Education Research

Across Multiple Paradigms

Royce Kimmons
This book is provided freely to you by

EdTechBooks.org

**CC BY**: This work is released under a CC BY license, which means that you are free to do with it as you please as long as you properly attribute it.
# Table of Contents

Introduction .......................................................................................................................... 4
Acknowledgements .................................................................................................................. 7
1. Groundwork ..................................................................................................................... 8
   1.1. Research .................................................................................................................. 10
   1.2. Knowing .................................................................................................................... 19
   1.3. Theories ..................................................................................................................... 33
   1.4. Ethics ........................................................................................................................ 51
2. Paradigms .......................................................................................................................... 66
   2.1. Inferential Statistics ................................................................................................. 73
   2.2. Sampling .................................................................................................................... 78
   2.3. Qualitative Rigor ...................................................................................................... 91
3. Learning Theories ............................................................................................................. 102
Appendix A. Supplements .................................................................................................... 103
   Glossary .......................................................................................................................... 104
   References ...................................................................................................................... 105
Back Matter .......................................................................................................................... 106
   Author Information .......................................................................................................... 107
   Citation Information ......................................................................................................... 109
Introduction

What's the purpose of this book?

Being able to understand and interpret research findings is an essential skill for all education practitioners, and being able to conduct a research study from start to finish is part of what we expect from our education scholars. For all of these professionals, finding reliable, helpful guidance on how to effectively approach research can be difficult — a serious challenge in an age when research is ever-more-loudly praised as the hopeful solution for education's and society's afflictions.

Don't get me wrong, there are several great education research books on the market, and novice scholars would greatly benefit from reading them. These include such titles as *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* by Creswell & Guetterman.
(2018) and Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods by Mertens (2020), along with others. Many of these books are in their fifth and sixth editions and have guided generations of scholars through their first steps of becoming prolific education researchers. They are great books that should be highly valued.

Yet, reliable, openly-licensed materials on the topic of education research remain surprisingly limited. This presents access barriers to students wishing to learn about this topic and who are unable to afford an expensive textbook, while also limiting the reusability, interactivity, up-to-dateness, and usefulness of such materials for those of us involved in advanced education on research methods.

This textbook attempts to remedy this problem by providing a high-quality learning resource that documents modern approaches to education research and seeks to bring together all of the most important aspects of textbook quality — accuracy, usability, readability, helpfulness — into a single, free package.

As such, I’ve written this textbook with new scholars in mind, attempting to break down difficult concepts into understandable language and providing both explanatory examples and real-world connections to published research articles. In other words, I’ve written this textbook for my students in an attempt to drive down their costs and to improve their learning opportunities both while they are in my class and as they move beyond it. Furthermore, as an open, living work, my hope is that this textbook will continually evolve and be refined to better address the changing needs of education researchers in a world where the students, institutions, and societies we serve are ever-changing and always in need of professionals who are well-trained in this difficult craft.

To do this, the book is broken into the following sections: Groundwork, Paradigms, Methodology, Data Collection, Data Analysis, and Reporting Results.

In Groundwork, I lay necessary foundations for moving forward by defining a host of terms and grappling with age-old problems impacting how we approach research enterprises, such as what it means to know. This ensures that readers will understand the vocabulary that I use in subsequent chapters and sets the stage for exploring diverse approaches to research that might have fundamentally different assumptions on foundational issues.

In Paradigms, I then highlight some of the dominant approaches to doing education research today, exploring what the foundational assumptions of each seem to be, how this influences what researchers within each paradigm will do, and what this means for us as consumers and producers of such paradigm-directed work.

After exploding the reader's vision to these various approaches, I will then in Methodology begin shifting to more concrete matters that influence research design and the day-to-day activities of the researcher, such as establishing rigor, designing protocols, and ensuring ethical behavior. As I do this, I will continually circle back to the various paradigms discussed in the previous section to help the reader to understand how such considerations would either be approached differently or uniformly between paradigms.

From that point on, chapters will remain concrete and will focus on what good practice in education research looks like in the areas of Data Collection, Data Analysis, and Reporting. By starting off in the abstract and becoming increasingly concrete, I hope to give the reader both a birdseye, philosophical view of education research and also very practical, actionable guidance in how to do it well.
Also, I think it's worthwhile here to point out (in case you haven't noticed) that this book takes a less formal tone than may be normally used in other textbooks. For instance, I will use colloquialisms, contractions, and the first-person pronoun "I" with abandon. I do this to make the treatment of difficult concepts more approachable and to shed the deceptive cloak of the third-person, which academics often don to prevent the careful scrutiny of their ideas by making their very biased opinions look unbiased as they filter them through unnamed, third-person others (e.g., "one might conclude" vs. "I think"). If such frankness and lack of ceremony feel unnerving to you, dear reader, then I apologize for my impropriety at the outset, but I will nonetheless proceed in this fashion, believing that clarity and directness are more important for learning than is putting on airs of inhuman objectivity.

It is also important to acknowledge here that this book heavily reflects a U.S.-centric (or at least a North American-centric or Anglo-centric) exploration of the presented topics. Whenever I use historical examples or trends, for instance, I will generally draw upon those that have occurred in the U.S. Readers in other contexts may find such examples less useful for understanding the presented content, but hopefully the learning objectives and content will still generally be understandable to non-U.S. learners despite such culture-specific trappings.

Let's get started!

**References**


Acknowledgements

The gear mind icon used throughout this book was provided by Vecteezy.com.

Various other images used throughout this book were provided by Pixabay and Pexels.

All images in this book are either in the public domain or are licensed under a Creative Commons license, which will be noted by the image if applicable.
Groundwork

What fundamental understandings do we need to talk about education research?

Many research textbooks will jump right away into the nuts and bolts of doing research. That might seem appealing — after all, you're probably reading this because you were required to take a class, you want to do a specific study, you want to get published, or you just want to graduate, and all of those are good goals — but it's somewhat akin to embarking on a hike only to be blindfolded and dropped off by a stranger halfway up a mountain. Sure, it might be easier to get to the top, but you don't know where you are, how you got there, or even if you're heading up the mountain you had planned on.

So, in order to make sure that you don't waste your time, don't harm anyone, and don't wake up months or years down the road realizing that you have been short-sighted, misdirected, or unaware in your research, I think it's important for us to lay some groundwork first in a few key areas. Let's get our bearings, identify our goals, and make sure that we're getting there together.

In this section, we'll explore some fundamental aspects of research that may initially seem basic but are nonetheless essential to uncover in a thoughtful manner. This includes such topics as what is the nature of knowledge, what constitutes ethical behavior, and what socioeconomic factors have influenced education research throughout history and even today. I'll approach each of these (and other) topics in a fairly cursory manner but will do so to operationalize important terms and concepts.
that will be used throughout the rest of the book. I will then close this section with some structured guidance on how to select a research problem and how to write effective research questions, both of which are essential for moving forward with a study.

By going through this process, I hope to spark a desire in you to be self-aware and to grapple with some of these deeper issues in your work (such as the ethics of what we do) so that you will help encourage the field toward becoming more critical of our assumptions and thoughtful about our impacts on the world.

No one wants to slow you down. If you want to do research, then let's get you doing research as soon as possible. But an ounce of prevention is worth a pound of cure, and by laying just a bit of groundwork here at the outset, my hope is that you can approach your research in a more reflective manner from day one and avoid some of the pitfalls that even seasoned researchers occasionally fall into.

**Section Contents**
1.1

Research

What is education research, and what makes a good researcher?

Royce Kimmons

Imagine you are at a dinner party and begin chatting with someone you have not seen for many years. Trying to be a good conversationalist, you ask him, “So, what have you been up to lately?” To which he responds, “Oh, lots of stuff, but the coolest thing I’ve been doing is a ton of research into dark matter.”

Pause the scene. What do you think he meant when he said “research”? Did he mean that he was reading articles on Wikipedia? Did he mean that he was spending his nights at the library, poring over physics textbooks? Or did he mean that he regularly found himself deep underground working on a particle accelerator? In our vernacular, “research” could reasonably be interpreted in any of these ways. Literally, the word only means to look diligently for something (search) and then to look again (re-search). Yet, each of these interpretations is quite different and implies different processes, different types of effort, and different levels of expertise.

Key Terms

Imagine now that you are an elementary school teacher and your principal swings by your office for a word. She is very excited because she just heard about a new reading program and thinks it will drastically help your students improve their comprehension and fluency. “And do you know what the best part is,” she asks energetically. “It’s research-based. So, it should be easy to get the district to approve us to purchase it for every classroom.”
Pause the scene again. What do you think she meant by “research-based?” Did she mean that the program was designed upon literacy principles found in research literature? Did she mean that the company that made the program also wrote a white paper about how a group of teachers they gifted the program to found it to be useful? Did she mean that a market analyst had contacted previous clients and asked them how much they liked it? Or did she mean that an unbiased, outside organization had conducted randomized classroom implementations to see how the program influenced standardized reading test scores? And if that was the case, then how valid was the test, how drastic were the student improvements that were attributable to the program, and were all students benefited equally?

