Task 4:

Design your experiment

Deliverables for this Task

Step by step

Overview

Terminology

Experimental design, design

An experimental design is a plan for collecting and analyzing data. The initial version of your experimental design will specify the factors, their levels, how they are arranged, and how they will be tested. A more complete design will also specify the measures, the tasks, the materials, and other details about data collection.

By this time you have chosen a research question to study and have reviewed the research literature to determine what is already known about your question. Now it's time to start thinking through the details of how you will collect and anlyze your data.

At the end of Task 1, you started to think ahead about how you will formulate the details of your experiment in such a way that the research will be feasible or doable with the limited time and resources that are available. In this Task, you'll spell out some of the details that will help you plan data collection and data analysis, as well as evaluate how feasible the experiment is.

Goals

For this Task, you need to do these three things:

1

Choose the *levels* for your factors

For every factor that you investigate, you will be contrasting two or more *values* of that factor. For example, if you're studying the effects of Gender, you may compare people with a value of "male" for Gender and people with a value of "female" for Gender. **These different values (male, female) are called the *levels* of the factor Gender.**

2

Identify the *conditions* in your experimental design

Each combination of the levels of your factors is called an *experimental condition.* For example, if you have males and females as the levels for the factor Gender and young (<30 years old) and old (>40 years old) as levels for your factor Age, then there are four possible combinations of these levels, or *experimental conditions*. In this example, one condition will have young males, another old males, a third condition will have young females, and the fourth will have old females.

3

Decide whether to test your factors "between subjects" or "within subjects"

Testing a factor "between subjects" means that you will use different individual participants for each level of the factor. Testing a factor "within subjects" means that you will use the same ndividual participants for different levels of the factor.

Deliverables

For this Task, you need to hand in your experimental design.

Below is a sample experimental design. It includes the following information:

a) for each factor – factor name, which levels, kind of factor (between subjects or within subjects)

b) for the table – factors, level names, sample size for each condition. For now, estimate that you'll need 15 participants in each condition. The arrangement of the table shows whether you will test each factor between subjects or within subjects.

c) for the design – how many experimental conditions, how many participants total

Sample:

		Time Limit (between-subjects)	
		No Limit	Time Limit
TV Exposure (between-subjects)	No TV	*No limit; no TV* $n = 15$	*Time limit; no TV* $n = 15$
	With TV	*No limit; with TV* $n = 15$	*Time limit; with TV* $n = 15$

Total: 4 experimental conditions, N = 60 participants

Why do you need to do this?

Developing an experimental design is part of the broader process of planning your research. Before you proceed with your experiment, you need to plan in great detail exactly how your data collection will happen: you need to *design* your experiment to decide on the details well before collecting data. This planning will help you to do several things:

- Think more about the details of your research problem;
- Produce the right data for your analyses;
- Produce more relevant and more reliable data;
- Avoid stress and confusion during data collection and analysis;
- Use your time more effectively;
- Finish on schedule;

Planning is the key to successful research (and to other complex activities). It is particularly important to plan your analyses **before** you start to collect data.

Your experimental design is an important part of a plan for how you will collect and analyze your data. This plan becomes the Method section of the research paper. For many scientists, the clarity and precision of the Method section helps them decide whether or not they will believe your results.

Keep in mind that you are specifying your experimental design so that you know *exactly* what you will do at every step of your data collection and analysis. As with other plans, however, when they're put into practice, reality has a way of forcing you to change them. Planning ahead helps prepare you to react to these changes.

<u>Statistical analyses</u>. You should build your experimental design with a particular statistical analysis technique in mind. The experimental design is also a plan for how you will do your statistical analyses after the data has been collected. This way, you plan ahead so that you collect the right data to fit the statistical analyses that you will do.

This course focuses on the statistical technique called **Analysis of Variance (ANOVA)**. It's been chosen ahead of time to simplify parts of the teaching process. It's also one of the most flexible and widely used statistical techniques in experimental psychology. ANOVA, however, is only appropriate when you want to study the effects of discrete or categorical grouping independent variables (ex: male vs. female; old vs. young) on continuous dependent variables (ex: number correct). You need to use other techniques, such as regression analysis, when you study the effects of *continuous* independent variables (ex: age, IQ, etc.) or when you have categorical dependent variables.

> ANOVA is used to study the effects of discrete or categorical independent variables on continuous dependent variables.

