Introduction.............................................................................................................................................. 4

Methodology and the Scientific Research Approach ............................................................................... 6

Some words about methodology .................................................................................................................. 6

The world of scientific research .................................................................................................................. 8

The major coordinates ................................................................................................................................. 8

Exploring, describing, and explaining the world ......................................................................................... 12

Being objective ........................................................................................................................................... 12

Deduction and induction ............................................................................................................................... 13

The Essence of Measuring and Related Statistical Concepts ................................................................. 18

Measuring means representing .................................................................................................................... 18

Measuring the invisible: psychological constructs and the research model .............................................. 19

Status versus process diagnostics ............................................................................................................... 22

Measurement scales .................................................................................................................................. 23

Errors and quality of measurement: The classic test theory ..................................................................... 26

Population and Sample ................................................................................................................................. 30

Description of univariate measurements: seeking the normal distribution .............................................. 34

Standard error of the mean .......................................................................................................................... 42

Exemplifying the information so far ........................................................................................................... 45

Research Methods ....................................................................................................................................... 48

Quantitative and qualitative research ........................................................................................................ 48

Reactive and non-reactive research strategies ............................................................................................. 52

The observation ............................................................................................................................................ 55

The interview and questionnaire .................................................................................................................. 58

Grounded Theory and Ethnography ............................................................................................................ 61

Grounded Theory ......................................................................................................................................... 62
Ethnography ........................................................................................................................................ 71

The experiment ...................................................................................................................................... 73

Field, simulation, or laboratory experiment? ..................................................................................... 74

Experimental scenario and transparency .......................................................................................... 75

Independent and dependent variables .............................................................................................. 77

Design .................................................................................................................................................. 81

Control and counterbalancing ........................................................................................................... 87

The role of the experimenter ................................................................................................................. 88

When preparing experiments: ........................................................................................................... 88

When organizing the experimental session: ....................................................................................... 89

When running the experiment: ......................................................................................................... 90

Your key ethical responsibilities: ..................................................................................................... 91

A Few Final Remarks ......................................................................................................................... 92

References ........................................................................................................................................... 93
Introduction

*Research and Statistical Methods I* is the first in a series of working papers discussing relevant issues in user psychological research (Saariluoma, 2004). The series is intended for students, but also researchers and teaching personnel, active in areas that by tradition have not been, but are increasingly concerned with human beings. The current paper is written in the style of a study reader and provides a brief introduction and overview of some important method-related issues, which shall be continued in *Research and Statistical Methods II*.

A good example for beneficiary research fields are those related to the study of the nature and impact of information technology (IT). Especially the discipline of Information System Science (IS), by nature situated and intrinsically related to the disciplines of Computer Science on the one side and Human or Social Sciences on the other side, is progressively confronted with the need to study human beings, as individual and collective users of technology. Hence, it is the comprehension of human experiences, judgments, feelings, and actions that eventually make the very essence of the technological artifacts themselves understandable. These very same issues are naturally also being tackled in user psychological research, an applied branch of psychology: How do human beings make sense of technology and interact with them in every day contexts, and how does this reflect back on the ways technology is conceived of, designed, engineered, and advocated?

Naturally, it may be somewhat misleading to speak of user psychological research methods, because the methods and their foundations have not been developed under the label of user psychology itself. Rather, the foundations have been laid within the social sciences and especially psychology for more than 100 years, and are now
being applied to the specific questions that are of interest when studying human technology issues. It is my firm belief that the expertise in conducting research involving human subjects is of great contemporary value and needs to be made available to novel disciplines that emerge in closely associated areas, such as IS. Hopefully this reader will contribute to this endeavor.

As is evident from the table of content, the current paper does not intend to address all method-related issues relevant to conducting user psychological research. This is true in general, as well as in particular - considering the delimitation of the field. In its current version the *Research and Statistical Methods I* (a) portrays the foundations of scientific thinking and the nature of research relevant to the field, (b) discusses the essence of measuring, and (c) provides a brief introduction to the core strategies of examination.

The current paper does not essentially proceed to the discussion of issues of analytical models and statistical analysis, nor does it provide a detailed description of the multitude of concrete research techniques used in the field of user psychology - especially those pertaining to usability investigations. It also leaves aside questions about research in a greater context as well as the issues related to the communication and reporting of research findings. All of these concerns shall be covered in future working papers concerned with method-related issues in user psychology.

Finally, being ‘working papers’ it is essential that the readers take notice of the version they access. The texts are updated in infrequent intervals, without further notice.
Methodology and the Scientific Research Approach

Some words about methodology

Before we plunge into more technical issues we need to address the subject of methodology and its relation to method. Methodology is far too often used carelessly and interchangeably with the term method, i.e., “my methodology was to use interviews”. Methodology, however, does not refer directly to the practical issues of conducting research, but rather to its epistemological underpinnings. Being a compound of the Greek terms “methodos” (i.e., the pursuit or ways to reach a goal) and “logos” (i.e., word, reason, or discourse), methodology refers to a meta-theoretical and deeper reflection about method. In doing methodology a researcher is concerned with “why”-questions about his or her research rather than “how” and “what”. E.g., “Why do I believe that these questions are the right ones to ask about my matter of interest?”; “Why do I usually choose this kind of technical approach in my inquiries?”; “Why do I think that this kind of data will reveal essential aspects about the mental processes I want to investigate?” etc.

Of course there are no easy answers to these questions, and we shall here not confuse ourselves too much with these rather philosophical issues. On the other hand, we shall also not deceive ourselves into thinking that answering these questions, explicitly or implicitly, really can be avoided. In fact, methodology is what really binds all the chapters included in this reader together. Starting from very general issues about conducting scientific research down to very concrete statistical procedures, they are all built on a wide range of logical assumptions and scientific conventions, to which, in principle, we can or cannot subscribe - but which we cannot evade.
For instance, reflecting upon our intrinsic idea about human nature, we all have quite clear ideas about whether some people are just born the way they are, whether people really change, whether we can generalize from hearing news about one terrorist to all terrorists, whether our own performance in an intelligence test is really representative for how smart we are in real life, whether there can be one single event changing a person’s life, whether parents are to blame when their children have no manners, etc.

As researchers, we do generally not create such beliefs as a consequence of choosing certain research topic or methods. Quite to the contrary, it is our methodological standpoint that shapes the kind of questions we ask and the research we are doing. Commonly, this fact becomes clearer to us only along the way of growing as a researcher. At onset it is the saliency of research fields and topics that provide us with an identity. Later on we often find ourselves in influenced and collaborating with other researchers that share similar methodological views.

Bastalich (2005) provides on her website this summary of why awareness of the own methodological approach is essential. It influences:

- the research questions you ask;
- the type of research you do;
- the method and mode of analysis you use;
- what you extrapolate from your data set;
- your claims to ‘intellectual authority’.

And to avoid common misconceptions it is important to understand that:
• methodology is NOT determined by your method, or your choice of qualitative or quantitative data (e.g.: ‘My methodology is qualitative, I am doing interviews’);
• methodology is NOT something you choose based on your topic or research question;
• methodology is NOT something you can easily mix and match (e.g.: ‘My research is grounded in a number of methodological approaches’).

So, whenever you read something about methodological approaches or whenever you find yourself asking “why”-questions about your research, take them serious. Established experimental techniques and statistical procedures are not the holy truth of scientific research. They are all grounded in numerous prior assumptions the researcher community has made – however, we are usually only limitedly aware of these.

The world of scientific research

The major coordinates

Before getting lost in a space of methodological vagueness, we shall therefore reiterate in the current chapter the major assumptions that underlie conventional research practice. Hence, what is commonly understood as the scientific research approach?
Let us start out by making matters simpler. In order to do so, I list here a few general intuitions that most of us will find supportable:

- We believe that there is a world of real things out there, such as the Nokia N-Gage and a friend of ours that bought it last week and is absolutely fond of it.
- We usually also believe that we all have a shared basic awareness of reality, i.e., the world of things themselves (although we may experience, judge, and interact with the N-Gage, for instance, in very different ways).
- We agree that human thinking and behavior displays certain regularities within and across individuals and that these suggest the existence of general and universal psychological laws.
- We interpret the world as a gigantic causal web of if-then relationships. We can engage in searching a reason or cause for anything and our models of the world and ourselves are constructed of multitudes of such causal relations. Hence, we also believe that based on our research findings we will be able to purposefully intervene in a world with events.
- We believe that human beings do not function in a pure mechanistic way and that there are certain degrees of freedom to every causal relation (e.g., we believe in free will).
- We believe that our theories and models are imperfect approximations of these laws and that careful testing of our models with sufficient numbers of carefully selected participants will reveal where we have to make changes to our present assumptions.
These statements taken together, we can say about the nature of most modern research activities that it is:

- **positivistic** (as opposed to idealistic and interpretivistic);
- **empiristic** (as opposed to radically rationalistic);
- **inferential** (as opposed to being simply descriptive and correlational);
- **nomothetic** (as opposed to idiographic); and
- **stochastic or probabilistic** (as opposed to deterministic).

In adopting a positivistic view psychologists rejected living in a chaotic and purely speculative world. This was a very important step to take, because it opened to the path away from the philosophical discourse alone and allowed for the application of empirical research models borrowed from the field of natural sciences. Psychologists started to engage into carefully constructing cycles of research and development where theoretical predictions were compared to empirical observations. However, there remained the problem about how much meaning may be interpreted into the collected data. Does the data about a person’s behavior describe only behavior to be translated into laws of behavioral regularities, or does it tell us something about the person’s mental constituents? Today, the dispute between the former (Behaviorists) and the latter (Cognitive Psychologists) has been largely decided and we believe that carefully devised empirical research allows us to make inferences about psychological processes that are otherwise hidden from direct observational access.

As a drawback to the adoption of positivism and the worshipping of the natural sciences (especially physics) social scientists sometimes tend to neglect the special kind of research subjects and environment they are dealing with. Unlike in the world of
physics, where twice an object’s mass renders twice the gravitational force, human systems do not conform to our models in quite such a stringent way. This means, our measured relationships are usually of stochastic nature and our predictions probabilistic.

However, in spite of this restriction and also in spite of the self-evident uniqueness of each and every mind, we believe that all individuals function generally in very similar ways. Therefore the experiences with a few can, within limits, be generalized - and insights derived from measuring the mass be applied to a single person. This nomothetic commitment means also that we do not need to develop methods for each individual separately.
Exploring, describing, and explaining the world

Now, what are we really after when doing research? In what ways do we strive to enhance our knowledge?

Research either wants to find out what there is (exploratory research) describe what was found (descriptive research), or explain why things were found the way they are (explanatory research). Most of the time we describe and explain – indeed unprejudiced observation has become very rare. We chose a research topic, scan the current theoretical models, adopt a popular empirical paradigm and continue on a well-defined research path. Especially the final issue (i.e., explaining) is the one that really tickles us; and we must therefore be careful not to underestimate the value each of the three aspects of scientific inquiry.

A real problem is for instance that human beings have all too often ready explanations for world affairs without thoroughly examining the phenomena and its contexts in the first place. Careful observation takes time and skill and is very critical, because we easily tend to see only what we believe to be true and neglect many surrounding issues.

Being objective

Now this is a tricky one. A popular credo is that research can be measured by its degree of objectivity. This would mean that the research questions, the methods, and the findings are independent of the researcher. I.e., they tell only something about the world of affairs (mental or environmental) and are not part of the researcher’s own fabulations.

In fact, striving for objectivity in research is just as much an honorable virtue as it can be an act of self-deception or tragic illusion. This is simply because all the
peculiarities and laws of mental functioning that we uncover in the course of conducting research with human beings do not only apply to our Nokia N-Gage user, or to the student in his or her classroom, they apply to the same degree to the researcher itself.

As noted in the section on methodology, for instance, our whole research practice is heavily biased by our methodological leaning. And what makes matters worse, is that we are only seldom fully aware of all the implicit assumptions inbuilt into our research practice (see Saariluoma, 1997). Further, it is obvious that the data we collect is not the kind of knowledge we set out for. In order to convert empirical data into research findings we must interpret those in the light of the theoretical models that we have chosen based on our methodological beliefs. We can do so in very transparent ways, or be very intuitive in our interpretations. Either way, they remain our personal interpretations of the matter.

Of much greater importance than the claim of impartial research findings is to reflect upon and emphasize the researcher as a key actor. It is the researcher’s obligation to point out what kinds of decisions were made, where, when, and why. The developed methods should be described in a transparent fashion so that the investigation can be reviewed, criticized, and, if desired, replicated by other researchers. It is this type of objectivity, which makes the research method itself part of the research object that enhances the quality of scientific inquiry. Not the one that tries to disguise the researcher’s involvement.

