1 Experimental Methods

As a hands-on approach to experimental methods, it is useful to go through the process of participating in a simple research study in order to provide an example.

You may have participated in research previously for class credit or being paid for your time. As a participant, you will likely not have had a view of the experimental design from the perspective of the researcher, who planned the procedure that you followed with attention to the basics of research methods and an intention to test an experimental hypothesis.

Our first example experiment is a simple design with two conditions that is run online. As a participant, you would be randomly assigned to one of the conditions and follow the experimental procedure associated with that condition. To get a general feel for this process, look at a clock or phone with the time shown and if the minutes digit is even, follow the description of the design as laid out on page 2 (left side). If the digit is an odd number, follow the procedure as described on page 3 (right side). The text will step through a portion of the design to illlustrate the study.

This is a simple method for creating what we will see is a powerful element of experimental design, **random assignment to conditions**.

This brief experiment will be used as an example for classwork this quarter.

You will be shown a series of words, one at a time and asked to report something about the word.

For each word, rate how much you like the word on a 1 to 5 scale.
5 means you like the work very much and 1 means you dislike the word very much.

dislike a lot	dislike a little	neutral	like a little	like a lot
1	2	3	4	5

Rate each of the words to the right.

POCKET	VEHICLE
PRISON	BRAIN
QUARTER	PALACE
CITIZEN	FACTORY

Then proceed to page 4.

This brief experiment will be used as an example for classwork this quarter.

You will be shown a series of words, one at a time and asked to report something about the word.

For each word, count the number of vowels in each word.
Report that number from 1 to 5

1	2	3	4	5

Rate each of the words to the right.

Then proceed to page 4.

POCKET

PRISON

QUARTER

CITIZEN

VEHICLE

BRAIN

PALACE

FACTORY

In the actual experiment methodology, you would now be asked to spend 3 minutes answering trivia questions such as the one to the right.

Lord Byron and William Blake wrote primarily in what genre of poetry?

1. Modernist
2. Renaissance
3. Romantic
4. Victorian

Next is a recognition test.
For each word shown, decide if you think you saw and rated the word at the beginning of the experiment.
If you think you saw the word earlier, choose Old.

old				new
1	2	3	4	5

CLOTH CITIZEN
PRISON PORCH
DRIVER MUSCLE
PALACE BRAIN

Thanks for participating in the demonstration experiment

Experiment 1

The experiment outline presented on the last few pages illustrates a very simple experimental design. This example will be used to illustrate the typical path from theory through experimental design, data collection and analysis.

Our first experiment is based on some fairly old ideas in memory research but which hold up well for a simple demonstration experiment. The underlying ideas are described in Craik & Lockhart (1972), which lays out a **framework theory** for thinking about memory. Craik & Tulving (1975) reports a series of experiments that establish that manipulations designed to vary the *level of processing* or *depth of processing* have robust and reliable effects on measures of memory. While the core terminology and theoretical framing presented in these older papers is slightly out of date by more modern theories of memory function, the procedure still serves as an excellent example of a simple design that consistently produces a measurable effect.

We will first review the design of this experiment as an example of how to systematically work through understanding a study in psychological science to see how the data collected from a simple design supports drawing conclusions about human cognition.

Experiment 1 Design

For the following questions it is a useful exercise to try to answer the questions yourself as you read on. This will help you assess how much of the basic terminology and experimental approach you are already comfortable with. The terms will be defined in this chapter as a general reference. The goal here will be to use the main terminology, the bolded terms, frequently enough that it simply becomes part of your understood vocabulary without need to look definitions up later on.

What was this experiment about?

The general temptation for the answer to this question is to give a lot of detail about your experience with the experiment and guesses about how this relates to the underlying hypothesis. However, after just going through the experiment, you actually do not know what the experiment is about because you have not seen enough of the design. This is a typical experience for a participant in an experiment that has an **independent variable** that is manipulated **between-participants**. You only experienced one of the conditions, so the underlying **hypothesis** is not visible to you.

However, when we consider and evaluate research with examples as short summaries or drawn from published papers, we will always start with this question and the answer we are looking for in this very basic question is the highest-level **construct** that gives the overall domain of the experiment. Here, that is simply *memory*.

