. Introduction

Qualitative data is non-numeric information, which is provided in in-depth interviews, diaries, anthropological field notes, answers to open-ended survey questions, audio-visual recordings and imaged. Qualitative analysis is the analysis of the collected language-based data. According to Thorne (2000), the analysis of such data is the most complex and mysterious of all the phases of a qualitative project, and the one that receives the least thoughtful discussion in the literature. A qualitative analysis, regardless of the type chosen, reduces the volume of the text collected, identifies and groups categories together, and seeks some understanding behind them (Bengtsson, 2016, p. 8). It is the understanding of various aspects of social life that are often subjective, such as people's opinions, attitude, motivations and reasons for doing something.

Qualitative data analysis takes much longer than quantitative data analysis because it required the researcher to read and re-read the free text before drawing on the insights. It provides the answers to the 'Why' and 'How' questions. It also requires the researcher to bracket himself/herself, and try to view things from another person's perspective. Although the researcher's personal beliefs and bias cannot be entirely eliminated, they can and should be minimized, and this can be done by the triangulation of data. Triangulation involves using multiple data sources, theories, methods, or investigators to enhance the validity and reliability of the research, but also to gain a more nuanced and holistic perspective (Figure 1).

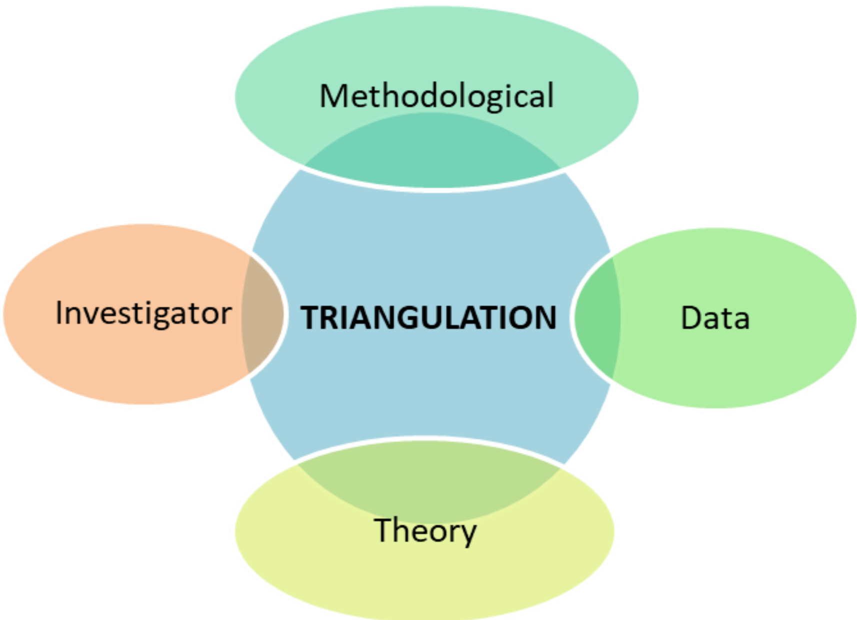


Figure 1. *Triangulation types*

- *The methodological triangulation combines research methods to tackle the same research question in order to avoid biases and flaws associated with using a single research method.*
- *The data triangulation implies multiple data sources, i.e. data collected across different spaces, times, and people.*

- *The theory triangulation implies using more than one theoretical approach in answering research questions, thus ensuring that the topic is understood from different perspectives, and the differences in the data reconciled.*
- *The investigator triangulation involves multiple researchers or observers who collect, process, and analyse data separately.*

Triangulation helps enhance the validity of the research, gives a more vivid picture or a complete understanding of the research problem, and ensures credibility, i.e. the fact that the data reflect reality. However, it has some disadvantages as well, as it is a time-consuming process, which may be costly because it involves working with an interdisciplinary team, and it might happen to obtain contradictory data from different sources.

There are different approaches to the analysis of qualitative data aimed at identifying patterns and trends, and at formulating insights. Their choice depends on numerous factors, such as: the type of data collected, the context in which data was collected, the researcher's background knowledge on the topic under research, whether the goal is to formulate a theory from the data, or apply the existing theories to the data. There are six main types of qualitative analysis approaches. They include:

- *Content Analysis,*
- *Narrative Analysis,*
- *Thematic Analysis,*
- *Framework Analysis,*
- *Discourse Analysis,*
- *Grounded Theory Analysis.*

Content Analysis

Krippendorff (2004, p. 18) defined content analysis as *'a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use'*. The goal is to link the results to their context or to the environment in which they were produced (Bengtsson, 2016, p. 9).

The purpose of content analysis is to organize and elicit meaning from the data collected, and draw realistic conclusions from it. In a qualitative content analysis, data are presented in words and themes, which makes it possible to draw some interpretation of the results. The researcher must choose whether the analysis should be of a broad surface structure (*a manifest analysis*) or of a deep structure (*a latent analysis).* In a manifest analysis, the researcher describes *what* the informants actually say, stays very close to the text, uses the words themselves, and describes the visible and obvious in the text. In contrast, a latent analysis is extended to an interpretive level, in which the researcher seeks to find the underlying meaning of the text: *what the text is talking about* (Berg, 2001; Catanzaro, 1988).

Content analysis comprises four main stages: *decontextualisation, recontextualisation, categorisation,* and *compilation* (Figure 7). However, each stage has to be performed several times in order to maintain the quality and trustworthiness of the analysis. It is the researcher's responsibility to maintain the quality of the process by assuring validity and reliability throughout the entire study, as the results must be as rigorous and trustworthy as possible. In a qualitative study, validity means that the results truthfully reflect the phenomena studied, and reliability requires that the same results be obtained if the study were replicated (Morse & Richards, 2002).

