



Theory, Literature, and Hypotheses

**In: Designing Experiments for the Social Sciences: How to Plan,
Create, and Execute Research Using Experiments**

By: Renita Coleman

Pub. Date: 2022

Access Date: February 23, 2022

Publishing Company: SAGE Publications, Inc

City: Thousand Oaks

Print ISBN: 9781506377322

Online ISBN: 9781071878958

DOI: <https://dx.doi.org/10.4135/9781071878958>

Print pages: 55-88

© 2019 SAGE Publications, Inc All Rights Reserved.

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

Theory, Literature, and Hypotheses

*Full-blown successful interventions never emerge from a brainstorming session. They are always suggested by theory, previous research, or extensive clinical experience.*¹

—R. Barker Bausell

Learning Objectives

- Explain the role of theory and literature in experiments.
- Prepare a literature review that connects previous work to your own and makes a theoretical contribution.
- Describe the role of hypotheses and research questions in experiments.
- Summarize when to predict differences or direction in hypotheses.
- Formulate original hypotheses using the diagram given.

To Bausell's assertion, we should add that experiments are not **exploratory** in the sense that they are used when problems are in the preliminary stage.² Rather, experiments presume a fair amount of well-developed theory and evidence from previous research. Thus, it is important to do the homework before starting any experiment, which involves a thorough search of the literature, including theory. This chapter builds upon the previous one by examining the role of what researchers have discovered before us, or Isaac Newton's concept of "standing on the shoulders of giants."³ While the information in this chapter also applies to other research methods, it is especially true that experiments are designed to create new knowledge by building upon what has been discovered previously. It also addresses two of the seven "attributes of a well-executed experiment."⁴ That includes theory **explication**, or explaining the theory being tested, and clear identification of research questions and hypotheses. It looks at the literature review section of the paper from the perspective of how to explain the theory being tested, relate it to the variables used in the experiment, and link all the hypotheses to that theory. Hypotheses are particularly powerful analytic tools in experiments, so knowing how to write a literature review and the hypotheses that arise from it is especially important for this method.



iStock.com/barammee2554

As covered in [chapter 1](#), the first step in designing an experiment is to formulate an idea of what causal

relationships are to be tested and clearly articulate that in the statement of purpose. But these relationships should not be based merely on hunches or curiosity. Preferably, the idea of what to test should grow out of an understanding of the domain and the literature and theories within it. Measuring causality requires thinking theoretically; otherwise, experiments would be a haphazard process of trial and error.⁵ (For more about theory, see More About box 3.1.)

One common pathway by which social scientific knowledge can develop is by starting with exploratory work using methods such as observations, interviews, and ethnography, among others. These types of qualitative methods are explicitly designed to be descriptive and generate theories.¹⁴ They reveal **latent** meaning—that is, not readily apparent on the surface—or help us understand the nature of phenomena and predict the conditions under which they occur.¹⁵ But because of their inherent qualities (small samples, not randomly drawn, lack of control, etc.), they are not well suited for testing or refining the theories they generate. That is one of the main purposes of experiments.¹⁶ After theory is generated by qualitative or other means, research then can progress to methods that test them. Surveys are good for eliciting opinions, attitudes, and self-reports about behavior. Content analyses can uncover the characteristics of various types of messages, text, or visuals. While these methods produce research that adds new knowledge and expands theory, they also suggest causes and provide correlational evidence that lays the groundwork for experimental work. Experimental studies are then able to determine if the phenomenon observed in a natural setting can be reproduced in a controlled environment, where all the many variables that could confound results or be responsible for plausible alternative explanations in the real world can be ruled out. Having arrived at the place where an experiment is the next logical method, the first step is to conduct a thorough search of the literature and particular theory.

More About . . . box 3.1 Theory

Theory is a generalized explanation for some particular phenomenon, or “an organized body of concepts and principles.”⁶ This is an academic way of saying that theory is about things that actually occur in the real world. To invoke Kurt Lewin, the father of gatekeeping, “There is nothing more practical than a good theory.”⁷ Although theory can seem mysterious and hard to comprehend, it really is nothing more than a tool for understanding what happens in life. Theories evolve when hypotheses are continually supported.⁸ A good theory should do more than describe; it should also explain and allow predictions about the phenomenon. For example, one aspect of framing theory describes how news stories are constructed as discrete episodes—a story about a crime, for example—or more generally as themes that add context—a story about crimes, not just

one in particular. It explains how people perceive these types of frames and also allows us to predict how people will react to them—for example, by holding individuals responsible when events are framed episodically but placing blame on social factors when events are framed thematically.⁹ Experiments build theory by testing assumptions and expanding upon our existing understanding.

However, there has been some discussion of “exploratory experimentation” in biology and related disciplines.¹⁰ What these researchers mean by an exploratory experiment is not an atheoretical, aimless data-gathering project. Rather, an exploratory experiment is one that is *directed* or *informed* by theory but does not necessarily *test* it via formal hypotheses.¹¹ These studies have a theoretical background, not a lack of theoretical framework altogether.¹² Theory is still “crucial to the success of exploratory experiments.”¹³ It is never appropriate to justify a lack of theory, low power, subjects not randomly assigned, uncontrolled variables, or other weaknesses by labeling the experiment as “exploratory.” In these cases, the researchers should consider it to be a pilot study, take what they learned, and mount another experiment that improves upon the weaknesses rather than attempt to publish it as an exploratory work.

All this is not to say that researchers can never come up with ideas for experiments by other means, such as by observing and becoming curious about why a particular phenomenon happens, as James Lind did for his study of scurvy. Ideas generated this way can certainly be tested with experiments. But there also should be some reason for testing them based on theory or evidence. When pursuing research questions generated from observations, it is important to find empirical evidence and theories that can help guide and refine them. This purpose can be served by correlational studies that link two variables, as can experiments that use a different population or different dependent variable (DV). For example, in a study to see if journalists’ moral judgment could be improved by showing them photographs,¹⁷ no work had been done using moral judgment as the DV and photographs as the independent variable (IV). However, photographs had been shown to improve other cognitive processes such as elaboration, empathy, and involvement.¹⁸ All of these DVs are related to moral judgment, as shown by other studies.¹⁹ These theoretical and empirical rationales provided links to the independent and outcome variables, albeit indirectly. This idea came about as a result of much speculation in essays and other nonempirical works about how journalists’ ethical reasoning improved when photographs were involved.²⁰

The Literature Review

The explanation of the theory being used and how it relates to the variables of the study is achieved in the

literature review section of the paper. A well-done literature review provides a strong foundation for the entire experiment, yet Bausell says the “single rule most often violated by beginning researchers” is not knowing the relevant literature thoroughly enough.²¹ Seasoned researchers may not feel the need to write a formal literature review in advance of running an experiment because they keep abreast of the latest findings and also work with theories they know well. But for others, before forming any testable hypothesis, a thorough search of pertinent literature and familiarization with relevant theories should be undertaken. The first step in this process is to perform a search of the literature, and the second is writing it up in the literature review section of the paper, which is the subject of this chapter.

Reading others’ research can be a great source of ideas for experiments. First, reading existing studies will show whether the idea has been done already, so the study is not just a simple replication.ⁱ It will also reveal gaps in what has been done and what still needs to be done to flesh out a theory. The statements of “limitations” and “future studies” at the end of every research article can provide even more ideas. This includes ideas for studies that need replicating and for extending and improving upon a theory. Reading literature also provides methodological help, including information about how others have measured the variables, how many subjects might be needed, and ideas for novel procedures.

The screenshot displays the SAGE Academic Search Complete interface. At the top, there are navigation tabs: New Search, Publications, Subject Terms, Cited References, and More. On the right, there are links for Sign In, Folder, Preferences, and Language. The search bar contains the query 'experiments AND social science'. Below the search bar, there are options to 'Select a Field (option...)' and buttons for 'Search' and 'Clear'. The search results are displayed in a list format. The first result is '1. Undertaking Experiments in Social Sciences: Sequential, Multiple Time Series Designs for Consideration.' by Phan, Huy; Ngu, Bing. The second result is '2. Field Experiments Across the Social Sciences.' by Baldassarri, Delia; Abascal, Maria. The left sidebar shows the 'Refine Results' section with 'Current Search' and 'Limit To' options. The 'Current Search' section shows the search criteria: 'Boolean/Phrase: experiments AND social science'. The 'Limit To' section has checkboxes for 'Full Text', 'References Available', and 'Scholarly (Peer Reviewed) Journals'. There is also a 'Publication Date' range from 1932 to 2018.

This book presumes that readers are familiar with how to conduct a literature search and write a review, perhaps even having conducted and written one or more already. Because an experiment is so much work on the front end of designing it, and it cannot be changed once the data have been collected, it is especially

important that the literature search for an experiment be performed thoroughly. I recommend actually writing a draft of the literature review before proceeding to collect data no matter how familiar you are with the topic. I have found that the act of writing slows me down and gives me time to think about the literature in a way I would not if I had just been reading it. It also forces me to synthesize the many different findings, which helps me better analyze the meaning and find gaps that may need filling. Making a chart or table can help show things more clearly, and sometimes these figures end up in the published study—for example, [table 3.1](#).

With experiments, it is necessary to isolate or control all things that could possibly affect the outcome or DV except for the one thing being manipulated, so it is especially crucial to know what all those things are. They can then be controlled, or used as mediators, moderators, or covariates. If a researcher does not know what they are and fails to measure or control them, the validity of the findings for the entire experiment is put in jeopardy.

There are many good books and articles on how to perform a literature search and write a literature review, so those will not be described in detail here. Instead, next is advice for avoiding some of the most common problems with literature reviews in experimental designs. These tips are not exclusive to this method, but the problems they address occur often enough in experiments that they are worth mentioning.

Table 3.1 ● Indicators of imaginative capacities (Liao et al., 2016)

Indicator	Definition	References
Novelty	The ability to generate unique ideas	Beaney (2005) and Vygotsky (2014/1967)
Productivity	The ability to produce thoughts using the extensive application of free association	Folkman (2010) and Gaut (2005)
Concentration	The ability to formulate thoughts through focus and immersion	Csikszentmihalyi (1996) and Folkman (2010)
Sensibility	The ability to evoke feelings during the creative process	Ricoeur (1978) and Scheffler (1986)
Intuition	The ability to generate immediate associations with a goal	Reichling (1990) and Townsend (2003)
Effectiveness	The ability to generate relevant and profound thoughts to attain a goal	Gilbert and Reiner (2000) and Shin (1994)
Dialectics	The ability to seek improvement by logically analyzing possibilities and alternatives	Cartwright and Noone (2006) and Reiner and Gilbert (2000)
Exploration	The ability to inquire about the unknown	Coiello (2007) and Thomas (2004)

Crystallization	The ability to visualize abstract concepts by using concrete examples	DeVries (1988) and Vygotsky (2004/1967)
Transformation	The ability to perform tasks by applying information acquired across multiple fields of knowledge	Kunzendorf (1982) and Liu and Noppe-Brandon (2009)

From: Liao, Kai-Hung, Chi-Chen Chang, Chao-Tung Liang, and Chaoyun Liang. 2016. "In Search of the Journalistic Imagination." *Thinking Skills and Creativity* 19: 9–20.

Tips on Writing the Literature Review for Experiments

It Is Not a Book Report

It is important to analyze and synthesize the findings of the various studies found in the literature search. Too often, literature reviews read like book reports. They are summaries of the findings of previous studies, but there is no analysis of those findings that shows the connections among or between them and the study about to be conducted. Do not assume that readers will be familiar with the topic; therefore, how the studies reviewed connect to each other and to the current study will not be as obvious to them as to you. Spell it out. Furthermore, experiments test hypotheses, which are proposed based on previous findings; therefore, an analysis of those findings is especially important for this method.

Rather than stringing together summaries of the abstracts (Smith and Jones found . . . ; White, Brown, Green, and Black showed . . . ; Tinker and Taylor studied . . .), a good literature review is a critical analysis of the studies, synthesizing the findings, and coming up with new ideas of your own. This approach also avoids what can look like shameless name-dropping or an attempt to bump up an author's citations or a journal's **impact factor**—a measure of how important it is.

As an example of analyzing and synthesizing, I relate a story of a colleague and I who were doing an agenda-setting study to discover whether issues in the news or attributes of candidates had more powerful effects on voters. We conducted a literature search and noticed that the correlations between the media's agenda and the public's agenda (the basic premise of agenda-setting theory) were consistently higher for attributes than for issues. That was the foundation for the resulting hypothesis, which predicted that attribute agenda setting would be more powerful than issue agenda setting.²² This was not something we found already written about but discovered by thinking critically about the theory and comparing the findings of all the studies.



iStock.com/wellesenterprises

Do Not Be Wikipedia

Too many literature reviews read like Wikipedia entries. Instead of reviewing everything ever written about a topic, a good review covers only the literature related to the study. For example, if the experiment is on second-level agenda setting, it might be important to briefly mention what first-level agenda setting is and say how it differs from second level, but it is not necessary to explain it in detail, summarize every study on it, and describe all the mediators, moderators, or contingent conditions—that is, when it works and does not. Instead, mention it and say how it differs from what this experiment will test, and move on. In other words, only focus on B if the study is about B, even if there is also an A. Do be sure to cover all the *relevant* literature; some authors go too far in the opposite direction and leave out important work. Key is finding balance—cover just the right amount and do not leave out any of the truly seminal studies, or the classics.

This also applies to the theories written about in the literature review. There are many theories that overlap and can be used to inform an experiment, but a study should typically use only one or two. For example, there are at least ten theories that can be used to explain aggression from watching television.²³ In studies using other methods, reviewing many similar theories may be more appropriate than in an experiment designed to test a specific theory. Find the one theory that is best suited to what the experiment is trying to do, and write a solid rationale for why it is used and not others rather than reviewing every possible theory. Sometimes, two theories will lead to opposite predictions, and so those can be juxtaposed in a single study.²⁴ In this process, a compelling link should be made between independent, dependent, and **intervening variables**, or those that come between the treatment and the outcome.

Study Spotlight 3.2

Shoemaker, P., J. Tankard, and D. Lasorsa. 2003. *How to Build Social Science Theories*. Thousand Oaks, CA: Sage.

Novice researchers frequently have a difficult time understanding not only what theory is but especially how to build it. It is not uncommon for journal reviewers to ask authors: “What is your theoretical contribution?” “What theoretical statements does this study

offer?” and “What does this study do to build theory?” Finding ways to answer these concerns keeps many a budding social scientist up at night. When this book on how to build theory in the social sciences came out, it spoke directly to my inner experimentalist in its logical, step-by-step approach. It delivered on its promise to teach “the challenging activity of theory building with the minimum amount of difficulty.”²⁵ This book begins by assuring readers that they are not alone and that there is no shame in admitting a lack of understanding of how to build theory and provide theoretical contributions.

The chapters build logically, beginning by describing hypothesis writing. It then explains theoretical concepts, which it calls “the building blocks of theory.”²⁶ It illustrates ways to combine concepts into theoretical statements, which broadly describe relationships among variables. It shows how to synthesize disparate parts and make linkages, both theoretical and operational. Then it shows how to describe and explain the relationships among concepts in ways that offer explanatory power. Chapters deal with how to relate two variables, three, four, and more, and how to build models in advance of a full-blown theory.

The book offers examples that social science scholars can relate to. The development of cultivation theory by George Gerbner and colleagues is one, showing how it began as a way of explaining the effects of television on viewers.²⁷ One hypothesis proposed that watching television shapes viewers’ beliefs, ideologies, and worldviews. As studies were conducted and the hypothesis supported, it grew to include other concepts with labels such as “resonance” and “mainstreaming,” which specified the conditions necessary for the effect to occur. More research resulted in more revisions and additions to the theory, including first- and second-order effects that delineated perceptions based in reality and beliefs.