*Deduction and induction*

The classic research paradigm is *deductive* in its nature. This means:

- it starts out from accepting a certain theoretical model (given or developed) as its basis;
• then formulates research questions that are within the scope of the theoretical model (actually this point partly precedes the previous one);
• makes predictions based on the model (i.e., research hypotheses);
• operationalizes the proposed relationships and processes;
• conducts measurements;
• tests the degree to which they seem to be in line with the proposed hypotheses;
• and makes certain refinements to the theory it started out with and/or continues to ask additional research questions.

There is a considerable danger that researchers get too hung up in deductive research only. Critical research continuously needs to question whether the selected phenomena and methods do not promote the findings of artifacts and whether the observations strictly exclude alternative explanations. Otherwise they must be considered as well. There are several ways to do this – *inductive* research is one of them.

Induction, as I want to advertise it here, is not simply the reasoning step leading from the data we collected back to suitable explanations within the pre-chosen theoretical frameworks and again forward to consistent conclusions. Generic inductive research necessitates that we frequently broaden our view and become naïve observers of our surroundings and the phenomena we are interested in, only to see whether we develop new ideas and assumptions that might affect our theories in much more fundamental ways than deductive research alone does. In this way inductive reasoning reaches beyond the set of premises we accepted initially. And it is this set of alternative
explanations, which in turn needs to be explored again in deductive research. In this sense, deduction and induction are the “yin and yang” of all well founded scientific progress.

_Causality – there is a “explanation” for everything_

As we have already noted before, just knowing “what is” does mostly not provide us with the kind of knowledge we are thirsty for. As designers, engineers, advertisers, retailers, technical supporters, customers, users, or whoever in the chain of HCI participants, we would like to understand how and why things come about. Causal knowledge allows us to influence or excise control over events, to predict and prepare and prevent, or simply find peace of mind through understanding. The core of knowledge construction is therefore concerned with finding connections, links, and associations between things. We do not experience the world as a great puzzle of detached events and facts. Indeed, we mostly overemphasize the relations between incidents, by having an explanation for just about anything that occurs.

Assertions about causality are the driving force in knowledge construction, and the core ambition of scientific research. Our explanations are incorporated in functional interpretations and _cause-end_ beliefs. And although we are all innately familiar, comfortable, and exceptionally quick in drawing causal conclusions, the scientific pursuit of causality remains probably the most intricate of all issues.

There are at least three logical reasons for this:

- There is probably no single effect that has only one possible cause.
- There is probably no single effect that is brought about by a single cause alone.
• There is probably no cause that has only one specific effect.

There are also many psychological reasons, e.g.:

• Human judgment is subject to certain biases.
• Human judgment is vulnerable to false impression.
• Correlation between processes or paired-occurrences of events is easily mistaken to signify causality.

In reality things are always multi-determined by a network of causes. And it is a very tedious process to single out the actual causes and their orders of impact inside the causal chain-reaction. Figure 1 illustrates this idea in a very simplistic way. The key question to ask is “What’s the cause for the train to end up in point B?” This question is logically equivalent to the majority of causal research questions, such as why did user X press the wrong button, why did customer Y prefer device P over device Q, etc. As we will discuss later in this reader, most research is concerned with isolating causes that decide between different alternatives, but not so much with a comprehensive explanation of why something comes about.

**Figure 1: Tracking causality**
In the train example, it is for instance obvious that there are literally thousands of causes (optional and necessary ones) for the train to progress to point B: e.g., there are train tracks leading from the train's current location to point B, the steam engine has been invented in the 18th century, because the locomotive is working properly, there is someone operating the locomotive long enough for the train to reach point B, the train is headed into the right direction. However there is only one causal entity because of which the train should end up in point B, instead of A. This is, because there is a track switch that effectively decides the train’s path and has been set into position B. Why it is set to this position is yet another question.

Returning to logical propositions we usually test these premises in order to assume an exclusive causal relation between two events when A and B:

- Whenever A, B must follow.
- Whenever not A, there must also not be B.
- In any case of ‘A then B’, there must not be simultaneously C.
- A change in A coincides with a change in B.

What do these propositions come down to? On the one hand, if there shall be a connection between A and B, it is obvious that A and B need to be related to each other in time and space. For instance, they may be concurrent events, or follow each other in certain regular chronological order. The other important aspect of a causal link, is that it is present between some facts and events, but not between others. If a certain effect takes place no matter what, it is hardly of interest to investigate its causes. In this case we turn to fatalism. Often we also expect the events to be correlated, in the way that
only a small amount of A brings about a small amount in B, whereas huge amounts of A intensify also B.

The Essence of Measuring and Related Statistical Concepts

*Measuring means representing*

No matter what our inquiry is about, we need data, i.e., some kind of information about some kind of world affair, which we can further process. Acquiring this data means that we need to create a representation of the events we are interested in. Measuring, then, is one step in the process of transforming the event into data form representation. In measuring we focus and interpret what we perceive in a priori defined ways. That means, when measuring, we purposely omit a wide range of what is really taking place, and we usually alter the information in some way or another.

Data can be of various formats and degree of abstraction. It can be for instance verbal reports, audio and video footage, human behavioral traces, or classic numeric and string codes. Whatever your data looks like, it is important to realize that even in the rawest form of measurement, i.e., simple recording of events, one is poised to make choices about what is represented and what not, as well as how. For instance, if we videotape a set of user actions, we will have to decide upon the camera focus, resolution, automatic lighting corrections, etc. The measure (or raw recording in our case) will never fully comprise all aspects of the original event. The same is naturally also true for a researcher’s simple observations, which are subject to all kinds of cognitive processing.
Usually, however, we mean by measuring something more invasive than just the recording of events. This means, measuring can mean anything on a continuum from pure (i.e., analog type) gathering of data, to highly abstracted forms of event coding. In the following chapter we will discuss these and other issues related to measuring and measurement.

Measuring the invisible: psychological constructs and the research model

As stated in the opening chapters, as human researchers we are usually interested in measuring and explaining more than just that was is visible: user psychological research is therefore highly inferential.

![Diagram of measuring the invisible: Theory, data, and facts](image)

Figure 2: Measuring the invisible: Theory, data, and facts

In fact, the majority of the affairs we are interested in as psychologists are by nature not directly perceivable, or for that matter, measurable. To be even more precise,
of most variables that we set out to measure, we do not even know for certain that they exist in the form we conceive them. All psychological constructs belong to this world, e.g., human values, intelligence, emotions. That means that measuring can never be independent from the theory of the concepts that we include in our research. As more data is collected in the course of empirical research progress, not only will the measurements change, indeed the whole idea of what we measure, i.e., the constructs will change. Figure 2 depicts the general idea of psychological investigation.

Coinciding with the choice of a particular research phenomenon, we usually also generate a psychological theory (naïve or research-based) about the affair we are interested in. This theory provides us with a model about what might be happening beneath the surface (e.g., in the mind of the user) and it allows us to selectively attend to some aspects and disregard others during our investigation (step [a] in Figure 2). When constructing a measurement environment for our research, the same psychological theory guides our process in creating a test situation (i.e., a set of test materials and tasks) and it helps us in preparing the necessary observational criteria and measurement instruments (steps [b] and [c]).

Finally, when running the investigation, we will confront the participant with the test situation we have devised, which comprise the stimuli from the participant’s perspective, and after processing the data the participant will show some kind of behavior, of which we interpret a part as the participant’s response to our test situation. Again, part of this response will be recorded or measured by us during the process of test observation.
Based on our psychological theory, the specifics of the test situation and the observed behavior, we will then analyze and interpret the data we have collected. This step can lead to the output of the measure we were after, granted the theory is unchallenged or confirmed (option [e1]) and/or it can result in necessitating an alternation or adaptation of the psychological model (option [e2]).
It is should be obvious from this description of the research cycle that throughout our investigations we live in a model world that abstracts from many aspects of realities. Figure 3 picks up on this ‘real world’ vs. ‘model world’-idea. It shows that the real phenomenon we would like to study ceases to exist as such in our research as soon as we can formulate it, and our research questions concerning the issue. From this point on forward research is guided by the theoretical, conceptual, and empirical models we adopt. They guide the generation of hypotheses, the selection and development of constructs, the preparation of the test situation, the measurement processes, and the collection and interpretation of the data. Only thereafter we generalize and project our findings back into the “real” world.

*Status versus process diagnostics*

So far we have looked at the basic role and idea of measuring in scientific research. We have cared little about *what* we measure. Hence, we will do so in this and the following section.

There are in principle two distinct broad measuring focuses. One is *status*-oriented, the other *process*-oriented diagnostics. According to the status model human behavior is the product of relatively stable characteristics or *traits*. Typical examples are intelligence, personality, values, etc. Instruments that are based on this view are also called psychometrical measurement tools. Process-oriented measuring models on the other hand put the actual behavior of the human being in a certain context into the center of attention, i.e., we measure psychological *states* instead of traits. Here I refer especially to interactionist approaches (i.e., the study of situation-reaction dynamics)
and leave other types of process-oriented models aside because they have less application value in user psychological research.

Obviously, we commonly need to mix both of these perspectives in our actual research. That means we might be interested in certain (personality) types of users, and study their actual reactions to particular devices in an experimental environment. Or we interview them to get their story about how and why they use a device in their everyday life.

Monitor yourself in your private life as well as in your research whether you register a certain behavior that someone displays simply as a sequence of interaction between person and context, or how fast and how far you attribute it to stable underlying person characteristics.

Measurement scales

As stated in the previous section, when measuring we assess people with respect to whether or to which degree they are something (e.g., technophobe) or they behave in a certain way (e.g., avoid the use of technological devices). In any case, we assess humans on certain attributes that we are interested in. Measuring is therefore equal to the assignment of a value on a particular attribute dimension. Attributes themselves are hierarchically nested, so that a value on one attribute dimension may be an attribute with own values on a more fine graded approach. E.g., ‘English’, ‘French’, and ‘German’ may be the values for the measurement of which type of foreign languages a particular Finnish student speaks, and ‘Beginner’ ‘Intermediate’, and ‘Proficiency Level’ might be the values to characterize the skill level in each of the languages by

Sacha Helfenstein
themselves. When passing a proficiency test in French we would finally also be able to assign a numeric value to the question how well the skill is developed.

Usually we do not only know what attribute we want to assess but we also have a rather clear idea about the kinds and range of values that are to be assigned. In the example above the possible values for the attribute ‘Foreign Language’ for a Finnish speaking person could be all languages except Finnish. For the attribute ‘Skill Level’ the values are arbitrary labels, such as ‘Beginner’, and grades may finally be awarded to describe the degree of expertise within a certain skill level.

This being said, it is obvious that attribute values have different formats (verbal expressions and numbers), a certain range, and also an order component. Traditionally, we distinguish between four different types of orders for attribute values: so called scales.

The most basic data level is the one of nominal order. That means that there is no more than the name of the value itself that identifies its place within the attribute dimension. Typical examples are gender, nationality, and brand: Some people are male, Finnish, and use Nokia mobile phones; others are female, Korean, and use a Samsung phone.

Ordinal scales represent measurements in more dimensionally ordered form than nominal scales because they imply that some values are more or less than other values. Typical examples might be level of education, type of mobile phone, and user expertise: Some people have gone through basic education, use an old NMT phone, but know the whole phone by heart (i.e., they are experts); others have visited tertiary education institutions, own a 3G model, and have no clue how it works (i.e., they are novices). We can therefore say that the latter individual has enjoyed higher education, owns a more
sophisticated phone, and displays *inferior* user skills (the words in italics emphasize the ordinal character of the attributes).

Yet more orderly types of scales are those where each successive value is equally distant from the previous one. These are called *interval* scales, and if the scale has an absolute and logically valid zero point, *proportional* scales. Temperature or light are intuitive examples for interval scales. We can say that 30 degrees Celsius is 10 degrees warmer than 20 degrees the same number of degrees colder than 40 degrees Celsius. It is however irrational to say that it was twice as hot on a day with 40 degrees Celsius compared to a day with 20 degrees Celsius. It is also senseless for someone with an IQ of 120 to argue that he is twice as smart as someone with an IQ of 60 (indeed one should subtract at least 40 IQ points for such a statement, but increase it by the same amount if someone with an IQ of 60 argues analogically).

When data is represented proportionally, however, such ratio inferences are valid. Someone that owns four mobile phones, is 20 years of age, and has no children possesses not only two phones more than someone with two mobile phones, two children, and 40 years of age. The former, indeed, owns twice as many mobile phones, is half the latter’s age, and has infinitely less kids (you get the point).

Scale types and data levels are not always as intuitive as they may appear here. It is nevertheless absolutely essential that you are well aware of the kind of scale level you accept or assume for each the attributes you measure. This information essentially affects statistical analysis of the data, because every chosen procedure incorporates a series of (mathematically-related) assumptions that is based on the data level premise.