Given that you're going to use ANOVA, you have to know several things to use the statistical analysis software:

- which factors are in the design;
- how many *levels* there are of each factor;
- how you want to test each factor;

You need to identify which experimental conditions you will use to help you understand this information.

> It is very important to note that specifying your experimental design will sometimes make you **reformulate or modify your initial research problem** if it turns out to be too difficult to study. This means that you will also have to make changes to your Literature Review so that it matches the changes in your research problem.

Background: Kinds of Research

You suspect that each of your factors will have an effect on the process or sub-process that you're studying. During your review of the literature you will have found other researchers who have studied the same factors and who found evidence that they do in fact (or do not) have an effect on your process. You also want to discover whether your factors interfere with each other or show some sort of synergy or additive effect. You probably chose factors like Gender, Music, Major, Amount of Sleep, etc. [Note that it's a standard convention to capitalize the name of each factor.]

The methodological choices that you make now will determine what *kind* of research you will carry out. This Research Methods course focuses on experimental research but there are other options to consider.

The first thing to observe is that experiments must involve data, but

> Not every study that uses data is an experiment.

Why not? Because experiments have special characteristics that other kinds of research do not have. To put this in context, consider some types of non-experimental research, described below.

A first important point to make is that non-experimental research is perfectly valid and is often the *only* kind of research possible in a particular field or for a particular problem. Non-experimental research, however, generally studies different questions and provides weaker evidence than experimental research.

> It is extremely important **not** to think that *non-experimental* means *non-scientific* or *invalid.*

Non-experimental research

Here are a few kinds of non-experimental research.

Observational studies. In *naturalistic* or *observational studies*, the researcher observes directly or makes recordings of what people or animals do in a particular situation. They're called *naturalistic* exactly because the researcher decides *not* to control any factors that might influence the outcomes. Another difference is that these studies try to describe *observable* behavior, where experiments usually try to identify the *un*observable mechanisms that theories postulate. *Ethnographic research* is one well-known and systematic approach to observational research.

A specific kind of observational study is one where the researcher asks people questions – this is called *survey-based research*. These studies are very common before elections: hordes of researchers ask people in person and by phone a list of questions (a *questionnaire*) about the candidates and their positions.

Small-*N* studies like *case studies* are another kind of observational study. The researcher may only have access to one or two patients with a certain kind of brain damage or rare disease. This is a real-life limitation that makes experimental research impossible. Rather than not study the problem at all, researchers try to accumulate weaker observational data, even though they know that they can't reliably generalize the conclusions to other people.

Of course, people are very variable, especially when they know they're being observed. Sometimes they don't feel well or are distracted; sometimes they answer questions according to what they think the researcher wants to hear; sometimes they intentionally try to fool researchers. Just observing them, without measuring and taking into consideration these other factors, leads to less reliable results than a carfully controlled experiment would produce.

Correlational studies. In *correlational studies*, the researcher measures two (or more) factors of interest and evaluates statistically what their *strength of association* is, i.e., when one factor changes, does the other change, too? Do they change in the same direction (a positive correlation)? Do they always or very frequently change together (a strong correlation) or not?

Correlation is *not* the same as causation. It might be true that in a particular sample, the people with long noses have bigger shoe sizes or salaries. This does ***not*** mean that nose length *makes* their shoes or salaries bigger! It just means that these things appear together. The tobacco companies used this argument for very many years to say that the research did not show that smoking causes cancer. They were quite correct because most of the research was correlational, so by definition it did not show cause. Later (experimental) research, of course, showed conclusively that the tobacco companies are criminals because smoking in fact creates chemical dependencies and causes cancer.

The technique called *meta-analysis* uses other researchers' statistical results as raw data, most often to find correlations among those results. One limitation of meta-analysis is that this kind of research uses only the data that is already available to test a given hypothesis. The available data may be more or less directly related to the specific question that a researcher wants to answer. Experimental research, on the other hand, generates special data that is custom designed to be as relevant as possible to the hypothesis under study.

Correlational studies are particularly useful in establishing strong candidates for important determining factors. They ask the question (rather than answer it): are these factors strong candidates for the things that are *really* causing what's going on?

Advantages of non-experimental research. The strength of non-experimental research is that very often it looks at broader, less controlled situations; so, many phenomena appear that experimental research has not yet dealt with. Non-experimental research does an excellent job of separating illusions and coincidences from systematic phenomena: it shows that *x* does (or does not) in fact occur systematically. The next question, of course, is why or how does *x* occur, and that's an issue for experimental research. Correlations provide excellent hints for study: they're good candidates for determining factors, if they fit in with existing theories and data.