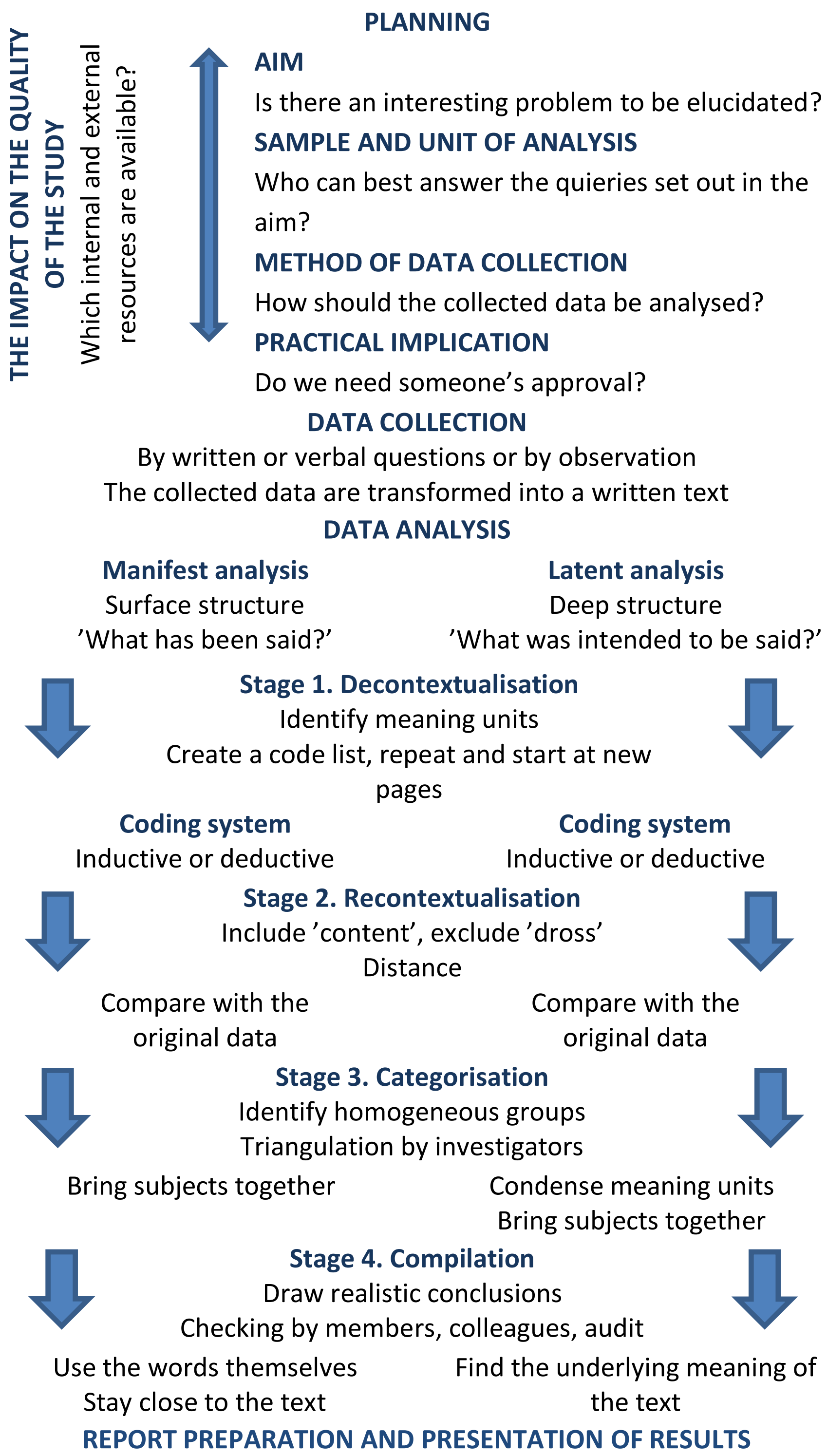


Figure 7. *Overview of Content Analysis from Planning to Presentation (adapted from Creswell, 2009)*

1.1. Decontextualisation

The researchers must familiarize themselves with the data, and have to read through the transcribed text to obtain the sense of the whole, i.e. to find out what it is about, before it can be broken down into smaller *meaning units*. A meaning unit is the smallest unit that contains some of the insights the researcher needs, and it is the constellation of sentences or paragraphs containing aspects related to each other, and answering the question set out in the aim (Catanzaro, 1988). Each identified meaning unit is labeled with a code, which should be understood in relation to the context. This procedure is known as the '*open coding process'* in the literature (Berg, 2001). During the analysis, codes facilitate the identification of concepts around which the data can be assembled into blocks and patterns (Catanzaro, 1988). The researcher should use a coding list, including the explanations of the codes, to minimize the cognitive change during the process of analysis in order to secure reliability (Catanzaro, 1988). Codes can be generated inductively or deductively, depending on the study design. If the study has a deductive reasoning design, the researcher has to create the coding list before starting the analysis. Otherwise, the list can be created in the course of the process (Catanzaro, 1988). Codes created inductively may change as the study progresses, as more data become available. The interpretations of the meaning units that seemed clear at the beginning may become obscured during the process. Therefore, the coding process should be performed repeatedly, starting on different pages of the text each time to increase the stability and reliability (Downe-Wambolt, 1992). However, it is much easier to obtain high reliability with code lists generated deductively rather than inductively (Catanzaro, 1988). There are also computer programmes which can be of help. Though their use is not imperative, they may facilitate the process. Although these programmes do not analyse the data, they do speed up the process by locating codes, and grouping data together in categories. Nevertheless, it is up to the researcher to decide what constitutes the themes, and what conclusions can be drawn from the results.

Recontextualisation

After the meaning units have been identified, the researcher has to check whether all the aspects of the content have been covered with regard to the aim (Burnard, 1991). The original text is re-read alongside the final list of meaning units. Coloured pencils are useful to distinguish each meaning unit in the original transcript. After this process has been performed, some unmarked text nearly always remains. The researcher must then consider whether or not the unmarked text should be included. If the unmarked text gives some answers to the research question, it should be included in the analysis (Burnard, 1995). When the researcher is deeply involved with the data, everything seems to be of importance. Nevertheless, the process of distancing is necessary, and the researcher must allow themselves to let go of the unimportant information that does not correspond to the aim of the study.

Categorisation

Before the researcher can begin to create categories, extended meaning units have to be condensed, which means that the number of words is reduced without losing the content of the unit (Graneheim & Lundman, 2004). The depth of the meaning units determines the level at which the analysis can be performed. This process of condensation is often needed when data are based on interviews, and when the latent content analysis is to be carried out. To extract the sense of the data, the coded material can be divided into *domains* – broad groups based on different focuses of the study. Graneheim and Lundman (2004) prefer the concept *content area*, since, in their view, this elucidates a specific, explicit area. For example, the material can be divided on the basis of the questions used when the data were collected or on theoretical assumptions from the literature (Bengtsson, 2016, p. 12).

In the categorization process, themes and categories are identified. However, in the literature there is no consensus for which headings or concepts are to be used in content analysis. Sub-categories, which Burnard (1991) terms *sub-headings*, are the smallest units based on meaning units. In the manifest analysis, sometimes these are the same as the codes of the meaning units. Sub-categories can be sorted into broader categories. The concept *sub-theme* can be used in the latent analysis instead of the concept *categories*. Identified themes and categories should be internally homogeneous and externally heterogeneous, which means that no data should fall between two groups nor fit into more than one group (Krippendorff, 2004). A *theme* is an overall concept of the underlying meaning on an interpretative latent level, and it answers the question 'How?'

All categories must be rooted in the data from which they arise. Moving meaning units back and forth between categories ensures the progressive development of the category outcome. Initially, several categories are often generated, but the number is later reduced (Burnard, 1991). How the researcher knows when the categorization is good enough depends on the aim of the study, and the categorization is finished when a reasonable explanation has been reached (Bengtsson, 2016, p. 12).