Again, even in this professional education setting, the term "research" might mean any number of things, allowing interested parties to pass just about anything off as "research-based" in some form or fashion — potentially equating the most thoughtful, rigorous experimental project with the most superficial word-of-mouth check into a product’s value. You can probably guess why this is a problem.

What is research?

If the term "research" is applied to education in too broad a fashion, then suddenly anything can count as research. Reading a blog post is research, asking a friend is research, noticing a headline is research, and all of these might be implied in the same breath as a deep, multi-year, ethnographic study exploring the experiences of an underserved population. Such imprecision and willy-nilly use of the term makes the rigor and clout of research lose its persuasive sway and confuses us into thinking that we are talking about the same thing when we may not be.

Conversely, if we apply the term “research” too narrowly, we also might miss out on truly meaningful and insightful examples of research that are necessary for moving forward. If, for instance, we assume that randomized controlled trials performed in a sterile lab are the only activities that should really count as research, then where will we develop the theories that are used to justify those trials, and how can we be sure that the results of those trials will work in vibrant, messy (non-sterile) classrooms? So, though there does seem to be a clear need to define research more narrowly than it is used in the vernacular, we need to be careful not to describe it so narrowly that we lose its key benefits.

Whenever education researchers use the term "research," we might sometimes have different meanings in mind (favoring certain methods or focusing on particular problems; cf. Phillips, 2006), but it seems fair to suggest that we would generally agree on at least four things:

First, research operates as a form of inquiry (Mertens, 2010). The process of doing research is one of asking questions and seeking answers (Creswell, 2008). We might do this by audibly asking a question of a learner, relying on their own self-reports to find the answer, or by framing our question into an experimental test, using our own observations to find it. We might ask a broad question, like "what's going on in Zavala Elementary," or we might ask a narrow, specific question, like "does using Technology X significantly improve performance among 6th graders on the Super Valid Measure of Visible Learning (SVMVL) test?" We might ask a single question in a research study or we might ask many. Researchers operate differently in this regard, but we all ask questions for the purpose of finding answers.

Second, research is systematic (Mertens, 2010). While trying to answer our questions, there are specific, intentional steps we take. We don't just shout against the wall "Are learning styles real?!"
and expect the universe to answer. Rather, we embark on a step-by-step process to ever-more-certainly approach the answer, such as proposing "if learning styles were real, then designing instruction for a student's reported style should help them perform better" and then performing necessary steps for checking to see if this is the case.

Third, research is auditable. As we systematically move toward answering questions, we document the steps that we are taking and how they either align with or deviate from the systematic approaches others have taken in the past. This makes the process we followed, from start to finish or from question to answer, visible to anyone else who might also be interested in our question, allowing them to interrogate any step we took and to determine for themselves whether the answer we found should be trusted.

Fourth, research is empirical. Unlike other forms of inquiry that operate exclusively in the internal mind of the inquirer, such as artistic expressions of one’s subconscious or mathematical proof generation, education research relies upon evidence observable in the external physical and social world. This means that research involves the observation or analysis of people or things, not just ideas, and though theory generation, literature synthesis, and critical reflection might all be valuable activities in their own right, they can only be called education research as they are combined with empirical analysis, such as of people or social institutions.

Merging these pieces together, we are able to construct a fairly simple and straightforward definition of education research as "systematic, auditable, empirical inquiry" that is done within the realm of education, such as to address problems impacting learners, educators, or their educational institutions.

**Learning Check**

Which of the following are empirical forms of inquiry?

a. Observing classroom teachers  
b. Reviewing existing literature  
c. Testing student performance  
d. Interviewing parents

**For whom do we conduct research?**

By emphasizing systematicity, auditability, and empirical verification, this definition implies that as we undertake research we are doing it with a goal of influencing others' thinking. That is, if we find an answer, we want others to trust our answer and for them to find it compelling or persuasive based on the process we followed. This may further complicate matters in education, because the imagined audience that we conduct research for might represent a variety of people who have a stake in education, including teachers, students, parents, policymakers, administrators, or the general public.

Problems in education are complicated partially because they influence such different stakeholders, but also because they are found in complex and unique social institutions (e.g., schools, universities, communities) and involve various disciplinary knowledge domains (e.g., behavioral psychology, sociology, teacher practice, curriculum design). So, attempting to solve even a single problem in education might involve drawing upon several knowledge domains, implementing interventions in
many unique social institutions, and attempting to influence diverse stakeholders. For researchers, this means that upon completing a study, we might write up our results in a journal article for other researchers to learn from, might summarize our findings for policymakers in a brief, might post a short practitioner guide on a parenting blog or teaching magazine, or might be interviewed by reporters about the implications of our work for the public. Such complexity makes education research quite different from research in many other fields because the expected audiences for our research are quite varied, and this influences what questions we will ask, how we will go about answering them, and where we will report our answers (Floden, 2006).

It turns out that each of these audiences has very different values, expectations, and practices when it comes to research, and as researchers, we will often find ourselves doing research for mixed audiences to achieve our goals. For instance, other researchers in our audience will value work that exhibits "careful design, solid data, and conclusions based on cautious and responsible inference" (Plank & Harris, 2006). Researchers also relish nuance and recognize the importance of context for their results to be valid, meaning that they may be reserved in providing sweeping policy suggestions or strategies for reform that are intended to work in every situation. Researchers will also generally target scholarly journals as publishing venues both to make their work available to other researchers and to meet scholarly rigor requirements placed upon them for tenure and promotion by their educational institutions.

Practitioners, on the other hand, such as teachers, instructional designers, curriculum developers, and other education professionals, will value ready applicability of research findings to their current needs and contexts. Teachers will not care whether STEM achievement is lagging nationwide if their own students are performing acceptably, and they will not adopt purported solutions to achievement gaps in their own classrooms if they do not think the solutions will be effective or can be implemented in their specific contexts. Because this context piece is so vital to teachers, they are more likely to trust their colleagues (who are closer to their contexts) than professional researchers (who might be more informed but also more contextually distant). This does not mean that practitioners are antagonistic toward emerging research or are resistant to change but rather that applicability is their top, guiding priority, and research that ignores context or reports in abstraction will generally be ignored.

As a third group, policymakers tend to be intensely interested in particular research questions that reflect issues of current concern to the general public, but such interest is typically fleeting, and "as public attention shifts, they quickly move on to other issues" (Plank & Harris, 2006). Policymakers will value research that is compatible with their prior beliefs and commitments, that are timely, that are simple, that have clear action-oriented implications, and that are generalizable beyond specific contexts (Plank & Harris, 2006). This leads to very blunt or simplistic interpretations of research findings that may lack nuance or engineered precision and also reflects limited trust in scientific findings wherein "Aunt Minnie's opinion may carry equal or greater weight [to a published research study] because of her 30 years as a teacher or parent" (Plank & Harris, 2006).

And finally, the general public is a fourth potential audience for research that both shapes what policymakers care about and also what practitioners are required to do (e.g., as parents make mandates upon school boards). Like some of the other groups, the general public has limited understanding of research methodology and limited interest in and access to scholarly journals. This means that they will typically rely upon popular news outlets, such as television news hours, news sites, and popular blogs, to only highlight research findings that will be of broad interest and to summarize research methods and findings in ways that are readily understandable and that connect
to social issues and political topics of the day. Such outlets also will be heavily biased toward research studies with findings that seem to be novel, troubling, or contentious, because they will be more likely to garner readers, and will favor the dissemination of interesting pseudoscience and debunked beliefs to real science (cf., Willingham, 2012). A study that considers student STEM preparation, then, might only be reported by popular news outlets if it contradicts current wisdom, connects to a popular political agenda, or suggests a crisis (e.g., “the U.S. economy is becoming less competitive due to poor STEM preparation for the workplace”). This means that research will have a highly mediated relationship to the general public, where most people will only read research results that have been intentionally sifted, collated, and rebranded through multiple layers of editorial bias, thereby influencing general misunderstanding or misrepresentation of research results, which will, in turn, influence policymaking and practice through public pressure.

Such divides between potential research audiences are often lamented but may be necessary and deserving of respect in democratic societies. Elected officials, practitioners, and the general public cannot reasonably be expected to study and evaluate the research literature at the same depth that scholars do, and so the responsibility for bridging any such gaps will fall mainly on scholars to “make their work more accessible and useful to [other] audiences” (Plank & Harris, 2006). Thus, researchers cannot study for the sake of a journal article publication alone and be done with the research process, but interpreting and disseminating results to diverse and broad audiences should also be an integral part of the research process. How this is done via a variety of publication venues will be explored more deeply in later chapters on research reporting.