Be also aware that different types of theoretical bases and research interests can change the scale level for one and the set of data. Whereas employment status or work
title labels may suffice to be interpreted as being of nominal nature when investigating
the humor displayed by the individuals, the same values are of clear ordinal nature when
you investigate salaries, prestige, etc. (in any case, be careful about publishing a
discovered negative correlation in the former case).

Errors and quality of measurement: The classic test theory

Ok, now it starts to get progressively trickier. We have said that measuring
means representing real world events in a model world. This representation is not only
different from the real world due to its being part of a model, there are also other factors
why it deviates from the “original”, which we intend measuring. Being part of human
and social sciences, psychology is not an exact science. Hence, there are always
different kinds of uncertainties involved in user psychological measurement. These
pertain partly to the inaccuracies of the measuring instruments (including the
researcher), and partly to the object of measurement (i.e., the user).

As a consequence of this, the classic test theory states that every measurement
(i.e., datum [D]) is a composition of a true measure (T) and an error term (E).

\[ D = T + E \]

The error term again is composed of a systematic error, e.g., the systematic
flaws of an instrument we use in measurement, and a random error (e.g., human
imperfection). The smaller therefore the error term the better we are off.

The relative contribution of T and E to D is captured by the concept of
reliability. Reliability of a measure or measurement instrument expresses the degree of
accuracy of the datum, i.e., to which degree the measured value is representative for what has been measured and not for how it has been measured. Hence, reliability is inversely related to the size of the error term.

The other famous term in this context is validity. Validity describes the degree to which our datum is not only representative for what has taken place, but also for what we intended to measure. As is easily inferred from the previous sentence, validity is dependent on reliability, but not the other way around. If a measure is completely flawed, e.g., the background noise on the tape recording is so strong that we have difficulties to decipher the original words from the interview of a participant, we can hardly expect that we can truly find out more about what we were trying to investigate: our transcription of the interview will be unreliable and any interpretation of the transcribed text largely invalid.

However, if we are able to understand and transcribe all what has been said with high accuracy, but do not realize that the negative emotions the participants is talking about are reactions to the fact that he was obliged to participate in the experiment as part of a university course, and not specific reactions to the IT device we confronted the participant with, then we have a validity problem. I.e., the emotions are true (reliable measure), but not the type of emotion we intended to measure.

Validity as well as reliability has many faces and a series of logical and procedural tests can be applied to argue in favor or against the quality of a particular measurement. For us it is important that these are the two core criteria with which we can judge the quality of our research measures. And it is important to realize that measurement accuracy alone is not a sufficient basis for the assertion that we have measured something senseful. Our measurements can be completely reliable and still
have little validity. Hence a more complete version of Formula 1 is depicted in Formula 2.

\[
D = T_V + T_I + E_S + E_R
\]

Each datum consists of a valid component of the true measurement \((T_V)\) and an invalid component \((T_I)\), as well as of systematic error \((E_S)\) and a random error \((E_R)\). \(E_S\) and \(E_R\) affect reliability. In addition to this, validity is also affected by \(T_I\).

Now, from what has been said up to now, we might conclude the worst with regard to the quality of our research with human beings. Can we at all make any statements about the real world based on our measurements? Yes, of course. The reason for this is simply that we usually have more than one participant that we examine, and the fact that there are a series of normative assumptions that we can use to enhance our measurements.

For once, we believe to have usually quite a good idea about the systematic error included in our measurement, so that we can account for it or at least discuss it. Further, classic test theory comes to help with another axiom that states that errors are overall distributed in such a way that individual measurements are with equal probability either too large or as too small. This means that errors are distributed symmetrically around zero. The type of distribution assumed is the one typical for the classic test theory, namely the bell-shaped curve (see Figure 4).
This is a very important assumption, and has wide-ranging consequences on data processing and statistical analysis as we will later learn more about. One of the most important effects is that the arithmetical mean \( M \) (i.e., average) of our measurements across a large number of participants is equal to the actual mean in of the measured event in the real world. This is because the mean of the errors included in our measures is equal to zero. The only way in which aggregated representation of our measures differ from those of the object of measurement is due to it having a greater variability \( V \).

**Formula 3:** \[ M_D = M_T \]

**Formula 4:** \[ V_D = V_T + V_E \]
Population and Sample

“What you want is not what you get, and what you get is not what you want.”

This latest axiom of the classic test theory did not sound too bad, did it? If we just take averages of our measurement, we do not have to bother about unknown errors. However, just when we seemingly solved one problem we run into a next one. For instance, if we want to know how many mistakes people make when working with a certain interface, we could assume that an average quantity gives us a good and reliable measure, because it is supposedly free of random errors. Nevertheless, this assumption is absolutely consistent with the principles of the classic test theory only when measuring infinitely large number of people. For reasons of practicality we can say the whole research population. As we all know, however, we scarcely measure more than a few dozens, maybe hundreds, and sometimes thousands of participants. This means our findings are based on measurements on population samples, or, “what we want is not what we can get”. This fact has certain implications, which are shortly discussed in this section.

First, what is my population? The research population comprises all potential measurement units or events that display a certain characteristic: e.g., all users of broadband internet connection, or all occurrences of user frustration with using MS Windows. Obviously, as researchers we live in a world of limited time and financial resources and we can not really set out to measure every single instance where our event of interests occurs. We will have to do with a sample.

A sample therefore comprises all theoretically desirable and, within economic reason, accessible measurement units or events needed to fulfill basic statistical
requirements (e.g., all subscribers to the local cable internet service provider, kanetti.fi). The questions then remain, how shall we draw samples and how good of an approximation of the population is our particular sample in the end? The latter part is important because in research we usually want to make statements about affairs in the population and not only about the people in our sample, which are part of our model world. Hence, “what we get is not readily what we want”.

There is a wide range of sampling techniques and indeed, it there is a whole philosophy of its own behind it. Here, I will make a distinction only between three different groups of sample, or sampling techniques: *random sample, judgment sample, convenient sample.*

The fully random sample is usually the ideal small version of the population, because – some size issues taken into account – it behaves in almost identical manner as its big sister. In other words, it is statistically the best approximation of the population we can get.

In random samples, measurement units and events are chosen completely randomly (“surprise surprise”) with a known probability. Choosing randomly is per se not difficult, but to get the entire population as the pool where to draw from is usually already beyond our possibility. Further, to get all the chosen people to respond to our investigation request is another difficult nut to crack.

Hence, we usually settle for one of the other two sampling techniques. In judgment samples, the measurement units or events are chosen according to the theoretically-based judgment of someone who is familiar with the relevant characteristics in the population. The key issue is representativeness of the people in the sample for the people in the population (the attentive reader will have noticed that this
same theme of how representative the research model of the “real” world of affairs emerges over again throughout this reader; compare also Figure 3). Thus, we might decide that for some research question it is enough to study only this lot of people, because all others will most probably behave in similar ways.

On the other hand, we just might want to be careful that all types of users of users, based to some criteria (e.g., age, gender, use history), are represented in the draw of our sample. An example for this are stratified samples where we explicitly, for instance, select X number of users of the age below 18, Y number of user of the age class 19-25, Z number of users of the age class 25-35, etc. Doing so we base our sampling technique on clearly reportable considerations, i.e., (pre-)judgments.

The final sample type I mention here is the convenient sample, which is, as the name says, the most convenient and therefore also rather popular one. As in judgment samples, in convenient samples measurement units or events have unequal probabilities to be selected. Different from the judgment sample, however, these differences are not really based on theoretical considerations, but usually occur simply for research economy reasons. Probably the internationally and historically best studied convenient sample in the fields related to human sciences (e.g., Human-Computer Interaction) is the various teaching institutions’ psychology students. Students, and especially psychology students, are usually easy prey and are examined in relation to a variety of research projects.

Hence, participants in convenient samples simply happen to be reasonably suitable and easily accessible for a particular research project. This is not to say that convenient samples can not in some ways also be judgment samples, where the
researcher implicitly or explicitly argues that the research findings would be largely identical regardless which sampling technique is chosen.

Let us now return to the question of adequacy. Herefore we must remember that whatever our sampling technique and final sample composition may be, the bottom line is that our data will differ in some way from the data we would have obtained when measuring the whole population. This is as true for the individual level, i.e., running a

![Histograms showing population distribution and samples](image)

**Figure 5: Approximation of the population measures with numerous samples**
test with Anna does most probably not yield the same result as when running the test with Hanna, as it is true for aggregated data.

Luckily, however, samples often tend not to be very bad approximations of populations, if we steer around certain problems of sampling biases. In Figure 5 this belief is visualized in the way that samples and population have roughly the same forms of data distribution and, if large enough, the samples start to represent the population data rather well.

Now, what we need next are some instruments or criteria with which we are able to compare different data. This is provided in the next section.

*Description of univariate measurements: seeking the normal distribution*

“Description of univariate measurements” sounds maybe rather intimidating, but it means nothing else than what we have talked about all the time so far. As said near the beginning of the reader, measuring works in the way that we decide upon a characteristic or attribute, and on which dimension we assign to individuals or events a certain value. Naturally, we are usually interested in more than just one attribute but, for sake of simplicity we shall start with describing what we found out about people in our sample with respect to one characteristic only. This means, we are interested in a single variable, hence univariate statistics.

After we have for each person in our sample assigned an individual value on the chosen characteristic, it seems sound that we set out to see whether several people have the same value and which value is most common and so forth. This means we make counts and create a frequency table: Value X so many times, value Y so many times,
value Z no one, etc. This can be done no matter what scale level we have, normal, ordinal, or interval.

The charts in Figure 5 display nothing else than such frequency distribution, and there is now a range of distribution parameters that help us to further characterize the data distribution:

- **Basic distribution**
  - Counts/Frequencies

- **Central tendency**
  - Mean (M)
  - Modal value
  - Median (Md)

- **Dispersion**
  - Variance
  - Standard Deviation (SD)
  - Percentiles
  - Range

- **Normality**
  - Skewness
  - Kurtosis

Counts and Frequencies tell us how many measurement units were assigned a certain value, this yields the distribution chart. *Central tendency* parameters tell us something in the direction of which values were more popular or significant than others: The average $M$, is the arithmetic mean of all measurement points; the modal value is the
value that was measured most frequently; and the median Md is the value that is
surpassed by exactly 50 percent of the measured units (i.e., the other 50 percent were
assigned a value smaller than the median). Of this group actually only the modal value
is of any use for data coded at nominal level. Medians can also be used for ordinal
scales; means are reserved for interval scales.

Obviously all three central tendency parameters have their own distinct value.
The very popular mean M, for instance, gives quite a good idea about the core value in
the case of an ideal distribution as the one depicted in Figure 4 (we will talk more about
this distribution type below). However, M is very sensitive to outliers and it tells us
little about the case where more than one core value or value group has been popular. If
we take the cases displayed in Figure 6 below it is obvious that all four of these
measurement sample examples display the same average ($M = 3$), but in fact, the data
speak quite a different language if examined at face value.

How often do you become frustrated when using Windows?

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>Sample 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) 1000005</td>
<td>a) 1000005</td>
<td>a) 1000005</td>
<td>a) 1000005</td>
</tr>
<tr>
<td>b) 1000005</td>
<td>b) 1000005</td>
<td>b) 1000005</td>
<td>b) 1000005</td>
</tr>
<tr>
<td>c) 1000005</td>
<td>c) 1000005</td>
<td>c) 1000005</td>
<td>c) 1000005</td>
</tr>
<tr>
<td>d) 1000005</td>
<td>d) 1000005</td>
<td>d) 1000005</td>
<td>d) 1000005</td>
</tr>
<tr>
<td>e) 1000005</td>
<td>e) 1000005</td>
<td>e) 1000005</td>
<td>e) 1000005</td>
</tr>
<tr>
<td>f) 1000005</td>
<td>f) 1000005</td>
<td>f) 1000005</td>
<td>f) 1000005</td>
</tr>
</tbody>
</table>

1=never
5=constantly

Figure 6: Variety of distribution with equal mean M but different message.
 Whereas, in example 1, we have a very homogeneous sample, it appears that in example 2 there are two very distinct groups of users, one that gets never frustrated and the other being constantly frustrated. The former distribution is an extreme example of unimodal data set (data with a single peek), whereas the latter is an analog example for the bimodal distribution. Sample 3 suggests that we should maybe investigate more closely the situation of user \( d \), since he or she deviates clearly from the rest of the sample (i.e., outlier). And finally, sample 4 suggests yet another situation, namely that frustration tendency is evenly distributed, which may hint at some other variable that is correlated with the leaning to get frustrated.