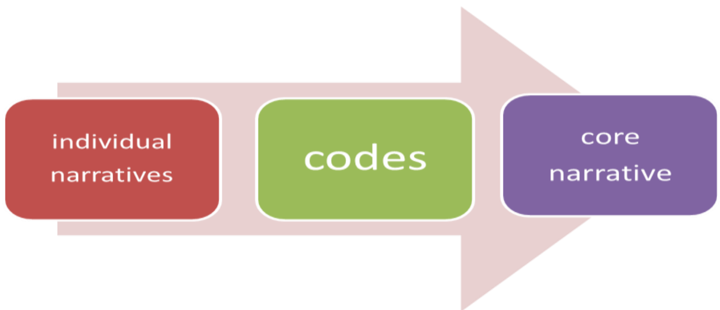
Compilation

Once the categories are established, the analysis and the writing process begin. One difference between the various qualitative analysing methods is how the researcher relates to the analysing process itself, and adapts to the results. When performing qualitative content analysis, the investigator must consider the data collected from a neutral perspective, and consider their objectivity. However, the researcher has a choice between the manifest and latent level, and the depth of the analysis will depend on how the data are collected. In a manifest analysis, the researcher works this way gradually through each identified category, and in a latent analysis through the themes. In a manifest analysis, the researcher often uses the informants' words, and they remain aware of the need to refer back to the original text. In this way, it is possible to stay closer to the original meanings and contexts (Burnard, 1991). In contrast, a latent analysis invites the researcher to immerse themselves to some extent in the data in order to identify hidden meanings in the text. For each category or theme, the researcher chooses appropriate meaning units presented in the running text as quotations. Regardless of the form of the analysis, the researcher can present a summary of themes, categories/sub-themes and sub-categories/sub-headings as a table to allow the reader to get a quick overview of the results. In addition, it is appropriate to present one example of the analysis process. There is also the possibility to add information by performing some quantification in which sub-categories and categories are counted. This is not normally done in other qualitative research methods. However, nearly everything can be counted in written messages – such as words, characters, paragraphs and concepts – depending on the focus of the study. By combining the quantification with a qualitative approach, the magnitude of the individual phenomena studied appears more clearly (Berg, 2001). However, the variables cannot be ranked, since not all informants have had the opportunity to discuss all the phenomena that the researcher finally counts.

Finally, the researcher must consider how the new findings correspond to the literature and whether or not the result is reasonable and logical (Burnard, 1991; Morse and Richards, 2002). To validate the outcome and to strengthen the validity of the study, the researcher can perform a respondent validation, a member check, which means that the researcher goes back to the informants and presents the results in order to achieve agreement (Burnard, 1991; Catanzaro, 1988). However, there is a time-delay between the data collection and analysis. This approach, therefore, constitutes a risk for various reasons, one of which might be the possible unreliability of the informants' memory. Another risk is that informants have a tendency to deny less attractive aspects of their behavior. In addition, as the researcher often creates a deeper holistic understanding of the studied phenomenon, the informants may not recognise how the data is presented. Keeping this in mind, it is better for the researcher to obtain some confirmation on the content from the informants in connection with the data collection (Catanzaro, 1988). Another way to increase validity is for a colleague not involved in the study, or an inquiry auditor, to read the original text and results and then judge whether they are reasonable or not (Burnard, 1991; Catanzaro, 1988). However, it is obviously difficult for an independent person to familiarize themselves with another person's coding (Bengtsson, 2016, p. 13).

Narrative Analysis

Researchers use narrative analysis to gain insight into how research participants construct stories and narratives based on their personal experience. People give meaning to their lives through the stories they tell, and their stories help to shape other people's lives. The goal of narrative analysis is to transform prople's individual narratives into data that can be coded and organised so that researchers can easily understand the impact of a certain event, feeling, or decision on the involved people, i.e. they can reveal how humans experience their world (Connelly and Clandinin, 1990, p. 1). The result of the narrative analysis is a core narrative of experience. This process involves a two-stage interpretation process. First, the research participants themselves interpret their own lives through the narrative they create. Then, the researcher interprets the participants' narratives.



Narratives can be obtained from various sources, such as journals, letters, conversation, autobiographies, transcripts of in-depth interviews, focus groups, or other forms of qualitative research. These narratives can be individual or collective, and pertain to various life aspects, such as experiences, identities, values, attitudes, or social contexts. Narratives serve as the basic units for exploring and interpreting a phenomenon or problem.

The analysis of narratives typically involves several steps. The first step is collecting narratives or stories from relevant individuals or groups. After collecting the narratives, the next step is coding or categorizing. This entails identifying key themes, patterns, or elements that emerge from the narratives. Coding can be qualitative, where researchers manually identify and categorize key elements, or quantitative, where computer programmes are used to analyse large datasets. Coding is followed by the analysis and interpretation. Researchers analyse the collected narratives to identify relationships, contradictions, trends, or deeper meanings that can be extracted from the stories. This phase may also involve linking the narratives to a theoretical framework or conceptual model to gain a deeper understanding of the phenomenon under study.

The narrative methodology provides contextually rich and in-depth information about individual or group experiences, perspectives, and identities. It also allows researchers to explore subjective experiences, and get first-hand information on the complexity of human life. However, it is important to be aware that the narrative methodology has its limitations, such as the subjectivity of the collected stories, the possibility of selective reporting, or the researcher's biased interpretations.

Narrative analysis offers valuable insights into the lived experiences of individuals and groups, shedding light on their perspectives, beliefs, and social contexts. According to Bruner (1990), the primary way individuals make sense of experience is by casting it in the narrative form, which is especially true of difficult life transitions and trauma. Researchers must respect informants' ways of constructing meaning, and analyse how it is accomplished, because their stories do not mirror the world, but are creatively authored, rhetorical, replete with assumptions, and interpretive (Reissman, 1993, p. 5). So, narrative analysis has to do with 'how protagonists interpret things' (Bruner, 1990, p. 51), whereas the researcher systematically interprets their interpretations. Investigators do not have direct access to another's experience, but have to deal with its representations – text, talk, interaction and interpretation, which are impossible to be neutral and objective. So, in telling about an experience, there is inevitably a gap between the experience as one lived it and any communication about it. How a story will be told depends on the listeners, too. Narratives are inevitably self-representations. However, individual narratives also reveal a lot about the social life, making it possible to examine gender inequalities, racial oppression, and other practices of power that may be taken for granted by individual speakers (Reissman, 1993, p. 5).

Narratives are usually taped and then transcribed for the purpose of research. Transcribing discourse is not easy, and there is always a dilemma as to how detailed transcriptions should be, how they could best capture the rhythm of one's talk, if they should include silences, false starters, discourse markers, etc., and it can be said that there is no single, true representation of the spoken language, whereas the choices as to what to include and how to arrange the text have serious implications for how a reader will understand the narrative (Reissman, 1993, p. 13). Then, the researcher analyses the transcript, edits, and reshapes what was told, creating a hybrid story, influenced by their values and theoretical commitments. Then the text reaches readers, and each text is open to several readings and constructions, even for the same reader, but in different historical contexts (Reissman, 1993, p. 14). Therefore, theoretical levels of abstraction or generalisation are difficult to reach when working with personal naratives, which require comparative work (Reissman, 1993, p. 70). Instead, the aim is to offer insights into how a given person, in a given context, makes sense of a given situation., i.e. to produce an account of lived experience in its own terms rather than one prescribed by pre-existing theoretical preconceptions (Smith & Osborne, 2015, p. 53).

Narrative analysis emphasizes the importance of researchers' being prepared for unexpected consequences, and taking appropriate action when they arise during the research process (Smythe & Murray, 2000). This evolving approach of the narrative methodology contributes to the development of knowledge in a meaningful and sustainable way, informing future practices on socially significant issues (Bruce et al., 2016). What makes the narrative research unique is its emergent design, which involves evolving from data collection to analysis, and generating new knowledge through inductive reasoning from participants' accounts (Bruce et al., 2016). In narrative analysis, how a story is told is as important as what is said in order to understand the psychological and social life.