As we will see, designing an experiment in psychology generally starts with something we are trying to learn about. In psychology, that will be a concept like memory, perception, anxiety, relationships, language, identity, etc. One of the specific challenges of experimental methods in psychology, as opposed to other areas of science (chemistry, physics, biology), is that while we intuitively understand each of those concepts, there is a significant amount of effort needed to turn that idea into things that can be used in research. That process is called identifying the **operational definition** of the **construct**, which is essentially, how are we going to capture that idea in a controlled study.

Answering the next questions will require being familiar with some technical terms that you may have encountered in prerequisite classes. If you are unfamiliar

Key Terms

The bolded terms in the answers are key concepts in experimental design that will be used daily in class and throughout the text. A glossary of definitions is provided below for general reference.

with the terms, they are defined below for your reference. A glossary of the bolded terms is provided later in this chapter, however, these terms will be used so frequently throughout the text and class that they should become part of your vocabulary rather than something you need to look up.

What was the independent variable?

If you followed the instructions carefully then to answer this question, you would need some additional information. There were two different conditions used in this experiment described on pages 2 and 3. If you were in this study, you would have only experienced one of the conditions and would have no knowledge of the other one. Half of the participants in the study would be given instructions to rate how much they like each word as on page 2 and the other half would be given the instructions to count vowels as on page 3.

The **independent variable (IV)** is the conditions created by the experimenter and applied to the participants. Here it is the instructions given for how to read and engage with the list of words. A more interesting question is what **construct** is this **independent variable** an **operational definition** of? What is the construct that the experimenter is manipulating in this study? The answer is *depth of encoding* which refers to how much engagement the participants have with the meaning of the words in the study list. Understanding why this is an interesting factor to manipulate will require some background reading to become familiar with the theory (which we will get to later).

Here, *depth* is an **experimental operational definition**, which refers to turning this **construct** (concept) into conditions that can be applied to a research experimental design. Rating liking creates a higher level of depth by encouraging semantic engagement with the words. Counting vowels creates comparatively lower depth by focusing the participant on surface features of the word instead of meaning. The experiment is about how these conditions affect memory, which raises the next question.

What was the dependent variable?

The **dependent variable (DV)** in this experiment is a measured operational definition of memory, as in, how much memory did participants have of the word list after engaging with the work list in either of the experimental conditions. A measured operational definition turns a concept/construct into a quantitative number used to measure outcome. Here, the answer will be a numeric measure of performance on the recognition test that came at the end of the experimental protocol.

Although the examples here only presented a short list of 8 words, the design this experiment is based on had participants view and rate a list of 30 words initially. Then after a short delay/distraction task based on answering trivia questions, a recognition memory test was given. In that test, participants were presented with 60 words, the 30 seen previously and 30 new wordsnot seen before. Performance on this 60-item test is the outcome data to be evaluated in this experiment. Note that you might be tempted to answer the question of "what is the dependent variable?" with "the number of studied words you responded *old* to on the test." Here that is not quite correct as answering *old* to all 60 words would not reflect good memory (because you called all the new words old). More accurate is to describe the DV as score on the recognition test, which we can count as the number of test items responded to correctly, that is, old words called old and also new called new). We can also convert this number of correct items to simply consider a percent correct score on the recognition test. Higher scores on the test reflect more memory of the rated words.

Measuring Memory

If you are familiar with memory research, you might be familiar with more sophisticated ways to measure memory. A simple percent correct measure is enough for our simple study but not for all memory research.

State a hypothesis *relating the* independent variable *to the* dependent variable*.*

This is the first question that engages with the psychological science of the research study. The;previous questions identify the key elements of the design to provide a basis for figuring out what the study might tell us about human thought or behavior. Stating **hypotheses** about experimental variables is a deceptively tricky task. It requires that the stated hypothesis be testable or falsifiable, which is not the same as correct.

Any statement relating the levels of the independent variable (IV) to scores on the dependent variable (DV) are correct answers to a prompt like this. The hypothesis relating the experimental variables is: rating liking of words will lead to higher scores on the recognition test than counting vowels. Stating the opposite, that counting vowels will lead to higher recognition scores compared with rating liking is also an equally valid hypothesis, although we will see that it is false. That is, it is not supported by the data.