How to Build Social Science Theories stresses the importance of creativity to building theory, an idea that is also important in this book. It offers concrete, practical ways to make contributions to theory. How to develop theory never really “clicked” for me until I read it.

Make a Theoretical Contribution

Far too many studies of all kinds will summarize a theory and then go on to do whatever the authors intend to do but never circle back to the theory at the end. They often fail to explicitly say what the study intends to do to contribute to that theory or theories. With experiments, this is an especially egregious oversight given that one of the main reasons for doing an experiment is to develop, test, refine, or otherwise contribute to theory. It is never enough to lay out a theory as the foundation and then abandon it. It should be explicitly stated what this experiment will do for the theory or theories, and that should be expanded on in the discussion and

conclusion of the paper. (For a synopsis of a book on how to build theory, see Study Spotlight 3.2.)

Toward this end, it is important to link all the variables—dependent, independent, causal mechanisms, and individual difference variables—clearly explicating the relationships among them.²⁸ Start by stating how the IV will affect the DV, and then explain through what theoretical processes this will occur. If it helps, make a figure that models this and use it to visualize the process (for an example, see [table 3.1](#)). As an example, the study on photographs' effects on moral judgment²⁹ mentioned earlier was interested in establishing a causal relationship between journalists who were exposed to photographs and their levels of moral judgment compared to journalists who were not exposed to photographs. The study theorized this would occur because photographs had been shown to increase empathy, which intensifies involvement and also encourages central route processing, all of which were found to increase moral judgment. The theoretical contribution was not only that photographs had the ability to improve moral judgment—an intervention that had not been tested previously—but also the demonstration of the theoretical explanation for *why* photographs worked.

Connect the Dots

Finally, another important feature of a good literature review is to relate other studies to your own, what I call connecting the dots. After reviewing every batch of studies on a common topic, explicitly say how they are related to the study to be conducted. Will it be measured the same way? Do they contain gaps that this one will fill? What connections in these studies will this experiment build upon? Always relate the findings of other studies to your study, locating it within the context of the larger body of work. A reader should never have to go more than a paragraph or two to find out why specific studies are being reviewed and how they are connected to the current one. For example, in the study mentioned earlier that examined the correlations from various agenda-setting studies, the paper explained how they were relevant to the current study by saying, "All these second level correlations are similar to the highest correlations at the first level; none reaches the lowest levels or even the mean level of .53 in meta-analysis of the first level effects."³⁰ That is an example of a sentence that connects all the studies just reviewed to the current one.

How To Do It 3.3 Examples of Connector Sentences

In a literature review, it is important to relate studies to each other and to your own, locating your work in the context of others. I offer here some examples of how various researchers have made this explicit in their experiments.

- After reviewing the three sequences in the Hierarchy of Effects Model, this study said, "We are not particularly concerned with the sequence but rather with the strength of the first two components of these models—knowledge and attitudes—on the behavioral outcome, voting. In other words, we seek to determine not which came first, but which has a stronger effect: cognition about issues (first level) or feelings about attributes (second level)."³¹
- After reviewing celebrities' Twitter use and source credibility, the authors say, "Acknowledging

the aforementioned novelty and importance of celebrity-generated messages embedded in social media, this research tested the effects of celebrities' Twitter-based electronic word-of-mouth on consumers' perceptions of source credibility, intention to spread electronic word-of-mouth, brand-related outcomes, online bridging social capital, and social identification with celebrities."³²

- Before reviewing attribution styles, this study said, "This study delves into the role of consumers' different attribution style, specifically how different attribution styles lead to the differential impacts in evaluation of the negative celebrity information."³³
- In a study of math learning, the authors reviewed the literature and summarized the findings, and then tied it to the question this study was designed to answer this way: "Simply put, as math anxiety increases, math achievement declines. A possible inherent relationship between anxiety and achievement poses an obvious question, however: Is a poor performance on a math assessment/problem due to math anxiety or due to lack of mastery of the content?"³⁴
- After reviewing studies on different methods that were effective in helping students learn English, this study summed it up this way: "An examination of the salient characteristics and benefits of a technology-enriched curriculum for English Language Learners underscores the pivotal role Computer Assisted Language Learning can play in second language teaching and learning."³⁵
- Before summarizing literature, the authors of a study testing businesses that drop out of overcrowded markets note the similarities among the studies and point out the gap their work fills: "Although wars of attrition have an important place in the game theoretic literature, there are surprisingly few experimental studies directly relating to them."³⁶

These connections can never be too obvious. Good connecting sentences include: "This is relevant to our study because . . ." "We use this finding in our study by incorporating it as a mediator . . ." "These studies raise questions about . . . , which is the purpose of this research," and similar phrases. Many good examples of connector sentences exist; some are presented in How To Do It box 3.3. As you read other studies, look for them and adapt to your own work.

Conceptual Definitions vs. Operationalizations

The literature review is also where all the important concepts and variables in a study are conceptually defined. A **conceptual definition** is an abstract, theoretical description of something using general qualitative terms. Students often confuse conceptual definitions with **operational definitions**, or operationalizations, which take a concept and define it in the specific, concrete ways it is measured. A **concept** is a general idea about something that has many specific characteristics. For example, credibility was conceptually defined as "the public perception of news quality"³⁷ in one study, and then it was operationalized with thirteen specific characteristics that included how fair, complete, accurate, believable, credible, informative, interesting, likeable, in-depth, important, and well written a news story was, plus how trustworthy the sources and information were. These are two different kinds of definitions, which appear in two separate sections of

a paper. It is not adequate to provide a list of terms to be used to measure the concept as the conceptual definition. The conceptual definition should instead describe and explain the concept based on theory. Another example is the conceptual definition of moral judgment as the reasons “that people use to decide that a course of action is ethically right or wrong.”³⁸ The operationalization given in the methods section is a list of twelve statements about things that were important to the subject when deciding what to do about a particular dilemma (for examples, see the footnoteⁱⁱ).

The credibility study cited previously also provides a good example of a concept that has been defined differently by various researchers, including as “attitude toward a source of communication,” “global evaluation of the believability of the message source,” a combination of trustworthiness, expertise, and goodwill, and as the combined public perceptions of media’s approach to “quality, profit making, privacy, community well-being and trustworthiness.”³⁹ It is not uncommon to find a lack of consensus about the conceptual definition of some concept. When that is the case, it is important to acknowledge that in the literature section, review the different conceptual definitions, and then defend your choice of definition and explain why it is best suited for your study. To repeat: Every variable used in a study must appear in three places:

- conceptually defined in the literature review,
- incorporated in the hypothesis or research questions, and
- operationalized in the methods section.

Trace each variable throughout the paper and be sure that each one appears in all three places. There should be no orphan concepts or variables.

A good way to make sure that all bases are covered is to think of the literature review in terms of an outline that lists all the DVs, IVs, intervening variables (mediators, moderators, causal mechanisms), and covariates, if any. Use subheads that mirror the outline to help clarify it for readers and also to help organize the writing. This ensures that everything is defined, the empirical evidence is covered, and all the discussion about one variable is in the same place. Experienced readers expect to find all the information about a particular concept together and can become annoyed when something pops up in different places throughout the paper.

Literature Reviews With Multiple Experiments

Many researchers report more than one experiment in a single article for reasons that range from replicating a finding to ruling out plausible alternative explanations. Sometimes this is planned beforehand, and other times the need arises only after the first study results are known. This phenomenon will be explored in [chapter 5](#) on validity. When multiple experiments are reported in one paper, there may be one section that covers all the literature for both studies, or there may be two literature reviews, the first one the longest and the second that includes only new literature that pertains just to the second study. Usually, these will be labeled “Study 1” and “Study 2,” or more if there are more experiments. The general pattern is to report the literature

for the first study, followed by the methods for it under a subhead such as “Study 1,” then the results and discussion of it. The second study will follow under a subhead “Study 2” or “Experiment 2,” starting with a description of it and a small literature review if necessary to explicate any new variables. For example, if new IVs, mediators, or moderators were introduced in order to rule out confounds or plausible alternative explanations that may have arisen in the first study, they obviously were not covered in the first literature review. It is not necessary to repeat all the previous information from the main literature review, only to add the new concepts and put them into context with the previous literature and theory. A good illustration comes from a study of cognitive processing in agenda setting.⁴⁰ The second experiment was conceived of after the first one showed the effects the authors expected on some subjects but not others. They speculate this could have been affected by the content of the stories they used, and write: “We decided to conduct a follow-up study to test the assumption that people who process information more centrally than others could be influenced by content aspects of the articles such as journalistic evaluations of an issue’s importance.”⁴¹ They follow with the subhead “Study 2” and begin with a one-page literature review just of studies that deal with content of news stories.

Finally, a few miscellaneous but still important points for the literature review:

- *Include current sources.* In addition to the seminal or classic studies, be sure to include up-to-date works. This is especially important if a study was started some time ago. Go back and see what has been done recently and include it.
- *Use different search methods.* Searches using key words in databases do not always uncover all the important literature, even when using multiple databases. To avoid missing important studies, go through the last two or three years of journals that are likely to publish articles on the topic to see if anything has been missed. Include the journal you plan to submit to in this process.
- *Include conflicting findings.* Very few programs of research always produce consistent findings. Look for studies that *did not* work as expected. Not only is it important to be honest, it is important to know what might cause your experiment to not work as planned so you can control for it as much as possible. Review the studies with findings that do not support your predictions and explain why they might have come out this way.
- *Use quotes sparingly.* Quoting others at great length or too often raises red flags for readers because it can appear as if you did not understand what the author was trying to say. Paraphrasing someone else’s writing also affords the opportunity to write it in a clearer way. Make it a habit not to use direct quotes unless there is absolutely no better, clearer, more concise way to say it. There are very few sentences in academic papers that meet this criterion. When direct quotes are used, cite them appropriately with page numbers.
- *Be careful not to plagiarize.* See preceding point. Much plagiarism results from careless practices rather than bad intentions—for example, taking notes by copying and pasting from the original article and then forgetting whether these were already rewritten in your own words. Repeating your own writing from a different article is also a form of plagiarism. Develop a system to know what is

paraphrased and what is not, for example, by highlighting, using quotes, or putting them in a different type font.

- *Always read the paper.* It can be time consuming to trace back and read the original papers cited in others' literature reviews, but relying on another author's interpretation of a paper rather than reading it yourself can be perilous. Scholars have been known to uncritically report the same summary of a study over and over, perhaps even wrongly.⁴² I find this to be the case particularly when books or articles are written in another language and not translated.

Hypotheses and Research Questions

The literature review is also where the hypotheses and research questions are typically incorporated. One approach is to list all hypotheses at the end of the literature review. This style involves a block of hypotheses before the methods section, like this:

"Hypothesis 1 (H1): Individual donors will donate less when they know that a nonprofit receives some amount of government funding. (Categorical crowding-out hypothesis)

Hypothesis 2 (H2): Individual donors will donate less if a nonprofit receives a greater share of its funding from government. (Continuous crowding-out hypothesis)

Hypothesis 3 (H3): Knowing that government funding comes from a competitive merit-based program will increase individuals' willingness to donate. (Crowding-in hypothesis)"⁴³

For the purpose of illustration, in these hypotheses I have slightly edited the second one to keep the wording parallel; the original paper uses "give" in H2, while the other hypotheses use the term "donate." As "give" and "donate" could conceivably be defined differently, it is important not to use synonyms in writing hypotheses. Instead, use parallel wording for hypotheses whenever possible.



To keep readers from having to refer many pages back in the literature review in order to recall the evidence leading up to each hypothesis, some journals use a short summary of the literature that led to each prediction just prior to or after the hypothesis. As it can be inefficient to have the full description of studies in the literature review section and then to repeat a summary of it at the end of the literature review with the hypotheses, another style embeds hypotheses into the literature review at the point where the topic of each hypothesis is

covered. So, for example, literature on topic A is followed by a hypothesis related to A; literature on topic B is followed by a hypothesis related to B. That approach allows the empirical evidence to be fresh in readers' minds. For example, in a study on the climate of opinion judgments comparing the relative weight of explicit cues to implicit cues,⁴⁴ the authors first reviewed the literature on explicit cues and gave this hypothesis:

"H1: Surveys, as explicit media cues, influence recipients' judgments of the climate of opinion in the direction of the survey results."⁴⁵

They next reviewed implicit cues and predicted:

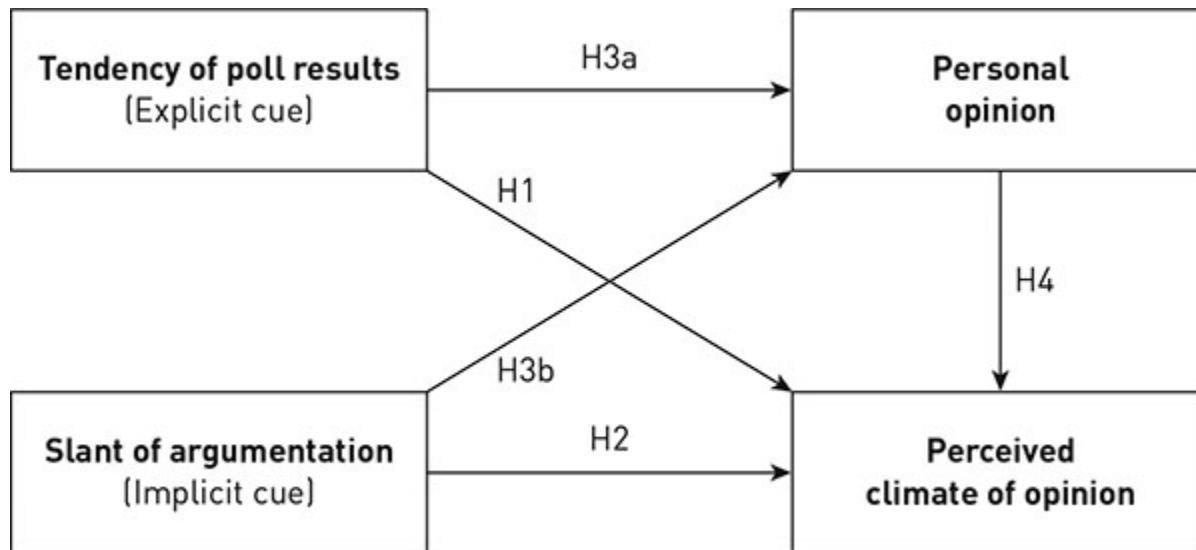
"H2: If arguments in a media report support (oppose) a certain opinion, the perceived public agreement to this opinion increases (decreases)."⁴⁶

They then reviewed literature on persuasive effects of both kinds of cues in media and gave this hypothesis:

"H3a: If survey information on an unknown issue is presented, recipients tend to follow the majority opinion."⁴⁷

There were several more hypotheses, with it all visually summarized in a model (see [figure 3.1](#)).

Figure 3.1 ● A Model Showing the IVs, DVs, and Process Proposed in Hypotheses



Source: Zerback, Thomas, Thomas Koch, and Benjamin Kramer. 2015. "Thinking of Others: Effects of Implicit and Explicit Media Cues on Climate of Opinion Perceptions." *Journalism and Mass Communication Quarterly* 92 (2): 421–443.