There are many different types of the narrative analysis. Smith and Sparkes (2007) introduced a typology (Figure 8):

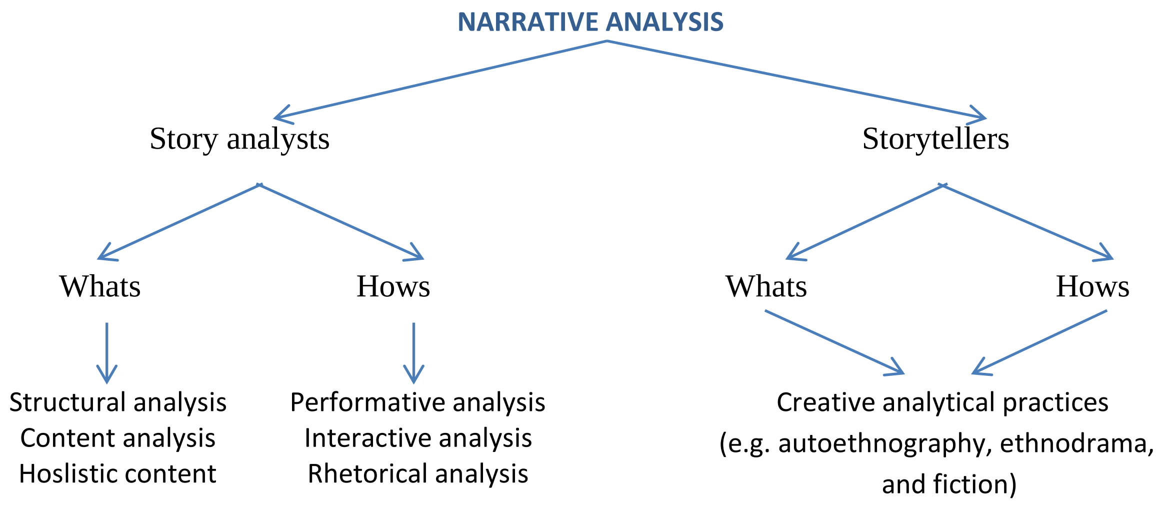
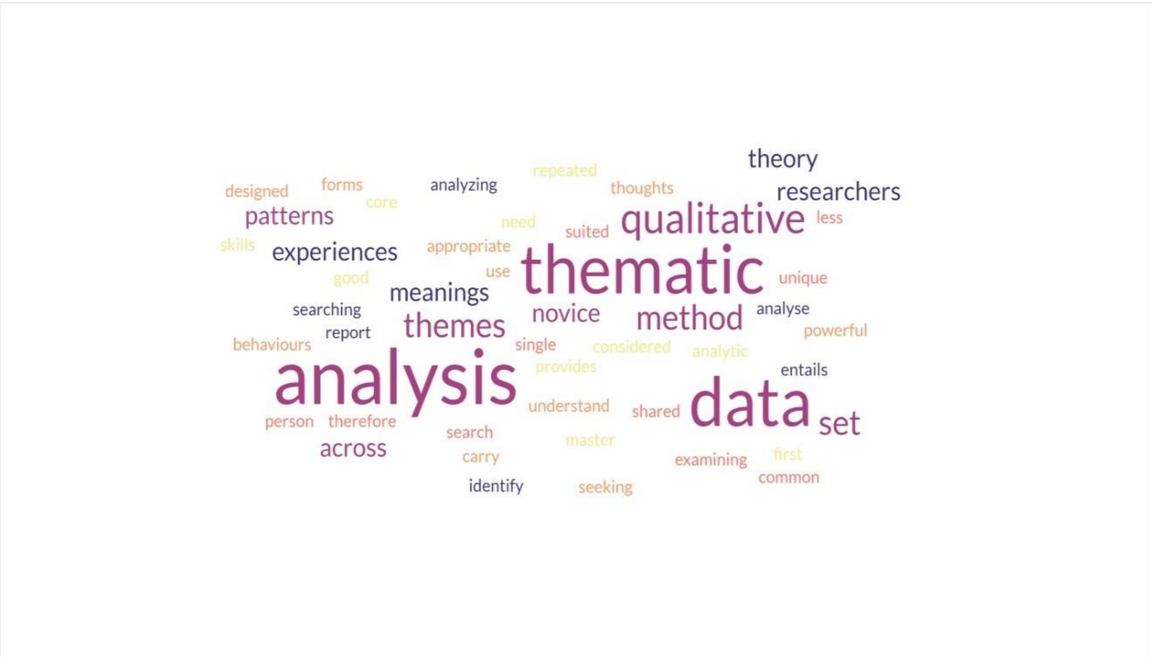


Figure 8. *Typology of Narrative Analysis (Smith and Sparkes, 2007).*

Within this typology, two contrasting viewpoints on the narrative analysis (the storyteller and the story analyst) are illuminated, along with three specific methods (structural, performative, and autoethnographic creative analytic practices) that each viewpoint can employ to analyse the content and characteristics of stories. Story analysts are the researchers who conduct narrative analysis by stepping outside from the story, and using analytical procedures, strategies and techniques in order to abstractly analyse, explain and think about its certain features, theorising it from a disciplinary perspective. On the other hand, storytellers move away from abstract theorising and explaining towards the goal of intimate involvement, engagement and embodied participation with stories. For them, stories are analytical, because when telling stories, people employ analytic techniques to interpret their words (Smith & Sparkes, 2007, p. 21).

Thematic Analysis

Thematic analysis is a method for analysing qualitative data that entails searching across a data set to identify, analyse, and report repeated patterns (Braun & Clarke, 2006). It is considered a good first analytic method for novice qualitative researchers to master. It is an appropriate and powerful method to use when seeking to understand a set of experiences, thoughts, or behaviours across a data set (Braun & Clarke, 2012). It is designed to search for common or shared meanings, and therefore less suited for examining unique meanings or experiences from a single person or data item (Kiger & Varpio, 2020, p. 2). Through a thematic analysis, the researcher constructs themes to reframe, reinterpret, and/or connect elements of the data, which means that themes are not merely organisational tools, used to classify and label data, and that the thematic analysis goes further into the interpretation and data transformation processes, but not to the point of developing a theory, which is the main goal of the grounded theory (Glaser & Strauss, 1967).



In thematic analysis, the theme is a patterned response or meaning derived from the data that informs the research question (Braun & Clarke, 2006, p. 82). It is more abstract than a category, and researchers can identify themes irrespective of the number of times a particular idea or item related to that theme appears in a data set. So, according to the above-mentioned authors, a theme might be given a considerable space in some data items, and little or none in others, or it might appear in relatively little of the data set. Therefore, it is upon the researcher to determine what a theme is, and researchers have great flexibility when identifying themes. They can use an inductive or deductive approach to theme identification (Braun & Clarke, 2012, p. 12). The former derives themes from the researcher's data, and is not necessarily reflective of the researcher's theoretical interests or beliefs on the subject. On the other hand, a deductive approach uses a pre-existing theory, framework, or other researcher-driven focus to identify themes of interest (Braun & Clarke, 2012, p. 12).