Non-experimental research is also an excellent option when experimental research is not possible, practical, or ethical.

Qualitative "vs." Quantitative research. It's really unfortunate that there's a widespread and very misleading belief that research comes in two basic (stereo)types: numbers-only research that misses the "essence" of what it studies (also called "hard" science) and more detail-oriented, intuitive research that focuses on the qualities of what is studied without using much math (also called "soft" science). This extremely ill-informed opinion arose from discussions about what science "really" is with the appearance of the social sciences in the late 1800s. Social scientists who neither had the training nor the interest in mathematical methods worked with problems that were so new that it was not at all clear how to mathematize the things that they wanted to study.

On the other hand, the social scientists also wanted the prestige of doing "real research", so they "adjusted" the notion of research to fit their needs. However, they did a really poor job of it. "Quantitative" research uses numbers to measure amounts of *qualities*, so it, too, is qualitative. Otherwise the numbers would be meaningless. Only mathematics studies numbers without paying attention to some form of measurement. On the other hand, *quali*tative research also routinely uses mathematical methods: quantities, frequencies, relative amounts, etc. even if they are simpler. More recently, the social sciences have seen an amazing increase in the use of advanced statistical modeling to the point where the distinction between quantitative and qualitative research becomes totally meaningless. Forget you ever heard about it.

Experimental research

With these kinds of non-experimental research in mind, it's easier to characterize experimental research. In this Research Methods course, the emphasis will be on *experimental* research so that you have a clearer idea of what you lose when you do non-experimental research.

Experimental research focuses on different, deeper questions than non-experimental research. Non-experimental research tends to focus on *what*'s happening, and experimental research focuses on the more theoretical or profound question of *why* or *how* it's happening. So, when experimental research is a good option, researchers avoid non-experimental research.

Data collection in experimental research is more controlled. The key thing that makes experimental research different from other kinds of research is this fact:

> The experimenter systematically *controls* a range of factors that might influence the observed outcomes.

Experimental control is essential to generate tightly focused or "uncontaminated" data that is immediately and directly relevant to the hypothesis under investigation.

The Method section has to make clear not only what the experimenter is studying, but also what is *not* being studied and how the experimenter is controlling the effects of these other factors.

1 Choose the levels for your factors

Terminology

Levels A factor usually refers to a characteristic that can have different values. Age is a factor that can have numeric values (in years or months) or categorical values (like old and young). Text Type is a factor that can have values like story, news, instructions, etc. When researchers choose which values to focus on, they give them names and call them *levels*.

Factors vs. independent variables
Terms like *factors* and *independent variables* which often seem to be referring to the same thing. When speaking informally, many people in fact use them as synonyms. But when researchers are being careful, they distinguish between factors and independent variables as follows.

Factors are theoretical entities or parts of theories: a researcher suspects or hypothesizes that they "are a factor" in how a given process happens. The researcher can't call them "causes" because he/she doesn't know yet whether the factors are in fact causes; they are still under investigation. For most experiments, you'll limit yourself to two or three factors.

Independent variables are mathematical entities: they're the numerical values of some *indicators* of the factors and are the result of how you measure the factors. For example, if Weight is a factor of interest, then researchers have several options for independent variables to indicate a participant's weight: number of kilos on the day of the experiment, mean number of kilos of the last two months, kilos relative to some group (a z score, for example), body mass index, etc.

Independent variables, in sum, are indicators or measures of factors.

For every factor that you investigate, you will be contrasting two or more values of that factor. For example, if you're studying the effects of Gender, you my compare people with a value of "male" for Gender and people with a value of "female" for Gender. In other words, you'll focus on male and female as *levels* of the factor Gender.

However, you may have noticed that simply choosing "male" and "female" for Gender is really quite a simplification: unusual genetic combinations and lifestyle choices sometimes lead to people who cannot easily be classified as clearly male or clearly female. Male vs. female, then, is a simplification of reality, and this simplification adds some error or distortion to your experiment.

> Simplification adds some error or distortion to your experiment, but it is an absolute necessity.