Learning Check

When considering the results of a research study, what will teachers likely consider to be most important?

a. Rigor  
b. Consensus  
c. Applicability  
d. Broad Impact

When considering the results of a research study, what will researchers likely consider to be most important?

a. Rigor  
b. Consensus  
c. Applicability  
d. Broad Impact

When considering the results of a research study, what will policymakers likely consider to be most important?

a. Rigor  
b. Consensus  
c. Applicability  
d. Broad Impact
What makes a good education researcher?

Research serves several important functions in our society, such as adding to general knowledge, improving practice, and informing policy debates (Creswell, 2008). Unlike other areas of inquiry where research might be undertaken for the sheer purpose of learning about the world (e.g., naturalists like Thoreau experiencing nature), education research tends to start with some specific assumptions and goals, and at the heart of these lies a central, golden premise that we go about doing research for the direct purpose of improving education.

What "improving education" means may vary between people, projects, or contexts, and it may include a host of goals including improving curricular quality, improving equity, reducing cost, improving people's lives, or improving the economy. However, no matter what these underlying goals may be, education research is unique as a discipline, because it attempts to connect systematic data-gathering and analyses to better achieve improved educational outcomes either by informing what needs to be done or by adding legitimacy to what may already be happening. This typically takes the form of researchers tackling social problems and issues that can be very complex and highly politicized — such as school choice, racial segregation, economic disparity, individual learner differences, standardized testing, and so forth — and requires researchers to both rigorously grapple with problems in thoughtful, insightful ways via critical thinking and also to socially influence the world toward becoming a better place via advocacy.

This two-fold mission of education researchers may at first seem odd to some who are new to the discipline, because Western narratives of research have historically positioned the researcher as an objectivist outsider who observes the world at a distance, such as an ancient white man in an ivory tower peering down on the masses through a telescope. Such monastic views of the researcher have emphasized his (yes, his; we will discuss the role of gender in education research in a later chapter) role as a dispassionate observer who must be physically separated and free from the biases and problems of the day to provide dispassionate analysis of the state of the world from a privileged position. The reasoning for such a view is that in order to provide rational, valid guidance to the world, the researcher must prevent passions, connectedness, and proximity from skewing his rational judgment and unbiased analysis of the facts as they are. Through this separation, such narratives attempt to frame the researcher as the quintessential critical thinker, a purely analytic machine, an inhuman demigod, or a brain in a vat, so that we can confidently trust his judgments as thoughtful, unbiased, and fair.
Yet, though critical thinking is indeed a necessary requirement of the researcher, treating critical thinking as an unequivocal virtue and monastic approaches as the only or even best way of knowing seems dubious. In subsequent chapters, I will explore ways of knowing in more detail but will close this chapter with a brief explanation of why critical thinking and advocacy must work in tandem as the most useful and meaningful way forward for education researchers.

All humans face problems that require critical thinking on a daily basis. Critical thinking is characterized by thoughtfully weighing alternatives and evidence, resisting premature conclusions, and attempting to understand how our own biases and assumptions about the world may be influencing what we are experiencing (cf., Fig. 1). Such processes require concerted intellectual effort and can place us in a state of paralytic limbo where we willingly withhold action or judgment until we can fully weigh our options between which grocery item to purchase, which class to take, which book to read, which movie to watch, which person to talk to, or which words to use. Such activities are essential for our wellbeing because, without them, we would regularly purchase expired produce, speak to people in languages they cannot understand, and generally waste our time on efforts that have little or no benefit. Yet, we engage in critical thinking for the purpose of coming to a reasonable conclusion, and if we fail to come to a conclusion, then have we done it right?

Suppose I have a few hours to spare. I switch on a video streaming service and begin browsing the list of movies. If I spend the entirety of my time browsing the options but never settle on one to watch, then have I successfully employed critical thinking? I might have weighed my options, read the synopses, looked at ratings, and checked out the cover photos, but if this process doesn’t produce a decision that I then act upon, then was the effort worth it? If you and I start a conversation, and I stare at you blankly for 10 minutes, because I am trying to decide on the best words to use, then would you praise me as being a great critical thinker?

Critical thinking exists to serve a purpose: to inform action. As Reber (2016) explains, “thinkers can deliberate too much," leading to worse outcomes and suggesting that we need "a criterion regarding when to stop" (p. 29). Properly employed, then, critical thinking consists of temporary inaction that is undertaken to provide better-informed action. If an action does not follow, as in the cases of extreme intellectual skepticism, cynicism, and the paralytic social behaviors mentioned above, then such paralysis should not be lauded as expert critical thinking but as failed critical thinking. That is, critical thinking only serves its purpose if it leads us toward becoming advocates for something, by choosing which movie to watch, which words to say, or which educational initiative to champion.

By shifting from critical thinking to advocacy, we take an (again) temporary stance of confident action toward the problems we are trying to solve and act upon the guidance that our critical thinking has provided. This might mean adopting an intervention, advocating for a policy change, or writing a provocative op-ed. We boldly act on what we have learned and attempt to put it into action. Once we do this, however, the essential element that keeps us true to form in our roles as education researchers is that we always step back to critical thinking in the next movement, being willing to critically analyze new evidence, evaluating how our forward movements worked to influence our goals, and so forth.
The imagery that emerges from this constant stepping forward in advocacy and back in critical thinking is that of a pair of dancers aesthetically moving about a stage. At the macro-level, movement is occurring, change is happening, and the dance progresses to an eventual beautiful finale, even though the micro-footsteps of the dancers skirt back-and-forth in every direction. This is quite different from the regimented forward-marching of advocacy alone or the backpedaling and directionless navel-gazing of critical thinking alone in that each of these two essential components of research work together in tandem to lead us to the desired result.

We engage in this dance not to try to announce to the world from our tower that we understand the masses better than they understand themselves, nor to shout back up toward the tower "No thanks, we’ve got everything figured out!" Rather, the best education researchers find their homes in this constant dance of stepping back to think critically about evidence, stepping forward to advocate for improvement, stepping back to think critically about new evidence and assumptions, and so forth ad aeternum.

References


1.2

Knowing

What does it mean to know, and how do we discern truth in education research?

Royce Kimmons

If the goal of research is to come to know something about the universe, then we need to begin by having a clear understanding of what it means to know anything. On first exposure, this may seem like an absurd question. "Of course I know what it means to know ... I just know." Such a common-sense approach to knowledge seems reasonable until we encounter other people who don't just know the same way we do — which we constantly do — at which point we find ourselves needing to resort to providing evidence, making arguments, or sharing stories in order to either justify our own knowledge or to try to influence others to know what we know.

This problem of knowing is ancient because people have always valued the ability to know but have also always claimed to know things that they had difficulty proving, defending, or explaining. Some of the earliest evidences of this problem date back to ancient Greek writings, such as Plato's recorded dialogue between Socrates and Meno, wherein Socrates, amongst other things, presents an apparent paradox of knowledge: "[A man] cannot search for what he knows — since he knows it, there is no need to search — nor for what he does not know, for he does not know what to look for." This is a problem for us, because if we cannot search for knowledge, then research is impossible, and if we only search for what we already know, then research is useless.

Epistemology, or the study of knowing (Steup, 2005), has been of philosophical interest for thousands of years precisely because the concept of knowing is ingrained in our day-to-day lives, but there seem to be different types of knowing and different types of certainty that are sometimes difficult for us to understand and articulate. The situation for education researchers is particularly poignant here
because, in addition to grappling with our own knowing, we are also in the business of knowing about learning, which is itself a process of coming to know. So, we must grapple with knowing at multiple levels: for our students, for ourselves, and for our research audiences.

This chapter will attempt to provide some groundwork regarding knowing that will be drawn upon in subsequent chapters. I will begin by highlighting three types of knowing and then explore processes that we often follow to come to know something. I will then operationalize some terms that are commonly used in the realms of knowing, truth-seeking, and science so that we can have a common language for delving into important epistemological questions in subsequent chapters.

**Key Terms**

**Types of Knowing**

The history of epistemology can broadly be summarized as a dialogue between two different types of knowing: objectivity and subjectivity. In this dialogue, thinkers have grappled with what we can actually know about things in the world (objects) when any knowing that occurs can only happen in the mind of the knower (subjects). In more recent years, sociologists, historians, and others have further complicated this issue by suggesting that focusing only on the lone observer/thinker in this process is limited because it ignores cultural or shared aspects of knowing. I will now explore each of these types of knowing — objectivity, subjectivity, and intersubjectivity — in more detail.

**Objectivity**

Objectivity assumes that people can directly observe the world, or facts or truths in it, and have an accurate knowledge of what they are observing (Reiss, 2014). In a simple case, a person (the subject) might observe a flower (the object) and be able to tell you what it is: “It is a pink rhododendron with five leaves.” If this is objectively true, then anyone else seeing the object would draw the same
conclusion: “Yes, it is, in fact, a pink rhododendron with five leaves.” As long as we can make sure that the object doesn’t change and that anyone can observe the object under similar conditions with the same result, we can conclude that this is a truth embedded in the world: “This is a rhododendron, and it is pink.” In other words, objectivity assumes that the world can be accurately understood through unbiased observation.