 For the same reasons it is also very important to carefully consider the additional distribution parameters, and to employ a visual examination of the data. Dispersion parameters tell us then something about the variability of the data points (i.e., homogeneous vs. heterogeneous). The variance and its square root, the standard deviation SD, tell us in average, how far away from the mean value (M) the values of the other observed units are. The variance is the main dispersion parameter and it is obviously of relevance for interval and proportional scaled data only.

 Percentiles are value ranges between which always 10 percent of the observations fall. The second percentile marker tells us, for instance, that 20 percent of the measurements scored below this point and 80 percent above it. Hence, the median (MD), mentioned earlier, is indeed identical to the mark of the 5\(^{th}\) percentile, because below it are 50 percent of the data and above also 50 percent. The range is nothing else than the area within which observations were made. Sample 1 in Figure 6 displays a range of 1, with only one value assigned to each user; Sample 3 has a range of 4 (values 2 to 5); and Samples 2 and 4 have a full range of 5 (values 1 to 5).
Finally, there are the distribution parameters that check for normality of the data dispersion: skewness and kurtosis. In order to better understand their essence we should know first what is meant by normal distribution. The normal distribution is the holy grail of measurement. The bell-shaped curve in Figure 4 already illustrated us this type of distribution: it is unimodal, symmetric around its mean, it has two tails that converge asymptotically to zero when values progress to $-\infty$ and $+\infty$, and the area sum under its curve is equal to 1 (see Figure 7). I will not comment on these facts in more detailed fashion, so just “swallow” them.

![Figure 7: The standard normal distribution](image)

The normal distribution depicted in Figures 4 and 7 is not just any type of normal distribution, it is called standard normal distribution (also z-distribution) because it possesses a distinct mean (M = 0) and standard deviation (SD = 1). Otherwise it is equivalent to all other normal distributions and all values represented in the latter can easily be transformed into referring z-values of the standard normal distribution (see Formula 5).
Normal distributions have also some other neat properties. For instance, we know in advance how many measurements lie between certain values, not only that 50 percent will fall below the mean $M$. Figure 8 shows this fact for a normal distribution with mean $M$ and standard deviations $SD$.

Formula 5:  

$$V_z = (V_N - M_N) / SD_N$$

$V_z$: Value in z-standard distribution (i.e., z-value)

$V_N$: Value in some other normal distribution

$M_N, SD_N$: Mean and standard deviation of the normal distribution

Figure 8: Value observation probabilities for standard deviation intervals

If our data is normally distributed then something over two-third (68.2%) of all our measured units will display values that lay one standard deviation or less away from the average. About 95% will fall into the interval defined by two standard deviations off
the mean, and nearly all observations (more exactly 99.7%) will fall into the interval
defined by three standard deviation units off the mean. The rest (i.e., 0.3%) will lie
outside of this interval. This is handy to know, and in fact I suggest to every one to bang
these figures into one’s head.

As scientists we usually have a rather firm belief that if we could measure
infinite numbers of participants or events, the distribution would end up looking like a
normal distribution. Every true normal distribution can be sufficiently described by its
mean $M$, and its variance $s^2$ (or standard deviation $SD$). However, as we could expect,
in reality our data sets will not readily produce true normal distribution. Instead they
will look something like the examples in Figure 9.

![Figure 9: Examples of different distributions](image)

Curve (a) is probably as good as it gets in terms of attaining a normal
distribution. In contrast, curve (b) is too flat, curve (c) too peaked, curve (d) has the
problem of having more than one peak in addition to being too flat, curve (e) is
asymmetrically leaning towards to left (i.e., too fat on the right side and too steep on the left), curve (f) has significant bumps on its tails (i.e., outliers), and curve (g) is in contrast to curve (e) asymmetrically leaning to the right.

The problems with curves (e) and (g) are easily detected by testing were normal skew, which should be zero (i.e., fully symmetric) but is negative in the former and positive in the latter case. The issues with curves (b) and (c) are on the other hand a case of their kurtosis. Too small kurtosis (by definition, a kurtosis value below 3) tells us that the curve is too flat, too high kurtosis hints at a curve which is too thin and overtly peaked. To be precise, kurtosis is not so much a test of flatness vs. peakedness, as it is a measure for the length (or weight) of its tails. Peaked curves usually have a tail that merges only very slowly towards zero, which means that we have unusually many measurements that are extremely high and/or low (i.e., outlier-problem). In contrast, flat curves display too short tails. Skewed curves are as a consequence of this usually a combination of a too long tail on one side and a too short tail on the other.

So why should we be worried by all of this? The answer to this is very simple. Every attempt to measure reality, e.g., in an experiment, is achieved by employing a standardized method that assesses the behavior in a sample of the actual population. Usually these measures are of interest to us because we can compare them to the same peoples’ behavior prior or later in their development, to different peoples’ behavior, or because we can relate the measured attribute to other attributes. Whatever we do, our analysis will be based on statistical norms, and usually it is very essential that we can assume that our data set behaves in the same way as a normal distribution, because it enables us to run a great variety of statistical tests. In any case, the decisions about the
normality of our measurements will be crucial in deciding which analytical method we should use.

Statistical packages such as SPSS offer therefore standard procedures by which we can test whether normal distribution can be assumed for a particular empirical data set and also allow us to automatically “modify” our data in such ways that it will fulfill the requirements.

*Standard error of the mean*

Before moving ahead to the discussion of actual methods of research, i.e., ways to measure, we want to look at one of the key concepts in measurement: the *standard error of the mean* (SEM or sometimes just SE). Herefore we need to remind ourselves of the fact that samples are and remain only approximations of the population. No matter how big your sample is, as long as it is smaller than the population your findings will differ from the reality you set out to investigate.

Probably there are a lot of ways in which your data misrepresent the actual state of affairs (indeed it will differ with regard to all distribution parameters discussed in the previous section), but one which is particularly significant, is the deviation of your sample’s mean from the population’s average value.

Let us imagine we have a human behavioral characteristic, whose values in population X are distributed normally around a mean $\mu$, with a variance $\sigma$ ($\mu$ and $\sigma$ are the population parameters equivalent to M and s used for samples). A particular data distribution that we attain from measuring a sample of population members will, so we can readily assume, not have exactly the same mean as the one in the population (i.e., $\mu \neq M$). So, if I go out to argue that men have in average a shoe size of 46, this statement
will depend heavily on whether, by unfortunate coincidence, I measured a group of hobby basketball players or not. But even, if the case is not so obvious, there will be some inaccuracy in my statement, because I have not asked all men around the globe. Figure 10 shows this general idea for three samples and their means in comparison to the mean in the according population.

![Figure 10: Sample and population means](image)

So in effect, if I had time and money to test a huge number of samples I will frequently overestimate the actual mean, frequently underestimate it, and frequently hit the nail right on its head. Mathematicians tell us now that, in the case of drawing an infinite number of samples (of reasonable size [> 30 measurement units]), the means of the samples will themselves be normally distributed around the actual population mean (see Figure 10). And just to be consequent, the same will be the case for the standard
deviations. They too will distribute normally around the actual standard deviation that could be obtained from measuring the whole population.

Obviously these two normal distributions (i.e., the one around the population mean, and the one around the population’s standard deviation) will not only have a mean value (i.e., the population mean and the population standard deviation), but each also a standard deviation. The one for the distribution of the samples’ standard deviations is called the standard error of the standard deviation (SESD), and the one for the distribution of the samples’ mean values is called – you guessed it – the standard error of the mean (SEM) (Formula 6 and 7 give the mathematical calculation for these standard deviations).

Formula 6: \[ \text{SEM} = \frac{\sigma}{\sqrt{n}} \]

Formula 7: \[ \text{SESD} = \frac{\sigma}{\sqrt{2n}} \]

n: Sample size

Now we know it, but what do we do with it? Well, the SEM is a very important value because it tells us something about the accuracy of our measure. We are all easily familiar with illustrations of some group’s average behavior as displayed in Figure 11. The thin T-shaped lines around the top of the bar indicate to us how trustworthy the measurement is. It tells us that, in average, we might just as well have gotten a sample from which our estimation of M would have been M plus SEM, or M minus SEM.
In principle, we have two ways to trim down the SEM, one is to have a measuring method that is as exact as possible (i.e., reliable instrument and representative sample); the other is to work with large samples (compare Formula 6 and 7). But just as a reminder, small SEM alone does not immediately make your measurement more valuable – there is always also the validity issue (see section on quality of measurement).

Exemplifying the information so far

The following example shall get a pre-taste of the principles underlying basic statistical analysis.
The question is, what else can we do with the SEM? Well, for this we need to remember what we know about normal distributions, namely how many observations tend to be within a certain range of the data dispersion (e.g., one or two standard deviations off the mean; see Figure 8). If we for instance know what should be the observed average in the population, we can now measure whether the calculated mean in our sample is moderately off this mark or quite significantly.

For the sake of an example we may assume that a device is acceptable when, in average, users do not make more than 3 minor mistakes while using it for the first 10 minutes, and the rest of the users’ number of faulty operations is normally distributed around this mean of 3 errors with a standard deviation of SD = 2. This, let us assume, has been derived from long term user research with products that were well adopted in the community (attention, the example is totally fictive). After inviting 36 users to our test lab we find out that they made in average 5 mistakes. So, what is our conclusion based on this?

In order to draw any conclusion, we need to have more precise question at hand. What we actually want to know is whether the sample participants that we tested in our lab, are representative for the population of all those users that positively adopt the use of the tested new device because it encourages not more than 3 errors in average (this is a little more complicated question than the one whether the device will be accepted or not)? If not, then we might have tested participants that belong to a different population, namely those that actually have been exposed to a device that encouraged significantly more errors, causing them to reject the device. Hence, our participants would then not belong to the population of positive adopters and the new device would be substantially different (i.e., worse) from the one we intended to produce.
Taking Formula 6, we can calculate the SEM ($\frac{2}{6} = \frac{1}{3} = 0.333$). We now know that, if our participants are representative for the population we intended to measure (i.e., $M = \mu$), our samples’ mean values should be distributed normally around the population mean ($\mu = 3$) with a standard deviation of (SEM =) 0.333. Indeed, we measured however an empirical mean of $M = 5$. So the question is what’s the chance to get such a sample parameter?

Also from earlier discussions we know the probability for certain values to be observed within a normal distribution (see Figure 8). For instance, more than two-third of all values fall within the range of two standard deviations around the mean. In the case of the z-distribution this means a value between -1 and +1. So, let us calculate the z-value for our empirical mean ($M = 5$) based on the assumption that it was drawn by coincidence and therefore is part of the normal distribution with $M = 3$ and SD = .333. Applying Formula 5 we get a z-value of about 6.

Figure 8 tells us that the chance to get a z-value above 3 is 1.5‰ (i.e., half of 3‰). Our value is actually 6 and therefore the probability will even be much smaller, let us say 0.5‰. Whatever it is exactly, everybody will be able to agree that it is rather small. Because it means that if we would have tested 2000 different samples of users (each comprising 36 users) that make potentially an average of 3 mistakes only, we would have not more than once gotten such a large average of errors. Hence, we can safely conclude that our sample is unlikely to be representative for this population; rather we have drawn our sample from a population of user-device interactions that encourage more than 3 errors. Therefore the device we tested will most likely find little acceptance within the community, because it is genuinely more error-prone than devices that do get accepted.
This example might not have been very easy to follow. However, this is less an issue of the example or the explanation provided, and more an issue of the true complexity of measuring and decision making. We will therefore return to these concerns in later.

Research Methods

Quantitative and qualitative research

So far we have discussed how data represents real world affairs. The question therefore remains: How do we get to data? In this we refer to strategies and techniques of gathering data, in short research methods.

There is of course a wide variety of research methods and they can be categorized in different ways. The distinction between descriptive, correlative, and experimental (i.e., inferential) approaches has been mentioned at the beginning. There is also the well-known distinction between qualitative and quantitative research, about which I shall make a few very general remarks in this section.

The notion of qualitative research has in recent years become a very fashionable, and the actual differences between what is called quantitative and what is understood by qualitative research have been at times exaggerated, often neglected, and frequently simply misunderstood. Just because it contains numbers, does not make your research quantitative. And just because one claims to do a qualitative study does not automatically increase the quality of the research. Indeed, there does not even need to be such a huge difference between the two empirical territories.
In principle, every research starts out qualitatively and is qualitative up to some stage; just as well as all qualitative data can eventually be quantified. The real difference lies with the development and the role of the research model in investigating the “real” world (see Figure 3). As discussed earlier, researchers are forced to distort (i.e., simplify) reality in the course of their research. This means that they observe what is going on through the lens of some model, and an important question is how invasive or how dominating this model is during research.

Quantitative research favors very strong, model-driven research, where the model is mostly defined a priori. In qualitative research, on the other hand, we usually attempt to capture more authentic details about the actual affairs that are being examined. In doing so we are trading in pure mass of measurement units and sometimes also representativeness of the data against the attainment of a more comprehensive data set and format. This I would call a more “real world”-driven approach in contrast to model-driven research.