Simplification is an absolute necessity: you'll have to simplify different things at many points in research. Researchers have to be aware of the distortions that simplification can add and try to minimize them where the distortions can have important effects. In other situations, researchers just have to accept this consequence of simplification and keep in mind that research (by definition) can't be perfect. This is one of the reasons why researchers don't actually "prove" anything: they know that every experiment involves many simplifications and some margin of error.

> Researchers don't actually "prove" anything.

To choose levels for your factors, you need to think about two issues:

a) **Simplicity**. You want to maximize the simplicity of your experiment so that you can execute it reliably. In most experiments, you'll limit yourself to two or three levels for each factor. One common notation for this is the form <factor> (<value1>, <value2>, etc.), where you use the names for your own factors and values. For example:

- Gender (male, female);
- Music (presence, absence);
- Amount of sleep (less than 6 hours, more than 8 hours)

On the other hand, you also want to minimize the errors and distortions that simplification can introduce into your results. Researchers try to balance these conflicting goals by repeating similar experiments in different ways and by focussing on one or two factors (with less simplification and more accuracy) at a time so that they can simplify the other, (temporarily) less important factors.

In the clearest case, the two levels of the factor are *present* and *absent*. For example, some participants might read a text *with* a picture, while others read it *without* the picture. In this case, you'd be studying the effects of the factor Picture, which has two levels: *with Picture* and *without Picture*. You might also want to study the same factor with three (or more) levels, for example: *with Picture*, *with partial Picture*, and *without Picture*, for the case where you also show only a part of the picture.

b) **Contrast**. You want to choose levels of a factor so that the difference or *contrast* between them is very clear. Your experiment will focus on the effects of the difference between the levels of each factor. The contrast between these levels is very important. If the levels are too similar, then it will be harder to measure the differences in their effects.

One way of emphasizing the contrast between levels is to ignore some possible intermediate values when you define your levels. For example, Age might be defined with levels Young (<30 years old) and Old (>30 years old). Notice, however, that many of the participants in the two groups will be quite similar: the 28-, 29-, 31- and 32-year-olds. A more effective way of defining the contrast would be to skip some of the ages and define Young (<20 years old) and Old (>40 years old), leaving a gap in the middle of ages that you will ignore. This way the contrast between the two groups becomes much clearer.

Control groups are the simplest way of establishing a contrast: one group gets no experimental treatment (the control group) and another gets the experimental treatment (the experimental or treatment group). Talking about control groups usually doesn't work for the more sophisticated experiments that you will do in this course: what's the "control" group if you compare males and females? There is no control group in this case. When you investigate the effects of two or more factors, the notion of control group gets even more confusing: if you have males and females reading a story with or without music, which is the control group? It's difficult to say. Avoid talking about control groups for this course.

"Baptize" your factors and levels

You will have to talk about your factors and levels more and more as you develop your experiment and write it up. It's best to give your factors and levels specific names right now and use these names consistently. You don't want to have to say "the differences between having a time limit and not having a time limit" each time you mention your factor; you can define this as one of your factors and call it Time.

> Do *not* ever use general, uninformative names like Group A, Factor 1, etc. for your factors or levels. No one will understand anything that you're saying!

Do you understand a sentence like "Group A recalled significantly more information than Group B"? Neither does anyone else. And experimenters who use these names end up confusing themselves, too. "The no-time-limit group recalled significantly more than the time-limit group" is much more informative. "No time limit" and "time limit" are the names that one experimenter chose for the *levels* of his factor, which he called Time.

Notice that factor names are capitalized and level names are usually not capitalized.

"The Time factor had two levels: time limit and no time limit."

"The first factor was Time (time limit, no time limit); the second factor was Gender (male, female)."

When you have a name that has more than one word, like "no time limit", remember to put in hyphens when you use the name as an adjective.

"In the no-time-limit condition, participants…"

"The no-time-limit group recalled…"

Exercise 1: Identify the levels of your factors

For each of your factors, write the name that you will use to talk about it (start the name with a capital letter), which levels you will study, and explain how you will emphasize the contrast between the levels.

Factor 1 name:

Level names:

Contrast:

Factor 2 name:

Level names:

Contrast:

2 Identify your experimental conditions

Terminology

Experimental conditions, conditions

Each combination of the levels of your factors is called an experimental *condition.*

For example, if you have males and females as the levels for the factor Gender and young (<30 years old) and old (>40 years old) as levels for your factor Age then there are four possible combinations of them. One condition will have young males, another old males, a third condition will have young females, and the fourth will have old females. If you only have one factor, then there is an experimental condition for each level of your single factor.