For the purpose of this question, stating the hypothesis in terms of the constructs would not be correct here. At some level, the experiment is about the hypothesis that deeper encoding of items being studied leads to better memory later. This is a perfectly valid hypothesis but in our analysis process we first focus specifically on how the experimental design tests a hypothesis about the experimental IV affecting the experimental DV.

An important part of analyzing an experiment is to find problems or errors in methodology. When we design studies, we need to consider our design critically to see if any errors have crept into our approach. And when we review research reports that we encounter and ask the question “how do they know that?” we should be looking for potential problems with the conclusions.

By explicitly framing the question in terms of the variables as asked here, we focus our attention on how the constructs of *deep processing*, *shallow processing*, and *memory* are implemented in this specific design. For example, memory here is operationally defined as a recognition test for the list of words. A statistically reliable result for this study allows us to make a confident statement about how this independent variable affected this dependent variable. However, extending the idea from this study to all other ways we might study memory is an additional step that we should consider carefully.

One of the important and unique aspects of psychological science is being aware of the difference between the experimental design and data, which are based on operational definitions, and the theoretical conclusions, which are based on constructs. In this design, the operational definitions led us to use lists of words as the things to be remembered and one specific approach to what we mean by *depth of encoding*. These might be important **limitations** to consider about our conclusions, for example, do they apply to non-word stimuli, or how does depth influence other kinds of ways to measure memory?

The data obtained will tell us about the relationships of the variables we used in the experiment, pending the appropriate use of a statistical test to evaluate the reliability of any effects observed. Following this, we hope to draw a theoretical inference about the constructs as the scientific conclusions about the study. Critically evaluating research requires being able to identify methodological issues that might limit those conclusions that arise at any step in the research process.

Psychology is a STEM field!

Research Methods is about using the scientific method to understand human behavior, attitudes, cognitive processes, social interactions, personality and mental health. It is fundamentally quantitative even though advanced math skills are not strictly necessary for basic design

What statistical test would we use to establish a reliable relationship between our independent *and* dependent *variables that would allow us to test our* hypothesis?

In virtually all psychological science, we are going to collect or consider data collected from a group of participants in our study. The people in our study are considered the **sample**, who are drawn from the larger **population** of all people. We want to make a broader statement than simply that the people who happened to be in this study showed better memory after deep encoding, we want to infer that deep encoding would likely improve memory for all people. Statistical analysis is the method for drawing that broader inference that deep encoding generally improves memory and future uses of deep encoding by anybody would most likely improve their memory for the studied material. This basic idea should be familiar from your prior study basic statistical methods from a prerequisite class. However, statistics will be used here in a potentially different manner than in prior classes. Here they will be the bridge from your numeric, quantitative data to statements about the conclusions and meaning of your study.

Since this is a simple two group design with participants randomly assigned to one condition or the other, the most appropriate statistical test would be a **two independent samples t-test.** While other more powerful approaches could certainly be used, it is generally most effective to use the simplest test that effectively communicates the main findings.

For simple experimental design, where participants are randomly assigned to one or two conditions of one or two independent variables, questions of reliability are generally simple and often relatively uninteresting. Our use of statistics here will therefore be streamlined. We will focus on identifying the correct test to use from a constrained set of options and extracting the key values from typical programs used to carrying out these analyses. The result of the analysis will be reported in standard format (based on the American Psychological Association; APA) as part of the process of writing up the results of a study. While a strong foundational grasp of the underlying mathematics is always helpful, we will primarily focus on how statistics are used to test research hypotheses and how to report these in a result that is complete and comprehensible to other scientists.

In carrying out a research project, statistics are used to establish the **reliability** of the effect of your IV on your DV. As we will see over the next several chapters, this is a separate question of the **validity** of your conclusions drawn from the study.

In general, the review of basic experimental methodology will focus more on validity of experimental design and data than reliability. Details of the statistical approach become more important as experimental design becomes more complex. Here as we review how to design research studies, we will focus on procedures aimed to obtain reliability that are assessed through statistics. When we review descriptions of published research findings, we know that the peer-review applied to these findings before publication generally establishes statistical reliability. By asking How do they know that? and applying an understanding of experimental design, we will identify questions of validity and alternate interpretations the findings that might differ from the experimenters' stated hypothesis.