Qualitative data are usually of a rawer format and are not immediately accommodated to some pre-defined model. In this way, they can be fed into a mental incubator and a more generic, self-sustaining model construction or theory development is enabled. That does not mean that it will actually take place. Quite often, the implicit theoretical assumptions are so strong that they will govern data interpretation no matter whether the original approach was more qualitative or directly quantitative. However, qualitative researchers have by tradition been more reflective concerning their own role in research. The concept of research as a subjective endeavor, not objective, has been embraced to a much greater degree compared to conventional quantitative researchers. This distinction, I believe, does however not so much be one between qualitative and
Quantitative research per se, and much more one between communities of researchers and their methodological doctrines.

Qualitative research usually also attempts to get the pig picture about some situation, and not a great number of microscopic accounts of some specific aspect of behavior, as in quantitative research. In this way, qualitative research is by nature also more holistic in its research ambition. Synthesis of research findings is hereby often regarded as more vital than analysis. Again, however, it would be unfair to conclude that quantitative research is not aimed at understanding meaningful wholes. Indeed, I believe, that both approaches need each other desperately in order to cross-validate their findings and data interpretations.

The two other key issues in the comparison of quantitative and qualitative approaches lie therefore with the observational model and the coding model. In a qualitative investigation we might collect extensive data about the individual user and his or her interactions with a device. In describing the individual we may want to stay very loyal to the actual personal characteristics the user displays. For instance, in a firm we may describe an employee on managerial level in very elaborate fashion by the way he or she leads, communicates, organizes the work environment etc., and relate this information to the interaction we recorded. Using a quantitative method, on the other hand, we might just assign the code “3” for “Employee on managerial level”, nothing more.

Quantitative methods are usually more coding-laden and coding takes place much earlier in the research process. And in coding - this is very important to notice - we always lose information. The codes “3”, for instance, suggest that all those who have
been assigned this code are equal with regard to the issues concerning employment and managerial qualities. Qualitative data may easily proof otherwise.

Nevertheless, it is easily to conceive that we could apply a much more fine-grained coding system that codes most of the dimensions that we have described in our qualitative data. Namely, in addition to the code “3” for “Employee on managerial level”, we assign a code “15” for “Democratic type of leadership”, a code “2” for “Poor communication skills”, etc. In this way we manage to get a step closer to the qualitative data, however, we are again this one important inch away from qualitative data because we still need to interpret our observations (based on our research model) in order to assign them a value out of a finite number of value alternatives on an according attribute dimension.

Even before issues of coding arise, there is an important difference with regard to the manner of observation in qualitative research. Instead of being essentially model-driven, qualitative observation is to a much greater extent guided by the object and dynamics of the observed events. There is usually a less strict agenda for what is measured, and in which way. Instead, actual circumstances and already collected information continuously influence the proceeding inquiry. It is however important to realize that this distinction too, is not as absolute and there is rather a continuum of method-related differences. There is no qualitative investigation that is free of method and theoretical assumptions, just as there is no quantitative research that is totally detached from actual observation of the “real” world affairs.
Reactive and non-reactive research strategies

Yet another fundamental way to distinguish among methods is with regard to the degree of *reactivity* vs. *non-reactivity* of the research technique. Figure 12 displays a collection of eight important research strategies and orders them according to the degree of reactivity and the degree of universality vs. context-dependency of the aimed at research findings.

![Figure 12: Research strategies](adapted from Stroebe, Hewstone, Codol, & Stephenson, 1992; see also Runkel & McGrath, 1972)
Reactivity of the research method refers to the degree to which the observed behavior of the participant is a reaction to some stimuli that was purposefully induced by the researcher. Classic experiments (i.e., laboratory experiments) are an ideal example of a reactive research method. When confronting a participant with a slightly modified IT-device and observe his or her interaction with the technology, we are explicitly interested in reactions to the selected and prepared device. On the other hand, wherever there is observation involved, the behavior of a participant will never be completely free of influences by the researcher, because the researcher can not make him- or herself completely invisible. This is essentially also true for questionnaires. Not only are the question contents and form themselves special types of stimuli to which we like the participant to react (e.g., different kinds formulations of the same question content can yield very different results), but also the context of the questionnaire, e.g., the person that interviews, may have a critical influence on the answers. Judgment tasks are usually special kinds of questions, such with a higher degree of desired control over the way the question is presented and a more precise problem-focus. Hence, the link between research stimulation and a participant’s reaction is intended to be stronger.

Formal theory and computer simulations are in this sense actually non-empirical methods because the do not involve observation of research participants. In using formal theory a researcher attempts to construct a symbolic web of theory-based postulations and tries to deduce logical consequences from it. Hence, formal theory involves an analysis of the model world, and only indirectly of the real world (compare Figure 3). It is in essence a researcher’s mental simulation of relations and events.

Computer simulations are obviously very similar. The only difference is that, here, the model is instantiated in a computer program, which can be fed with
information and whose output can be compared to theoretical predictions or empirical data collected from human research. Computer simulations, although still in use, had most relevance in HCI research in the 70s and 80s, when the model of the human information processor (Card, Moran, & Newell, 1983) and also artificial intelligence research were particularly en vogue. Models like Card et al.‘s GOMS, and other cognitive architectural derivates set out to simulate human behavior. Cognitive walkthrough procedures (Lewis & Wharton, 1997), for instance, are on the other hand very applied forms of the formal theory approach, i.e., thought experiments where “experts” instead of actual users cognitively go through every step of an interaction and try to imagine what will occur, and what the outcome will lead to next, etc.

Finally, there are those two strategies that are most keen to preserve as much authentic contextual information as possible: field study, field experiment and experimental simulation.

A field study usually concerns a systematic observation of a phenomenon of interest in its native context, i.e., real-life settings: e.g., the actual use of SMS-messaging in classrooms. In the case where the researcher introduces a relevant and purposeful change to the natural situation, we speak of a field experiment. As experimental simulations, we label those types of observations that do not take place in a coincidental natural situation, but rather employ a well-controlled imitation of well-chosen real-life settings. The practice drills that are used for educational purposes with medical and rescue personal (e.g., authentic-looking simulation of a road accident), fire fighters (e.g., fire houses) and soldiers (e.g., combat training in the some military zone) are good examples for such kind of simulations. In very similar forms we can also construct experiments concerning issues of use. In the preceding examples, for instance,
we may be interested in the efficiency and effectiveness of the radar and search technology use by rescue troops, communication technology employment by medical personal, or weaponry handling by soldiers.

*The observation*

Of the multitude of research methods we can of course discuss only a small selection within a reasonable frame of this reader. I find it particularly important to achieve some familiarity with three very classic techniques of data collection. These are the *observation*, the *interview and questionnaire*, and especially the *experiment*, which has been the most important empirical method in psychology for quite a while. This selection also tentatively reflects the basic distinction between descriptive, correlative, and inferential research.

In the previous section we have already learned about a particular label given to natural observations, i.e., field studies. Observation as an investigative practice, is however a skill that is of utmost relevance not only when doing field studies, but in other types of research as well, e.g., the interview or the experiment. For this we must recall that there are in principle not too many ways how we can find out something about humans: We can observe them doing something and we can ask them about what they did and why. And being precise, we can actually only observe them doing something, because even their answering to our question is just some observed behavior influenced by the specific context of our questioning. This remark may seem odd but it is important to consider ones in a while.

A well trained observer tries to be as unbiased and discreet as possible during his or her observation. What we are really interested in is what takes place, and not what we
implied or implicitly induces, or what we thought should have taken place. However, interpretation of the observation material is a natural consequence of measurement, and data processing in general. Indeed, unbiased (i.e., purely objective) observation is simply not possible, as we have noted in the beginning of this reader already. Hence, wherever the observer makes implications or “fills in the blanks”, this should be according to a well spelled-out theory or model, so that these data supplementations can be tracked and evaluated by other researchers.

There is a variety of names for research strategies that put the context-sensitive observation of human behavior into the center of their approach. No matter whether you associate these with anthropology (e.g., ethnography), ethology (i.e., the study of behavior, usually animals) or social psychology, they are concerned with one and the same thing: the comprehensive description of human functioning in different environments. In the last section we have already introduced one label that is most frequently used in social psychology, namely field study.

A field study can vary along various dimensions of research practice, e.g.:

- Systematic vs. non-systematic
- Participative vs. non-participative
- Informed vs. non-informed

In most cases the researcher has a clear idea and focus on a certain phenomenon of his or her interest. Observation can in this case be planned systematically, i.e., what shall I concentrate on, what can I ignore, in what form shall I record the data, etc. Frequently it is, however, necessary or even advisable to regress to the status of a very naïve observer. In such non-systematic types of observation a researchers sets out to
simply see what is happening, what kind of an appeal the observations make, and what
ideas are generated.

In field studies the researcher quite generally attempts to be as unobtrusive and
neutral as possible in order not to provoke any reactions that are not part of the natural
repertoire of behavior, but, instead, specific distortion caused by the investigation itself.
There are special cases where the researcher actually becomes part of the field, e.g., in
order to get in closer contact with the studied systems. This is then called a participative
observation. However, as soon as the researcher purposely induces some relevant
changes to the natural setting, we enter the domain of experimentation (i.e., field
experiment). Obviously, the two concepts are ranges on a continuum of possible field
investigations.

Finally, the researcher can choose to inform the observed subjects, or to disguise
the observation. The former carries the problem of affecting the findings, e.g., in an
informed observation of classroom SMS messaging, nobody might use the phone
anymore. The latter is subject to ethical issues because it involves the recording of
personal data and their use in research that has not been approved of by the concerned
individuals.

Apart from live observation of individual behavior, the researcher can of course
also refer to other sources as alternative or in addition. Content analysis is one such an
example where we investigate documents that are themselves already transcripts of
behavior, e.g., navigation logs from web pages, analysis of called phone call reports,
chat discussion archives, etc. The problems with this type of observation are that we
often lose information about the causal chain of events, as well as the mere bulk of data
that needs to be processed.
Interviews and questionnaires belong to the group of self-evaluation techniques. Here we generate information through a special instrument of natural situations, namely language: speech and dialogue. This also hints to a very basic problem of these approaches, namely language-based barriers and problems, such as misunderstandings.

After having defined our population, and selected our sample, most conceptual work usually goes into the development of the questions and their formulation. In this stage it is advisable to run many informal pilot tests with our current sets of questions in order to get as much feedback as possible, and to resolve ambiguities, for instance. We also need to consider that different user communities and users from different socio-economic backgrounds utilize different language and terminology. We have to come to a decision about the time-frame or chronological target of our questions. Do we chose a retrospective approach (e.g., “Why did you chose this mobile phone over the other?”), prospective approach (e.g., “What mobile phone would you chose if you were to buy a new one right now, and why?”) or moment-oriented approach (“You are choosing a new mobile phone at the moment. Can you tell me what goes through your mind?”). We will also have to decide whether we chose a more intimate, time-consuming technique of questioning (e.g., face-to-face interview), paper-and-pencil questionnaires sent by mail, or even web-based online questionnaire forms. And we need to settle for some type of questioning and answer options (see Figure 13).
Questions can be standardized in their formulation and location within the questionnaire and presented to all participants in exactly the same way. They can also be semi-standardized with only part of the question formulation being fixed or open location and the other part free for adjustment to individual requirements. Finally the questions can be fairly non-standardized, including a lot of improvisation on the part of the interviewer and each interview being different from the next one.

In terms of the forms of answer we allow for, it is usual to distinguish between closed and open questions. In closed questions individuals can usually choose from a list of presented answer alternatives, whereas in open questions they can formulate their own answers. These two types are often combined within a question, in that there are a set of fixed answer alternatives (i.e., multiple choice-type) and an option “other:”, or something equivalent to it.

The combinations of questioning forms as displayed in Figure 13 do naturally not cover all types of techniques used; they emphasize only a subset of rather common ones.

The direct-standardized questions-closed answers technique of questioning is the classic but rather expensive face-to-face interview. Its cheaper alternative is the
telephone interview. By use of written questionnaires mailed to individuals or available on the net one can reach even greater masses of people but the flip side of the coin is that they usually incorporate large difficulties related to motivating the individuals. Finally there is the direct-semi-standardized-open interview, also called narrative interview. When using this questioning technique the researcher not only interested in the actual answers but also in the individual’s own style and structure of the self-report or self-evaluation.

Interviews and especially questionnaires have become very widely used research techniques especially in combination with other research forms, such as the observation and the experiment. Questionnaires provide in rather economic manner valuable insights into aspects that are otherwise hidden from observation, because they are of purely mental nature of because they take place outside of the time-window of our investigation, i.e., earlier or later in the sequence of events. In return, observation data are usually essential in discovering distortions of questionnaire or interview data: You can ask a driver many times how he or she would react in a certain traffic situation, however, only the observation of the behavior I the actual situation will provide you with factual knowledge.