It can be a challenge not to introduce a concept in a hypothesis before having explicated it in the literature review. So, for example, do not introduce hypotheses that include the current or future state of something until the empirical evidence on that has been covered. Organizing the literature in such a way that it flows

naturally and does not assume reader knowledge of concepts before they are explicated is tricky. This makes it tempting to plop all the hypotheses in at the end, but then readers may have trouble recalling what was said about the concepts in H1 by the time they get to it. Instead, breaking hypotheses down into smaller statements, one IV at a time, in the order that the process occurs, is key.

The hypotheses in the previous examples are formally phrased and given the designation “H” plus a number (H1, H2, H3, etc.). The lower-case letters (H5a, H5b, H5c) indicate a subset of related hypotheses. But it is also quite common to see hypotheses phrased in a narrative style. For example, in a study of whether authoritarian governments are willing to include citizen preferences in the policies they enact, scholars in China posed their hypotheses this way:

Thus, based on previous research as well as statements by the CCP, we have reason to believe that when state–society relations are harmonious, leaders may be equally receptive to opinions expressed on the Internet and to those expressed through formal channels. However, when leaders believe that antagonism exists between the state and citizens, we expect that they will be less receptive to both formal and Internet channels.⁴⁸

Hypotheses should be written in future tense (e.g., “will have,” “may be”) rather than present tense (e.g., “is”) because they have yet to be tested; the present tense can sound like it is already a fact.

How many hypotheses are needed? Students often ask this question, but I am unable to give anything but the Goldilocks answer: Not too many, but not too few—just the right amount. One hypothesis may indicate a too-small study or the need to break a too-large hypothesis down into smaller bites. Hypotheses that run into the double digits may indicate a study that attempts to do too much and should be split into two publications. Let theory, logic, and reader attention decide. Have colleagues read the literature review to see if they can follow it clearly without getting lost or bored.

More important than how many hypotheses there are is how clearly they are written. If a process is specified, the hypotheses should build logically upon each other. Sometimes, experiments explore causal mechanisms that involve multiple steps in complex theoretical processes. When this is the case, it makes sense to break things down into a series of specific hypotheses that build upon each other in a cohesive model. For example, in the study of how photographs elevate moral judgment, the process was broken down into three hypotheses:

“H1: Participants who see photographs will have significantly higher levels of elaboration about stakeholders than those who do not.

H2: More elaboration about stakeholders will be significantly associated with higher ethical reasoning.

H3: Ethical reasoning will be significantly higher for participants who see photographs than for those who do not.”⁴⁹

The third hypothesis was what the study was most interested in, but it occurred last in the causal chain of events.ⁱⁱⁱ

Mutz offered a similar set of hypotheses, in narrative fashion, in a study of close-up camera perspectives and uncivil TV discourse (they are broken into separate paragraphs for easier reading here):

“The first hypothesis is that close-up camera perspectives and incivility will both increase levels of emotional arousal.

A second hypothesis following from the first is that heightened arousal will increase levels of recall.”⁵⁰

In this example, the first prediction involved close-up camera perspective and incivility leading to emotional arousal. The second prediction takes arousal from the first hypothesis and connects it to recall. The order is not chosen randomly; the researcher proposed them in this order because it was posited that they occurred in this order. With more involved processes, these pathways can be diagrammed out in figures to make it easier to visualize.

Hypothesis Basics

Null vs. Alternative

Before explaining how to write a hypothesis, this is a good place to review the concepts of null and alternative hypotheses, where a null hypothesis says there is no difference or direction, and the alternative hypothesis says there is. These were invented in order to reduce experimenter bias, or the tendency to find evidence that supports our ideas. A long time ago, scientists came up with the idea of basically arguing against your own position—that is, researchers should not try to show that a treatment causes effects but to *disprove* that it does not. In other words, to try to show that there is no relationship between two things, or no difference between two groups. The null hypothesis is written H_0 and the alternative H_a or H_1 . By rejecting the null, we say we have reason to think that there *is* a relationship or a difference, thus we can say there is a likelihood that our hypothesis is correct. This is technically the alternative hypothesis. In other words, we start from the position that the null (no difference, or no relationship) is true until evidence shows otherwise. If the null hypothesis is rejected, we say we have found support for our hypothesis. We never use the term *prove* because that cannot technically be done. Instead, the proper language to use is to *support* our hypothesis or *reject* the null; to reject the hypothesis or have it “*fail to be rejected*”; or to say that the hypothesis was “*not disconfirmed*.”⁵¹ This is rarely stated explicitly in research papers but is the thought process underlying them all.

A hypothesis is a prediction of how independent and dependent variables are related.⁵² They can be thought of as “if . . . then . . .” statements—for example, “If journalists see photographs, then they will use higher levels

of moral judgment.” Hypotheses should be stated in terms of concepts rather than operationalizations.⁵³ For example, in business, there is a concept called managerial trustworthiness, abbreviated MTW, and conceptually defined as “the trustworthiness attributed to supervisors.”⁵⁴ It is measured with nine items including capable, competent, concerned, and a strong sense of justice, among others. The hypotheses refer to managerial trustworthiness, not competence, justice, etc. For example:

“H1: Emphasis on internal management that relates to setting challenging but feasible goals has a positive effect on an individual’s perception of MTW.”⁵⁵

Difference vs. Direction

Hypotheses come in two flavors: those that predict a causal direction—for example, which of two variables will have larger effects than the other—and those that merely predict a difference but do not propose which variable will have a greater impact than the other. These are also called *nondirectional hypotheses*. An example of a hypothesis that predicts a difference is:

H: Journalists who see photographs will use significantly *different* levels of moral judgment than journalists who do not see photographs.

It predicts a difference between seeing and not seeing photographs, but it does not specify whether the outcome will be higher or lower. By this account, seeing photographs could result in moral judgment that is either better or worse. By contrast, this version of the hypothesis predicts a direction:

H: Journalists who see photographs will use significantly *higher* levels of moral judgment than journalists who do not see photographs.

The second hypothesis goes further than the first to say specifically what the difference between the two conditions—seeing photographs or not—will be.

Another example of a directional hypothesis is from the previous example of a study on managerial trustworthiness:

“H1: Emphasis on internal management that relates to setting challenging but feasible goals will have a positive effect on an individual’s perception of MTW.”⁵⁶

This hypothesis predicts a causal direction with the use of the word *positive*; the researchers expect *higher* levels of MTW in the goal-setting condition. A nondirectional way to state this hypothesis would be to say that goal setting would have an effect on MTW that was different from the other conditions, but not to specify whether that effect would be positive or negative.

Whether to make a hypothesis that predicts a direction or a difference depends on the theory and empirical evidence available; if there is enough to suggest a causal direction, then it is appropriate to make that

prediction. If there is not enough evidence, or theory does not suggest a direction, then it is more appropriate to predict a difference but not specify a direction. For example, the first time the effects of photographs on moral judgment were tested, there were theoretical reasons to believe that photographs could either improve or worsen moral judgment. In that case, it is appropriate to propose that seeing photographs will cause subjects to behave differently without predicting whether their moral judgment would be better or worse than those who did not see photographs. After a study finds effects in a certain direction, from then on it is appropriate to use directional hypotheses.

If theory and prior evidence is in such short supply as to prevent a researcher from making a prediction at all, then a **research question** should be used. A research question represents a more preliminary state of affairs concerning theory and evidence.^{iv} So continuing the previous example, if there was even less theory and evidence to make the prediction that photographs would have some effect on moral judgment, it is appropriate to ask, “Does seeing photographs affect journalists’ moral judgment?” As a rule of thumb, it is always best to make the most specific prediction possible with the available evidence and theory, as that provides us with more powerful analytical tools.⁵⁷



Whether a researcher makes a prediction of direction or not is important in how hypotheses are tested statistically. What follows may not make sense to readers who have not taken an introductory statistics course; for them, the point is that the type of hypothesis used matters later in the data analysis phase. For those who are familiar with basic statistics, a one-tailed test is appropriate for testing a directional hypothesis. For a nondirectional hypothesis, a two-tailed test is used. (For simplicity, these examples assume two groups—one treatment and one control.) The graphic illustration of that is found in the normal distribution. A nondirectional or hypothesis of differences is tested with a two-tailed t test. Because no direction was specified, the hypothesis is supported if the t value falls in either end of the normal distribution, marked by the blue shaded areas in [figure 3.2](#).

A directional hypothesis is only supported if the t value falls in the end of the normal distribution that is specified in the hypothesis. In this example, the hypothesis is that photographs would lead to *higher* levels of moral judgment, so the t value must fall in the blue shaded area to be significant, represented in [figure 3.3](#). Practically speaking, if significance exists, it is easier to find it with a directional hypothesis; however, if the direction predicted is wrong, then a statistical test may find significance but in the opposite direction that was predicted, in which case the hypothesis is not supported.

Figure 3.2 ● Nondirectional Hypothesis

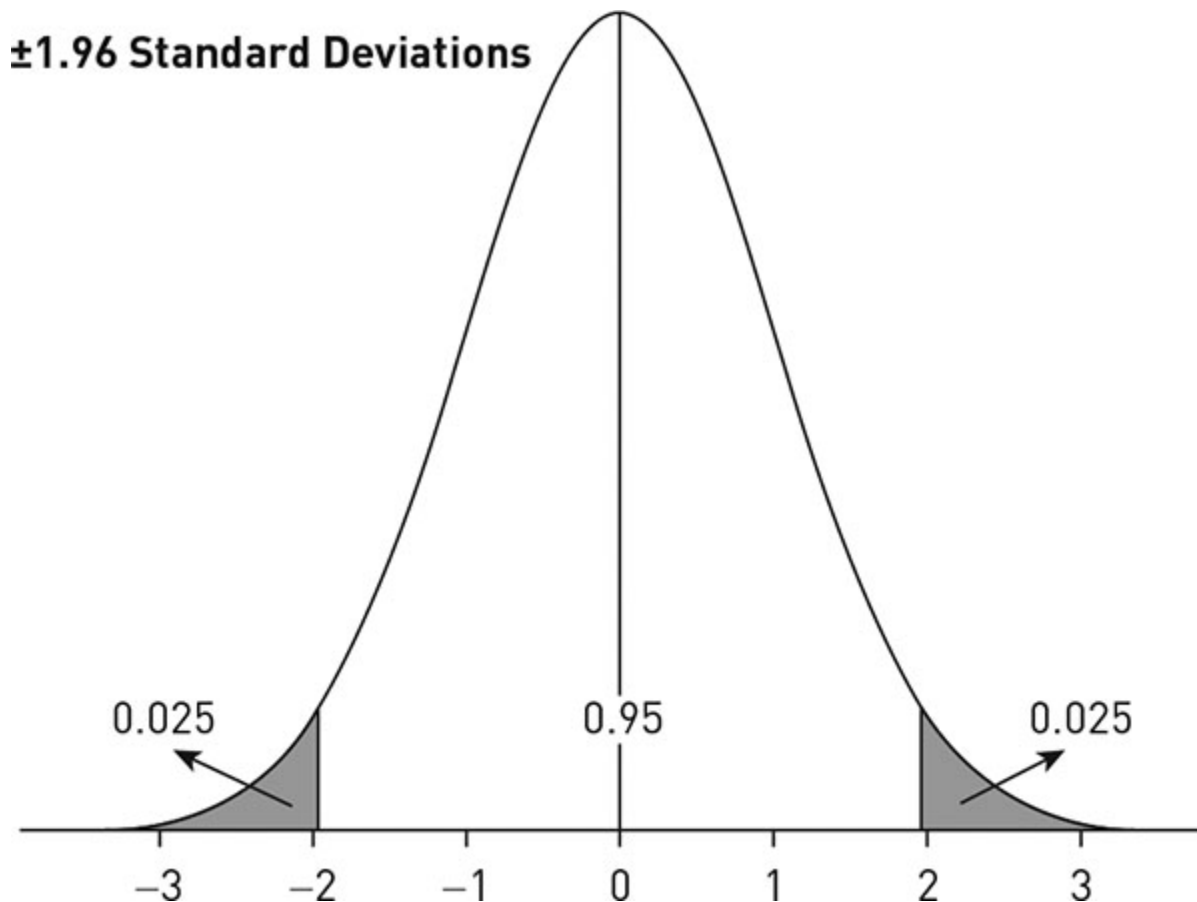
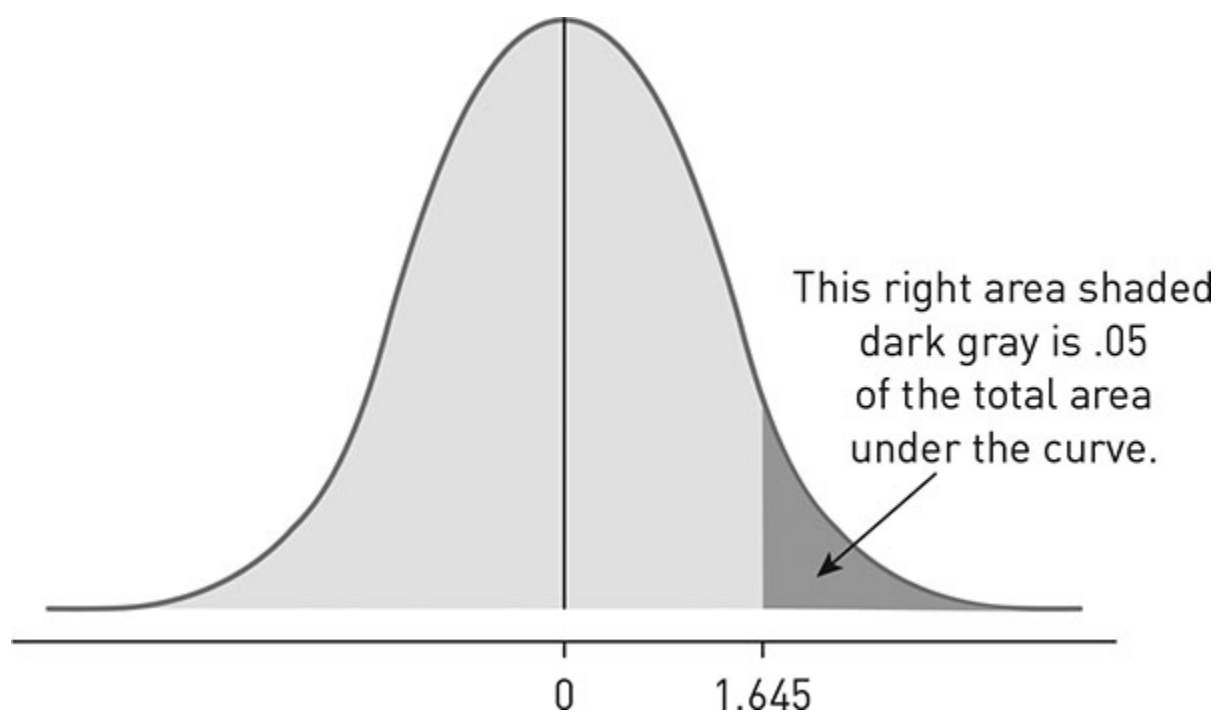


Figure 3.3 ● Directional Hypothesis



Hypothesis Writing Formula

Now that this textbook has covered when and why to propose a hypothesis that specifies a direction or not, I offer a basic formula to write clear and focused hypotheses. For each hypothesis, the formula includes:

- Prediction of either a statistically significant difference or direction of that difference (shown in underlines as follows).
- The independent variable (shown in **bold**).
- The dependent variable, stated in conceptual terms (shown in double underline).