Objectivity is the traditional goal of science and research because both assume that the world exists separate from the mind of the knower and that it can be universally known through observation and inquiry as a dispassionate (almost robotic) outsider. Like the lab-coat-clad scientist behind a one-way mirror, the objectivist researcher assumes that she can understand what she is observing from the outside-in and tries to separate herself completely from what she is observing. This is also called an etic perspective, wherein the researcher sees herself as a potential contaminant of what is being studied and intentionally separates herself physically, psychologically, and emotionally from it. For objectivist inquiry, the biases, emotions, intuitions, and rational limitations of the researcher are threats to the validity of the observation, because they have the potential to warp the observation or (even worse) to make the researcher influence what is being observed.

Thus, when seeking objective knowledge, the only proper role of the knower is that of a machine-like observer who is carefully controlled and separated from what is being observed. By treating the knower in this manner, it is hoped that the actual observation will not be influenced and that any other knower, no matter who they are, will derive the same knowledge from the object being observed.

**Subjectivity**

Subjectivity, on the other hand, applies to knowledge as it exists in the mind of the subject or as an interpretation of objects in the universe. In the simplest case, a person (the subject) might still observe a flower (the object), but what they actually perceive, learn, or know as a result of this observation will always be determined (or at least influenced) by the knower. In matters of aesthetics, this is clearly understandable, because some might think that the flower is beautiful while others might not, but it can also apply to more seemingly concrete aspects of knowing as well. In this case,
would you know a rhododendron from a pansy? Even making this distinction to categorize the flower would require certain prior knowledge on the part of the subject, which would dictate the categorization. Would you also see it as pink? Or would you see red? Or mauve? Or grey (in the case of deuteranopic people or animals such as dogs that have different photoreceptor cone structures)?

This latter point is important because when talking about objectivity, we assumed that the observer could objectively state that the flower was pink, but this assumes that color itself is something that is universal and exists in the world. However, color is itself only a physiological interpretation of wavelengths, which, as shown above, humans might interpret differently but, even more importantly, is nothing but an interpretation and is not a characteristic of the flower itself. The petals of the flower might be made of such a substance that they reflect certain wavelengths of light, but they do not objectively have a color any more than a politician can be said to have character or a painting can be said to be beautiful. Color is a perception and, as such, it can only ever exist in the mind. (This is similar to Berkeley's famous question that "if a tree falls in the forest and no one hears it, does it make a sound," because sound only exists as an interpretation in the mind of the knower.)

Thus, a weak claim for subjectivity would be that at least some knowledge is subjective in nature, meaning that it is interpreted and categorized entirely by the observer (subject), but many argue for a stronger claim that all knowledge can only ever be subjective. This calls into question the possibility of objectively knowing anything because no matter how carefully the observer is controlled, the very act of observation requires interpretation and may even be a form of influence (such as the Heisenberg uncertainty principle and "observer effects" in quantum mechanics). This can be interpreted to make any claims of objectivity uncertain and suggests that knowing cannot exist outside of or separate from the knower. When applied to research, subjectivist approaches to knowing would recognize and embrace the researcher's role in the research process. Also called an emic perspective, this would typically require the researcher to experience the world firsthand as a participatory insider and to articulate (and value) how their own biases, attitudes, expectations, and experiences might be shaping what they are learning through this process of inquiry.

**Intersubjectivity**
One clear problem with subjectivity is that if all knowledge is subjective, then how can we ever prove anything to one another or claim to know anything about the universe with any level of certainty? In response, some will double-down on objectivity as the only reasonable path forward, while others might propose an intersubjective approach, which seeks to find consensus, commonality, or shared experiences between subjective individuals. Similar to subjectivity in how it views the individual’s relationship to the world, intersubjectivity is nonetheless markedly different, because it assumes that people and the act of knowing is inherently social (not individualistic) in nature.

This means that intersubjective knowledge could only be gained as the process of meaning-making extends beyond the person-world (subject-object) relationship to encompass subjective experiences and knowledge between multiple, diverse people (subjects). An example of this would be a historian wanting to accurately understand the U.S. Civil War who studied the artifacts and experiences of various impacted peoples, including rich, poor, black, and white Northerners and Southerners. In so doing, the historian would not be seeking “the accurate” depiction of what happened (as with objectivity) or even one group or person’s limited perspective of what happened (as with subjectivity) but would be seeking to more deeply and fully understand the event by weighing the biases, attitudes, beliefs, and experiences of all involved to better perceive the event as a shared experience.

Like subjectivity, intersubjectivity is not capable of making claims on absolute truth or the world as it objectively exists, but by focusing on shared, multifaceted knowledge, it is sometimes used as a seemingly more responsible, trustworthy, or rich way of knowing than subjectivity. One research struggle with this approach, though, is that as multiple subjective experiences are meshed together, difficulties may arise for the researcher in determining whose perspectives to emphasize or highlight and whose perspectives to ignore or minimize. This typically requires some external standard of comparison (e.g., only including narratives that align with an existing assumption of objective truth) which can suggest overreach into objectivist aspirations, thereby subverting responsible subjectivist limitations.

**Forms of Reasoning**

The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge.

(Stephen Hawking)

Not only are there different types of knowing, but we also follow different processes of reasoning to develop knowledge. There are at least three common processes that we follow, including deduction, induction, and abduction. You have probably heard of deduction vs. induction, but abduction is not commonly used outside of formal logic to describe a form of reasoning. Furthermore, most definitions of deduction vs. induction are only loose applications of what the terms actually mean in formal logic. We engage in each of these processes every day, and each connects to the three types of knowing in specific ways. Furthermore, all are used by researchers in some fashion, so we will now explore each in more detail.

**Deduction**

You might have heard that deduction means moving from generalities to particulars, while induction is the opposite, but this is not always accurate. Rather, the core of the difference between deduction and induction is certainty (Douven, 2017). Deduction is always completely certain, while induction is an educated, reasonable guess.
To illustrate, suppose that we know that “all planets that are members of our solar system move around the sun” and that “Mars is a planet.” If these two premises are true, then we can determine with complete certainty that “Mars moves around the sun,” because moving around the sun is implicit in a planet’s being in the solar system. In other words, the definitional premise we are using for “solar system” and “planet” means that something cannot be a planetary member of our solar system and not move around the sun. Therefore, if Mars is a planet, then it must always, completely, certainly move around the sun. In this case, we did move from a general law (the definition of planets in the solar system) to a specific case, but that is only incidental to the central characteristic of deduction, which is that the logical step we took was completely certain. We could test this, of course, but if we found that Mars did not, in fact, move around the sun, then there was a problem with our first premise, and it is therefore not true. In such a case, the logical movement itself wasn’t wrong, we were just coming from a false beginning.

**Induction**

Induction, on the other hand, uses gathered evidence or existing understandings, which generally come in the form of observations, data, or laws, to draw a conclusion that seems likely (Hawthorne, 2004). For instance, suppose that we know that “all life on Earth requires water.” From this, we might conclude that in order to have life, other planets must have water also, and we embark on discovering whether or not other planets have water as a precursor to life. We might even go so far as to conclude that “water is essential for life.” At least in this current argument, we have not provided any evidence as to why or how water is essential to life, but we are operating on the very strong evidence that the millions of species of life on Earth all require water to survive, which is a massive, very compelling dataset. Based on this, it seems reasonable to conclude that we would not find any life where there is an absence of water and would, therefore, train our search for life only on planets that have the possibility of water.

However, what if we did find life on another planet, and it did not have water? Would this invalidate our premise the same way it did in deduction? Of course not. The premise would still be true — all life on Earth would still require water — it is just our conclusion that was false. The difference here is that the inductive argument was making a calculated leap that seemed reasonable based on existing evidence (millions of species supporting the conclusion without a single contradicting case), but based on how the argument was structured, there could always be the opportunity for a counter-instance to arise where the conclusion did not necessarily follow from the premises.

Another characteristic of induction is that it operates off of likelihood rather than from an underlying rationale. We might, for instance, find that women are more likely to suffer from a specific disease than men at a ratio of 50-to-1 and inductively use this evidence to suggest that women should be screened for the particular disease even if we don’t understand why it might be more prevalent in women. If we then screened a particular woman and found that she did not have the disease, this new evidence alone wouldn’t invalidate our premise nor our conclusion that women should be screened for the disease. All that it would show was that our argument was inductive and, therefore, might not be accurate in every instance.

**Abduction**

By clarifying this difference between deduction and induction, we can now understand why abduction is also necessary. Abduction is a special case of induction where we not only draw a conclusion that seems likely based on the evidence but also attempt to explain or justify the conclusion by reference...
to some explanatory mechanism (Douven, 2017). For instance, suppose that we know that “the Earth is teeming with life and is in a Goldilocks zone (not too hot, not too cold).” When asked “Is there life on Neptune?” we might reasonably conclude “no” and proceed to explain that “based on what we know about life, no organisms could evolve or survive outside of a Goldilocks zone.” In this case, it seems reasonable to inductively conclude that there is no life on Neptune from the simple fact that it is too dissimilar to the Earth in its location, but this argument goes one step further by imputing a causal mechanism to explain why the Goldilocks zone is essential to life.