10. *According to Boyatzis (1998), there are two levels at which themes can be identified: semantic or manifest, addressing more explicit or surface meanings of data items, and* ***latent*** *or interpretative, reflecting deeper meanings, assumptions or ideologies. So, semantic themes are identified without looking for anything beyond what a participant has said or what has been written, whereas latent themes are the result of examination of underlying ideas, assumptions, ideologies that shape or inform the semantic content of the data.*

The most widely adopted method of thematic analysis consists of six steps (Clarke and Braun, 2017). It is a recursive, rather than linear process, in which subsequent steps may prompt the researcher to circle back to earlier steps in light of new data or newly emerging themes that merit further investigation (Kiger & Varpio, 2020, p. 3). The steps are as follows:

- *Familiarizing oneself with the data – with the entire data set, which requires repeated and active reading through the data. Though time-consuming, the transcription of audio recordings is an excellent way of getting familiar with the data.*
- *Generating initial codes – a code is the most basic segment of raw data that can be assessed in a meaningful way regarding the phenomenon (Boyatzis, 1998, p. 63). A code must be sufficiently well-defined in order not to overlap with other codes, and should fit logically within a larger coding framework/template. Once the coding template is defined, researchers apply the same codes to the entire data set by labeling data extracts with relevant codes, making notes of any potential patterns or connections between items that might inform subsequent theme development. A single extract can be labeled with multiple codes if relevant (Braun & Clarke, 2006).*
- *Searching for themes – the coded extracts are examined in order to find potential themes of broader significance. Braun and Clarke (2012) offer an analogy that, if an entire analysis is seen as a house, the individual codes are bricks and tiles, and themes are the walls and the roof. So, themes do not simply emerge from the data – they are constructed by the researcher through analysing, combining, comparing, and even graphically mapping how codes relate to one another. In an inductive analysis, researchers derive themes from the coded data, and they are reflective of the data set, whereas in a deductive analysis, the theme development is informed by predefined theories, and these themes focus more on a particular aspect of the data set or a specific question of interest (Braun & Clarke, 2006). Thematic maps help to visually demonstrate cross-connections between themes and sub-themes. The researcher should be inclusive at this point, and make note of any theme of potential significance regardless of whether it is directly related to the research question, and regardless of the quantity of data that falls under them (Kiger & Varpio, 2020, p. 5). Researchers can even create a miscellaneous theme to incorporate the codes that do not fit well within the theme template.*
- *Reviewing themes – the researcher looks at coded data placed within each theme to ensure that they are coherent in supporting the theme, that they have enough*

commonality, but are different enough to merit separation. So, data extracts can be re-sorted at this point, and themes can be modified, added, combined, divided, discarded, to better reflect the coded data (Kiger & Varpio, 2020, p. 6). The researcher should keep detailed notes regarding their thought processes and decisions made on how themes were developed, modified, removed. When the researcher decides that the thematic map properly covers all of the coded data, they start checking whether individual themes fit meaningfully within the data set, and whether the thematic map accurately represents the entire data set (Braun & Clarke, 2006). The thematic map should clearly demonstrate how themes interrelate. The researcher has to re-read the entire data set to re-examine the themes and re-code for additional data newly created or modified in this phase, and then revise the thematic map accordingly (Braun & Clarke, 2006), thus confirming the recursive nature of the thematic analysis (Kiger & Varpio, 2020, p. 7).

- *Defining and naming themes – a narrative description of each theme is created. The names of themes are then reviewed to ensure they are brief and sufficiently descriptive. The overlapping areas of themes are identified, as well as sub-themes. Data extracts that illustrate the key features of themes, and are to be presented in the final report, should be selected at this stage, and narratives about them should be created (Braun & Clarke, 2012).*
- *Producing the report/manuscript – writing up the final analysis and description of findings, which is the continuation of the analysis and interpretation that have already been done (King, 2004, p. 267). Both narrative descriptions and representative data extracts (direct quotations from participants) should be used. The discussion section can broaden the analysis by relating themes to larger questions, discussing implications of findings, and questioning the assumptions and preconditions that gave rise to the themes (Braun & Clarke, 2016). Referencing the related literature can also add to the strength of the analysis by building support for why particular themes were selected, and situating findings within the existing body of literature.*

Thematic analysis is simple to master and apply. It is a powerful method for analysing data, which allows researchers to interpret a wide range of data sets. The flexible nature of this analysis can make it difficult for some researchers to determine which aspects of data to focus on, and which theoretical frameworks to use for their analysis.

Framework Analysis

The framework method is becoming an increasingly popular approach to the management and analysis of qualitative data. It is appropriate for use in research teams even when not all members have previous experience of conducting qualitative research. The key terms used in this analysis are thoroughly explained by Gale et.al (2013), as follows:

- *analytical framework - a set of codes organised into categories that have been jointly developed by researchers involved in the analysis, and can be used to manage and organise the data. The framework provides a new structure for the data (rather than the full original accounts given by participants), which is helpful to summarize/reduce the data in a way that can support answering the research questions;*

- *categories - during the analysis process, codes are grouped into clusters around similar and interrelated ideas or concepts. Categories and codes are usually arranged in a tree diagram structure in the analytical framework. While codes are closely and explicitly linked to the raw data, developing categories is a way to start the process of data abstraction;*
- *charting - entering summarized data into the framework method matrix;*
- *code - a descriptive or conceptual label that is assigned to excerpts of raw data in a process called 'coding';*
- *data - qualitative data usually needs to be in the textual form before the analysis. These texts can either be elicited texts (written specifically for the research), or pre-existing texts, such as meeting minutes, policy documents, or can be produced by transcribing an interview or focus group data, or creating 'field' notes while conducting the participant observation or observing objects or social situations;*
- *indexing - the systematic application of codes from the agreed analytical framework to the whole dataset;*
- *matrix - a spreadsheet containing numerous cells into which summarised data are entered by codes (columns) and cases (rows);*
- *themes - interpretive concepts or propositions that describe or explain aspects of data, which are the final output of the analysis of the whole dataset. Themes are articulated and developed by examining data categories through the comparison between and within cases. Usually a number of categories would fall under each theme or sub-theme;*
- *transcript - a written verbatim (word-for-word) account of a verbal interaction, such as an interview or conversation.*

Like thematic analysis and content analysis, this approach identifies commonalities and differences in qualitative data before focusing on relationships between different parts of the data, thereby seeking to draw descriptive and/or explanatory conclusions clustered around themes. Its defining feature is the matrix output: rows (cases), columns (codes) and 'cells' of summarised data, providing a structure into which the researcher can systematically reduce the data in order to analyse it by case and by code (Richie & Lewis, 2003). Most often a 'case' is an individual interviewee, but this can be adapted to other units of analysis, such as predefined groups or organisations. While in-depth analyses of key themes can take place across the whole data set, the views of each research participant remain connected to other aspects of their account within the matrix so that the context of the individual's views is not lost. Comparing and contrasting data is vital to the qualitative analysis, and the ability to compare with ease data across cases, as well as within individual cases, is built into the structure and process of the framework method.