For experiments that use two groups, a treatment and control, hypotheses use the term *between*. Future tense is also preferred. Written out as if we were filling in the blanks, this looks like:

- There will be a significant difference in credibility ratings (the DV) between subjects who see the story on **Twitter** and those who see it on a **website** (the IV).⁵⁸
- There will be a significant difference in importance ratings (the DV) between subjects who see **long or short stories** (the IV).⁵⁹
- Audiences exposed to the **advocacy frame** (IV) will rank crime as significantly different in importance (DV) than those exposed to the **objective stories** (IV).⁶⁰

These are all nondirectional hypotheses, as they specify a difference but do not say whether that difference will be higher or lower, large or small. They could be written directionally, for example:

- Credibility ratings (the DV) will be significantly higher for subjects who see the story on **Twitter** than for those who see it on a **website** (the IV).
- Importance ratings (the DV) will be significantly lower for subjects who see a **short story** than for those who see a **long story** (the IV).
- Audiences exposed to the **advocacy frame** (IV) will rank crime as significantly more important (DV) than those exposed to the **objective frame** (IV).

If an experiment involves three or more groups—for example, two treatments and a control group—the term *among* is used instead of *between*, like this:

- There will be a significant difference in credibility ratings (the DV) among subjects who see the story on **Twitter**, a **website**, or in a **newspaper** (the IV).

Notice this one has simply changed the IV to include three conditions—Twitter, a website, and a newspaper—instead of two.

In all these examples of basic hypotheses, each is precisely worded. They use clear, simple, single cause-and-effect predictions. The same words are used throughout rather than synonymous terms. These should

be the same terms used in the literature review. Each predicts a clear causal relationship between specific IVs and DVs. Directional hypotheses are preferred if they can be supported by theory and/or prior evidence.

If an experiment includes covariates or statistical controls for individual differences that have been shown to affect the DV but are not equivalently distributed during random assignment, the phrase *controlling for age, education, and gender*, or whatever the variables are, is added to the end of the hypothesis:

H1: There will be a statistically significant difference in credibility ratings among subjects who see the story on Twitter, a website, or a newspaper, controlling for age, education, and gender.

Hypotheses With More Than One IV

So far in this book, we have talked about manipulating one independent variable at a time—for example, the format of the message in the preceding hypothesis is whether subjects see it on Twitter, a website, or in a newspaper. In the previous example, the independent variable is the frame: advocacy or objective. These are known as *single-factor designs* and will be explained in detail in a later chapter. In reality, many experiments manipulate more than one independent variable at a time—for example, the format of the message *and* its frame, in the same study. When more than one independent variable, or **factor**, is manipulated, separate hypotheses should be written for each factor, also called a *main effect*, and for the **interaction** of the two factors or the effects of the two factors considered together. So, for example, in an experiment that manipulates the message format (Twitter, website) and the frame (advocacy, objective), there should be two hypotheses, one for each independent variable or factor's main effect:

H1: Credibility scores will be significantly higher for subjects who read stories on Twitter than on a website.

H2: Credibility scores will be significantly higher for subjects who read objectively framed stories than advocacy framed stories.

Finally, there should be either a hypothesis or a research question about the interaction between the format and the frame, such as:

H3: Credibility scores will be higher for subjects who read objectively framed stories on a website than for any of the other combinations of message format and frame;

OR

RQ1: Is there an interaction between message format and frame?

Naturally, the choice of whether to make a prediction or ask a research question depends on theory and evidence; use a hypothesis whenever possible. This next example of an interaction hypothesis shows the two experimental manipulations or IVs—disgust and harm—interacting with subjects' preexisting political views. The author predicts different effects of the IVs depending on whether the subject supports or opposes a political practice:

“H4: Incidental disgust and harm associations will increase moral conviction and lead to a harsher moral judgment among opponents of a political practice but lead to the opposite effect among supporters.”⁶¹

These hypotheses for studies with more than one independent variable, or factor, also build in a logical progression, from one IV to another, then to the interaction of the two IVs. If intervening variables are proposed, then they too are taken one at a time in a logical order. To think more clearly about these more complicated hypotheses, it may help to create a table listing the independent variables or factors and their levels, the intervening variables, and dependent variables. Use these to visualize the hypotheses, as shown in How To Do It box 3.4.

Here is an example of a hypothesis that includes an intervening variable: the concept of powerfulness. This experiment studies spokespeople who respond to an organizational crisis of contaminated food. The IV in this experiment is voice pitch, with the two levels being high pitch or low pitch. The intervening variable is powerfulness, which has been perceived in low-pitched voices. Thus, the hypothesis says:

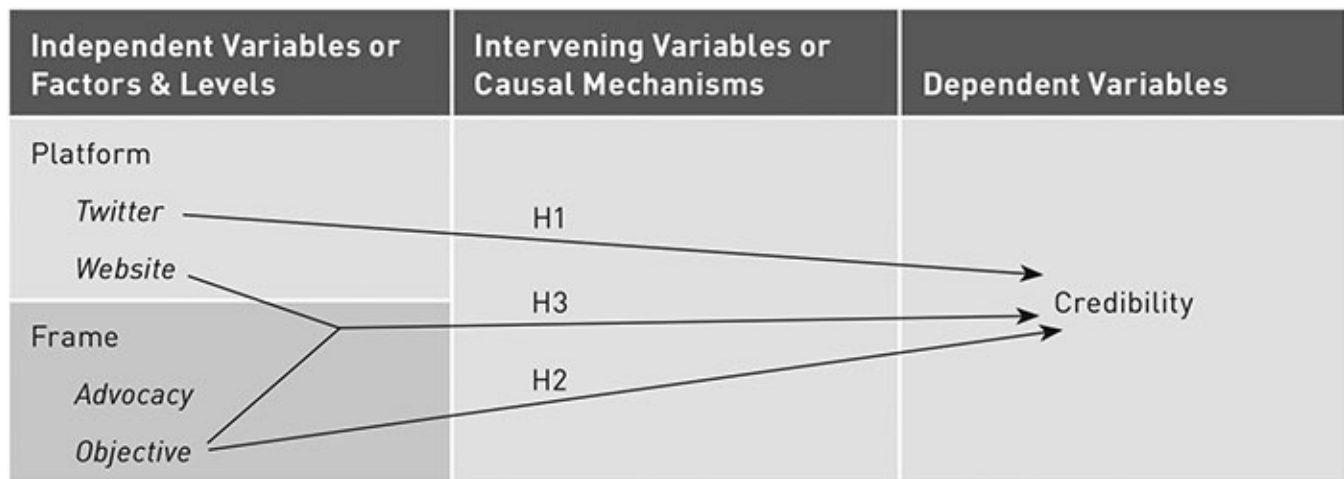
“A lowered voice pitch will result in greater perceptions of competence compared to a raised voice pitch due to an intermediate effect of perceived powerfulness (H2).”⁶²

There are many more variations on hypotheses in published journal articles; this is meant to introduce the beginner to the basics of hypothesis writing. As you read published experiments and conduct your own, adapt your hypothesis writing to the needs of each study.

How To Do It 3.4 Hypothesis Writing Table

To more easily diagram out the steps for writing hypotheses, it may be helpful to visualize the variables with a table, as follows. List the independent variables or factors in a column on the left, with the levels of each. List the dependent variables in the column on the right. If there are any intervening variables or causal mechanisms, list them in the middle column. Then, draw lines to demonstrate the process and predictions that will become the hypotheses. The tables that follow are simplified but can be expanded to accommodate as many factors, intervening variables, and dependent variables as needed.

For an Interaction Effect When There Are Two or More Factors

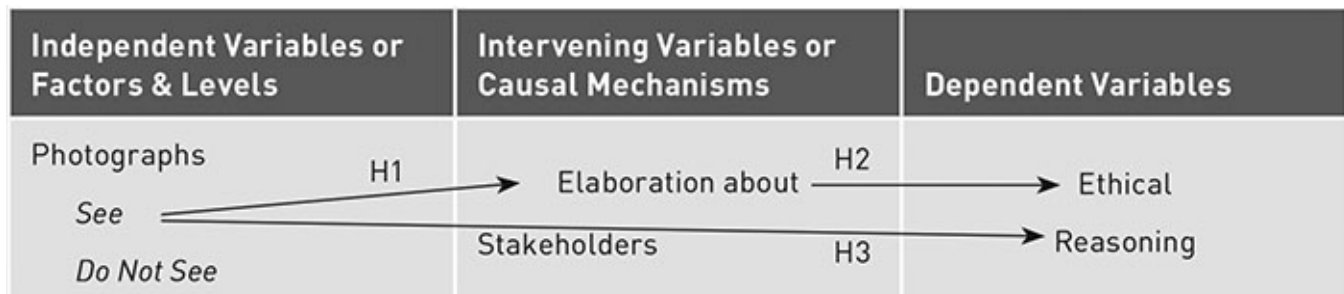


H1: Subjects who read stories on Twitter will have significantly higher credibility scores than subjects who read stories on a website. (Main effect of Platform factor.)

H2: Subjects who read objectively framed stories will have significantly higher credibility scores than subjects who read advocacy framed stories. (Main effect of Frame factor.)

H3: Credibility scores will be higher for subjects who read objectively framed stories on a website than for any of the other combinations of message format and frame. (Interaction of Platform and Frame factors.)

For Hypotheses With Causal Mechanisms That Build on Each Other



H1: Participants who see photographs will have significantly higher levels of elaboration about stakeholders than those who do not. (Factor leads to causal mechanism.)

H2: More elaboration about stakeholders will be significantly associated with higher ethical reasoning. (Causal mechanism leads to DV.)

H3: Ethical reasoning will be significantly higher for participants who see photographs than for those who do not. (Factor leads to DV.)

One final piece of advice on writing hypotheses is to think about the importance of each one. Ask if the hypotheses are interesting, meaningful, or important. Are they self-evident or already well tested? Try not to propose what I call “duh hypotheses”—that is, those whose answer is obvious. For example, the basic agenda-setting hypothesis that the issues the media cover most are the same ones that the public says are

most important is well tested. There is simply no earth-shattering revelation to be had by testing this as if it were the first time anyone had asked it. Many agenda-setting studies never ask it. When they do, it is frequently used to establish that a media agenda existed as a basis for more specific predictions, or to see if untested contingent conditions make the outcome different than what has been shown repeatedly for forty-plus years.

Also, ask what it would mean if one of these three things happened:

- What would it mean if the hypothesis was supported?
- What would it mean if it was not? (No significance)
- For directional hypotheses, what would it mean if significance was found in the opposite direction than predicted?

This exercise is designed to help you think about all the possible outcomes and what it would mean if each occurred. This will help you think about the meaningfulness of your study and also how you might explain the results in the discussion section. If you cannot imagine the conclusions under any of these situations, or they seem obvious or unimportant, it is time to rethink the study. Also, think about what might be included in the study as evidence to back up the explanations. For example, if data fail to support the hypothesis or find significance in the opposite direction than predicted, why might that have happened? This exercise also addresses a phenomenon I find frequently among novice experimentalists—feeling they have “failed” if their hypotheses are not supported. Even in studies that fail to support the hypotheses, something valuable has been learned. In some cases, not supporting a hypothesis is more interesting than supporting it. This, after all, is the purpose of conducting research in the first place, not merely to support all hypotheses. As Campbell and Stanley say, “The task of theory-testing . . . is therefore predominantly one of rejecting inadequate hypotheses.”⁶³ Rest assured, failing to support the hypotheses does not make a study unpublishable.

Research Questions

In addition to being used to probe interactions between two independent variables, research questions are sometimes found for main effects. These are made when there is not enough evidence or theory to support a prediction, and also when models are being tested. Writing research questions for experiments is no different than for other methods. In addition to the formal research question style shown earlier, research questions can also be written in narrative style, for example:

In sum, there are four primary questions addressed by this study. First, do the two teaching interventions differ in their overall effectiveness with respect to CVS learning? Second, do the two teaching interventions differ in the degree to which learning transfers to new domains? Third, do students learn different things from the two interventions? Specifically, do they develop a better understanding of the need to control irrelevant variables from the intervention based on invalid designs? Fourth, do the two teaching interventions differ in their relative effectiveness for students

in different learning environments (i.e., higher vs. lower achieving schools)?⁶⁴

The process of designing an experiment does not necessarily work in the same order as the chapters in this book. Sometimes, researchers need to revisit something they have already done. Writing good, clear hypotheses is one of those things. In addition, it is usually necessary to write, rewrite, edit, rewrite, edit, and rewrite again before being satisfied with the precision and clarity of the hypotheses.

The next chapter will discuss experimental designs, including the classic versions of the true laboratory experiments this book focuses on, as well as quasi, natural, and field experiments. As with hypothesis writing, once you have thought through your design, it might be necessary to go back and rework the hypotheses and research questions. This is normal, and research is messy.

Common Mistakes

- Not knowing the relevant literature thoroughly enough
- Orphan Variables—not having all the variables conceptually defined in the hypotheses or research questions, and operationally defined in the methods sections
- Not explaining the posited relations among variables and how they relate to theory

Test Your Knowledge

1. Which of these is an example of a conceptual definition?

- a. Authoritarianism is defined as a personality trait strongly related to aversion to uncertainty and threat sensitivity.
- b. Authoritarianism is measured by valuing respect for elders, having good manners, being obedient, and being well behaved.
- c. Authoritarianism is measured on a 7-point scale.
- d. Authoritarianism is defined differently by various researchers.

2. Which type of hypothesis does the following represent?

“Sources with high levels of expertise will positively affect perceptions of credibility.”

- a. Directional hypothesis

- b. Nondirectional hypothesis
 - c. Interaction hypothesis
 - d. Null hypothesis
- 3. Whether you make a directional or nondirectional hypothesis is determined by how much risk you are willing to take that your hypothesis will be supported or not.
 - a. True
 - b. False
- 4. When you have more than one hypothesis, always start with the one that represents the final outcome of the process you are testing.
 - a. True
 - b. False
- 5. The purpose of a literature review is to _____.
 - a. Develop theoretical contributions
 - b. Explain how variables are measured
 - c. Provide enough detail that other researchers can replicate the study
 - d. Build upon what has been discovered previously
- 6. The idea for an experiment should come from:
 - a. Hunches and conventional wisdom
 - b. A process of trial and error
 - c. Theory and existing evidence
 - d. None of these
- 7. Hypotheses should _____.
 - a. Be described as proven or not
 - b. Be stated in terms of operationalizations
 - c. Go in the methods section
 - d. Build logically upon each other
- 8. Research questions should be used instead of hypotheses when there is not enough evidence to make a prediction
 - a. True
 - b. False

9. "Journalists who see photographs will use significantly different levels of moral judgment than journalists who do not see photographs." This statement represents which of the following?

- a. A directional hypothesis
- b. A nondirectional hypothesis
- c. A research question
- d. A null hypothesis

10. Whether to make a directional or nondirectional hypothesis depends upon _____.

- a. Theory
- b. Evidence
- c. The statistical test being used
- d. Both A and B

Answers:

- 1. a
- 2. a
- 3. b
- 4. b
- 5. d
- 6. c
- 7. d
- 8. a
- 9. b
- 10. d

Application Exercises

1. Choose one of your three ideas from the assignments in [chapter 1](#) and add to it at least five pages that review the theory it will develop, test, or extend, and the literature related to it. Use at least twenty-five articles. Be sure to analyze and synthesize the literature, not just summarize it, and to connect the literature to your study.
2. Write two to three clear and concise hypotheses to go with the experiment you are developing. Use the formula in this chapter. If you have more than one factor (IV) in your experiment, write a hypothesis for the main effect of each factor and a hypothesis for the interaction effect. Have two colleagues read it to see if it is clear.