Each combination of the levels of your factors is called an experimental *condition*. Each experimental condition describes a situation in which you'll collect data.

Most researchers like to make a table like the one below to think about their experimental conditions. The four grey cells in the table correspond to the four experimental conditions in the experiment. In this case, the data collection situations are described in terms of the characteristics of the participants. But in other experiments, the experimental conditions can vary in terms of the materials, the setting, the materials, the tasks, or other aspects of the factors.

		Age	
		Young	Old
Gender	Male	Young males	Old males
	Female	Young females	Old females

Two factors: Age, two levels; Gender, two levels = 4 experimental conditions

In this example, the factors are in blue letters and the levels are in black letters.

To calculate the total number of conditions in your experiment, you just have to multiply the number of levels for each factor. For example: for Gender (male, female) and Music (presence, absence), you have 2 (two) levels of Gender and 2 (two) levels of Music. 2 x 2 = 4 (four) experimental conditions. Remember that experiments with more conditions are more difficult and time consuming to carry out, so for this course, stick to factors with only 2 or 3 levels.

Here are some examples of similar tables for other possible experiments, with different numbers of factors or levels. Again, the grey cells in the table correspond to the conditions in each experiment.

One factor: IQ, three levels 1 x 3 = 3 experimental conditions

	Low	Medium	High
IQ			

Two factors: IQ, two levels; Text, two levels 2 x 2 = 4 experimental conditions

		IQ	
		Low	High
Text	Story		

Manual		

Three factors: IQ, two levels; Gender, two levels; Order of presentation, two levels 2 x 2 x 2 = 8 experimental conditions

		IQ	
	Order 1	Low	High
Gender	Male	1	2
	Female	3	4

		IQ	
	Order 2	Low	High
Gender	Male	5	6
	Female	7	8

In this last, more complex exmple, think through what is going on in each of the numbered conditions. In condition #1, for example, there will be low-IQ males who will be exposed to the stimuli in Order 1.

In these examples, ALL of the possible conditions (multiplying out the levels of all of the factors) are in fact used in the experiment. These are called fully *factorial designs*: all of the possible combinations of levels are included in the experiment.

Exercise 2: First draft of your experimental design

- How many conditions will you have in your experiment?
- Describe all of your experimental conditions, one by one. Does it make sense to study *all* of your experimental conditions?
- Draw a table for your experimental design.

Background: Other design options

Incomplete designs. In some situations, it may be the case that not all of the conditions are relevant or interesting. For example, consider an experiment with two, two-level factors: Task (listen, translate) and Expertise (expert, novice). A fully factorial design would yield four conditions:

- Expert, listen
- Novice, listen
- Expert, translate
- Novice, translate

But it makes little sense to compare "expert" and "novice" listeners: native speakers are all pretty much experts at listening. You're interested in how listeners compare (as a baseline) to expert and novice translators. So, you're only interested in conditions a, c, and d. To run participants in condition b would be a waste of effort. Collecting data for only some conditions of an experiment is referred to as using an *incomplete design*, if it's planned that way. If it's unplanned, it's called "a big mistake".

There usually has to be a strong theoretical motivation for this option because the statistical analyses are often more complex than for a complete design and lots of information is lost. This also why you need to plan data collection carefully: not collecting data for a given condition, for example, can be a very costly mistake.

Single-*n* designs. In other situations, the participants are so unique or the task so complex that it becomes impractical to study groups of randomly selected participants and the researcher has to focus on individuals. Of course, generalizability is limited, but these designs are often the best option available. To ensure internal validity, a common strategy is to use an **A-B design** in which the same subject is tested once (the A condition) without any treatment to establish baseline performance and then again (the B condition) after the experimental treatment. Treatment is seen as a within-subjects factor and the participant serves as his/her own "control group". See Ch. 13 in Mitchell & Jolley, 2005 for more information.

Longitudinal designs. Many research problems investigate how a particular skill or process – for example, reading ability – develops over time. One strategy is to sample different participants at different stages of the development period, for example, 4th, 6th, 8th, and 10th graders. This strategy has well-known disadvantages because of the variability from group to group and the need to infer that differences between the groups are representative of individual development. Another strategy is a design in which an additional within-subjects factor (Time) is added: the same experiment is repeated with the same participants at (usually) regular intervals to chart the evolution in participants' performance. This is called a longitudinal design. Internal validity is an issue with these designs because as time goes on, there are more possibilities for important non-experimental influences to happen.