The framework method provides clear steps to follow and produces highly structured outputs of summarised data. It is therefore useful where multiple researchers are working on a project, particularly in multi-disciplinary research teams, where not all the members have experience of the qualitative data analysis, and for managing large data sets, where obtaining a holistic, descriptive overview of the entire data set is desirable. However, caution is recommended before selecting the method as it is not a suitable tool for analysing all the types of qualitative data or for answering all the qualitative research questions, nor is it an 'easy' version of the qualitative research for quantitative researchers. Importantly, the framework method cannot accommodate highly heterogeneous data, i.e. data must cover similar topics or key issues so that it is possible to categorize it. Individual interviewees may, of course, have very different views or experiences in relation to each topic, which can then be compared and contrasted. The framework method is most commonly used for the thematic analysis of semi-structured interview transcripts, although it could be adapted for other types of textual data, including documents, such as meeting minutes or diaries or field notes from observations. The framework method, however, is not aligned with a particular epistemological, philosophical, or theoretical approach. It is rather a flexible tool that can be adapted for use with many qualitative approaches that aim to generate themes.

The development of themes is a common feature of a qualitative data analysis, involving the systematic search for patterns to generate full descriptions capable of shedding light on the phenomenon under investigation. In particular, many qualitative approaches use the 'constant comparative method', developed as part of the grounded theory, which involves making systematic comparisons across cases to refine each theme. Unlike the grounded theory, the framework method is not necessarily concerned with generating a social theory, but can greatly facilitate constant comparative techniques through the review of data across the matrix.

The framework method can be adapted for use with deductive, inductive, or combined types of the qualitative analysis. However, there are some research questions where analysing data by case and theme is not appropriate, and so the framework method should be avoided. For instance, depending on the research question, life history data might be better analysed using the narrative analysis, and documentary data using the discourse analysis (Hodges et al., 2008).

Gale et al. (2013) go further and provide a detailed explanation of seven stages of the framework method procedure:

- *transcription – a good quality audio recording, and ideally a verbatim (word for word) transcription of the interview is needed. Transcripts should have large margins and adequate line spacing for later coding and note taking. The process of the transcription is a good opportunity to become immersed in the data, and is to be strongly encouraged for new researchers.*
- *familiarisation with the interview – becoming familiar with the whole interview using the audio recording and/or transcript and any contextual or reflective notes that were recorded by the interviewer is a vital stage in the interpretation. It can also be helpful to re-listen to the whole audio recording or its parts. In multi-disciplinary or large research projects, those involved in analysing the data may be different from those who conducted or transcribed the interviews, which makes this stage particularly important. One margin can be used to record any analytical notes, thoughts or impressions.*

- *coding – after familiarisation, the researcher carefully reads the transcript line by line, applying a paraphrase or label (a 'code') that describes what they have interpreted in the passage as important. In more inductive studies, at this stage 'open coding' takes place, i.e. coding anything that might be relevant from as many different perspectives as possible. Codes could refer to substantive things (e.g. particular behaviours, incidents or structures), values (e.g. those that inform or underpin certain statements), emotions (e.g. sorrow, frustration, love), and more impressionistic/methodological elements (e.g. the interviewee found something difficult to explain, the interviewee became emotional, the interviewee felt uncomfortable) (Saldaña, 2009). In purely deductive studies, the codes may have been pre-defined (e.g. by an existing theory or specific areas of interest to the project) and therefore this stage may not be strictly necessary, and one can just move straight onto indexing, although it is generally helpful even if a broadly deductive approach is taken to do some open coding on at least a few of the transcripts to ensure important aspects of the data are not missed. Coding aims to classify all of the data so that it can be compared systematically with other parts of the data set. At least two researchers (or at least one from each discipline or speciality in a multi-disciplinary research team) should independently code the first few transcripts, if possible. It is vital in inductive coding to look out for the unexpected, and not just to code in a literal, descriptive way. So, the involvement of people from different perspectives can greatly aid in this. As well as getting a holistic impression of what was said, coding line-by-line can often alert the researcher to consider that which may ordinarily remain invisible because it is not clearly expressed or does not 'fit' with the rest of the account. In this way, the developing analysis is challenged, whereas to reconcile and explain anomalies in the data can make the analysis stronger. Coding can also be done digitally using CAQDAS, which is a useful way to keep track of new codes automatically. However, some researchers prefer to do the early stages of coding using paper and pen, and only start to use CAQDAS once they reach Stage 5.*
- *developing a working analytical framework – after coding the first few transcripts, all researchers involved should meet to compare the labels they have applied, and agree on a set of codes to apply to all subsequent transcripts. Codes can be grouped together into categories (using a tree diagram, if helpful), which are then clearly defined. This forms a working analytical framework. It is likely that several iterations of the analytical framework will be required before no additional codes emerge. It is always worth having an 'other' code under each category to avoid ignoring data that does not fit; the analytical framework is never final until the last transcript has been coded.*
- *applying the analytical framework – the working analytical framework is then applied by indexing subsequent transcripts using the existing categories and codes. Each code is usually assigned a number or abbreviations for easy identification (and so the full names of the codes do not have to be written out each time), and written directly onto the transcripts. The Computer Assisted Qualitative Data Analysis Software (CAQDAS) is particularly useful at this stage because it can speed up the process and ensure that, at later stages, data is easily retrievable. It is worth noting that, unlike the software for statistical analyses, which actually carries out the calculations with the correct*

instruction, putting the data into a qualitative analysis software package does not analyse the data; it is simply an effective way of storing and organising the data so that they are accessible for the analysis process.

- *charting data into the framework matrix - qualitative data are voluminous (an hour of interview can generate 15–30 pages of text), and being able to manage and summarize (reduce) data is a vital aspect of the analysis process. A spreadsheet is used to generate a matrix, and the data are charted into the matrix. Charting involves summarizing the data by category from each transcript. Good charting requires an ability to strike a balance between reducing the data on the one hand, and retaining the original meanings and feel of the interviewee's words on the other. The chart should include references to interesting or illustrative quotations. These can be tagged automatically if you are using CAQDAS to manage your data, or otherwise a capital 'Q', an (anonymized) transcript number, page and line reference will suffice. It is helpful in multi-disciplinary teams to compare and contrast styles of summarizing in the early stages of the analysis process to ensure consistency within the team. Any abbreviations used should be agreed by the team. Once members of the team are familiar with the analytical framework, and well practised at coding and charting, on average it will take about half a day per hour-long transcript to reach this stage. In the early stages, it takes much longer.*
- *interpreting the data – it is useful throughout the research to have a separate notebook or computer file to note down impressions, ideas and early interpretations of the data. It may be worth breaking off at any stage to explore an interesting idea, concept or potential theme by writing an analytic memo for subsequent discussion with other members of the research team. Gradually, characteristics of, and differences between the data are identified, perhaps generating typologies, questioning theoretical concepts (either prior concepts or ones emerging from the data) or mapping connections between categories to explore relationships and/or causality. If the data are rich enough, the findings generated through this process can go beyond the description of particular cases to the explanation of, for example, the reasons for the emergence of a phenomenon, predicting how an organisation or other social actors are likely to instigate or response to a situation, or identifying areas that are not functioning well within an organisation or system. It is worth noting that this stage often takes longer than anticipated, and that any project plan should ensure that sufficient time is allocated to meetings and individual researcher time to conduct the interpretation and writing up of findings.*

While the framework method is amenable to the participation of non-experts in data analysis, it is critical to the successful use of the method that an experienced qualitative researcher leads the project. The qualitative lead would ideally be joined by other researchers with at least some prior training in or experience of qualitative analysis. The responsibilities of the lead qualitative researcher are: to contribute to the study design, project timelines and resource planning; to mentor junior qualitative researchers; to facilitate analysis meetings in a way that encourages critical and reflexive engagement with the data and other team members; and finally to lead the write-up of the study.