Hence, it is easily anticipated, that there is a whole bunch of problems involved in developing a good questionnaire and collecting data with it. Here are a few:

- Adequacy of language (e.g., interviewing children and elderly people)
- Ambiguity of expressions (e.g., what is “state-of-the art” technology)
- Order of questions (i.e., earlier answers provide a frame for the consideration of later questions)
• Positive and negative formulations (e.g., “I try not buy products from the USA or Asia” vs. “I try always to buy product from the EU-market”)

• Human tendencies to answer according to social desirability and majority views (e.g., “Would you steal something?”)

• Return rate and compliance (e.g., do those that returned the questionnaire about satisfaction with Volvo automobiles belong to the same population as those that did not return the questionnaire?)

• Interviewer skills (e.g., polite vs. arrogant)

• Context-dependency of answers that are conceptualized as de-contextualized (e.g., are answers about future confidence the same when asked in autumn as spring)

• Memory lapses and distortions (e.g., “Was there more snow when you were young, products better built, and people generally happier?”)

*Grounded Theory and Ethnography*

Observation and inquiry-type of investigation play a crucial role in such environments as information system research. Their key contribution is to generate theories, i.e., to process raw phenomena into scientific constructs and models of how these contracts are interrelated with each other. Two research notions have received much attention recently, and have indeed become rather fashionably. They are Ethnography, an investigative approach and set of techniques geared at discovery, and Grounded Theory, an analytical approach and set of techniques geared at distilling discoveries in order to reveal underlying regularities and systematicities. We will in the
forthcoming shortly characterize these two methods (text composed by Panagiotis Kampylis).

*Grounded Theory*

**What is meant by Grounded Theory Research?**

The term Grounded Theory (GT) refers to a theory that is grounded in the data and emerges inductively from it (Cohen, Manion & Morrison, 2000). Strauss and Corbin (1990) define Grounded Theory as “… a qualitative research method that uses a systematic set of procedures to develop and inductively derived grounded theory about a phenomenon” (bolds from the original).

The Board of Scientific Affairs of the American Psychological Association in its Task Force on Statistical Inference Initial Report (APA, 1996) points out “the need for theory-generating studies”. The centrepiece of GT research is the development or generation of a theory closely related to the context of the phenomenon being studied (Creswell, 1998). Generally speaking, theory does not go before research but follow it; Strauss and Corbin (1990) explicitly state that in a GT study the researcher first gathering and analyze data and after develop the theory.

Grounded Theory research goes beyond existent theories and preconceived conceptual frameworks in search of new understandings of social processes in natural settings. The basic idea beyond GT is that research that reveals the complexities of the real world should derive from theory generated from that world (Hutchinson, 2001).

Grounded Theory is introduced by the sociologists Barney Glaser and Anselm Strauss in their book *The Discovery of Grounded Theory* (Glaser & Strauss, 1967) but later on they disagreed on methodological and practical issues. Grounded Theory
according to Glaser should emphasize induction or emergence, and the researcher’s creativity within a clear frame of stages, while Strauss is more interested in validation criteria and a systematical approach (Wikipedia, 2006).

Creativity is a vital component of the GT. Grounded Theory is designed to allow the creative interpretation of data and the invention of theory. Its procedures drive the researcher to make hypotheses and to create new order out of the old. As Strauss & Corbin (1990) state: “Creativity manifest itself in the ability of the researcher to aptly name categories; and also to let the mind wander and make the free associations that are necessary for generating stimulating questions and for coming up with comparisons that led to discovery”. The creative interpretation and analysis of data develops a GT that is unique; it depends on the interaction between the researcher and the data and even with the same corpus of data, two different researchers would probably develop different theories.

In this point, I would like to stress that the generation of a new theory that grounded in and emerged from data, offers a new and -hopefully- creative perspective on a given situation. Afterwards, this theory could be tested and verified by other research methods, qualitative and/or quantitative. Qualitative research such as GT research should not be regarded as opposed or incompatible with quantitative methodologies. As Hutchinson (2001) asserts, qualitative research is a necessary and useful precursor to quantitative one; “both approaches need each other desperately in order to cross-validate their findings and data interpretations” (Helfenstein, 2005).

Grounded Theory research can be classified as applied research and offers a systematic method to study complex human actions, phenomena and structures such as education. The final “product”, the emerged theory, should have practical
implementation. In addition, the method can also be used in the evaluation of educational programs and policies (Hutchinson, 2001). I believe that especially in schools and classrooms which are very complex social environments, we need data-based theory that explains the “real world” within pupils, teachers, parents and administrators live and act. Grounded Theory research offers to teachers the freedom to explore specific aspects of the complex educational “puzzle”. According to Glaser and Strauss (1967) the practical application of GT requires developing a theory that contains four highly interrelated properties:

- **Fitness.** It should directly be induced from diverse data and fit the situation that researches.

- **Understanding.** It should be understandable and make sense both to the participants in the study and to those practicing in that area.

- **Generality.** It should be sufficiently general to be applicable to a variety of contexts interrelated to the substantive area.

- **Control.** It should offer its “user” enough control in everyday situations to make its application worth trying.

*Procedures and key concepts of Grounded Theory Research*

The GT method, especially the way Strauss develops it, consists of a set of stages or procedures whose cautious implementation secure a suitable theory as the outcome (Borgatti, 2006). Strauss and Corbin (1990) propose that Grounded Theory should be evaluated by the process by which it is constructed; can be evaluated only if its procedures are sufficiently explicit to the reader and he/she can judge their suitability.
**Theoretical sensitivity**

The term appears first in the title of a Glaser’s book which is published at 1978. It refers to the aptitude to distinguish what is important in data and to give it meaning; it is the researcher’s ability to perceive variables (categories, concepts and properties) and their relationships.

Theoretical sensitivity constitutes an important creative characteristic of GT because represents the researcher’s ability to use creatively his/her experience (personal and professional) and the literature. Theoretical sensitivity allows researcher to formulate theory that is faithful to the reality of the phenomena under study (Glaser, 1978 as cited in Strauss & Corbin, 1990).

Theoretical sensitivity has a number of sources (Strauss & Corbin, 1990).

1. The literature
2. Personal experience
3. Professional experience
4. Analytical process

**Data gathering and recording**

Data gathering starts as soon as the researcher has identified a researchable situation and goes for the first time into the field (Hutchinson, 2001). The researcher gathers, codes and analyzes simultaneously the data; it is an ongoing and spiral process during which the researcher can change focus. According to Creswell (1998), data gathering in a GT study is a zigzag process: out to the field to collect information,
analyze the data, back to the field to gather more information, analyze the data and so forth.

Interviews are commonly the main source of information in GT but not the only one (Dick, 2006). The researcher also collects and analyzes observations, documents and other “pieces of information” such as informal conversations, individual or group activities, recording and so on. After several visits to the field the researcher conducts 20-30 interviews in order to collect sufficient data to saturate the categories (Creswell, 1998).

**Coding the field notes**

Qualitative coding is an open-ended, creative, emergent, developmental and inductive procedure (Hitchcock & Hughes, 1995). Researcher creates categories through interpretation of the corpus of data. This procedure differs from quantitative coding which calls for preconceived, logically deduced codes into which the data are placed (image 1).

A category represents a unit of information composed of events, happenings and instances (Strauss & Corbin, 1990). The category that appears central to the study is
referred as *core category*. This category emerges with high frequency and it is
connected to many of the other categories. The core category may be more than one.

The process of data analysis in a GT study is a systematic procedure with the following steps (Creswell, 1998):

- **Open coding**: the researcher forms initial categories
- **Axial coding**: the researcher assembles the data in new ways using logic
diagram in which he/she identifies a central phenomenon
- **Selective coding**: the researcher identifies a “story line” and presents
hypotheses.
- **Conditional matrix**: the researcher develops a conditional matrix that clarifies
the social, historical, and economic conditions influencing the central
phenomenon.

*Constant comparative method*

There are several procedural tools for analysing qualitative data such as analytic
induction, constant comparison, typological analysis and enumeration. *Constant
comparison* is used widely in GT because it combines the elements of inductive
category coding with simultaneously comparing these with the other events and social
incidents that have been observed and coded over time and location. This enables social
phenomena to be compared across categories, giving rise to new dimensions, codes and
categories

Constant comparison can start from the beginning of data gathering, in search of
key topics and categories and can continue up to the writing process that is a rather
continuous process in Grounded Theory research. Through constant comparison, emerges the theory for the phenomenon that is researched (Bogdan & Biklen 1992).

Glaser and Strauss (1967) propose that the constant comparison method involves four stages:

1. Comparing incidents and data that are applicable to each category, comparing them with previous incidents in the same category and with other data that are in the same category.
2. Integrating these categories and their properties.
3. Bounding the theory.
4. Setting out the theory.

In constant comparison data are compared across a range of situations, times, groups of people, and through a range of methods. The process resonates with the methodological notion of triangulation namely the “testing one source of information against another to strip away alternative explanations and prove a hypothesis” (Woods, 1986).

Memoing

Memoing occurs in parallel with data gathering, analyzing and coding. Memo is a note about some hypothesis the researcher does about a category and mainly about connections between categories. The researcher through memoing records his/her ideas in order to capture the initially impression and shifting connections within the data quickly (Hutchinson, 2001). As Glaser and Strauss (1967) put it “… the second rule of the constant comparative method is: stop coding and record a memo on your ideas. This rule is designed to tap the initial freshness of the analyst's theoretical notions and to
relieve the conflict in his thoughts. In doing so, the analyst should take as much time as necessary to reflect and carry his thinking to its most logical (grounded in the data, not speculative) conclusions”. Memos also act as the starting point for extra coding of the field notes, and for returning to the field or library to accumulate more data.

**Theoretical sampling**

Glaser and Strauss (1967) define theoretical sampling as “the process of data collection for generating theory whereby the analyst jointly collects, codes, and analyzes his/her data and decides what data to collect next and where to find them, in order to develop his theory as it emerges”. Sampling decisions are made during the entire grounded theory research process. The researcher seeks appropriate data to fill in the evolving categories and interacts with the data in order to create directions for further sampling. The idea behind the sampling process is to maximize comparability (Hutchinson, 2001).

**Sorting**

When the researcher chooses the core category (or categories) he/she starts sorting and attempts to discover the relationship of the different levels of codes to the core category. An outline emerges from the sorted memos which are the basis for writing the theory. During sorting procedure, the researcher may illustrate and re-illustrate visual schemata such as diagrams, tables, charts and concept maps. These visual representations are especially useful in the development of the theory. In addition, during sorting new ideas can emerge which in turn are recorded through new memos.
Saturation

As the researcher notices similar instances over and over again, when all new data fit into one of the already formed categories, the researcher ultimately have a sense of closure. Glaser & Strauss (1967) used the term saturation for this feeling namely that no additional data are being found whereby the researcher can develop properties of the category. Hutchinson (2001) define saturation as “…the completeness of all levels of codes when no new conceptual information is available to indicate new codes or the expansion of existing ones”.

Review of literature

In a Grounded Theory study the researcher first develops or generates a theory based on corpus of data and then turns to the literature to find relevant studies or texts which may support, illuminate or extend the proposed theory. In many cases the Grounded Theory is supported by the literature but in other cases the proposed theory goes beyond the existing theories and contradicts with the literature. Connecting the emergent theory to existing literature enhances the internal validity but Dick (2006) makes an interesting note that the literature in the Grounded Theory has the same status as other data.

Reliability, Validity and ICT

In Grounded Theory, through constant comparison and coding, data are compared and contrasted many times. In addition, the multiple data collection methods (interviews, observations, documents…) increase the value of information. The reliability and validity augment when there are several observers and data collectors.
Information Technology artefacts can assist in the development or generation of grounded. Through IT artefacts the researcher can enhance:

- Reliability: by retrieving all the data on a given topic, thereby ensuring trustworthiness of the data
- Validity: by the management of samples

In addition, IT artefacts can assist in the generation of Grounded Theory through coding, constant comparison, linkages, memoing, use of diagrams, verification and, ultimately, theory building.

**Ethnography**

*Ethnography* is another qualitative research method used by social scientists to study human behaviour and it has its roots in *cultural anthropology*. In grounded theory the focus is on producing a theory grounded in the collected data; in ethnography the focus is to a set of incidents as a critical event that offers an opportunity to see “culture at work” (Creswell, 1998). Ethnography has a holistic character (based on the idea that a system's characteristics cannot be truly understood independently of each other) and aspires to give a detailed description of the relationship between all the characteristics of a single human group. But ethnographer must not stop at description; the basic goal of his/her research is the development of theory (Woods, 2001).