Suggested Readings

From the University of Washington's Psychology Writing Center, this white paper on how to write a literature review:

https://depts.washington.edu/psych/files/writing_center/litrev.pdf

Shoemaker, P., J. Tankard, and D. Lasorsa. 2003. *How to Build Social Science Theories*. Thousand Oaks, CA: Sage.

Chapter 2, "Manuscript Structure and Content," in the *Publication Manual of the American Psychological Association*, 6th ed. (2010). Washington, DC: American Psychological Association.

Chapters 2, 3, and 4 in Bausell, R. Barker. 1994. *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist*. Thousand Oaks, CA: Sage.

Notes

1. R. Barker Bausell, *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist* (Thousand Oaks, CA: Sage, 1994), 32.
2. Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 3rd. ed. (Chicago: University of Chicago Press, 1996); Earl Babbie, *The Practice of Social Research*, 11th ed. (Belmont, CA: Thompson-Wadsworth, 2007).
3. H. W. Turnbull, ed. *The Correspondence of Isaac Newton: 1661–1675, Volume 1* (London: The Royal Society at the University Press, 1959).
4. Esther Thorson, Robert H. Wicks, and Glenn Leshner, "Experimental Methodology in Journalism and Mass Communication Research," *Journalism and Mass Communication Quarterly* 89, no. 1 (2012): 112–124.
5. Rebecca B. Morton and Kenneth C. Williams, *Experimental Political Science and the Study of Causality: From Nature to the Lab* (New York: Cambridge University Press, 2010).
6. Paul D. Leedy and Jeanne Ellis Ormrod, *Practical Research: Planning and Design* (Boston: Pearson Education Inc., 2010), 5.
7. Kurt Lewin, *Field Theory in Social Science: Selected Theoretical Papers by Kurt Lewin* (London: Tavistock, 1952), 169.
8. Leedy and Ormrod, *Practical Research*.

9. S. Iyengar, *Is Anyone Responsible?: How Television Frames Political Issues* (Chicago: University of Chicago Press, 1991).
10. L. R. Franklin, "Exploratory Experiments," *Philosophy of Science* 72, no. 5 (2005): 888–899; C. Kenneth Waters, "The Nature and Context of Exploratory Experimentation: An Introduction to Three Case Studies of Exploratory Research," *History and Philosophy of the Life Sciences* 29, no. 3 (2007): 275–284.
11. Franklin, "Exploratory Experiments"; Waters, "The Nature and Context of Exploratory Experimentation."
12. Franklin, "Exploratory Experiments."
13. Ibid.
14. Babbie, *The Practice of Social Research*.
15. Bausell, *Conducting Meaningful Experiments*.
16. Ibid.
17. Renita Coleman, "The Effect of Visuals on Ethical Reasoning: What's a Photograph Worth to Journalists Making Moral Decisions?" *Journalism and Mass Communication Quarterly* 83, no. 4 (2006): 835–850.
18. H. B. Brosius, "The Effects of Emotional Pictures in Television News," *Communication Research* 20, no. 1 (1993): 105–124; David Domke, David Perlmutter, and Meg Spratt, "The Primes of Our Times? An Examination of the 'Power' of Visual Images," *Journalism* 3, no. 2 (2002): 131–159; A. Friedman, "Framing Pictures: The Role of Knowledge in Automatized Encoding and Memory for Gist," *Journal of Experimental Psychology: General* 108 (1979): 316–335; Doris Graber, "Seeing in Remembering: How Visuals Contribute to Learning From Television News," *Journal of Communication* 40, no. 3 (Summer 1990): 134–155; George E. Marcus, W. Russell Neuman, and Michael MacKuen, *Affective Intelligence and Political Judgment* (Chicago: University of Chicago Press, 2000); Richard E. Petty and John T. Cacioppo, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* (New York: Springer-Verlag, 1986).
19. Elinor Amit and Joshua D. Greene, "You See, the Ends Don't Justify the Means: Visual Imagery and Moral Judgment," *Psychological Science* 23, no. 8 (2012): 861–868; J. D. Greene et al., "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44 (2004): 389–400; J. D. Greene et al., "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293 (2001): 2105–2108; Joshua Greene and Jonathan Haidt, "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6, no. 12 (2002): 517–523; D. Kahneman and S. Fredrick, "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," in *Heuristics and Biases*, ed. T. Gilovich, D. Griffin, and D. Kahneman (New York: Cambridge University Press, 2002), 49–81; Emma Rodero, "See It on a Radio Story: Sound Effects and Shots to Evoked Imagery and Attention on Audio Fiction," *Communication Research* 39, no. 4 (2012): 458–479; Adam B. Moore, Brian A. Clark, and Michael J. Kane, "Who Shalt Not Kill? Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgment," *Psychological Science* (0956-7976) 19, no. 6

(2008): 549–557.

20. Vicki Goldberg, *The Power of Photography: How Photographs Changed Our Lives* (New York: Abbeville, 1991).

21. Bausell, *Conducting Meaningful Experiments*, 21.

22. H. Denis Wu and Renita Coleman, “Advancing Agenda-Setting Theory: The Comparative Strength and New Contingent Conditions of the Two Levels of Agenda-Setting Effects,” *Journalism and Mass Communication Quarterly* 86, no. 4 (Winter 2009): 775–789.

23. James Potter, *On Media Violence* (Thousand Oaks, CA: Sage, 1999).

24. Thorson, Wicks, and Leshner, “Experimental Methodology.”

25. Pamela Shoemaker, James W. Tankard, and Dominick Lasorsa, *How to Build Social Science Theories* (Thousand Oaks, CA: Sage, 2003), 11.

26. Ibid., 11.

27. George Gerbner et al., “Living with Television: The Dynamics of the Cultivation Process,” in *Perspectives on Media Effects*, ed. Jennings Bryant and C. Zillmann (Hillsdale, NJ: Erlbaum, 1986), 17–40; George Gerbner et al., “Growing Up with Television: The Cultivation Perspective,” in *Media Effects: Advances in Theory and Research*, ed. J. Bryant and C. Zillmann (Hillsdale, NJ: Erlbaum, 1994), 7–14.

28. Ibid.

29. Coleman, “The Effect of Visuals on Ethical Reasoning.”

30. Wu and Coleman, “Advancing Agenda-Setting Theory,” 778.

31. Ibid., 777.

32. Seung-A Annie Jin and Joe Phua, “Following Celebrities’ Tweets About Brands: The Impact of Twitter-Based Electronic Word-of-Mouth on Consumers’ Source Credibility Perception, Buying Intention, and Social Identification with Celebrities,” *Journal of Advertising* 43, no. 2 (2014): 183.

33. Nan-Hyun Um and Wei-Na Lee, “Does Culture Influence How Consumers Process Negative Celebrity Information? Impact of Culture in Evaluation of Negative Celebrity Information,” *Asian Journal of Communication* 25, no. 3 (2015): 329.

34. Elena Novak and Janet Tassell, “Using Video Game Play to Improve Education-Majors’ Mathematical Performance: An Experimental Study,” *Computers in Human Behavior* 53 (2015): 125.

35. Horacio Alvarez-Marinelli et al., “Computer Assisted English Language Learning in Costa Rican

Elementary Schools: An Experimental Study,” *Computer Assisted Language Learning* 29, no. 1 (2016): 105.

36. Ryan Oprea, Bart J. Wilson, and Arthur Zillante, “War of Attrition: Evidence from a Laboratory Experiment on Market Exit,” *Economic Inquiry* 51, no. 4 (2013): 2019.

37. Miglena Mantcheva Sternadori and Esther Thorson, “Anonymous Sources Harm Credibility of All Stories,” *Newspaper Research Journal* 30, no. 4 (2009): 56.

38. James R. Rest, Lynne Edwards, and Stephen J. Thoma, “Designing and Validating a Measure of Moral Judgment: Stage Preference and Stage Consistency Approaches,” *Journal of Educational Psychology* 89, no. 1 (March 1997): 5–28.

39. Sternadori and Thorson, “Anonymous Sources Harm Credibility of All Stories,” 56.

40. Kristin Bulkow, Juliane Urban, and Wolfgang Schweiger, “The Duality of Agenda-Setting: The Role of Information Processing,” *International Journal of Public Opinion Research* 25, no. 1 (Spring 2013): 43–63.

41. Ibid., 52.

42. Tara Halle, “A Cautionary Tale: Have You Checked That Citation?” *Covering Health* (2017); Lisa Marriott, “Using Student Subjects in Experimental Research: A Challenge to the Practice of Using Students as a Proxy for Taxpayers,” *International Journal of Social Research Methodology* 17, no. 5 (2014): 503–525.

43. Mirae Kim and Gregg G. Van Ryzin, “Impact of Government Funding on Donations to Arts Organizations: A Survey Experiment,” *Nonprofit and Voluntary Sector Quarterly* 43, no. 5 (2014): 913.

44. Thomas Zerback, Thomas Koch, and Benjamin Kramer, “Thinking of Others: Effects of Implicit and Explicit Media Cues on Climate of Opinion Perceptions,” *Journalism and Mass Communication Quarterly* 92, no. 2 (2015): 421–443.

45. Ibid., 423.

46. Ibid., 424.

47. Ibid., 425.

48. Tianguang Meng, Jennifer Pan, and Ping Yang, “Conditional Receptivity to Citizen Participation,” *Comparative Political Studies* (2014): 8.

49. Coleman, “The Effect of Visuals on Ethical Reasoning,” 839.

50. Diana C. Mutz, “Effects of “In-Your-Face” Television Discourse on Perceptions of a Legitimate Opposition,” *American Political Science Review* 101, no. 4 (2007): 624.

51. D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. (Chicago:

Rand McNally, 1963), 35.

52. Thorson, Wicks, and Leshner, "Experimental Methodology."

53. Ibid.

54. Y. J. Cho and E. J. Ringquist, "Managerial Trustworthiness and Organizational Outcomes," *Journal of Public Administration Research and Theory* 21 (2011): 53–54.

55. Mogens Jin Pedersen and Justin M. Stritch, "Internal Management and Perceived Managerial Trustworthiness," *American Review of Public Administration* (2016): 5.

56. Ibid.

57. Thorson, Wicks, and Leshner, "Experimental Methodology."

58. Adapted from Mike Schmierbach and Anne Oeldorf-Hirsch, "A Little Bird Told Me, So I Didn't Believe It: Twitter, Credibility, and Issue Perceptions," *Communication Quarterly* 60, no. 3 (July-August 2012): 317–337.

59. Ibid.

60. Adapted from Sean Aday, "The Framesetting Effects of News: An Experimental Test of Advocacy Versus Objectivist Frames," *Journalism and Mass Communication Quarterly* 83, no. 4 (Winter 2006): 767–784.

61. Pazit Ben-Nun Bloom, "Disgust, Harm and Morality in Politics," *Political Psychology* 35, no. 4 (2014): 500.

62. An-Sofie Claeys and Verolien Cauberghe, "Keeping Control: The Importance of Nonverbal Expressions of Power by Organizational Spokespersons in Time of Crisis," *Journal of Communication* 64 (2014): 1162.

63. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research*, 35.

64. Robert F. Lorch et al., "Using Valid and Invalid Experimental Designs to Teach the Control of Variables Strategy in Higher and Lower Achieving Classrooms," *Journal of Educational Psychology* 106, no. 1 (February 2014): 18–35.



Notes

iIn some disciplines, replications that do not extend theory or empirical knowledge in some relevant way are frowned upon.

iiThe twelve statements vary according to each dilemma, but an example is: (1) A chance like this photo comes only a few times in a career; (2) The kids will grow up having this horrible, graphic reminder of what happened; (3) This is a family newspaper, children might see this photo at the breakfast table; (4) You'll probably get a lot of angry calls and people will cancel their subscriptions; (5) How these people and their families will feel when they see this; (6) Your competition is working on a similar story. If you don't run the photo, your competition will just run something similar; (7) Publishing this photo would help your paper's reputation for investigative reporting; (8) Whether it is our duty as journalists to show all the news, regardless

of the circumstances; (9) Whether the public has a right to know all the facts about drug use and its effects on people, especially children; (10) What would best serve society; (11) Photos that are painful to some have to be shown so others will benefit; (12) If I don't run this photo, I may prevent these children from being taken from their parents, but the conditions leading to situations like theirs will persist.

iii This study also predicted an interaction of photographs and involvement, but I leave the discussion of main effects and interactions for later.

iv Other uses of research questions will be covered later in this chapter.

<http://dx.doi.org/10.4135/9781071878958.n3>



Types of Experiments

**In: Designing Experiments for the Social Sciences: How to Plan,
Create, and Execute Research Using Experiments**

By: Renita Coleman

Pub. Date: 2022

Access Date: February 23, 2022

Publishing Company: SAGE Publications, Inc

City: Thousand Oaks

Print ISBN: 9781506377322

Online ISBN: 9781071878958

DOI: <https://dx.doi.org/10.4135/9781071878958>

Print pages: 89-110

© 2019 SAGE Publications, Inc All Rights Reserved.

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

Types of Experiments

Truth has nothing to do with the conclusion, and everything to do with the methodology.

—Stefan Molyneux

Learning Objectives

- Summarize the different types of experiments using Campbell and Stanley's typology.
- Recommend when to use each of the three true experimental designs.
- Critique the strengths and weaknesses of designs with pretests.
- Describe quasi, natural, and field experiments.
- Explain how quasi, natural, and field experiments differ from each other and from true experiments.

With the literature review written or at least in draft form, as discussed in the previous chapter, the next step is to decide what basic type of experiment to conductⁱ—that is, the methodology or system of the experiment itself referred to in this chapter's opening quote. One authoritative word on this topic is the classic *Experimental and Quasi-Experimental Designs for Research* by Donald T. Campbell and Julian C. Stanley.¹ This work is only seventy-one pages long without the references. It started life as a chapter in a research handbook in 1963 but is so popular that it continues to be published as a monograph. It describes sixteen different ways to do an experiment, all still valid today, although some designs are more popular than others. It was written for the teaching field, so examples are from education research. Next is a summary of Campbell and Stanley's typology of experimental designs, highlighting six designs popular in social science along with a critique of each. Most experiments actually conducted today are slight variations of the designs described here. In addition to these designs in Campbell and Stanley's typology, this chapter will also briefly describe quasi experiments, natural experiments, and field experiments.

Campbell and Stanley's Typology of Experiments

Campbell and Stanley use graphic shorthand to describe experiments that should be familiar to football fans but have different meanings here.

X = an exposure, treatment, manipulation, or intervention. Usually an independent variable (IV).
O = an observation or measurement of an outcome variable. Usually a dependent variable (DV). These are the data that are recorded, either by the researcher, with technical instruments such as a heart rate monitor, or a self-report by the subject. If observations are made or data recorded more than once, that is indicated by a subscript—O₁, O₂—meaning the first and second observation.

If there is a space (____) and no X or O, that means no treatment was given or no observation made.

R = random assignment of subjects to conditions. This will be discussed more in [chapter 7](#). Basically, this is the element that qualifies an experiment as a “true” experiment versus a “quasi” experiment.

The first three design types are classified as “pre-experimental designs” and are “of almost no scientific value,” according to Campbell and Stanley.² These designs tend to be reported as “exploratory” in articles I review, but “pre-experimental” is a more accurate term. They also qualify as quasi experimental designs because of the lack of random assignment. Here, I briefly review them but recommend using other designs as described.