Discourse Analysis

Discourse analysis (or critical discourse analysis) is a research method for studying written or spoken language in relation to its social context. It examines how language functions and how meaning is created in different social contexts in order to understand how language is used in real life situations, not to understand the language itself as a system. The focus is therefore placed on the purposes and effects of different types of language, on cultural rules and conventions in communication, on how beliefs and assumptions are communicated, and how the language use relates to its social, political, and historical context.

Discourse analysis can be applied to any instance of written or oral language, but also to non-verbal aspects of communication, such as tone or gestures, because discourse is the complete system by which people communicate, it is the widest interpretation of what we call 'language'. It includes both written, verbal and non-verbal communication, as well as wider social concepts that underpin what language means, and how it changes. For example, it can be revealing to look at how some people use a particular word, or terms from a particular local dialect. This can reveal their upbringing and life history, or influences from other people and workplace culture.

It can also be interesting to look at non-verbal communication as people's facial expressions and hand movements are an important part of the context of what people say. But language is also a dynamic part of culture, and the meanings behind terms change over time. How we understand certain terms tells us a lot not just about the times we live in or the people using those terms, but also about groups that have power to change the discourse.

Language is more than a neutral way of communicating, it is deeply connected with actions and personal identity, and can even shape the way we think about and understand the world (Lucke, 1996, p. 12). Who we are, what we do, and our beliefs are all shaped by the language we use. This makes it a very rich field for analysis. By analysing different types of discourse such as books, newspapers, brochures, advertisements, websites, social media posts and comments, interviews and conversations, researchers try to understand social groups and how they communicate.

Discourse analysis focuses on social aspects of communication. It aims to discover how people use language to achieve certain goals (e.g. manage conflicts, build trust, create doubt). So, it focuses on larger chunks of language, on entire conversations, texts or collections of texts in an interpretive manner, and interpretations are made based on both the content of the material and on contextual knowledge.

Van Dijk (2001) states that the primary aim of critical discourse analysis is to study 'the way social power abuse, dominance and inequality are enacted, reproduced, and resisted by text and talk in the social and political context'. Critical discourse analysts tend to understand, expose and resist social inequality by addressing social problems (Fairclough & Wodak, 1977, p. 271), but the objects under investigation need not necessarily be related to negative or exceptionally serious social or political events, i.e. the term 'critical' should not be confused with 'negative' (Wodak, 2014, p. 302). A discourse analysis includes four different steps:

- *defining the research question and selecting the content of analysis – there must be a clearly defined research question, and then a range of material that is appropriate to answer it is selected (large volumes or smaller samples, depending on the aims and timescale of the research);*
- *gathering information and theory on the context – the social and historical context in which the material was produced and intended to be received. Facts about when and where the content was created, who the author is, who published it, and whom it was disseminated to. Also the literature review on the topic should be done, and a theoretical framework constructed to guide the analysis;*
- *analysing the content for themes and patterns – various elements of the material are closely examined and related to attributes, themes and patterns relevant to the research question;*
- *reviewing results and drawing conclusions – once particular attributes are assigned to the elements of the material, the function and meaning of the language used are examined in relation to the broader context established earlier, and conclusions are drawn that answer the research question.*

Grounded Theory Analysis

Grounded theory analysis is a systematic process consisting of several flexible strategies for constructing theory about social behaviour through analysing qualitative data systematically collected (Glaser & Strauss, 1967, p. 1). So, the theory is grounded in the actual data. Therefore, it is an inductive analysis developed in social sciences, which emphasises the importance of developing an understanding of human behaviour through a process of discovery (Mohajan & Mohajan, 2022, p. 17). It is typically used when there is no available theory to explain a process that occurs over time, but instead of making assumptions, it ensures a more natural view of human actions within a social framework (Creswell, 2007). Scientific theories need to be conceived, elaborated and checked, which means that induction (the discovery of hypotheses), deduction (drawing implications from hypotheses) and verification (checking if they are a total, or a partial qualification or negation) take place throughout the research project.

Charmaz (2016) defined the grounded theory as follows: *'GT begins with inductive theory, relies on comparative analysis, involves simultaneous data collection and analysis, and includes strategies for refining your emerging analytic categories'.*

Grounded theory analysis is guided by 7 basic principles:

- *research should start with a broad research focus, which means that there should not be some predetermined research questions (Charmaz, 2006).*
- *literature review should be delayed until later stages of the research, so that the researcher remains as neutral as possible towards the existing theories (Glaser, 1978). The existing literature is not used as a theoretical background in GT, but rather as the data to be used by the analytic strategies of the research (Creswell, 2014). The in-depth literature review takes place after the theory is discovered (Glaser, 1998).*

- *data collection and analysis should be conducted simultaneously. The data takes the form of verbatim interviews, and the researcher must ensure that the interview is transcribed exactly as recorded, because the theory evolves from the ongoing data collection and analysis, and not from concepts and ideas held by the researcher. The analysis should begin as soon as data is collected (Oliver, 2012).*
- *the constant comparison method should be used so that researchers constantly compare the emerging codes within the same data item, and across the same data set (Corbin & Strauss, 2015).*
- *memos should be kept, helping researchers to clarify what they meant or why they created certain codes in their analysis (Bryman, 2012).*
- *theoretical sensitivity – being open to what emerges from the data, and identifying possible connections between the emerging findings and literature (Glaser, 1978).*
- *theoretical sampling – sourcing new data with which the emerging concepts can be further explored, i.e. the theory decides where and what the sample is (Glaser & Strauss, 1967).*

It is this cyclical motion of data collection, immediate analysis, and further collection to produce concepts from which a theory evolves that makes the grounded theory unique (Pulla, 2016, p. 79).

Grounded theory analysis is time-consuming, and difficult to conduct. It requires in-depth interviews from many individuals, who are familiar with the phenomenon being studied, and therefore there are large amounts of data to manage (Creswell, 2007).

Grounded theory researchers develop tentative interpretations about the data through constructing codes for data fragments, and categories for clusters of codes, and then they check and refine the major categories by using them to re-examine the data collected before defining the category, by returning to the field site and gathering more data, and by building focused questions for later data collection (Charmaz, 2017, p. 2).

Grounded theorists go back and forth between collecting and analysing data, because grounded theory is an iterative process, where data collection sparks analysis, and analysis directs subsequent data collection with an aim of refining and checking the analysis (Charmaz, 2017, p. 2). They focus on what is happening, rather than on preconceived general topics, making comparisons throughout the research process, thus discovering new properties, dimensions, visible and hidden processes. They check the developed theoretical ideas with subsequently collected data, and finally state the implications for professional practice and public policy (Charmaz, 2017, p. 3).

Coding in grounded theory analysis

Coding is an essential procedure in a qualitative analysis, which highly influences the excellence of qualitative research (Strauss, 2003, p. 27). Strauss (2003) provides a thorough explanation of three different types of coding: open coding, axial coding, and selective coding.