Much of your ability to identify strengths and weaknesses scientific methodology will come from your understanding of human behavior as human. In this class, will augment this with some practice applying critical thinking skills systematically to these questions.

Psychology is the science of human behavior.

As a human, you have pretty good intuitions about how humans behave. Personal knowledge and experience can be a good starting point for the tricky problem of coming up with the operational definitions for psychological constructs.

Review of key experimental design terms

- **Experimental research**: The experimenter manipulates an independent variable and measures a dependent variable to test if the manipulation has an effect.
- **Construct**: The high-level concepts we aim to do research about. Typically, these things we that have an intuitive understanding of but need to be translated into specific experiment elements.
- **Operational definition**: Turning an intuitive but imprecise concept into something that can be measured quantitatively, or controlled categorically.
- **Measured operational definition**: A quantitative measure of a construct, essentially turning an idea into something that can be characterized as a number. For example, Experiment 1 operationally defines "memory" as percent correct on the test, a quantitative measure of the amount of memory obtained. A similar process might turn other constructs like anxiety, impulsiveness, attention into numbers that could be used as dependent variables in experimental design.
- **Experimental operational definition**: A controlled method of implementing a specific definition of a construct into levels or categories that can be manipulated by an experimenter in order to create the independent variable(s) for an experiment protocol.
- **Independent variable (IV)**: Often referred to by the acronym IV, this is the element manipulated by the experimenter to see if or how it affects the measure being collected in an experimental design. Controlled manipulation of the IV is the defining feature of experimental research.
- **Dependent variable (DV)**: Frequently referred to by the acronym DV, this is the measurement collected by the experimenter. The core idea in experimental research is to see how the scores on the DV change across the manipulation of the IV. If they do, we can conclude that the IV affected the DV.
- **Experimental Hypothesis**: A statement about the relationship between experimental variables that can be tested and importantly, falsified. If there are no data that would render a statement false, then it is not a falsifiable statement and is typically a description rather than a hypothesis. Typically the hypothesis is that the IV affects the DV, and we use statistics to reject the **null hypothesis**

(that the IV does not affect the DV). Note that hypotheses can be stated about the specific IV and DV used in an experiment but also stated separately about the constructs from which the IV and DV were operationally defined. Experimental data gives us confidence to make statements about the specific IV affecting the implemented DV but the goal of research is to draw inferences about the relationship among the constructs.

- **Limitations**: Concerns that conclusions about the underlying constructs might not be true in all cases and conditions other than the specific operational definitions used in the experimental design. Generally these are not issues with the fundamental **validity** of the experiment (Chapter 3), but questions about how widely the results can be applied. Identifying what limitations should be considered often requires some knowledge of the underlying theoretical ideas for a research study and can also indicate directions for future research. Using Experiment 1 as an example, we have data about memory for word lists measured with a recognition test a few minutes later. We might wonder if deeper encoding similarly affects memory for pictures, or if the effect might change with another measure of memory like recall. Studies examining those questions would reflect different operational definitions of memory, using different DV's and/or different operational definitions of deeper encoding as IV.
- **Statistical reliability**: We will evaluate whether the IV has a robust effect on the DV using standard statistical tools. Our focus here will be selecting the correct tool and reporting the use of the tool accurately. Statistics are often presented as a simple binary outcome: did the IV affect the DV reliably, can we reject the null hypothesis, was the probability of the null less than the criterion of .05 (these three statements are essentially synonymous). However, we will see that Psychological Science is moving towards a model of reporting **effect size** rather than relying on these binary descriptions. The effect size is helpful both with understanding the reliability of the statistics and also communicating the results. For Experiment 1, we might want to be able to say not just that deep encoding improved memory, but how much did this study approach increase our measure of memory?
-

Experimental vs Non-Experimental Research

A useful approach for understanding the definition of something complex, like experimental research, is to define what isn't experimental research. In non-experimental research, we also look for a relationship between an independent variable and a dependent variable, but the independent variable is not manipulated or controlled by the experimenter. For example, we could look for a correlation between your GPA and the score on the memory test in the demonstration experiment.

Non-experimental research is a powerful tool for psychological science as well as fields such as epidemiology, economics and sociology. However, the methods of the design of research studies and tools for analysis of data for non-experimental methods are quite different. The current approach focuses on experimental methods first, followed by some discussion of contrasting these methodologies for general reference in Chapters 9-11.