This step may involve a certain level of creativity or imagination to provide the explanation and is likely why Albert Einstein famously quipped that “Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution.” Whenever scientists move beyond what is known to try to construct theories that explain the “why” or “how” of what is happening, they are involved in abduction, and this requires an imaginative leap to make connections that have not previously been made.

Could it still be possible that life exists on Neptune? Perhaps, but the argument for why we shouldn’t expect to find life on Neptune is still strong because it inductively relies on current evidence (in the form of cases of life and non-life) and also abductively provides a rational explanation for why we would expect Neptune to conform to the non-life cases we know about. If we did, however, miraculously find that life, in fact, existed on Neptune, it would merely mean that our explanatory mechanism was flawed and, again, would not invalidate our original premises (facts) we were reasoning from.

**Connections to Knowing**

As thinking, experiential beings, we use all three forms of reasoning daily, and researchers also rely upon each when doing their work. We use deduction to categorize the world into pre-existing conceptual chunks with absolute certainty. We use induction to predict or infer a conclusion that seems reasonable based on existing evidence without understanding the underlying mechanisms. And we also use abduction to try to explain those inferences in accordance with logical, aesthetic, or otherwise meaningful theories or narratives we have about the world.

Recognizing these different forms of reasoning is paramount for responsible practice because doing so allows us to see the affordances, limits, and dangers of our reasoning approaches. If we are using deduction to apply our predetermined theoretical categories to the world with certainty, then we need to recognize that doing so opens us up to the possibility that counter-instances or contradictions must necessarily call into question or altogether disprove our assumptions. If we are using induction, then we need to recognize that any conclusions we draw will always only ever be our best guesses based on provided evidence and be open to the possibility that we might be wrong in certain cases. And if we are using abduction, then we must also recognize that there might always be other reasonable explanations for phenomena and relationships we are observing that might be equally justifiable given the evidence we are operating from.

In any case, recognizing each of these limits to our reasoning approaches should engender some level of intellectual humility on the part of knowledge-seekers and should help us recognize that reason is not an infallible master to be served but is rather a tool that is applied in various, messy, contextualized ways. In particular, this can make claims of objectivity dubious and always open to refutation and also help us to recognize where subjectivity may be influencing our conclusions.
Learning Check

The primary difference between deduction and induction is that:

a. Deduction moves from a law to particulars, while induction does the opposite.
b. Induction moves from a law to particulars, while deduction does the opposite.
c. Deduction yields a certain conclusion, while induction only yields a likely conclusion.
d. Induction yields a certain conclusion, while deduction only yields a likely conclusion.

The primary difference between induction and abduction is that:

a. Induction is more certain.
b. Abduction is more certain.
c. Induction provides an explanation for the mechanics of why the conclusion is likely.
d. Abduction provides an explanation for the mechanics of why the conclusion is likely.

Scientific Terms for Truth-Seeking

Pilate saith unto him, What is truth? (The Holy Bible, John 18:38)

So far, we have only discussed the nature and processes of knowing and have wholly ignored the all-important question of whether what we claim to know is actually true. Like Pilate, many cynics or skeptics have essentially thrown up their hands in surrender when confronted with the difficulty of ascertaining truth, and others have intentionally used the difficulty of nailing down truth to surreptitiously short-circuit any truth-seeking endeavor.

Yet, at the end of the day, each of us does seem to care about what is true and what is false, and we spend much time and energy trying to distinguish between the two. Knowing alone isn’t enough, because we can certainly know things that are false, so we rely upon mechanisms of truth-seeking and a variety of notions related to truth to help us (hopefully) arrive at it in reliable ways.

True is an adjective that we use to characterize certain propositions or statements, and though people might differ on what they believe to be true or even what truth is, there seem to be some commonalities in how we use these words and what we expect them to mean. For instance, we wouldn’t call a proposition true if it didn’t reflect our observations or experiences with reality (e.g., “the newspaper says that the weather is very hot, but that can’t be true, because it is snowing”). Likewise, we wouldn’t call a proposition true if it changed every time we tried to test it (e.g., “this plant was a daisy yesterday and a rose today”) or if it was impossible to understand (e.g., “a garfump is a schnerf”). The terms truth and true, then, might mean many things, but they seem to at least suggest some essential characteristics of accuracy, consistency, understandability, and exactness.

In various scientific enterprises, we use many terms that we expect to have some level of truth to them or that we hope will help us to arrive at truth in some form or fashion, including facts, laws, theories, and hypotheses. These might also be considered types of knowledge or singular steps toward knowledge. In this section, I will differentiate what these terms actually mean and how they are used in scientific endeavors to help us to better understand how they relate to truth.
**Facts**

Facts are finite, specific, and observable things, and facts are considered true if they can be universally and consistently observed or measured. The word "fact" carries with it a connotation of truth. Thus, a statement like "it was 103 degrees Fahrenheit in Phoenix on March 20" would be a purported statement of fact that we could either prove or disprove by consulting records of temperatures or observing a reliable thermometer on the day in question. If proven untrue (or if different observations of the phenomenon conflict with one another), a statement ceases to be a fact. Thus, facts are a subset of truth.

Because facts are finite, specific, and contextual, however, they can be misconstrued or misinterpreted as evidence for things that may not be true. In the example above, one might use the single fact that Phoenix was 103 degrees to claim that the weather there is obscenely hot in March, ignoring other instances that place the average temperature at around 77 degrees. By looking at single facts or facts out of context, we can actually draw conclusions about the world that are clearly false, which means that though facts themselves may be readily discernible as true or false, how each fact in isolation relates to truths that are more universal or that actually matter may be difficult to discern. This means that facts alone have limited truth-telling power and that to be useful they often have to be considered in aggregate or as connected with other types of knowledge, such as laws or theories.

**Laws**

Somewhat similar, laws are finite, specific (often mathematical) descriptions of phenomena or their relationships to one another that are typically used for prediction. For instance, Newton’s first law of motion is a finite, specific description of the general relationship between an object and motion. The word “law” also carries with it a connotation of truth but represents a trickier case than “fact,” because laws may not be universal, may not be observable in every case, or may rely upon particular conditions to be true that are not explicitly stated in the law. For instance, Newton’s laws of motion do not apply to subatomic particles, and various scientific laws (e.g., Boyle’s law) require certain variables to remain constant (e.g., number of molecules, temperature). Thus, finding a counter-instance to a law doesn’t necessarily make it false but merely means that it is only contextually or conditionally true.

Additionally, one of the major reasons for having laws is that they allow us to make predictions about the world (e.g., “if I push this, it will move”), but if those predictions are not accurate, then the truth of the law is called into question. This means that laws can change or be refined over time, based on emerging evidence (via induction), and that truth claims about scientific laws are, therefore, conditional and contextual. It also means that laws are limited in explanatory power, because they merely describe and predict the observable aspects of phenomena (e.g., their motion) without helping us to understand the “how” or “why” of what is observed (e.g., why gravity works).

**Hypotheses**

Because we have to be able to evaluate the truth or falsity of purported facts and laws, observation and hypothesis testing are the methods of choice among scientists, constituting the backbone of the scientific method. Hypotheses are proposals that are neither true nor false but that can be tested for falsity. Suppose we want to know if a new reading curriculum influences student standardized test scores. We would start with two hypotheses: (1) the null hypothesis, which proposes that the
curriculum does not influence test scores, and (2) the alternative hypothesis, which proposes that the curriculum does influence test scores. Null hypotheses are always proposed to reflect received wisdom or current understandings of a topic. In the case of educational interventions, this generally means that the null hypothesis is that the intervention has no effect. Additionally, an alternative hypothesis is also proposed stating the opposite. With these hypotheses, we can then collect evidence or conduct experiments to try to disprove each. As we do this, our assumption is that we should keep the null hypothesis unless there is sufficient evidence to reject it. If we are able to do this, we reject the null hypothesis and accept the alternative, but if we are not able to do this with a reasonable amount of certainty, then we defer to maintaining the null hypothesis.

One important thing to realize about hypotheses is that they are never proven (or shown to be true) but can only be disproven (or shown to be false), which means that we end up accepting hypotheses not because we have proven them to be completely true, but because we have proven other hypotheses to be more false. In other words, our goal with any hypothesis testing is to try to disprove the null hypothesis. If we can disprove it, then we can tentatively accept the alternative hypothesis in its place (recognizing that others might attempt to disprove our alternative hypothesis in the future), but if we cannot disprove the null hypothesis, then we continue accepting it (recognizing that it could nonetheless be disproven with future evidence). In this way, hypotheses and hypothesis testing can only yield truth claims that are tentative (e.g., “we haven’t found any contradictory evidence yet”) about phenomena that are testable, and alternative hypotheses can always be proposed in an attempt to disprove the null hypothesis.