Ethnographer uses a variety of methods and techniques but *interviews* and *participant observation* being the most widely used. Ethnography research is used in many academic fields and not only in social sciences. An example of an ethnography research from the field of Computer Supported Cooperative Work is the study about

The ethnographer makes his/her research in the native environment to see people and their behaviour given all the real-world incentives and constraints (Fetterman, 1998). John Dewey, the pragmatic philosopher and educator, since the beginning of the twenty century declared that all inquiry arises out of actual, or qualitative, life. That is the environment in which humans are directly involved (Sherman & Webb, 1988). To study even a small fragment of the real world is in many ways more difficult than laboratory study. The extensive work in the “real-world”, in the field, is called fieldwork. It is the way most qualitative researchers collect data. The researcher goes to the subjects and spends time with them, in their environment (Bogdan & Biklen, 1992). As Creswell (1998) notes, in the field, the ethnographer observes what people do (behaviours), what they say (language) and what they made and use (artefacts).

Educational ethnography “examine the processes of teaching and learning; the intended and unintended consequences of observed interaction patterns; the relationships among such educational actors as parents, teachers, and learners; and the socio-cultural contexts within which nurturing, teaching, and learning occur” (Goetz, LeCompte, 1984).

According to Woods, (2001) educational ethnography can decrease the distance between theory and practice for the reason that is concerned with substantive issues that teachers recognize as their own, deals with their problems, points out their point of view, takes the implications of their actions in different situations into account and utilizes the concepts and language of school culture in drawing descriptions and spelling out theories.
Creativity is a vital component of the Ethnology as in any other method. As Woods (1986) phrases, “the ideal-typical circumstance in which ideas emerge is a mixture of, on the one hand, dedication to the task, scrupulous attention to detail and method, and knowledge, and, on the other, the ability to ‘let go’ of the hold of this rigorous application, to rise above it, as it were, and to ‘play’ with it, experimenting with new combinations and patterns”. Ethnography has to find the balance between “science” and “art” and only then will achieve its full potential.

The experiment

Let us now look a little closer at the experiment, which has probably been the most influential empirical method in psychology-related research. The experiment, especially the laboratory experiment, is frequently also called the royal way of investigation simply because it signifies the quality step from descriptive and correlative to inferential research, i.e., in using experiments we track down causal relationships between variables.

Much of the substantial gain in knowledge in all sciences has come from actively manipulating or interfering with the stream of events. In this sense, there obviously is more than just observation or measurement of a natural event. The key principles of experimental design and analysis are based on the very logic of causal inference. In experimental research, a selected experimental condition, i.e., a manipulative change (also called treatment) of some sort is introduced. This may be of many sorts, e.g., different kind of stimuli that are used on the same or different participants (e.g., two versions of a device), different kind of participants that are used
on the same stimuli (e.g., experts vs. novices), same stimuli and participants but
different contexts (e.g., unlimited time vs. rush).

Observations or measurements of selected participants’ behaviors are then later
analyzed in the light of being responses to these treatments. Because we usually have
more than one kind of treatment condition, or a treatment - next to a non-treatment
condition, specific effects of a particular manipulation are visible as differences between
conditions and treatment groups. It is easy understandable that a save attribution of any
measured behavioral effect to the induced manipulation is dependent on the
experimental conditions to differ with respect to the critical manipulation only. If we
confront young people with one device and old people with another, it is difficult to
draw conclusions as to the specific effects of the type of device on user interaction.

There are a few very critical issues when designing experiments, these are:

- Field, simulation, or laboratory experiment
- Experimental scenario and transparency
- Independent and dependent variables
- Design
- Control and balancing

*Field, simulation, or laboratory experiment?*

After having decided that we want to manipulate natural events and measure
effects of these manipulations, i.e., we have decided to run an experiment, we need to
decide whether we can “transplant” and recreate the behavior in an authentic way in our
laboratory environment. If we believe that there are too many factors of the natural
setting that influence the behavior we are interested in, we usually can not run a
laboratory experiment. E.g., it is not reasonable to investigate the organizational adoption of a new communication system as dependent on whether employees were involved in its selection or not in a laboratory context. We probably would need a field experiment for studying this, or, alternatively, a field study, if appropriate business examples are available. On the other hand, to investigate which kinds of telephone numbers people can remember easily in an emergency situation, we could go for a laboratory experiment by exposing participants first to cognitively and emotionally very demanding situations, but probably we would have to settle for an experimental simulation (see earlier section). However, just to see how many digits people can remember in correct order and groups to form telephone numbers, we are probably well off with a laboratory experiment concerning learning and memory issues.

For economic and practical reasons laboratory experiments have usually also a much stricter time frame. That means we invite a participant to the laboratory, run experiments for 30 minutes, one hour, or sometimes longer, and then we discharge him or her again. Field experiments can, and usually need to be run for much longer periods of time, because they are focused on slowly emerging continuous responses of participants, i.e., evolution of behavioral patterns.

*Experimental scenario and transparency*

In contrast to the field experiment where the whole idea is that all changes and treatment events are authentic and salient to the individuals or groups that we are observing - except maybe for the fact that they are part of an experiment - in laboratory experiments we usually want to disguise the real rationale of the investigation. The reason for this is simply that by taking part in the experiment, participants are already aware of being observed with regard to some behavior, which exerts by itself a certain
effect on behavior (i.e., the so called Hawthorne-effect; see origin with Roethlisberger & Dickson, 1939).

If we now even tell them what our concrete focus is, they will steer their full attention to our treatment and their responses and their behavior will most probably not anymore be of a kind that can be generalized to natural contexts outside the laboratory. However, this is exactly what we would like to do, i.e., we are not eager to present data about participants’ behaviors when operating some device, but we would like to talk about our data in terms of findings about how human beings act as users of the particular device.

In order to achieve a certain degree of “demand characteristics” (Orne, 1962, 1969) blindness with experiment participants, we usually use some experimental scenario, cover story, or minor deception of intention. Usually it is sufficient if we tell the truth or some truth about the experiment, but not the full truth. If, for instance, we investigate how pictorials on web-pages affect their judgment, we can say that we are interested on participants’ evaluation of different web-pages without saying what aspect we focus on. In some special cases, we need an actual cover story, where we disguise the real purpose of the experiment and create a kind of theater play. If for instance we want to investigate differences in users learning and emotional coping depending on whether they are being forced to use an obviously flawed program over a series of tasks compared to whether they can get a new bug-free program, we might introduce a manipulated raffle. By doing so we can disguise their selection to an experimental condition as a decision by Fortuna and do not need to explain the experimental idea openly. Naturally, this does not save us from the problem that individuals might respond differently to beliefs of destiny.
In very special cases, we even may need to consider whether it is necessary to run experiments in double-blind manner, i.e., where even the experimenter himself or herself does not exactly know what the true aim of the study is. Notably qualitative forms of observation can for instance easily be vulnerable to all kinds of behavioral artifacts and measurement distortions caused by beliefs and expectations on the side of the researcher: We usually like to see and hear what we hope to see and hear, and therefore findings are biased by our theoretical assumptions and hypotheses. Although these kinds of effects (summarized as Rosenthal-effect; see Rosenthal, 1966) do not usually necessitate drastic changes in the way we design experiments (as well as other instruments of investigation), it is important to be aware of them.

Independent and dependent variables

A variable refers to just about anything. There are two major kinds in every experiment. The variable that is manipulated, or changed, is known as the independent variable. The variable that is observed is called the dependent variable. Any variable that could have an effect on the dependent variable (our subjects' behavior), other than the independent variable (the stimulus or condition that we want to learn about), is known as an extraneous variable.

Now, variables are constructs in our research model, and, by themselves, have little application value in our experiment. For instance, what is meant by the effect of “presentation mode” on a mobile terminal on “user satisfaction”? Well, the independent variable is the presentation mode, the dependent the user’s satisfaction. However, as such, we can not measure the variables, we need to operationalize them. By operationalization we mean that we need to translate the essence of what the variable is
about into a concrete form of stimuli or behavioral responses that can be used, manipulated, observed, and thus measured in an experiment.

![Scene example: Events at a bathing beach](image)

<table>
<thead>
<tr>
<th>Degree of processing</th>
<th>Information to be mediated</th>
<th>Perceptional modality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital</td>
<td>Visual: Text, and/or abstract illustrations describing the scene</td>
<td>Auditive: Narration or explanation of the scene</td>
</tr>
<tr>
<td>Analog</td>
<td>Visual: Pictorials or video footage of the scene</td>
<td>Auditive: Sounds of the scene</td>
</tr>
</tbody>
</table>

Figure 14: Presentation modes of a natural scene

Hence, we may define presentation mode by perceptional modality that is involved and the degree of digitalization employed in its mediation. The grid in Figure 14 leaves us then with at least four distinct presentation modes and numerous combinations of them, all of whom we can now envisage in much more concrete fashion as to how they are to be operationalized (i.e., implemented in our experiment). We also do realize that a single variable (in our case the independent variable presentation mode) is not equal to a single treatment. A variable, as the name implies, can have an endless number of states or levels. If we chose to induce only one change
on our independent variable we end up with only two treatment condition, if we induce
two separate changes, we have three treatment conditions, and so on.

    In the simplest experiment, one would have a single independent variable with a
single change induced, and a single dependent variable - but there will always be many
extraneous variables. These extraneous variables must be controlled to keep them from
affecting the dependent variable. The logic is that if the only difference is in the
manipulation of the independent variable, then any differences in the dependent variable
must be due to the independent variable.

    Extraneous variables can be controlled in two ways: The first is to hold them
constant while the second is to allow random or controlled (representative) variation.
So, for instance, if we believe that gender may influence our dependent variable, we
may conduct the experiments with females only, or we may have 50% females and 50%
males in our sample.

    Two very special types of variables besides the independent and the dependent
variable are moderators and mediators (see Figure 15). A moderator is a variable that
affects the type and/or strength of the relation between an independent and a dependent
variable. E.g., alcohol has a detrimental effect on driving abilities, but this relation may
be more severe when it is dark than during daylight. Here, the lighting context
moderates the effect of alcohol on driving abilities.
A mediator, on the other hand, is much more difficult to identify, because it is often hidden and does not affect the reality of the relation between the independent and the dependent variable. A variable is called a mediator if it can account for part or the whole influence of the independent variable on the dependent one. In reality it is often the case that variable X does not directly affect variable Y, but the effect is mediated by M. E.g., everybody will find a relation between the socio-economic status of parents and the level of income of their children. But, of course, in most cases this relation must be seen as mediated by other variables, e.g., educational level. Personality and individual effort, on the other hand may again moderate this relationship because it can affect the mediating variable education level. This is then a moderated mediation, and the mediation effect is secured when the effect of the original independent variable X on Y disappears when controlling for the mediator M (e.g., holding M constant).

Mediation effects are often chains of mediations, i.e., also the educational degree does usually not directly influence one’s income level, but it is the occupational status one acquires, based on the education, etc.
Design

Experimental design has not much to do with the kind of fancy or sophisticated experimental setting you develop. These are part of your experimental material and the scenario. With experimental design we refer to the core issue about deciding what kind of manipulations you instigate and how you assign participants of your sample to different experimental conditions, so that you will be able to maximize the impact of your results. The experimental design is what determines your analytical approach and thus the kind of answer you will get to your research question. Its development will usually take most of your mental effort in planning, running, and analyzing your experiments and it is intimately related to all the issues discussed here concerning experimentation. Apart from discussing general issues we will come across a few classic design concepts in this section, e.g., randomized, within- and between-subject, crossed, nested, mixed, full-factorial.

Choosing the simplest of all designs (see Figure 16), we invite a person or a group to our lab and observe them all performing the same kind of task or responding to the same kind of treatment. This is a so-called a one-shot design, also non-experiment. The latter name may be surprising because the popular understanding of an experiment is just that we do something and see what happens. Indeed, most of our beliefs are based on such non-experimental observations and causal conclusions, but it is scary when scientists do the same. This is not to say, that non-experimental designs can not be part of research, but they can not provide sufficient knowledge to make causal inferences, which experiments are intended for.
The reasons for this have already been mentioned in previous sections. One problem is that the straightforward treatment-observation design does not really tell us if the thing that we observed would have taken place anyhow, no matter what our treatment. For instance, if we get a language trainer to teach our 1-year old kid to speak, and find out that after one year our child has made enormous progress and can already form several sentence fragments, we still have no clue whether this development would not have taken place anyhow, i.e., as part of natural growing up.

In some cases a non-experimental design does not even tell us whether anything happened at all. E.g., if you just hand out your new IT-gadget to a bunch of people and ask them how happy they are now, you actually do not know whether they were equally happy just before receiving your product. For this reason we actually need to observe our participants before and after the treatment.