Experimental research has a significant advantage in drawing conclusions about how a manipulated variable (IV) affects a measured variable (DV). If we manage the challenge of adequate experimental control (Chapter 3-4) we can be fairly confident that changes in our DV were caused by our manipulation of the IV. However, experimental design is limited by needing conditions where we can create effective and accurate operational definitions of the constructs we want to study so that we can implement a protocol for a well-controlled laboratory experiment. There are a lot of important and interesting questions in Psychology that rely on data collected from the world in imperfectly controlled conditions.

Non-experimental research typically fights against the "correlation is not causation" problem and frequently uses more advanced quantitative analytic tools to improve our ability to draw causation from these data.

Experimental research uses simpler methodology and simpler analytic tools, making it an effective introduction to the design of psychological research.

Experimental Analysis

The following questions will be asked regularly about example designs and findings from psychological research. These will train your intuition to identify strengths and weaknesses of designs from short research descriptions. Later we will see how to read and write formal research reports following APA guidelines. Most of the research that you encounter will be in more informal context, but you can still ask the question: **How do they know that?**

- What is the experiment about?
- What is the dependent variable?
- What is the independent variable?
- What is the hypothesis or finding about how the IV affects the DV?
- What statistical test is used to establish a reliable effect?
- What is the conclusion drawn by the researcher?
- Do we see any problems with this inference?

Trying to identify the hypothesis and potential problems with the inference are the hardest but most important questions from this list. If there was a tried-and-true approach to always identify inference errors, professional researchers would never make mistakes about their findings (spoiler alert: they do).

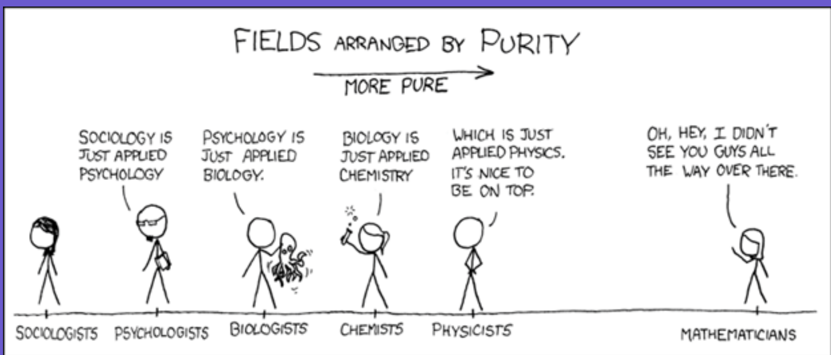
The first three questions depend on the operational definitions used by the researchers and how well they capture the intent of the research. When there is a mismatch, this often reflects differences in how people understand common phrases. For example, we might want to test a hypothesis related to an adage like “time flies when you are having fun.” One of the first challenges we would face is how to define the constructs of “time flies” and “having fun.” Different researchers would likely define these ideas in different ways and rather than saying that some operational definitions are right or wrong, it is important to understand that the different definitions reflect different design ideas. Experiments with different definitions might be quite properly constructed, but the conclusions drawn from carrying out the study might end up being very different.

Chapter 2 will discuss operational definitions as an example of Measurement Theory Mistakes in operational definition are one important source of error in experimental design. These can lead to studies where the results are quite robust, the IV clearly strongly affects the DV, yet the main conclusion of the study is inaccurately stated because the variables are ineffective operational definitions of the constructs they were intended to capture.

The question of what statistical test is appropriate for the research is necessarily more technical. As noted above, this class assumes background in basic statistics. In Chapters 5 and 10, we will review the process of selecting and carrying out the appropriate statistical tests for common experimental designs. The focus here is knowing which analysis to use, how to carry out the basic analysis procedure and most importantly, accurately state the inferences the analysis supports.

Understanding the hypothesis and conclusions that are tied to the IV and DV, the specific operational definitions used in an experimental design is the key to ensuring you understand how to read and interpret scientific findings. Being an effective reader of science and understanding what is confidently learned from the data obtained in a psychological study is a major goal of this class and text.