**Theories**

Theories are quite different from facts and hypotheses but are somewhat similar to laws in that they explain relationships between phenomena. The major difference between a law and a theory is that a theory is more expansive than a law and attempts to describe the “how” and “why” of some observed phenomenon or relationship. An apt, non-scientific synonym for theory would be *story, narrative,* or *model* because theories attempt to piece together disparate facts, experiences, reason, and musings in an understandable way. Like laws, theories are also conditional (i.e., a better explanation might arise based on new evidence) and contextual (e.g., evolution by natural selection may not apply in unnatural situations, as in the selective breeding of animals or humans hunting species to extinction), but unlike laws, theories also attempt to use logic and reference to other facts and laws to reasonably explain the processes by which things happen (via abduction). This means that theories, facts, and laws all have the same tentative, limited, and contextual relationship to truth but that theories also have a larger scope and goal of explaining rather than just predicting (as with laws) or verifying accuracy (as with facts).

This means that theories have a more complicated relationship to truth than laws or facts might. Because of their complexity and expansiveness, theories will always have counter-examples or contradictions that they cannot predict. This has led many philosophers of science to argue that theories are never true or false and can never even be proven or disproven, making them different from hypotheses. Rather, theories are adopted or rejected based upon their generative potential (e.g., Lakatos, cf. Musgrave, 2016), their relative advantage to other theories (e.g., Kuhn, 1996), or for a variety of irrational or unscientific factors, such as aesthetics or elegance (e.g., Feyerabend, 1975). We will delve into this issue more deeply in a subsequent chapter, but here it is mainly important to recognize (a) that the term *theory* does not suggest truth or falsity, (b) that a theory is inherently different from facts, laws, and hypotheses, and (c) that there is no movement between these terms (e.g., a hypothesis never becomes a theory, and a theory never becomes a law; a theory is always ever
exclusively a theory).

**Learning Check**

The primary difference between a fact and a theory is that:

a. Facts are more true.
b. Theories are more true.
c. Facts are broader and more explanatory.
d. Theories are broader and more explanatory.

The goal of hypothesis testing is:

a. To test if the null hypothesis is false.
b. To test if the alternative hypothesis is true.
c. To test which hypothesis is more true.
d. To develop a new theory.

Which of the following terms imply truth (i.e., if a purported _____ is proven wrong, it ceases to be a _____)?

a. Fact
b. Law
c. Hypothesis
d. Theory

**Truth and Error**

Half of a truth is often a great lie. (Benjamin Franklin)

Given these complexities with knowing, knowledge, and truth-seeking, it is not surprising that we all struggle with discerning, proving, and communicating truth. Historically, humans have addressed these problems in a variety of different ways, exemplified in different fields of inquiry, attitudes, faiths, ideologies, and professions.

As an example, the scientific method assumes that objectivity is possible and utilizes inductive and abductive processes to engage in hypothesis testing about the world, using observable facts to falsify or refine existing theories and laws. In contrast, though mathematics might also assume objectivity, it approaches knowledge-seeking in a much more deductive, non-empirical manner through the application of logic and theorems. Other areas of inquiry, however, might reject objectivist aspirations altogether and focus instead on subjectivist or intersubjectivist knowledge by cataloging facts as they relate to or are interpreted by humans.

In any case, these problems of knowing and truth are of paramount importance when moving forward with research because any method we employ will be influenced by our fundamental epistemological beliefs. If I observe a student in a class, then I am assuming that I can objectively understand what they are doing, while if I interview a student, then I am subjectively assuming that their perspectival
self-disclosure of their own behaviors, beliefs, and attitudes actually matters and has at least some truthful quality to it.

Each of us must also face the continuous conundrum both in life and in our research of determining when to believe a conclusion and when not to believe it. To help in this regard, scientists approach truth by focusing on errors and error likelihood, categorizing errors into two different types: Type I errors (or false positives) and Type II errors (or false negatives).

Type I or false positive errors occur when we incorrectly reject a null hypothesis that is actually true. Perhaps we concluded that a new game was helping students succeed at math, but in actuality, the students we tested simply had a natural propensity toward math. Or perhaps we concluded that a different group of students who were performing poorly on a history test didn’t understand history well, but in actuality, the students simply were struggling with the language used on the test. With Type I errors, we reject a null hypothesis (and accept an alternative hypothesis) when we really shouldn’t have.

Type II or false negative errors, on the other hand, occur when we fail to reject a null hypothesis even when we should have. Suppose I have an intervention that will help students read more quickly, but when I study its effects, I don’t include a large enough number of students in my sample to determine significance. I, therefore, might conclude that the intervention had no effect when it actually did. Or, let’s say that I’m studying the effects of school lunch on student performance but only study students from middle-class families. This would likely mean that I wouldn’t see negative effects from starvation on students in my sample and might therefore erroneously conclude that school lunch programs have no effect on student achievement.
Both types of errors prevent us from arriving at the truth, and different strategies are used to avoid each. To avoid Type I errors, a researcher might try to be very clear and specific in what they are studying, identifying all factors that might influence results, attempting to control for them, and only rejecting the null hypothesis if they had a preponderance of evidence compelling them to do so. To avoid Type II errors, they might target specific groups or affected contexts, increase their sample size, or refine instruments. Sometimes, these efforts can be conflicting. After all, if I only care about avoiding Type I errors, then I just need to create the most rigid process imaginable for determining truth (approaching every problem as a cool skeptic), but if I do that, then I might be rejecting many things as false that actually are true merely because my mechanism of discernment was too blunt or rigid.

Though the terms Type I and Type II error are normally only applied to scientific hypothesis testing, this state of affairs should be instructive to us in all forms of inquiry, because we can all make mistakes by either believing things that aren't true or failing to believe things that are true. We could attempt to radically address this by either believing everything, whether or not evidence is provided (and thereby avoid Type II errors), or by establishing unreasonable mechanisms for determining truth, essentially disbelieving everything (and thereby avoid Type I errors). But it seems like no matter what we are trying to learn or how we're trying to learn it, a balanced approach makes sense, and extremes should be avoided. Blind faith and unreasonable skepticism are both extremes, and being a responsible researcher means that we try to untangle truth from error but that we also allow truth to take its hold on us. Almost two thousand years ago Sextus Empiricus, one of the fathers of skepticism, explained clearly that skepticism is not capable of leading to knowledge or truth, only tranquility in the mind of the skeptic (Morison, 2019). Even the staunchest skeptic, upon entering a dark room for the first time, has enough faith in an untested light switch to flip it.

As we move forward in this book, we will continue to reference these matters more deeply within the context of individual paradigms of education research, as each paradigm approaches knowledge, truth, errors, and truth-seeking differently. Some will take an objectivist stance, while others will be more subjectivist. Some will err on the side of avoiding Type I errors, while others will err on the side of avoiding Type II errors. Through it all, it seems that every education researcher is in the business of discerning truth from error, but how we go about doing that may vary in important ways.

References


"Check out this study!" Your colleague excitedly smiles as she drops a dog-eared journal on your desk.

"This conclusively shows that learning styles don't exist! Finally, we can stop talking about them!"

Interested, you thumb to the marked article and begin reading.

The study is rigorous. Operating off of the proposed definitions of learning styles, the researchers had students self-report on their preferred styles (visual, aural, kinesthetic). Then, they assigned each student homework materials that taught the lesson's content in their preferred style. They then tested students on their comprehension via a standardized test and compared results to a control group.

Analyzing the results, the researchers found no performance differences between groups, meaning that though students believed they learned in a particular way, how they were presented the information actually didn't matter. The researchers concluded that though we might think we learn in a particular style, teaching to a preferred style actually has no impact on learning.

You close the journal and look back up at your colleague. What is your reaction?

Do you agree? "Yes, learning styles are definitely bunk." Maybe you jump onto Twitter and blast it to your likeminded colleagues. "Finally! Learning styles can die!"
Or do you disagree? "This doesn't completely disprove learning styles. I still think we do learn in different ways." Maybe you dig into the article further, nitpicking the researchers' methods and trying to find holes in their reasoning. "Maybe the homework didn't actually align with the students' styles. Or maybe people have difficulty determining their own styles. Or maybe the researchers have a conflict of interest in disproving learning styles." You grasp for whatever rationale you can. Learning styles are true, and it doesn't matter what this study says.

Whether you agreed or disagreed with your colleague in this fictional scenario, the most important question to ask is "why?" Why did you agree or disagree? Did you make your decision about learning styles based on your own experiences, on research evidence, on your own biases (e.g., aesthetic preference), or on what the common belief of your community happened to be?

**Key Terms**

**Abduction**
A logical inference of a conclusion (or special case of inductive reasoning) that likely follows (but doesn't necessarily follow) from the provided premises and provides a reasonable explanation of the mechanism by which the conclusion follows.

**Confirmation Bias**
A tendency to interpret new evidences as supporting or confirming existing theories and paradigms (often ignoring contradictory evidence or negative cases).