There are many other options; these are some of the most commonly used designs.

3 Decide how to test your factors

Terminology

[Contents](#)
[Overview](#)
[Kinds of research](#)
[Choosing levels](#)
[Exercise 1](#)
[Identify conditions](#)
[Exercise 2](#)
[Other designs](#)
[How to test factors](#)
[Exercise 3](#)
[Resources](#)

Between-subjects testing; between-subjects factor

Testing a factor "between subjects" means that you will use different participants for each level of the factor. In other words, individual participants will only be in *one* experimental condition. You can say that Gender was a between-subjects factor.

Between-subjects design

You can choose to use only between-subjects factors for a given experiment. This is called a *Between-subjects design* or *Factorial design*.

Within-subjects testing; within-subjects factor

Testing a factor "within subjects" means that you will use the same participants for each level of the factor. In other words, individual participants will be in *multiple* experimental conditions. You can say that Text was a within-subjects factor if your participants used more than one text.

Within-subjects design

You can also choose to use only within-subjects factors for a given experiment. This is called a *Within-subjects design* or *Repeated-measures design*.

Mixed design

Another option is to mix the kinds of factors, with some factors studied between subjects and others studied within subjects. Because there's a mixture of between and within-subjects factors in this kind of design, researchers call it a *Mixed design*. This is a particularly effective design for studying cognitive processes: often the between-subjects factor(s) reflect differences in groups of people and the within-subjects factors reflect (often subtle) differences in how cognitive processes happen. For example, this is a common design to study sex differences in cognitive processes.

It's very important to build your experimental design with a particular statistical analysis technique in mind. The experimental design is a plan for how you will do your statistical analyses after the data has been collected. When you think this way, you plan ahead so that you collect the right data to fit the statistical analyses that you will do.

Given that you're going to use ANOVA, you have to know which factors are in the design and how many levels of each. You already specified this information in the exercise above. The next step is to decide how to test your factors. This section describes two ways of testing your factors: with participants in only one experimental condition or with participants in more than one condition.

Decide, for each factor, whether you will test it between subjects or within subjects. The information below can help you choose.

Between-subjects testing

Testing a factor "between subjects" means that you will use different participants for each level of the factor. In other words, participants will only be in *one* experimental condition.

Some things simply require researchers to test them **between** subjects, like Gender and IQ. People don't have two different values for them *at the time of a given experiment* so they can't be tested twice for that factor. Of course, someone might have a sex-change operation or get very much smarter, but then you're talking about the same participants at very different points in time and this is another kind of ("longitudinal") study that won't be considered here.

One of the advantages of between-subjects testing is the fact that the statistical models are simpler and some people find them conceptually easier to understand.

One of the disadvantages of between-subjects testing is the fact that studying something as a between-subjects factor requires more participants and introduces more random variation (or error variance) in the data and so reduces the *statistical power* of the analysis.

Researchers classify a factor as a **between-subjects factor** if they test it in a way where individual participants only provide data for *one* of the levels of that factor. That is, the participants only appear in *one* of the experimental conditions for that factor. For example, one group of individuals appears in the young group for the factor Age, and a different group of individuals appears in the old group.

Example. In the example below, the researcher had to use different people in each condition. None of the participants can be tested as a male and then as a female or first as a low-IQ participant, and then as a low-IQ participant. So, the participants do not appear in more than one condition. Researchers call IQ and Gender ***Between-subjects Factors*** because the different **levels of a factor are split *between* different participants**: some participants are in one level, some in another.

		IQ	
	N = 60	Low	High
Gender	Male	***n* = 15**	***n* = 15**
	Female	***n* = 15**	***n* = 15**

Each individual participates in only one experimental condition in this design. There are a totoal of 60 participants for four conditions in this example. There are four groups of participants in this experiment: different people in each experimental condition.

Within-subjects testing

Testing a factor within subjects means that you will use the same participants for each level of the factor. If participants can experience multiple values for a factor in the course of a single experiment, then this factor is a candidate for within-subjects testing. So, for example the same participant might be trained or tested in two different ways, or may use two different materials.