Three Pre-Experimental Designs

The One-Shot Case Study

This can sometimes be seen reported in news stories when someone wants to attribute a cause to some effect—for example, how school absentee rates went down compared to previous years after some intervention. It might be something like a school that implemented a text messaging system that pinged students in time to get them to school by 9 a.m. This is graphically represented as:

____ X O

The text messaging system was the intervention (X) and absentee rates the observation (O). There are several problems with this, however—among them that the students who were observed were not randomly sampled, there is no control group of students who did not get the text messages to compare against, and there were no controls in place that would rule out alternative explanations. For example, some students’ parents could have started making waffles for breakfast, new construction outside others’ windows woke them up early, some could have gotten new cars and were excited to get to school to show them off, and maybe others saw a presentation on the importance of college and suddenly figured they needed to get to school in order to be able to support themselves. All these things—uncontrolled confounds—could have explained the lower absentee rates, not just the text messaging system. Because of these problems, the one-shot case study is not a good design for a true social science experiment.

One-Group Pretest–Posttest Design

Slightly better than the one-shot case study is when an observation or measurement is added before the treatment is given, called a **pretest**. The same observation or measurement is given after the intervention (the posttest), so researchers have a baseline measure to compare any changes against.

O1 X O2

For example, researchers might measure how much math a student knows in the pretest, then give the intervention, such as computer-aided tutoring, and then test the students again to see how much they learned. The difference between the pretest and posttest scores represents the effect. Sometimes, however, adding a pretest does not necessarily make things better, because people tend to learn how to do better on a test after they have taken it once or when they realize they are being watched. For example, those who take IQ tests more than once have been shown to get smarter on the second try.³ In research, this effect of testing is called “test, retest gain.”⁴ There could be other explanations as well. Reactivity is the idea that measuring something changes it; just by knowing they are being observed, people tend to do better⁵ (see More About box 4.1).

More About . . . box 4.1 The Hawthorne Effect

Hawthorne, Illinois, Works of the Western Electric Company, 1925



Western Electric Company

The Hawthorne effect is a specific type of reactivity, the idea that people react to being observed and change what they do. It is important in research because some people will not give true answers if they know they are participating in a study. This effect is not necessarily intentional; sometimes, people simply change without realizing it. The Hawthorne effect is the popular, although some say inappropriate, name for a demand characteristic that arises when subjects of study change their behavior because they know they are being observed.⁶

It developed from studies of worker productivity commissioned by the Hawthorne Works, a Western Electric factory outside Chicago from 1924 to 1932.⁷ They are also known

as the illumination studies because higher and lower levels of lighting were tested for changes in worker productivity; in reality, many other variables were also studied, including work hours, break times, the cleanliness of floors and work stations, among others.⁸ The truncated version of the study results is that when the researchers changed the workers' lighting and break times, their productivity improved. In spite of the changes that led to improvements staying in place, worker productivity fell when the study was over. The original conclusion that paying attention to workers would result in greater productivity was later reinterpreted to say that people change their behavior when they know they are being observed.⁹

Not all studies show a Hawthorne effect,¹⁰ and research still investigates it today,¹¹ especially in the health sciences, human relations, and organizational behavior.¹² The illumination studies, for there were many over several years, are more complex than usually presented. For example, it is a myth that improvement was continuous, and there were potential confounding variables such as learning, feedback, and incentive pay. Thus, the term *Hawthorne effect* to describe reactivity is diminishing in use.¹³

A similar phenomenon is known as the **demand effect**, or demand characteristics—the idea that some people change their answers or behavior in order to please an experimenter.¹⁴ In studying the experimental situation, Martin Orne discovered that study subjects tried to guess the purpose and altered their behavior to fit their interpretations.¹⁵ Similar phenomena include the *halo effect* or *social desirability effect*, where study participants try to portray themselves in a positive light.¹⁶

The Hawthorne studies make another important contribution to experimental design in that they show the importance of manipulation checks,¹⁷ which will be covered in [chapter 9](#). In the Hawthorne studies, it was not so much the manipulations that had an effect but the workers' interpretation of them, and understanding subjects' interpretations is the purpose of a manipulation check.

There could also be a change in the measurement standards if observers are recording the data. There is also a phenomenon known as *regression to the mean*, whereby those who score extremely well or poorly tend to go back toward the middle the next time they are tested.¹⁸ Moreover, if there is a time gap between the first and second observation, something else could have happened. For example, if the observation being recorded is a person's level of fear about flying, and if an airplane crashes and is reported in the news between O₁ and O₂, that could change the outcome.

Static Group Comparison

In this type of design, a second group that has not received the treatment has been added (O2).

X O1

___ O2

An example would be comparing people who saw the presidential candidates' debate versus those who did not, or comparing students who got antibullying training to those who did not. There are also problems with this, including that there is no way to tell if the people in the two groups are the same on important individual characteristics—for example, perhaps there were more Republicans in the debate-watching group than the nonwatching group, or more aggressive students in one group than the other. This will be discussed in greater detail in [chapter 7](#) on random assignment. People's political identification and students' innate aggressive tendencies could affect the outcome.

As I do not recommend these three designs, they will not be discussed in detail regarding their strengths and weaknesses. Instead, this chapter will concentrate on the next three, which are true experimental designs that I do recommend.

Three True Experimental Designs

Pretest–Posttest Control Group

This design is one of the most used in social science. It adds the crucial feature of randomly assigning subjects to either the treatment or control conditions, thus making sure the groups are equal on important characteristics that could otherwise cause any changes.

R O1 X O2

R O3 ___ O4

The drawback of this design is the same as in the one-group pretest–posttest design—that is, being observed or measured twice may cause changes in the subjects' performance, attitudes, or whatever else the outcome is. Because this threat has been shown to be so prevalent, the popularity of pretests has been declining, and pretests are actually not essential to true experimental designs.¹⁹

Solomon Four-Group Design

It is common for the Solomon four-group design to be described as the gold standard. And it is. It eliminates all the drawbacks described earlier, plus the researcher can actually tell if there are any effects of testing by having groups where no pretest is given.

R O1 X O2

R O3 ____ O4

R ____ X O5

R ____ ____ O6

But it is also time consuming, costly, difficult, and has statistical issues, so relatively few studies actually use it.²⁰ Bausell even calls it “wasteful.”²¹ In this design, there are four groups, with subjects randomly assigned to all of them. There are two control groups consisting of subjects who do not get the treatment or manipulation; they serve as the baseline for comparison with the groups that did get the treatment. One of the treatment groups is given a pretest and one is not, and one of the control groups is given a pretest and the other is not. This allows the researcher to compare not only the differences before and after the treatment, but also to see if the pretest affected the results. This provides what Levy and Ellis call a “defensible response to most rival hypotheses.”²² It also requires many, many more subjects to participate in the experiment, the costs of which sometimes outweigh the benefits. (See Study Spotlight 4.2 for an example of a Solomon four-group design.)

Posttest-Only Control Group Design

The final design in the Campbell and Stanley typology is one that includes no pretests and is the one I use most often. It is also the one that is now most recommended.²³

R X O

R ____ O

Study Spotlight 4.2 A Study Using the Solomon Four-Group Design



SAGE Journal Article: study.sagepub.com/coleman

Genç, M. 2016. “An Evaluation of the Cooperative Learning Process by Sixth-Grade Students.” *Research in Education* 95 (1): 19–32.

This study used a Solomon design to assess the effects of a particular teaching strategy on sixth graders’ science knowledge. The teaching strategy, called cooperative learning, has teachers organize students into groups who do research on their own to learn

information and solve problems together. Cooperative learning classrooms were the manipulation or treatment group; classrooms as they were organized already, with teachers presenting information to the students in the traditional way, were used as the control groups.

In this Solomon four-group design, two groups of students got the cooperative learning treatment, and two did not, making them the control groups. Pretests were used for one control group and one treatment group; the other groups did not get pretests. The author says that the reason the Solomon design was used was to be able to know if use of pretests caused effects because students could have learned simply by taking the test twice. Students were randomly assigned to all four groups. Here is how it is described:

“The Solomon four-group design is an attempt to eliminate the possible effect of a pretest. It involves random assignment of subjects to four groups, with two of the groups being pretested and two, not. One of the pretested groups and one of the unpretested groups is exposed to the experimental treatment. All four groups are then posttested. Although each group is put through post-experimental evaluation, the pre-experimental evaluations are performed in only two groups, one being an experimental group and the other being a control group. The first two groups are treated as the pretest–posttest control group design, and the other two groups are treated as the posttest–control group design” (p. 22).

The study does not go into whether the pretest affected the outcome; however, it can be seen from the mean scores that the groups that received pretests did score higher on the posttests than those that did not get the pretests. No significance tests are reported in this study to determine if the pretest alone had a significant effect, however. The mean posttest scores for those given the pretests were 27.70 (treatment) and 22.50 (control) versus 26.74 (treatment) and 21.91 (control) for students who were not given the pretest. I have reworked the author’s tables into the one that follows in order to illustrate the design and also facilitate comparison between posttest scores for the pretested groups compared to the non-pretested groups.

It should be noted that alternative forms of the tests were used for pre- and posttesting rather than identical tests for both occasions, which helps guard against learning from taking the test twice.

The cooperative learning treatment worked, as both pretested and unpretested students in the cooperative learning classrooms performed significantly better on the science tests given at the end of the experiment.

Some social science disciplines use the Solomon design more than others, so it is important to know where your colleagues stand on this.

Means and Standard Deviations for a Solomon Four-Group Design Experiment

Group	Pretest	Instruction type	Posttest Scores M (SD)
Treatment 1	X	Cooperative learning	27.70 (2.531)
Treatment 2		Cooperative learning	26.74 (1.797)
Control 1	X	Regular curriculum	22.50 (2.219)
Control 2		Regular curriculum	21.91 (2.327)

In a study comparing the pretest–posttest and posttest-only designs, Gorard²⁴ found the results of the posttest-only design to be “less misleading.” Even Campbell and Stanley²⁵ say that pretests are not essential to true experimental designs and explain that they are misunderstood. When the purpose of a pretest is to ensure equivalence of subjects assigned to conditions, random assignment is an adequate precaution, so pretests are not needed. They point out that almost all of Ronald Fisher’s agriculture experiments had no pretests. Pretesting is still preferred in some disciplines but not all; it is important to know the standard in the field. The interaction of a pretest with the treatment is not large in most cases, but in education, psychology, and sociology in particular, the effects are larger and should not be ignored.²⁶ Some studies straddle the line, so to speak, for example by conducting a pilot study that uses a pretest, and if the effects of testing are ruled out, conduct the actual experiment without a pretest.²⁷ (See How To Do It box 4.3 for examples of how to describe a posttest-only control group design.)

These designs represent a basic structure; actual experiments may use slight variations and still be acceptable. So far, this chapter has reviewed three experimental designs that are classified as “true” experiments and three that are “pre-experimental” from Campbell and Stanley’s typology. Campbell and Stanley go on to describe ten other designs they term “quasi experimental.” This book will not go into them in detail but will instead summarize the essential features of a quasi experiment and refer readers to the Campbell and Stanley book for more details.

Quasi Experiments

The key difference between true and quasi experiments is that in quasi experiments, subjects are not randomly assigned to conditions, the groups may not necessarily be independent—that is, some of the same people may be in more than one group—and, finally, that quasi experiments cannot control for all the extraneous factors that true experiments do. These features of random assignment, independence, and control are not always possible. When that is the case, a quasi experiment is considered a viable option, and even preferable in some cases. Some things simply cannot be “assigned.” For example, researchers cannot ethically assign someone to smoke and someone else to not smoke in order to have randomly assigned treatment and control groups. Experiments in business and education settings especially make it difficult to have complete control over the research.²⁸ Businesses involved in a study may want to handpick participants for some reason, or want everyone to be treated the same and not appear to be showing favoritism. In other cases, participants unwittingly self-select the group they are in. For example, education research frequently works with intact classrooms rather than classes whose students were randomly assigned. Educators may assign students to classes because they want diversity or a mix of boys and girls, do not want twins in the same class, or want to keep two children together. At the college level, it is obvious that there is something individually different about people who sign up for a class in experimental design than those who register for ethnographic methods. Students might choose one course over another because they heard a certain teacher is good. Or it might be a timing issue—some students do not like getting up early in the morning, have part-time jobs on Wednesday afternoons, or want to leave town on Fridays. Whatever is at work to create these individual differences could compromise the validity of an experiment.

How To Do It 4.3 Describing a Posttest-Only Control Group Design

As the most used and recommended of all the Campbell and Stanley true experiment types, the posttest-only control group design is likely one you will use often. Unlike experiments that use the Solomon four-group design, most studies that use this ubiquitous design do not announce themselves formally. Rather, you can determine they use this typology by looking for mention of a control group and random assignment, and no mention of a pretest.

Here are two examples; in the first, it does specify it is a posttest-only design, followed by the second, which does not.

Coleman, Renita, Paul Lieber, Andrew Mendelson, and David Kurpius. 2008. “Public Life and the Internet: If You Build a Better Website, Will Citizens Become Engaged?” *New Media & Society* 10 (2): 179–201.

“This study used a post-test, control group experimental design. The experimental stimulus for this study was a website on the topic of the state budget created by mass communication students in a class in website development . . .” (p. 188). “The control group website was the official state government website on the state budget. It was created without usability tests or knowledge of any of the issues described above, which guided the creation of the experimental website. It was important that the control site was on the same topic as the

experimental site in order to rule out the possibility of effects due to the subject matter rather than content, appearance or navigation . . .” (p. 189). “The 60 participants were randomly assigned to view either the control website or the experimental website” (p. 190).

Bennion, Elizabeth A., and David W. Nickerson. 2011. “The Cost of Convenience: An Experiment Showing E-Mail Outreach Decreases Voter Registration.” *Political Research Quarterly* 64 (4): 858–869.

“The design of the e-mail experiment itself was straightforward. Students were randomly assigned to one of three conditions: (1) a control group receiving no e-mail, (2) a treatment group receiving three e-mails from an administrator such as the university president or dean of students, or (3) a treatment group receiving three e-mails from a student leader—usually the student body president. The e-mails were brief, explaining why registration is important and providing a link to the Rock the Vote online registration tool” (p. 862).

When students in intact classrooms are used in experiments, these are known as *quasi experiments* because the subjects are not randomly assigned.



iStock.com/skynesher

Only random assignment is an antidote to such individual differences affecting results. Research can still uncover important knowledge from quasi experiments, but they bring different threats to validity than do true experiments, and these must be addressed and documented. It is important to say in the final paper that subjects could not be randomly assigned and to give the reasons why. The paper also should explain what

was done to minimize any issues with systematic differences—for example, were key variables measured and then used as covariates? Were statistical tests run to see if groups were equivalent on important variables? It is important to make a credible claim for being able to infer causality, or to generalize beyond the one case studied. Finally, quasi experiments cannot control all the extraneous factors that may cause an outcome the way true experiments can. Without random assignment, a quasi experiment cannot eliminate the possibility of uncontrolled variables confounding the results and impairing the ability to make causal claims. However, researchers should always attempt to measure possible confounding variables, then control for them statistically by using covariates. Every study has its weaknesses, including true experiments (see more about these in [chapter 5](#)). Limitations do not necessarily invalidate the importance of the findings, especially if this is the only way to study a particular phenomenon. Quasi experiments can feel less artificial than true experiments and are usually easier to conduct longitudinally than controlled experiments—all benefits. It is better to know what the study found than to not know anything, but readers should always be told about the limitations.

Triangulation, or confirming what has been found using one method with another, is also important. If different methods confirm the same finding—for example, if what is found in a true experiment in an artificial setting dovetails with what happens in the real world of a quasi experiment—then we can have more confidence that the findings represent something real.