Open coding is the initial, unrestricted coding performed by the close examination of data 'line by line, or even word by word' (Strauss, 2003, p. 28) in order to open up the inquiry and produce concepts that seem to fit the data. At this point, the concepts are provisional, and every interpretation is tentative - it may or may not work, or may be modified, and the coding is grounded not only in the data but also on the experiential data and knowledge of literature which the researcher brings into the inquiry. However, open coding forces the researcher to break the data apart analytically. It is important for the researcher to ask a set of questions continually aimed at the generation of a core category that will be at the centre of the theory, such as:

- *What study are these data pertinent to? – to remind the researcher that an original idea may not turn out to be that all.*
- *What category does this incident indicate? – to keep the researcher from getting lost in the rich data by forcing the generation of codes that relate to other codes.*
- *What is actually happening in the data? – to help the researcher define the main problem.*

The data should be analysed thoroughly, and carefully coded to ensure conceptual density, and minimize the overlooking of important categories, and when a code seems relatively saturated and nothing new is happening, the researcher will find himself/herself moving quickly through the data, scanning pages until something new catches the eye. If some portions of the total analysis are not satisfying, another open coding, and even new data collection may become necessary.

At the point of open coding, coding should be frequently interrupted in order to write a theoretical memo, and the accumulated memos will move the analyst further from the data and into a more analytic realm. The analytic relevance of common variables should not be assumed – they should force their way into the grounded theory.

It is important for the analyst not to be too committed to the first codes, and to ensure that individual codes are verified and saturated. Open coding proliferates codes quickly, but the continual verifying that each code really fits slows down the process (Strauss, 2003: 32).

Axial coding – the analysis revolves around the axis of one category at a time, i.e. the intense anlysis of one category at a time is performed resulting in cumulative knowledge about the relationships between that category and other categories and subcategories. According to Strauss (2003), axial coding is an essential aspect of the open coding, which alternates with looser kinds of open coding, especially when the analyst examines new aspects of the phenomena under study.

Selective coding – coding systematically for the core category, i.e. the analyst delimits coding to only those codes that relate to the core codes , and the core code serves as a guide to further theoretical sampling and data collection. The analytic memos become more focused and help to achieve the integration of the theory (Strauss, 2003: 33).

The findings of a grounded theory analysis are supposed to be unique, tightly anchored in the data collected for the particular research purposes, i.e. in the words and experiences of the research participants. It introduces a new phenomenon and emphasises its discovery, whereas the description and verification are not its primary concern. So, the success of the analysis to a great extent depends on the researcher's sensitivity and analytical skills, and it is important that the researcher constantly be aware of his/her own position in relation to understanding and conveying the messages of the participants.

References

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *Nursing Plus Open, 2*, 8–14.

Berg, B. L. (2001). *Qualitative research, message for the social sciences* (4th ed). Allin and Bacon, Boston, 15–35.

Boyatzis, R. E. (1998). Transforming qualitative information: Thematic analysis and code development. Sage.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association.

Bruce, A., Beuthin, R., Sheilds, L., Molzahn, A., Schick-Makaroff, K. (2016). Narrative research evolving: Evolving through narrative research. *International Journal of Qualitative Methods, 15*(1).

Bruner, J. (1990). *Acts of meaning*. Harvard University Press.

Bryman, A. (2012). *Social research methods*. Oxford University Press.

Burnard, P. (1995). *Learning human skills. An experiential and reflective guide for nurses* (3rd ed.). Butterworth-Heineman, Oxford.

Catanzaro, M. (1988). Using qualitative analytical techniques. *Nursing Research: Theory and Practice*. 437–456.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* Sage.

Charmaz, K. (2016). Constructivist grounded theory. *The Journal of Positive Psychology, 12*(3), 299–300.

Charmaz, K. (2017). The power of constructivist grounded theory for critical inquiry. *Qualitative Inquiry, 23*(1), 34–45.

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology, 12*(3), 297–298.

Connelly, M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher*, 19, 2–14.

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research*. Sage.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches.* Sage.

Creswell, J. W. (2009). *Research design qualitative, quantitative, and mixed methods approaches*. Sage.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). Sage.

Downe-Wamboldt, B. (1992). Content analysis: Method, applications, and issues. *Downe, 13*, 313-321.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In T. van Dijk (Ed.), *Discourse studies: A multidisciplinary introduction* (pp. 258–284). Sage.

Gale, N. K., Heath, G., Cameron, E., Rashid, S., Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology, 13*(1), 1–8.

Glaser, B. (1998). *Doing grounded theory issues and discussions.* Mill Valley, CA Sociology Press.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory strategies for qualitative research*. Mill Valley, CA Sociology Press.

Graneheim, U. H. & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, *24*, 105–112.

Hodges, B. D., Kuper, A., Reeves, S. (2008). Qualitative methodology: Discourse analysis. *BMJ, 337*(a879), 570-572.

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher, 42*(8), 846–854.

King, N. (2004). *Using templates in the thematic analysis of text*. Sage.

Krippendorff, K. (2004). *Content analysis an introduction to its methodology* (2nd ed.). Sage.

Lucke, A. (1996). Text and discourse analysis. *American Educational Research Association, 21,* 3-17.

Mohajan, D., & Mohajan, H. K. (2022). Constructivist grounded theory: A new research approach in social science. *Research and Advances in Education, 1*(4), 8-16.

Morse, J. M. & Richards, L. (2002). *Readme first for a user's guide to qualitative methods.* Sage.

Oliver, C. (2012). Critical realist grounded theory: A new approach for social work research. *British Journal of Social Work, 42*, 371–387.

Pešić, J. (1997). Analiza diskursa i njen odnos sa psihologijom. *Psihologija, 3*, 263-278.

Pulla, V. (2016). An introduction to the grounded theory approach in social research. *International Journal of Social Work and Human Services Practice, 4*(4), 75-81.

Reissman, C.K. (1993) *Narrative analysis*. Sage.

Ritchie, J., & Lewis, J. (2003). *Qualitative research practice – A guide for social science students and researchers.* Sage.

Saldaña, J. (2009). *The coding manual for qualitative researchers*. Sage.

Smith, J. A., & Osborn M. (2015). Interpretative phenomenological analysis as a useful methodology for research on the lived experience of pain. *British Journal of Pain, 9*(1), 41–42.

Smythe, W. E., & Murray, M. J. (2000). Owning the story: Ethical considerations in narrative research. *Ethics & Behavior, 10*(4), 311–336.

Sparkes, A. C., & Smith, B. (2007). Disabled bodies and narrative time: Men, sport, and spinal cord injury. In J. Hargreaves and P. Vertinsky (Eds.), *Physical culture, power, and the body* (pp. 158–75). Routledge.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory.* Sage.

Thorne, S. (2000). Data analysis in qualitative research. *Evidence-Based Nursing, 3*, 68–70.

Van Dijk, T. A. (2001). Critical discourse analysis. In D. Tannen, D. Schiffrin, & H. Hamilton (Eds.), *Handbook of discourse analysis* (pp. 352–371). Blackwell Publishers Ltd.

Wodak, R. (2014). Critical discourse analysis. In C. Leung, & B. V. Street (Eds.), *The Routledge companion to English studies* (pp. 302–317). Routledge.