**Consensus Fallacy**
A logical fallacy in which it is assumed that something is true because a majority of people believe it.

**Degenerative Science**
A research programme which either does not make novel predictions or which makes novel predictions that are systematically proven wrong (cf., Lakatos).

**Schema**
A mental organizing structure that people use to construct meaning and to make sense of information.

**Scientific Progressivism**
The belief that applying the scientific method has historically yielded positive, fairly-consistent improvement in human understanding and knowledge.

**Theory**
An explanation of the 'how' or 'why' of particular phenomena, typically referencing existing facts and laws.

I use this example because it's always polarizing among education professionals, and if you believe in learning styles, then you are fairly unlikely to change your stance even after being presented with compelling contradictory evidence (Newton & Miah, 2017). Education researchers generally treat learning styles as a myth, because studies like the one described above have failed to support the theory over and over again (e.g., Husmann & O'Loughlin, 2019; Kraemer, Rosenberg, & Thompson-Schill, 2009; Macdonald, Germine, Anderson, Christodoulou, & McGrath, 2017; Willingham, Hughes, & Dobolyi, 2015), and the strongest scientific evidence on the subject concludes that "there is no adequate evidence base to justify incorporating learning-styles assessments into general educational practice" (Pashler, McDaniel, Rohrer, & Bjork, 2008, p. 105) and that learning styles instruments "should not be used in education" (Coffield, Moseley, Hall, & Ecclestone, 2004, p. 118). Practitioners,
on the other hand, such as teachers, generally believe in learning styles and even develop lessons and entire courses around serving the needs of specific learners as defined by their preferred style. And as a go-between, teacher educators represent more of a mixed bag: some siding with the research and others siding with the received wisdom of classroom teachers.

This type of scenario unfolds daily and is not limited to learning styles. It also is not merely a theory-practice divide, because many fields have competing theories that researchers gravitate toward, bolster, and defend, much like political ideologies and parties. This happens, because we all struggle to know sometimes (a) when to believe a theory, (b) how far to apply a theory, and (c) how to deal with evidence that conflicts with our accepted theories.

I previously introduced the concept of a theory as a narrative, story, or model that researchers use expansively to explain what they observe and experience. Theories are necessary because they allow us to make sense of a messy, irrational, chaotic world, and as such, they serve as a kind of interpreter for our brains or lens for seeing the world in understandable ways, signaling to us what to pay attention to and what to ignore, what to emphasize and what to deemphasize.

Our brains need ways for organizing information, to make meaning, and to guide judgments. Consider a painting. A digital camera can copy a painting into a two-dimensional series of colored pixels, but being able to represent the painting digitally is not the same as being able to interpret it (to identify what is in the painting). Similarly, our bodies and brains can experience phenomena, but making any kind of sense of these phenomena requires something to compare it to or some organizing structure (or schema) to place it in. Today an artificial intelligence can take a colored pixel series created by a digital camera and identify faces because it has massive databases of images to compare to that a human has told them represent faces. In this way, the artificial intelligence is using a schema that allows it to make meaning of what it is seeing by comparing it to what it has seen before and using categories or methods of analysis provided to it. Without a schema, the digital representation of the painting would only ever be a series of colors. With a schema of what a face is, though, humans and artificial intelligence alike can begin to see faces in the colors, making meaning of the meaningless.

Theories provide these schemas. They give us meaningful structures and outlines to attach ideas, observations, and experiences to. Without them, we would just have decontextualized and meaningless bits of information, such as an infinite number of pixel colors, but with them, we can organize information in meaningful ways. This understanding is similar to constructivist learning theories, which pushed back against behaviorist and tabula rasa views of learning to recognize that any new learning will be dependent upon what we have already learned, influenced by our histories, beliefs, and attitudes toward what we are learning.

In addition, schemas and theories are not static but are expected to adapt, evolve, and grow to accommodate new information and new types of informational relationships to better and more fully understand the world. Science generally operates on this assumption of scientific progressivism, wherein we believe that through observation we can come to progressively know more and more about the world around us, and by applying the scientific method, we will gradually make our theories and schemas grow and expand, ostensibly until they encompass all truth.

As practitioners apply the scientific method, theories can be tested to see how well they perform, and through subsequent testing, theories can evolve to be more accurate or useful. Thus, scientists typically “view theories as provisional” and as progressing through failure, wherein “the best theory
we have now" is expected to "fail in some way and that a superior theory will eventually be proposed" (Willingham, 2012, p. 86). In this view, scientific progress is steady and linear, with subsequent discoveries building upon previous discoveries, ever-moving upward and onward to more accurate understandings of the world. "Good science is cumulative; ... [and] science is always supposed to move forward" (Willingham, 2012, p. 95).

**Theory Building and Revolutionary Crisis Cycles**

However, the actual history of theory building in science, and the concomitant adoption and rejection of theories, paints a messier picture than scientific progressivism generally acknowledges.

For starters, new, important, earth-shattering theories are often created by accident or as bolt-out-of-the-blue moments of inspirational abduction that come to scientists when they are grappling with scientific problems often in mundane settings. The story goes that Archimedes invented his principle of volume while he was taking a bath and noticed the displacement of water his body made, to which he shouted "Eureka!" About 1,900 years later, Newton shouted the same thing when an apple fell on his head, and he had the initial idea for the theory of gravity. In 1669, Hennig Brand accidentally discovered the first chemical element phosphorus while collecting bathtubs full of urine (for the purpose of creating gold), thereby paving the way for the discrediting of transmutation alchemy (and his own research program). And Einstein's theories of relativity evolved from thought experiments he had about how light would bend and be perceived if rooms and flashlights were accelerated.

In all of these cases, laws and theories developed out of abductive reasoning, imagination, and chance rather than pure empirical observation or rational proofs. In more recent years, attempts have been made to make theory-building more systematic (such as Grounded Theory, cf. Glaser & Strauss, 2017), but the history of science largely is not a history of one observation rationally stacking onto another. Rather, it represents changes in understanding that came in unexpected, irrational leaps and jarring revolutions rather than calculated, rational steps.

Recognizing this, Thomas Kuhn (1996) famously argued that the history of science is best understood as a series of scientific revolutions between dominant paradigms or theories. (I will explore paradigms more deeply in subsequent chapters, but for simplicity, I will equate paradigm with theory here.) In progressivist views, science increases knowledge incrementally over time (cf. Fig. 1), but in the revolutions view, knowledge may build slightly over time as scientists operate within current theories, but such progress is slow and necessarily comes to an end (once the theory reaches its explanatory limit). At that point, revolutions in theory are required to move knowledge-building forward.

Under this view, most of the day-to-day work of scientists operates firmly under the assumptions of preexisting, dominant theories, which Kuhn refers to as "normal science." But, such work always has a ceiling that it cannot pass, because the dominant theory will prevent scientists from learning anything beyond what the theory already allows for. At this point, for progress to continue, a radically different theory must be proposed.
Though Kuhn has often been mischaracterized as being a relativist, his explanation has a clear structure and pattern, which scientists have historically followed, and this applies to education researchers as well. Essentially, researchers will persist in using existing theories until a crisis to the theory arises. If the crisis is serious enough, then they will look for competing theories (or create new ones) only if the alternative is better able to deal with the crisis than the current theory. Such "paradigm shifts" to new theories are painful but necessary for scientific progress to continue.

To illustrate how the crisis cycle works, Figure 2 walks you through the questions that a researcher would ask while using a theory in their day-to-day activities of normal science. All goes well until an anomaly arises. In this case, the researcher must consider whether the anomaly is explainable by the current theory; if so, then she should stay with the current theory. If not, she should consider whether the anomaly is trivial or important; if it’s trivial, then she should stay with the current theory. If it’s important, then a crisis has arisen, and the researcher must decide whether the crisis has to be dealt with now or whether it would be okay to postpone dealing with it later (perhaps when better tools are available). If it must be dealt with now, then she must consider whether the original theory could be modified to address the anomaly; if so, then she should remain. If not, then she should consider whether any other theories might exist that better explain the anomaly. If so, then she should make a paradigm shift to the alternative theory. If not, then she should get in the business of theory generation and decide if she can abductively provide a better theory that accounts for the anomaly (and everything else that the current theory accounted for) in a better manner. If so, then she should shift to her new theory, but if not, then she should stay with the current theory.

By going through this process, researchers will generally adopt a conservative approach to
maintaining the status quo of current theories and will only shift to new theories if there is a clear, present, and vital need to do so (prioritizing the avoidance of Type I errors to Type II errors). But this also means that new ideas and theories are (at least in theory) met with skepticism unless there is a clear need for them. After all, why propose a new theory of learning, identity, curriculum development, or social interaction if our current theories are working fine? And why reject our current theories if there are no better alternatives to move to?

Though the use of theory is necessary for doing research, this state of affairs in how we adopt and reject theories leads to a number of problems that education researchers need to be aware of, including the consensus fallacy, confirmation bias, degenerative science, the Barnum effect, and irrationality. After addressing each of these problems, I'll then close the chapter by providing some guidance on how education researchers should responsibly adopt or reject theory and highlight what some key components of good theory might be.