Among the advantages of within-subjects testing is that the participants serve as their own control group – instead of having similar, comparable participants in two experimental conditions, you have *the same individuals* in both conditions. This helps control for possible effects of individual differences and gives the analysis more statistical power. Also, the number of participants needed for a within-subjects test is quite a bit smaller than the number needed for a between-subjects test. In the end, this means that you can detect much smaller differences or more subtle effects when you study something within-subjects. It's important to emphasize that within-subjects testing is *much* more effective for most experimental situations, when it's possible to use it.

Among the disadvantages of within-subjects testing are the fact that it requires more complex statistical models (called *repeated-measures analysis*) and the fact that some within-subjects testing will require the researcher to introduce an additional factor into the design, for additional experimental control. This additional factor is called a *nuisance variable* because it's not the researcher's main focus of study (so it's a real nuisance!).

Nuisance variable. An additional factor that an experimenter has to include only to double-check for its possible effects.

For example, if Text Type is studied as a within-subjects factor (or is "tested within subjects"), then each participant will read two texts, say a narrative and a dialogue. The nuisance variable would be Order of Presentation: participants have to read them in one order or the other and that may affect the outcomes. So, you add the nuisance variable to control whether reading one type of text facilitated reading the other, even though you may not be particularly interested in this question.

> **Statistical power.** The ability to reliably detect smaller or more subtle differences with statistical tests. Experiments are usually designed to maximize statistical power.

Researchers classify a factor as a **within-subjects factor** if they test it in a way where individual participants provide data for *all* of the levels of that factor. For example, all of the individuals in the narrative group (for the factor Text Type) ***also*** appear in the dialogue group, i.e., the same participants appear for all the levels of the factor. The same participants will provide data in two or more different experimental conditions.

Example. In the next design, each participant will read *both* a narrative and a procedural text. Some participants will read the narrative first (the N-P order), the others will read the procedure first (the P-N order). So, the participants are tested in two conditions each and there are two sets of measures for each participant. Order (of presentation) is another between-subjects factor: the levels are split between different participants. Text Type is a different kind of factor because you will observe both of the **levels** ***within*** **the same participant**: it's called a ***Within-subjects Factor.***

		Text Type	
	N = 30	Narrative	Procedure
Order	N-P	***n* = 15**	
	P-N	***n* = 15**	

Notice the change in the table: the same 15 participants in N-P order read a narrative then a procedure. A different group of 15 participants in N-P order read a procedure then a narrative. Each individual participates in two different experimental conditions in this design. There are a total of 30 participants in four experimental conditions in this example. The participants are in two groups: the group for N-P order and the group for P-N order.

Exercise 3: Update your experimental design

- Classify each factor in your design: is it a between-subjects factor or a within-subjects factor?
- How many groups of different people will you need? Describe what the individuals in each group will do.
- Update the table for your experimental design, if necessary.
- How many participants total do you estimate the experiment will require?

Your experimental design should look something like this:

		Time Limit (between-subjects)	
		No Limit	Time Limit
TV Exposure (between-subjects)	No TV	*No limit; no TV* $n = 15$	*Time limit; no TV* $n = 15$
	With TV	*No limit; with TV* $n = 15$	*Time limit; with TV* $n = 15$

Total: 4 experimental conditions, N = 60 participants

Further Resources on Designing your Experiment

Bunge, M. 1967. *Scientific Research* (2 vols.) New York: Springer Verlag.

Christensen, L. 2004. *Experimental Methodology*. Boston, MA: Pearson Education.

Jackson, S. 2006. *Research Methods and Statistics: A critical thinking approach.* Belmont, CA: Thomson Wadsworth.

Krathwohl, D. 1988. *How to prepare a research proposal.* Syracuse, NY: Syracuse University Press. [especially Chapter 4]

Kumar, R. 1996. *Research Methodology.* Newbury Park, CA: Sage.

Lammers, W. & Badia, P. 2005. *Fundamentals of Behavioral Research.* Belmont, CA: Thompson Wadsworth.

Leedy, P. & Ormrod, J. 2005. *Practical Research: planning and design [8th Edition]*. Upper Saddle River, NJ: Prentice-Hall.

Mitchell, M. & Jolley, J. 2005/2007. *Research Design Explained.* Belmont, CA: Thompson Wadsworth.

Rudestam, K. & Newton, R. 1992. *Surviving your dissertation.* Newbury Park, CA: Sage.

Salkind, N. 2006. *Exploring Research.* Upper Saddle River, NJ: Prentice-Hall.