The next, slightly more sophisticated experimental design is already the classic experiment (see Figure 17). It involves two groups, one which receives a treatment (also called the experimental group), and the other which receives no treatment (also called control group). The standard example for this is the discovery of the placebo-effect. Here we administer to the participants in the control a pill that looks the same as the one handed out to the participants in the experimental group, but contains no medical agent. If both participant groups get better, we have a placebo effect, probably caused by the expectations associated with taking the medicine. Actually, to be even more exact, we
would need a second control group, one that receives no pill or consultation whatsoever, just to make sure that health improvement is not generally inevitable.

![Figure 17: Experimental design (X: treatment; O: observation)](image)

The key issue in using the experimental design is that the two groups differ only in the type of treatment they receive (e.g., treatment vs. non-treatment, or treatment A vs. treatment B, or treatment A1 vs. treatment A2). This means, the assignment of our ideally randomly selected sample of participants to the two groups needs to be randomized itself. If this is not the case, we have a so-called quasi-experimental design. This is because the degree of certainty that any observed group differences are explained by the treatment variation is seriously lowered due to the fact that there are other differences between the groups.

A related term is the one of confounded variables, which refers to a very similar problem. Having confounded variables means that the difference in treatment applied to the experimental and control group is not just a variation in one single variable, but more than one.

A good example for this comes from esoteric circles. A group of people that believed in the magic powers of the prism run a test where they watered some indoor plants with water coming straight from the tab, and another group of plants with water that was filled from the tab as well, but then kept for 24 hours under a metal prism.
shape. Hence, they used a classic experimental design as depicted in Figure 17. After a few weeks it was noticed that the plants that had been watered with the prism-treated water flourished much better than the ones in the control group. So they concluded this to be a case of the power of the prism. Careful consideration, however, revealed the confounding of two treatment variables: The prism-treatment and the 24 hours that the water was kept at room temperature. A subsequent experiment run by a non-esoteric group showed that the improved condition of the plants is indeed due to the delay using the tap water, not the prism treatment.

As anybody can easily imagine, treatments that we use in experiments are usually always a combination of changes of very many, often trivial characteristics. It may, for instance simply be the case that we invited our experimental group participants one week before our control group participants and that there was in the mean time some news on TV affecting our experimental comparison. Hence, confounded variables are a constant threat to our research.

Let us return to the issue of design: There are of course a great number of variations to the classic experimental design. One such a variation is the introduction of a pre-treatment observation, to make sure that the two groups are really equivalent before the experiment. Again, a variation of this involves the combination of the design illustrated in Figure 17 and the variation explained just before. By doing so we get four groups of participants, (a) one that undergoes pre-treatment observation, treatment, and post-treatment observation, (b) one that is observed twice, but receives no treatment in between, (c) one group that undergoes the same procedure as group (a) without undergoing pre-treatment observation, and (d) a final group that is observed at the end of the experiment, but which receives no treatment. The reason for this so-called
Solomon Four Group design is to control for the possibility that pre-treatment observation sensitizes participants with regard to the demand characteristics of the experiment.

There is also another very important dimension of variation to the classic experimental design as explained in the examples so far. In addition to having been illustrations of the basic experimental plan, all examples up to now have been description of the archetypal between-subject design. In reality, however, we can just as well have one group of participants experiencing all treatment conditions (i.e., within-subject design; see Figure 18), as we can have separate groups of participants for each condition (between-subject design). Hence, if we test the relative effectiveness of two different treatments (e.g., interface A and interface B) we can have all our participants work with either interface, or one group with interface A and the other with interface B.

\[ \begin{array}{c c c}
X_1 & O & X_2 \\
\end{array} \\
\]

Figure 18: Within-subject experimental design (can be run with or without control group) $X_1$: treatment 1; $X_2$: treatment 2; O: observation

The advantages of using within-subject designs seem immediately obvious. We need half the amount of participants and the participants that are exposed to the various treatments are identical. However, both of these advantages have also their downside, one is that we expect a higher degree of commitment from our participants, which is especially true in the case of longitudinal studies. Longitudinal studies usually do not involve different treatments administered to the same participants, in the sense of the word treatment as it was discussed so far. In contrast it involves recurring instances of
measurement (which, on the other hand, may be seen as equal to treating people with subsequent intervals of time). Another downside to the within-subject design is that our participants are actually not the extent identical as we may believe them to be when administering several treatments to them. If our participants work first with interface A and then with interface B, they differ as participants of our experimental phase A with respect to the fact that, when working with interface B they have already been exposed to interface A. For this reason, the treatment sequence is usually counterbalanced: Half of the participants work first with interface A and then with interface B, the other half completes the experiment in reverse. This then leaves us again with challenges of randomization. And even after having solved that issue, we still can not escape the fact that all participants working with the second interface have already some experiences from being part of our experiment dealing with interfaces, which may influence their behavior in critical ways. These, and other considerations, usually cause us to use between-subject designs in user psychological research.

In many cases we use, however, mixed or nested designs, especially when our research involves more than one independent variable. Imagine the case where we want to test the visibility of two versions of an interface, both at home as well as in the car. In this case we have two variables: (1) the type of interface and (2) the use context. Both variables have for sake of simplicity only two levels: (a) interface version A and interface version B and (b) use context “home” and use context “car”. If we use a full-or complete factorial within- or between-subject design we need either four separate measurements with the same participants or four groups of subjects for each type of treatment (i.e., interface A at home, interface A in the car, interface B at home, interface B in the car). However, if we can decide in which of the two independent variables we
are more interested we may use a mixed design. For instance, we might decide that we have a good understanding of the differences between the interfaces, but really would like to know how each of them adapts to various use contexts. In this case, it is feasible to work with only two groups of participants: one group that uses interface A both at home as well as in the car, and another one that uses interface B in the two use contexts.

The difference between the three approaches is self-evident. Using a complete factorial, within-subject design (also called crossed design) each participant sees each experimental condition. Running the same experiment, using a between-subject design, one group of participant sees only one type of condition. And finally, in mixed designs, all participants see one type variation between the conditions, but half of them experience this variation in one context the other in another.

Control and counterbalancing

This section adds nothing substantially new to the discussion of experiments, but its purpose is to emphasize what has been said. The key issue in experimentation is control. Control can be achieved in many ways, through considerate use of the research model, careful operationalization of the constructs, and through design-related decisions. Hence, in order to examine the influences of one variable upon another, experimental manipulation has to be exact and any facts or events need to be measured very precisely. Understandably, control is also one of the most profound weaknesses of any experimental research: How well can we control events and do we control them only to such a degree that the results still can be generalized to non-controlled (i.e., natural) environments?

The purpose of the control condition in the classic experimental design is for instance to allow us to compare measurements of the experimental group's behavior,
with some other group or context that differs only with respect to the experimental treatment. Any differences we find between the behavior of our experimental group and our control group should be caused by our manipulation of the independent variable. This is how we establish cause and effect: the effect of \( X \) on \( Y \).

The control of extraneous variables is then absolutely critical. For instance if we believe that the influence of our treatment on the dependent variable is not the same for men and women, our results would be very difficult to interpret if almost all of the men were in the experimental condition and the control condition is made up mainly of women. Such things are issue to counterbalancing the sample and conditions. As a principle, always double check with another researcher whether your design is appropriate and you got the counterbalancing right.

The role of the experimenter

As experimenter or responsible person of your research you have a few core responsibilities and it is advisable to familiarize yourself with the concrete suggestions made in the following.

When preparing experiments:

- Think what tasks, interventions, and measures do you need to operationalize your variables and analyze your hypotheses?
- Develop the necessary materials and organize the equipment
- Carefully think through each step of the procedure, and pay attention to details that you will have to decide in an actual experiment (e.g., time, order, place of material, actions, etc.)
• Write down the instructions and comments that you will give, and think of how you will want to answer participants’ common types of questions (e.g., “What shall I answer here?”, “What does this mean?”)
• Think if any of the materials and/procedures may influence the measurement in a way you have not been considering before
• For experiments that require quiet conditions, pick a context where participant can complete experiment without being distracted.
• Decide how much you want/can give away about the contents/purpose of the experiment. You usually don’t want that people can prepare themselves, be suspicious, or influence the findings in any other way.
• Pilot experiments
  See how long it takes, whether people understand directions, and if you get data you want in an efficient way.
  (Pilot testing intends to eliminate all technical shortcomings and problems concerning clarity inbuilt into the materials and the procedure. After their elimination, all questions arising later during the actual experiment are part of your measurement, and shall not need your active assistance or further explanations).

When organizing the experimental session:

• Contact and invite participants to your experiment, if necessary, giving away some general aim of the experiments and a person/institution that holds supervising responsibility. Subtract some 25% of the actual time it takes – you don’t want to scare people off. Define a meeting place and exact time. Don’t invite them directly into your lab.
• Make sure the laboratory is tidy and looks the same as for the previous participant
• Have all materials ready at hand (Instruction sheet/running log. Informed Consent Forms. Answer Sheets/test booklets. Feedback.)
When running the experiment:

- Welcome the participant and thank him or her for coming. Give the participant some time to acclimatize. Don’t start to give instructions before you see that the participant is comfortable and not anymore distracted by the surroundings and his or her own belongings.
- Get informed consent, if necessary.
- Have a written copy of the exact instructions you will use, and read them aloud exactly as they are written each time (Try to take some pressure off, say e.g., “There is well enough time for completion of the task”, “I shall give you all the necessary explanations”, “this is not an intelligence test, just try do everything the way you can best”) If you feel weird reading the instructions aloud, tell the participant that you will do so in order to insure that all receive exactly the same information in the same way.
- Ask participants whether they are ready and whether they have any immediate questions before the actual task starts. If it is essential to clear out the question beforehand, do so. If you believe the participant has time and the chance to learn it while doing, tell him or her so.
- There are in principle two types of questions a participant might ask during the experiment: technical and procedural ones and content-related ones. The first type you either answer by restating the respective section in the written instruction or you answer in such a form that you can proceed with the experiment (if you need to answer the same technical question for different participants, your materials or your instructions are flawed). For the second type of questions you decline assistance and instruct participants to judge or behave in such way that they find it themselves most appropriate or meaningful.
- Follow exact protocol (timing, instructions). If you do something critical once (like changing the chair where you sit) do it every time. Remain quiet, and don’t consult your watch all the time in an obvious manner; it tends to make participants nervous.
- When everyone has finished, thank them, and give them some feedback/debriefing information if you decide to do so.
• Decline requests for personal or general results of the experiment. Tell them that
data are analyzed anonymously, and inform them where the results will be
used/published. If participants what happens to their personal information, tell
them that data about their identity is stored separately from their experimental
data.

Your key ethical responsibilities:

• Make sure that conditions are the same for all participants.
• Be polite and courteous at all times. Say “thank you for coming” at the
beginning and “thank you for participating” at the end.
• In using the element of deception try to stay truthful with regard to the nature of
experience participants are exposed to, and disguise only the purpose of
experiment.
• Ensure harm protection, and minimize risks involved for participants
• Privacy and confidentiality
• Debrief them, if necessary.
• Make sure all participants are participating by free will (e.g., informed consent)
• Never force someone to participate
• No participant should be made to feel bad for their performance on the task.
• If people complain about difficulty or stupidity of experiment, express your
understanding but don’t get involved (“nod and smile”).
• Keep all performance data completely confidential.
• Privacy and confidentiality: Never link the discussion of anyone’s performance
on any task to the actual person.
• Try to find the impossible balance between being indifferent (without appearing
unresponsive) and empathetic (without expressing active compassion).
Behave neutral and human – prevent becoming neither an accomplice nor an
enemy of the participant.
• I advise not to promise the sending of results to the participants. Inform them
where the data is being used and maybe where and when it might be published.
A Few Final Remarks

Whatever you do, remember that research work is not straightforward and linear. Things often appear very trivial and intuitive at the beginning and as soon as we look closer at matters, they turn out to be very complex. Some degree of “mess” is very normal in empirical research at early stages. Usually an enormous amount of decisions need to be made. These decisions pertain mainly to the selection of the phenomena and the research question, as well as the theories and methods you chose.

The key issue is to repeatedly run through all the method-related considerations in order to sharpen the research question and to develop a clean and robust design. When making decisions about theories and methods, it is most important to know what one did and why, and to be explicit about it. Everything else is subject to scientific discourse. This is very important to realize. You hold the authority for decisions made in your research, and what you do is in principle not so important as long as you have well-founded and communicable reasons for it: “Though this be madness, yet there is a method in it” (W. Shakespeare).

Also, never hide your research results in the drawer when the results did not confirm your expectations, and you find no method-related answer to this inconsistency. Nothing is more fatal to scientific progress than to be ignorant of findings that contradict our prejudices. This is part of your professional and moral responsibility as a researcher.
References