One example of best practices in a quasi experiment in social work education used two similar classes to study whether online or face-to-face teaching would result in greater learning.²⁹ In this quasi experiment, the researchers used intact classrooms and so were unable to randomly assign students to classes with different teaching styles. To attempt to control as much as possible for individual student differences, they used as covariates the students' ages and grade point averages adjusted for grade inflation. They say some of the limitations are the unaccounted for extraneous variables and the lack of generalizability.

In a business study, the researcher examined how formal mentoring affected workers' ability to build interpersonal networks.³⁰ Those who were mentored represented the treatment group; those who were not mentored were the control group. In that study, participants were not randomly assigned; rather, the management of the company chose the employees who would be mentored based on their perceived potential for advancement within the organization. This represents a serious threat to group equivalence, considering one group was chosen specifically because they were perceived as superior to the other, but the company would not agree to random assignment. To decrease the uncertainty, the researcher used a "matched pairs" design, where every worker in the treatment group was paired with a worker in the control group who was as similar as possible on important characteristics—they had the same salary, performance rating in the prior year, length of time with the company, and were in the same office. The author of that study provides statistical evidence that matching was successful, showing that there were no significant differences between the treatment and control groups on other characteristics, including age, education, and the number of networking contacts each group made before mentoring began. More network contacts was one of the outcome goals of the program, so this assured that people in the treatment group were not more likely to

network to begin with. Still, the workers in the treatment and control groups could have differed in other unknown ways that could have caused differences in outcomes that were supposed to be attributed to the mentoring treatment, so various other strategies were used to help overcome the lack of random assignment. However, because we can never think of all the other confounding variables, we can never be assured these were as successful as simple random assignment; more uncertainty always remains.

Another example of a quasi experiment is the study of an intervention where investigators interviewed abused and neglected children about their mental health and quality of life.³¹ This type of research typically only interviewed adults, as it was thought children were unreliable sources and did not need to be upset by these kinds of questions. In this study, the cases where only adults were interviewed represented the control group; the cases where both children and adults were interviewed represented the intervention. Random assignment to treatment or control group was not possible because the Dutch Medical Ethics Committee refused to approve it out of concern that the adult-only interview group might receive inferior care. The researchers were able to have treatment and control groups, but the participants could not be randomly assigned to them. The article notes the possibility of selection bias.

Beyond the pre-experimental, true, and quasi experimental designs in Campbell and Stanley's typology, researchers also employ natural or field experiments. These are briefly covered next, and there are many good books that treat these topics in depth (see the "Suggested Readings" section). Instead, this is offered as an introduction as food for thought about whether your topic is best suited for a true experiment, a quasi experiment, a natural or a field experiment.

Natural Experiments

Natural experiments are a subset of quasi experiments in that researchers do not randomly assign subjects to conditions or create the manipulation. Instead, natural experiments take advantage of some naturally occurring phenomenon that creates treatment and control groups. One group of people was exposed to something, and another group was not, in a natural setting. This is seen as approximating randomization as far as possible, what some researchers call *near random* or *as-if random*.³² Nature or society creates the treatment or exposure, and researchers discover it after it has already occurred, then conceive of it as an experiment. Crasnow says it is really more of an observational study.³³ Because researchers did not design the treatment or intervention, they cannot control all other possible factors that could have caused the observed outcome. This could lead to plausible alternative explanations, which need to be explained and accounted for as much as possible.

Evaluations of programs designed to improve some condition in society are frequently conducted as natural experiments. One example is a study evaluating a program developed by the Maryland Network Against Domestic Violence that provided social service advocates for victims.³⁴ In this study, the police and social service organizations designed their own intervention—providing advocacy, safety planning, and referral

services to women. The researchers came in afterward to study the efficacy of it.

At other times, researchers simply notice an “intervention” that takes place naturally and study it. For example, researchers compared criminal offenders who moved against those who did not to see if the locations of their crimes changed to be closer to their new homes.³⁵ The researchers did not design the intervention—assigning people to move—it just happened naturally. The researchers used a host of covariates to help keep the plausible alternative explanations to a minimum, including type of crime, time between crimes, where previous crimes were committed, and where previous homes were. Despite these precautions, a number of other plausible explanations are discussed, as they should be.

This is a good place to note that researchers sometimes use terms such as quasi experiment and natural experiment interchangeably; for precision of language, this book uses *quasi experiment* for studies where researchers design the intervention and *natural experiment* for studies that take advantage of some naturally occurring intervention. Neither of these uses random assignment. Articles that report true experiments or **laboratory experiments** with random assignment do not typically use the terms *true* or *laboratory* but are just called experiments, and refer to anything not conducted in its natural setting that uses random assignment.

Another example of a natural experiment, a classic in the field of communication, is the study of three Canadian towns: one that had no TV, one that had one TV channel, and another that had four channels. This was back in the days when television signals were broadcast over the airwaves through antennas, and small towns were the last to receive television. All three towns in the study were revisited three years after the no-TV town had gotten TV. This research added longitudinal knowledge about the short- and long-term nature of television’s effects on children’s aggressive behavior, reading skills, cognitive development, leisure activities, use of other media, sex-role attitudes, and other personality traits and attitudes.³⁶ In these experiments, people were assigned to conditions by forces other than researchers—that is, whatever led them to live in those towns. There is some treatment or exposure to something; in this example, it was exposure to television in varying levels—no TV, one TV channel, four channels—that leads to the ability to say that TV exposure caused any differences in outcomes. Because it was not a randomly assigned, in-laboratory experiment, there are many other things that could have caused the differences or confounded the outcomes,³⁷ but the researchers did everything they could think of to control for these.

Natural experiments are becoming increasingly popular in many different disciplines.³⁸ Researchers can take advantage of changes in the world to conduct natural experiments. For example, when business scholars studied the changes in Peru’s soft-drink market before and after the country entered into a free trade agreement with the United States, they used as a control group the country of Bolivia, which has no such agreement.³⁹ They selected the control country by matching important characteristics such as demographics, income growth, population, and economic trade indicators. While the comparison countries are not exact, they are as close as the researchers could get. The authors discuss the limitations and include suggestions for future studies that would include more controls for wage rates, economic and political stability, and using

more than one country as a comparison.

Another study in London looked at public health before and after the introduction of free bus rides for young people.⁴⁰ Potential confounds included the fact that other policies designed to change people's choice of transportation had also been introduced recently (e.g., higher charges for driving during congested times), a change in cultural attitudes in general (e.g., walking to prevent obesity increased), and that there was no control group (all people under eighteen were given free bus passes). They list other limitations of a natural experiment as including a weaker ability to make causal claims than true experiments and difficulty generalizing beyond the single case being studied. The paper explains well the trade-offs between a realistic setting and internal validity, and offers some solutions that include mixing designs and data collection. They say the results are "good enough" evidence and "as robust an evaluation as possible."⁴¹

Advance planning allows researchers to obtain approval to study human subjects from their **Institutional Review Boards** (IRBs; more on this in [chapter 11](#)). Not all situations that are ripe for natural experiments come with advance warning, however—for example, the study of how re-incarceration rates changed after Hurricane Katrina, forcing parolees to spread out around the state rather than be concentrated in particular neighborhoods,⁴² or how attitudes toward welfare recipients changed before and after riots.⁴³ These studies can only be done if researchers use data collected before the event, receive approval after the fact, or are in the enviable position of having already obtained IRB approval for a study on a similar topic before the event happens and can quickly get approval for an amendment. Some even make a distinction between natural experiments and "nature's experiments," but this book will not go into that.⁴⁴ It is never a bad thing to keep an eye open for opportunities such as this.

Field Experiments

While natural experiments typically do not have the benefit of random assignment because people are exposed to the experimental or control conditions based on natural factors outside the control of researchers, there is another type of experiment, the field experiment, that takes advantage of real-world settings but also employs random assignment to treatment and control groups.⁴⁵

The term *field experiment* makes the distinction between experiments conducted in natural settings versus laboratory settings, even though not many "lab" experiments are conducted in actual laboratories anymore. For example, experiments that use survey software (covered in [chapter 7](#)) can be conducted in the subjects' own homes. Field experimentalists consider any setting other than the environment under which something would naturally occur to be a lab. Anything conducted on a college campus is a lab experiment, unless the actual context is a college setting—for example, a study of cheating on tests. Another example would be having participants look at TV commercials on their home computer and evaluate them to be considered a lab experiment. Even though the subjects are in their own homes, they are not looking at the ads on TV and

they know that they are doing so for a research study. That is another key distinction for field experimentalists; subjects should not be aware of being studied. Field experiments “strive to be as realistic and unobtrusive as possible.”⁴⁶ The focus on realism and subjects being unaware of being studied stems from concerns about misleading results due to people behaving differently when they are being watched, known as the Hawthorne effect, and reactivity, or participants wanting to give the “right answer” or the answer the experimenter wants. Whether these are serious problems with lab experiments is unclear, as few studies replicate experiments in both field and lab conditions in order to estimate treatment effects.⁴⁷ Field experimentalists also note that what works in a lab setting might not work in the real world.⁴⁸ Some effects can be immediate and strong, so they show up in a lab experiment but decay over time and would show up weaker in a field experiment.

Another objection field experimentalists have to lab experiments is the potential lack of realism in the messages or stimuli created by researchers. These are all important concerns, and lab experimentalists should be careful to see that their stimuli are as realistic as possible, having practitioners in their field create or review the interventions. This is a topic of [chapter 9](#) on creating stimuli.

Field experiments should be authentic on four dimensions: the participants, treatment, context, and outcome measures.⁴⁹ Participants should be real voters, not students pretending to be, for example. The treatment should be a real political debate, not one fielded by actors. The context should have voters watching TV in their living rooms, not with a group of strangers in a university classroom. And the outcome measures should be their votes or donations to the candidate, not self-reports of their intentions to vote or donate.

Field experiments can be more expensive and difficult than traditional lab experiments, and also ethically challenged given that one hallmark is that subjects do not know they are participating in an experiment (see [chapter 11](#) on ethics). Gerber and Green discuss in detail three of the most common challenges:⁵⁰ Briefly, noncompliance is when subjects that were assigned to one treatment actually got something else; attrition is when outcome measures are not obtained for every subject; and interference occurs when subjects talk to each other, compare notes, or remember treatments. These issues are minimized or nonexistent in the environment of a lab experiment.

Whereas lab experiments are commonly employed to test theoretical propositions, where tightly controlled conditions are important, field experiments are better for applied studies—for example, evaluations of a program’s effectiveness.⁵¹

A few examples of field experiments:

To see if interacting directly with a politician could persuade people to change their attitudes about issues, their assessments of the politicians’ qualities, and how they vote, researchers used online town hall meetings, noting that the only studies up to that point had used laboratory settings that simulated only a few of the characteristics of personal interactions.⁵² Previous research was equivocal about whether real-life results were the same as those from hypothetical settings. So the researchers recruited U.S. senators and members

of the U.S. House of Representatives to interact with their constituents in a real-time online forum. The intervention in this field experiment was created by the researchers but was more true-to-life than a mock town hall with an actor playing a politician. Citizens who participated were randomly assigned either to participate in the online discussion with the politician or to the control group, which only received reading material with background about the issues. Those in the treatment group got the same reading material but also participated in the online session. You can see some of the same kinds of effort exhibited here as Philip Zimbardo did in recruiting the Palo Alto police to “arrest” his prison experiment subjects.

Other researchers did a randomized field experiment in the Netherlands on the effectiveness of an extended day program on elementary students’ math and language learning.⁵³ They noted that the research on program effects in education rarely used randomized experiments and that quasi experiments that did not have proper controls were the norm. In this study, the researchers randomly selected students and offered them the chance to participate in the program rather than allowing them or their teachers to select who participated. Like studies before them that used randomized experiments and found small to nonexistent effects, this one did not find any effects. It is important to know when different methods generate opposite results; for example, on this topic, quasi experiments without random assignment were likely to show effects, but more rigorous randomized true experiments were not. In this case, the field experiment, using random assignment, helped educators know which results to have more confidence in.

As with the terms quasi experiment and natural experiment, researchers also use the term *field experiment* to mean different things. In this book, I advocate the use of the term field experiment to mean an experiment that is conducted in a natural or real-life setting and also uses random assignment. I use quasi and natural experiments when subjects are not randomly assigned to treatment and control conditions.

This book is devoted primarily to true experiments, also called lab experiments, although the concepts covered here are applicable to other types of experiments as well. Those conducting a quasi, natural, or field experiment should also consult texts that address the specific issues associated with those designs.

In summary, there is no perfect experiment. One cannot usually have the benefit of a real-world setting and also have random assignment. The best researchers can do to understand a particular phenomenon is conduct many different studies of varying designs that have at least one of these features, using different contexts. Replication under different conditions affords more confidence that what has been found is real. In fact, being bold enough to understand when an opportunity presents itself to study something by an unconventional method is one of the hallmarks of a creative researcher.

This chapter is not an exhaustive compilation of all the many different experimental designs available. Many other options exist, and the creative experimentalist will be open to discovering new and better ways to study important causes and effects. The next chapter begins this textbook’s sole focus on true or laboratory experiments, starting with issues of internal and external validity.

Common Mistakes

- Using pretests when they are not truly necessary
- Reporting an experiment as “exploratory” because it contains flaws
- Failing to randomly assign subjects to conditions. Studies that aspire to be true experiments but fail to use randomization are rarely published.

Test Your Knowledge

1. From Campbell and Stanley’s typology of experimental designs, which one is considered the gold standard?
 - a. Pretest–posttest control group design
 - b. Solomon four-group design
 - c. Posttest-only control group design
 - d. Static group comparison design
2. From Campbell and Stanley’s typology of experimental designs, which one is most recommended today?
 - a. Pretest–posttest control group design
 - b. Solomon four-group design
 - c. Posttest-only control group design
 - d. Static group comparison design
3. A pretest is essential for an experiment to be considered a true experiment.
 - a. True
 - b. False
4. What is the *main* reason for using random assignment?
 - a. To make sure subjects in treatment and control groups are equivalent on important characteristics.
 - b. To make sure the same number of subjects are in the treatment and control groups.
 - c. To ensure that the same people are not in more than one group.
 - d. To make it fair to all subjects.
5. Which type of experiment uses random assignment in a naturally occurring

setting?

- a. Lab experiment
 - b. Quasi experiment
 - c. Natural experiment
 - d. Field experiment
6. Which of the following designs is recommended when a pretest would NOT threaten to change subjects' performance?
- a. Static group comparison
 - b. Pretest–posttest control group
 - c. One-group pretest–posttest design
 - d. Quasi experiment
7. Which of the following is a strength of designs with pretests?
- a. It gives a baseline measure to compare any changes against.
 - b. Being observed or measured twice may cause changes in the subjects' performance.
 - c. Pretests are essential to true experimental designs.
 - d. It ensures the groups are equal on important characteristics that could otherwise cause any changes.
8. A study used two similar intact classes to see whether online or face-to-face teaching would result in greater learning. This is an example of:
- a. A natural experiment
 - b. A field experiment
 - c. A laboratory experiment
 - d. A quasi experiment
9. An experiment that is conducted in a natural or real-life setting and also uses random assignment is:
- a. A quasi experiment
 - b. A natural experiment
 - c. A field experiment
 - d. A true experiment
10. Which of the following allows the researcher to compare the differences before and after treatment, and also to tell if there were any effects of a pretest?