XKCD



drawn by Randall Monroe, is an exceptionally sharp source of perspectives on science, https://xkcd.com/435/

Experimental Analysis Practice Examples

Practicing experimental analysis and learning the common types of research design will give you critical thinking tools to help strengthen your understanding of science. We will practice via example throughout class meetings with a daily example to evaluate and analyze.

Example 1

Time flies when you're having fun, but what is it about pleasant experiences that makes time seem to go by faster? In one experiment inspired by prior work (Gable & Poole, 2012), researchers tested the hypothesis that approach motivation causes perceptual shortening of time during pleasant experiences. That is, it isn't just positive affect (fun), time goes quickly when you are specifically motivated to obtain a reward. Thus, they predicted that time spent viewing pictures of "delicious desserts" would appear to go by particularly quickly if you expected to get to eat one of the desserts after the experiment.

Participants were randomly assigned to either be told they would get to eat a dessert after the experiment or not. Then they each looked at 36 pictures of desserts each presented for a 12s and rated a scale of 1 (time dragged) to 7 (time flew), how long the picture had been presented.

Go through the Experiment Analysis questions for this example

What is the experiment about?	The subjective experience of time passing
What is the dependent variable?	The numerical scale rating from 1 to 7 of whether time dragged or time flew
What is the independent variable?	Told they would get a dessert after the study or not

What is the hypothesis or finding about how the IV affects the DV?	Participants told they would get dessert would score higher on the DV reflecting a feeling that time flew
What statistical test would be used to establish a reliable effect?	Two independent samples t-test

If the data were consistent with the hypothesis such that scores on the time-passing rating scale were higher for the participants who expected a reward, the researchers would like to conclude that expecting reward makes time feel like it is passing more quickly.

We should always consider limitations of the broad level conclusion. We might note that the task is particularly dull but also intrinsically linked to the reward (both are related to eating dessert). We might also note that the conclusion does not argue against the idea that time flies when you are having fun, but only suggests time might also fly when you expect dessert.

Example 2

Martin hypothesizes that self-esteem affects snacking behavior. He thinks that low self esteem will leads to increased opportunistic eating. He conceives of the following experiment. A group of 50 participants is recruited. All are given the opportunity to play a game of chance. They are all told that the odds are in their favor and that 90% of the people who play win the game. However, they are really assigned randomly to two groups: half win and half lose. The winners are congratulated and the losers are told, “Wow, that’s really unlucky. You must be a really unlucky person. Do you lose a lot of games like this?” Afterwards, all participants are then left alone in a room with a full bowl of peanuts for 15 minutes. The average weight of peanuts eaten during this period is compared for the 2 groups.

What is the experiment about?	Snacking behavior, self-esteem
What is the dependent variable?	Weight of peanuts eaten

What is the independent variable?	Whether the participants were made to feel that they were lucky or not
What is the hypothesis or finding about how the IV affects the DV?	Being told they were unlucky would lead to lower self-esteem and increase the number of peanuts eaten
What statistical test would be used to establish a reliable effect?	Two independent samples t-test

If the data were consistent with the hypothesis, the group randomly assigned to lose and be told they were unlucky would have consumed more of the peanuts left with the participants. This result could be statistically reliable but we might still have concerns about the broader conclusions. We would want to be confident that the experimental manipulation really did affect self-esteem. An **alternate explanation** for the results would be that feeling unlucky leads to greater snacking, without involving perceptions of the self that incorporate self-esteem. The existence of this alternate account for explaining the result does not mean the conclusion is wrong, it simply means that there is more than one way of understanding the data from the experiment and we do not yet know which is correct. These situations are often good opportunities for future research with novel operational definitions of the underlying construct. Note that such an **alternate hypothesis** for the data do not imply the results were not reliable, but that there is a question or limitation about the validity of the conclusion about the constructs. To highlight this different, it is best to separately consider the results of the experiment both in terms of the actual variables (IV, DV) and then the inference in terms of the intended constructs.

We've dived into experimental design and analysis very rapidly here and introduced a fairly large vocabulary of critical terms and concepts very quickly. If that seems daunting, don't worry! We will be going back over the concepts in detail to ensure a solid foundation of methodology design principles across a range of common approaches and research areas. If that seems too easy because design is straightforward, don't worry! While simple designs are easy, it gets complicated fast. If it were really easy, trained and professional researchers wouldn't make mistakes in their research conclusions (spoiler alert: they do).