- a. Static group comparison
- b. Solomon four-group design
- c. Pretest–posttest control group
- d. One-group pretest–posttest design

Answers

- 1. b
- 2. c
- 3. b
- 4. a
- 5. d
- 6. b
- 7. a
- 8. d
- 9. c
- 10. b

Application Exercises

1. Use Google Scholar.com or your school library's database to find studies that use experimental designs in your discipline. Read three of the experiments that interest you most and identify the design of the experiment; is it a true or laboratory experiment, quasi experiment, natural or field experiment? Is it a pretest–posttest control group design, Solomon four-group design, or something else? What are some of the limitations, and how did the authors address them?
2. Examine the literature about your topic for the methodologies used. A grid-type chart will help you see which methods have been most used. Of the twenty-five-plus articles you read for the literature review assignment in [chapter 3](#), categorize them by method, including critical essay, focus group, interviews, ethnography, content analysis, survey, and experiment, among others. Which method has been most used? Is it appropriate for an experiment to be conducted now that establishes cause and effect?

Suggested Readings

Campbell, Donald T., and J. C. Stanley. 1963. *Experimental and Quasi-Experimental*

Designs for Research. Chicago: Rand McNally.

Crasnow, S. 2015. "Natural Experiments and Pluralism in Political Science." *Philosophy of the Social Sciences* 45 (4/5): 424–441.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: Norton.

Levy, Y., T. J. Ellis, and T. Cohen. 2011. "A Guide for Novice Researchers on Experimental and Quasi-Experimental Studies in Information Systems Research." *Interdisciplinary Journal of Information, Knowledge and Management* 6: 151–161.

Notes

1. D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. (Chicago: Rand McNally, 1963).
2. Ibid., 6.
3. David W. Catron and Claudia C. Thompson, "Test-Retest Gains in WAIS Scores after Four Retest Intervals," *Journal of Clinical Psychology* 35, no. 2 (1979): 352–357.
4. Ibid.
5. Brendon R. Barnes, "The Hawthorne Effect in Community Trials in Developing Countries," *International Journal of Social Research Methodology* 13, no. 4 (2010): 357–370.
6. John G. Adair, "The Hawthorne Effect: A Reconsideration of the Methodological Artifact," *Journal of Applied Psychology* 69, no. 2 (1984): 334–345; Ryan Olson et al., "What We Teach Students About the Hawthorne Studies: A Review of Content Within a Sample of Introductory I-O and OB Textbooks," *The Industrial Organization Psychologist* 41 (2004): 23–39.
7. E. Mayo, 1933; Chen-Bo Zhong and Julian House, "Hawthorne Revisited: Organizational Implications of the Physical Work Environment," *Research in Organizational Behavior* 32 (2012): 3–22.
8. Olson et al., "What We Teach Students."
9. Henry A. Landsberger, *Hawthorne Revisited. Management and the Worker: Its Critics, and Developments in Human Relations in Industry* (Ithaca, NY: Cornell University, 1958).
10. J. G. Adair, D. Sharpe, and C. Huynh, "Hawthorne Control Procedures in Educational Experiments: A Reconsideration of Their Use and Effectiveness," *Review of Educational Research* 59, no. 2 (1989): 215–227.

11. Barnes, "The Hawthorne Effect in Community Trials"; Zhong and House, "Hawthorne Revisited."
12. Baptiste Leurent et al., "Monitoring Patient Care through Health Facility Exit Interviews: An Assessment of the Hawthorne Effect in a Trial of Adherence to Malaria Treatment Guidelines in Tanzania," *BMC Infectious Diseases* 16 (2016): 1–9; Jim McCambridge, John Witton, and Diana R. Elbourne, "Systematic Review of the Hawthorne Effect: New Concepts Are Needed to Study Research Participation Effects," *Journal of Clinical Epidemiology* 67, no. 3 (2014): 267–277; Magnus Hansson and Rune Wigblad, "Recontextualizing the Hawthorne Effect," *Scandinavian Journal of Management* 22, no. 2 (2006): 120–137.
13. Olson et al., "What We Teach Students."
14. Martin T. Orne, "Demand Characteristics and the Concept of Quasi Controls," in *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*, ed. Robert Rosenthal and Ralph L. Rosnow (Oxford: Oxford University Press, 2009): 110–137; D. Steele-Johnson et al., "Goal Orientation and Task Demand Effects on Motivation, Affect, and Performance," *Journal of Applied Psychology* 85, no. 5 (2000): 724–738.
15. Martin T. Orne, "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications," *American Psychologist* 17 (1962): 776–783; "Demand Characteristics and the Concept of Quasi-Controls," in *Artifact in Behavioral Research*, ed. R. Rosenthal and R. Rosnow (New York: Academic Press, 1969), 143–179.
16. Barnes, "The Hawthorne Effect in Community Trials."
17. Adair, "The Hawthorne Effect."
18. David Salsburg, *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century* (New York: W. H. Freeman, 2001).
19. Campbell and Stanley, *Experimental and Quasi-Experimental Designs*.
20. Kaanan Butor-Bhavsar, John Witton, and Diana Elbourne, "Can Research Assessments Themselves Cause Bias in Behaviour Change Trials? A Systematic Review of Evidence from Solomon 4-Group Studies," *PLoS ONE* 6, no. 10 (2011): 1–9; Campbell and Stanley, *Experimental and Quasi-Experimental Designs*; Shlomo Sawilowsky and D. Lynn Kelley, "Meta-Analysis and the Solomon Four-Group Design," *Journal of Experimental Education* 62, no. 4 (Summer 1994): 361.
21. R. Barker Bausell, *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist* (Thousand Oaks, CA: Sage, 1994), 90.
22. Yair Levy, Timothy J. Ellis, and Eli Cohen, "A Guide for Novice Researchers on Experimental and Quasi-Experimental Studies in Information Systems Research," *Interdisciplinary Journal of Information, Knowledge & Management* 6 (2011): 154.

23. Stephen Gorard, "The Propagation of Errors in Experimental Data Analysis: A Comparison of Pre- and Post-Test Designs," *International Journal of Research & Method in Education* 36, no. 4 (2013): 372–385.
24. Ibid., 372.
25. Campbell and Stanley, *Experimental and Quasi-Experimental Designs*.
26. V. L. Willson and R. R. Putnam, "A Meta-Analysis of Pretest Sensitization Effects in Experimental Design," *American Educational Research Journal* 19 (1982): 249–258.
27. For an example of this, see Charles Boy Kromann, Morten L. Jensen, and Charlotte Ringsted, "The Effect of Testing on Skills Learning," *Medical Education* 43, no. 1 (2009): 21–27.
28. Levy, Ellis, and Cohen; J. W. Creswell, *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, 2nd ed. (Upper Saddle River, NJ: Pearson, 2005).
29. Ralph Woehle and Andrew Quinn, "An Experiment Comparing HBSE Graduate Social Work Classes: Face-to-Face and at a Distance," *Journal of Teaching in Social Work* 29, no. 4 (2009): 418–430.
30. Sameer B. Srivastava, "Network Intervention: Assessing the Effects of Formal Mentoring on Workplace Networks," *Social Forces* 94, no. 1 (September 2015): 427–452.
31. Froukje Snoeren et al., "Design of a Quasi-Experiment on the Effectiveness and Cost-Effectiveness of Using the Child-Interview Intervention During the Investigation Following a Report of Child Abuse and/or Neglect," *BMC Public Health* 13, no. 1 (2013): 1–16.
32. Sharon Crasnow, "Natural Experiments and Pluralism in Political Science," *Philosophy of the Social Sciences* 45, no. 4/5 (2015): 429.
33. Ibid.
34. Jill Theresa Messing et al., "The Oklahoma Lethality Assessment Study: A Quasi-Experimental Evaluation of the Lethality Assessment Program," *Social Service Review* 89, no. 3 (2015): 499–530.
35. Andrew Wheeler, "The Moving Home Effect: A Quasi Experiment Assessing Effect of Home Location on the Offence Location," *Journal of Quantitative Criminology* 28, no. 4 (2012): 587–606.
36. Tannis MacBeth Williams, *The Impact of Television: A Natural Experiment in Three Communities*, ed. Tannis MacBeth Williams (Orlando, FL: Academic Press, 1986).
37. Campbell and Stanley, *Experimental and Quasi-Experimental Designs*.
38. Crasnow, "Natural Experiments and Pluralism."
39. Phillip Baker et al., "Trade and Investment Liberalization, Food Systems Change and Highly Processed

Food Consumption: A Natural Experiment Contrasting the Soft-Drink Markets of Peru and Bolivia,” *Globalization and Health* 12 (2016): 1–13.

40. Judith Green et al., “Integrating Quasi-Experimental and Inductive Designs in Evaluation: A Case Study of the Impact of Free Bus Travel on Public Health,” *Evaluation* 21, no. 4 (2015): 391–406.

41. Ibid., 395–396.

42. David S. Kirk, “A Natural Experiment of the Consequences of Concentrating Former Prisoners in the Same Neighborhoods,” *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 22 (2015): 6943–6948.

43. Aaron Reeves and Robert de Vries, “Does Media Coverage Influence Public Attitudes Towards Welfare Recipients? The Impact of the 2011 English Riots,” *British Journal of Sociology* 67, no. 2 (2016): 281–306.

44. Mary S. Morgan, “Nature’s Experiments and Natural Experiments in the Social Sciences,” *Philosophy of the Social Sciences* 43, no. 3 (2013): 341–357.

45. Alan S. Gerber and Donald P. Green, *Field Experiments: Design, Analysis, and Interpretation* (New York: W. W. Norton, 2012).

46. Ibid., 9.

47. Ibid.

48. Ibid.

49. Ibid.

50. Ibid.

51. Ibid.

52. William Minozzi et al., “Field Experiment Evidence of Substantive, Attributional, and Behavioral Persuasion by Members of Congress in Online Town Halls,” *Proceedings of the National Academy of Sciences* 112, no. 13 (2015): 3937–3942.

53. Erik Meyer and Chris Van Klaveren, “The Effectiveness of Extended Day Programs: Evidence from a Randomized Field Experiment in the Netherlands,” *Economics of Education Review* 36 (2013): 1–11.

Notes

iNormally, experiments are not designed in such a linear fashion. Typically, researchers think about all the issues covered in this book simultaneously.

<http://dx.doi.org/10.4135/9781071878958.n4>

[statology.org](https://www.statology.org)

How to Perform t-Tests in Google Sheets - Statology

Published by Zach View all posts by Zach

4-5 minutes

Broadly speaking, there are three types of t-tests:

- One sample t-test
- Two sample t-test
- Paired samples t-test

This tutorial provides examples of how to perform each of these tests in Google Sheets.

Example: One Sample t-Test

Definition: A [one sample t-test](#) is used to test whether or not the mean of a population is equal to some value.

Example: A botanist wants to know if the mean height of a certain species of plant is equal to 15 inches. She collects a [random sample](#) of 12 plants and records each of their heights in inches.

The following screenshot shows how to perform a one sample t-test to determine if the true population mean height is equal to 15 inches:



The screenshot shows a Google Sheets interface with a formula bar at the top containing f_x . Below the formula bar is a table with 6 columns labeled A, B, C, D, and E. The first row of the table is highlighted in light blue. The first cell of the first row (A1) contains the number 1. The second cell of the first row (B1) contains the text "Plant height (inches)".

	A	B	C	D	E
1	Plant height (inches)				

1	Plant height (inches)				
2	14		Sample size	12	=COUNT(A2:A13)
3	14		Sample mean	14.3333	=AVERAGE(A2:A13)
4	16		Sample std. dev	1.3707	=STDEV.S(A2:A13)
5	13				
6	12		Hypothesized mean	15	
7	17		Test statistic t	-1.684847078	=(D3-D6)/(D4/SQRT(D2))
8	15				
9	14		Degrees of freedom	11	=D2-1
10	15		p-value	0.120145	=T.DIST.2T(ABS(D7),D9)
11	13				
12	15				
13	14				
14					
15					
16					
17					
18					
19					
20					

The two hypotheses for this particular one sample t test are as follows:

H_0 : $\mu = 15$ (the mean height for this species of plant is 15 inches)

H_A : $\mu \neq 15$ (the mean height is *not* 15 inches)

Because the p-value of our test (**0.120145**) is greater than alpha = 0.05, we fail to reject the null hypothesis of the test. We do not have sufficient evidence to say that the mean height for this particular species of plant is different from 15 inches.

Example: Two Sample t-Test

Definition: A [two sample t-test](#) is used to test whether or not the means of two populations are equal.

Example: Researchers want to know whether or not two different species of plants in a particular country have the same mean height. They collect a random sample of 20 plants from each species and record each plant height in inches.

The following screenshot shows how to perform a two sample

t-test using the **T.TEST()** function to determine if the two population mean heights are equal:

fx =T.TEST(A2:A21, B2:B21, 2, 2)							
	A	B	C	D	E	F	
1	Species 1 Height	Species 2 Height			E2		
2	14	15		0.530047	=T.TEST(A2:A21, B2:B21, 2, 2)		
3	15	17					
4	15	14					
5	16	17					
6	13	14					
7	8	8					
8	14	12					
9	17	19					
10	16	19					
11	14	14					
12	19	17					
13	20	22					
14	21	24					
15	15	16					
16	15	13					
17	16	16					
18	16	13					
19	13	18					
20	14	15					
21	12	13					
22							
23							
24							
25							

Note: It's also possible to perform a one-tailed two sample t-test with or without the assumption that both samples have the same variance. Refer to the [T.TEST documentation](#) to see how to adjust the assumptions for the test.

The two hypotheses for this two sample t test are as follows:

H₀: $\mu_1 = \mu_2$ (the two population means are equal)

H₁: $\mu_1 \neq \mu_2$ (the two population means are not equal)

Because the p-value of our test (**0.530047**) is greater than alpha = 0.05, we fail to reject the null hypothesis of the test. We do not have sufficient evidence to say that the mean height for this particular species of plant is different from 15 inches.

Example: Paired Samples t-Test

Definition: A [paired samples t-test](#) is used to compare the means of two samples when each observation in one sample can be paired with an observation in the other sample.

Example: We want to know whether a study program significantly impacts student performance on a particular exam. To test this, we have 20 students in a class take a pre-test. Then, we have each of the students participate in the study program for two weeks. Then, the students retake a test of similar difficulty.

The following screenshot shows how to perform a paired sample t-test to compare the difference between the mean scores on the first and second test:

fx =T.TEST(B2:B21, C2:C21, 2, 1)								
	A	B	C	D	E	F	G	
1	Student	Pre-test Score	Post-test Score			F2		
2	1	88	91		0.011907	=T.TEST(B2:B21, C2:C21, 2, 1)		
3	2	82	84					
4	3	84	88					
5	4	93	90					
6	5	75	79					
7	6	78	80					
8	7	84	88					
9	8	87	90					
10	9	95	90					
11	10	91	96					
12	11	83	88					
13	12	89	89					
14	13	77	81					
15	14	68	74					
16	15	91	92					
17	16	94	93					
18	17	95	97					
19	18	88	90					
20	19	84	84					
21	20	82	80					
22								
23								
24								
25								
26								

Note: It's also possible to perform a one-tailed two sample t-test with or without the assumption that both samples have the same variance. Refer to the [T.TEST documentation](#) to see how to adjust the assumptions for the test.

The two hypotheses for this paired samples t test are as follows:

H₀: $\mu_1 = \mu_2$ (the two population means are equal)

H₁: $\mu_1 \neq \mu_2$ (the two population means are not equal)

Because the p-value of our test (**0.011907**) is less than alpha = 0.05, we reject the null hypothesis of the test. We have sufficient evidence to say that there is a statistically significant difference between the mean pre-test and post-test score.