**Learning Check**

The revolution view of scientific progress proposes which of the following:

- a. There is no such thing as scientific progress
- b. Science progresses over time
- c. Science progresses through steady, incremental improvements
- d. Science progresses through paradigm shifts

According to Kuhn, which of the following must happen before a theory is rejected?

- a. An anomaly (or conflicting evidence) to the theory must arise
- b. The theory must enter a state of crisis
- c. The scientific community must reach a consensus regarding the theory’s ability to account for emerging evidence
- d. A better theory must be proposed

**Consensus Fallacy**

A consensus fallacy simply means that people can mistakenly believe that a conclusion is true based on its popularity or that given two options we should go with the option that is more popular. The problem with the consensus fallacy is obvious: the truth of any claim has nothing to do with its popularity, and major advancements in science historically happen when alternatives rise up in opposition to popular theories (such as Copernican heliocentrism challenging geocentrism).

Popularity itself is not evidence, but though this truism is obviously reasonable in its finite applications to logic, in the messy, complex real world of research we do often rely upon consensus for moving forward. For instance, many arguments used in popular media supporting the theory of anthropogenic climate change use premises like "97% of climate scientists agree that human carbon production is causing glacial melting." Or, in the case of learning styles, "most teachers believe in learning styles, so they must exist." The structure of such an argument is that if most experts or practitioners agree on something, then it must be true. However, we wouldn't make the same argument in other settings, such as "65% of Americans believe in the supernatural, so the supernatural must exist."
One rationale for this discrepancy is that in highly technical fields popularity amongst experts is used as a proxy for evidence when non-experts might struggle to know and weigh the evidence themselves. Not everyone has access to all the data points and not everyone has the training necessary to interpret the data in reasonable ways, which means that we may at times feel justified in relying on consensus among experts to guide us when the field is too esoteric (e.g., not everyone sees what goes on in a teacher’s classroom) or complicated (e.g., not everyone can weigh historical temperature data points across the globe).

But, this still isn’t a completely satisfying answer, because we wouldn’t think that every argument made on expert consensus would be valid. “Most theologians believe in God, so God must exist,” is as ridiculous as “Most mathematicians believe that trigonometry is essential for leading a fulfilled life, so all kids should be taught it.” Another similarly absurd argument would be “Most evolutionary biologists do not believe in God, so God must not exist.” If someone made these arguments to me, I’d probably respond with something like this: “Of course theologians believe in God, and mathematicians believe in the importance of math! That’s their jobs! That’s what they’re paid to do!” That is, even in the case of expert opinion, we don’t always believe that consensus is an appropriate proxy for evidence, especially when the expert’s position or theoretical assumptions conflict with their ability to give unbiased guidance on a particular topic or when experts overstep the limits of their expertise (such as evolutionary biologists making metaphysical claims or mathematicians making claims about social needs and wellbeing, cf. Feyerabend, 1975).

Thus, if we are relying upon consensus in research, we should at least ensure (1) that the experts we rely upon are qualified to draw the conclusions they are drawing (and are not overstepping their bounds), (2) that the theoretical assumptions, biases, and conflicts of interest they are subject to in their positions are not negatively influencing their ability to draw reasonable conclusions, and (3) that we only treat consensus as a shorthand proxy for evidence and not as evidence itself (meaning that counter-evidence should be carefully considered).

In practice, researchers rely upon the expertise of others to make complex projects possible. If, for instance, I am trying to make a smartphone app to teach kids how to read basic words, I should of necessity rely upon the expertise of early childhood literacy colleagues to inform its pedagogical design. However, I would not rely on their expertise to supplant my own in matters that draw upon my own realm of expertise, such as privacy, usability, database architecture, and so forth.

This all means that in our own fields of expertise, consensus should never be a determining factor of whether to accept or reject a theory, because (as experts) we should be aware of and be able to weigh evidence regarding the theories before us, and we should never treat any theory as being beyond critique merely because our colleagues believe it. Yet, we should also exercise some level of intellectual humility with regard to the limits of our own expertise and recognize that soliciting and relying upon consensus among experts in other fields may be a practical necessity.
Learning Check

Which of the following would be examples of the consensus fallacy?

a. 98% of climate scientists agree that anthropogenic climate change is occurring, so global warming is real
b. Most people believe in God, so God must exist
c. Microbiologists study single-cell organisms and agree that organisms act for their own genetic survival, so all of human action, morality, and decision-making is the result of self-interest and genetic survival
d. Most teachers believe in learning styles, so learning styles must be an accurate explanation of how people's brains work

If we are going to commit the consensus fallacy to accept a conclusion (without fully weighing and understanding the evidence ourselves), what should we ensure?

a. That the conclusion agrees with our biases
b. That the group making the consensus has legitimate expertise
c. That the conclusion does not overstep the limits of the group’s expertise
d. That the conclusion is not merely an expression of a fundamental bias of the group

Confirmation Bias

Confirmation bias means that if we already believe a theory, then we will tend to collect and interpret evidence that agrees with the theory while ignoring evidence that contradicts it. Kuhn (1996) explained that scientists fall into this trap with theory when they "take the applications [of the theory] to be the evidence for the theory, the reasons why it ought to be believed" (p. 80). They do not recognize that what they observe will be dictated by what they are expecting to observe. This prevents the emergence of "alternative interpretations" or the discussion of "problems for which scientists have failed to produce theories," meaning that whole swathes of phenomena are ignored or pigeonholed in inappropriate ways (p. 81).

This has been a major and ongoing problem in theoretical psychology and has reverberations in the education literature, as experimental results that did not conform to behaviorist models were typically ignored, or tests were modified until results conformed with them (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; MacKay, 1988). In response, both psychology and education produced competing models (e.g., cognitivism, constructivism, social constructivism) to address important phenomena being ignored or inappropriately explained by behaviorism (e.g., higher-order reasoning, individual differences). This problem of confirmation bias inherent in operating from any dominant theory suggests that practitioners will only see in the world that which their theories allow them to see and also thereby suppress new ideas as being "necessarily subversive of [the accepted theory's] basic commitments" (Kuhn, 1996, p. 5).

To illustrate, let's circle back to learning styles. If I’m a teacher that believes in learning styles, I will begin to classify my students (at least in my mind) according to these styles and will use observations in my classroom to solidify both my classifications and the theory itself. If I notice that Juan does better at a test that has illustrations, I might think "hmm ... Juan must be a visual learner." If I then give a lecture, and Juan falls asleep, I might think "hmm ... Juan is definitely a visual learner because
he was just not drawn into the aural modality of the lecture." If I show a video, and he falls asleep again, then I’ll definitely be convinced: "Yes, Juan is a visual learner. All the evidence proves it!"

If a colleague then shows me an article debunking learning styles, I’ll protest the findings on the basis that I know that learning styles exist, because Juan is an obvious example of a student who has a clear learning style. Never mind the fact that these observations could have been explained by any number of other theories and facts, like that Juan’s language proficiency is struggling, that he had less sleep than his classmates the night before, or that I’m simply a boring lecturer. I have seen a learning style manifested in Juan because that is what I was looking for. The theory has, therefore, manipulated my observations and confirmed itself through that very manipulation.

This phenomenon is not unique to learning styles or to education. We all generally tend to find examples of whatever it is we’re looking for. If you believe in God, you’ll see evidence for God. If you believe in phrenology, you’ll see evidence of how people who have wronged you seem to have similarly-shaped heads. If you believe in astrology, you’ll witness how your zodiac sign affects your life (and ignore how it doesn’t). If you believe the world should be just, you’ll see evidence of injustice. If you believe the world is inequitable, you’ll see evidence of inequity. In this way, theories are powerful lenses that let us find whatever it is we are looking for, just as a microscope allows us to find a bacterium or a telescope allows us to find a star.

It’s not bad that theories allow us to see things; that’s what they’re supposed to do. What is problematic, though, is when we rely upon what the theory is showing us as the guiding evidence for why we should believe the theory. Theories that are self-confirming have a circular logic that doesn’t make them amenable to outside verification or critique.

In response to this problem, Popper (1959) and other critico-rationalist philosophers of science argued that the scientific enterprise must critically evaluate its relationship to the world and the limits of empirical observation alone. Turning science on its head, Popper argued that the testing of a theory can only occur after the theory has been accepted and that such testing must rely upon logical falsification rather than evidence-building.

Unlike non-scientific methods of inquiry, Popper explained that scientific theories must be amenable to the rules of logic and that logical tests may be applied to specific statements implied from a theory to "falsify the whole system" (p. 56). In this way, singular anomalies contradicting theories could be used to test a theory's strength and either disprove the theory or limit its universalizability. From this perspective, to be scientific, a theory must reveal testability (i.e., provide predictions that could be proven false), and the less a theory provides testability, the less it should be trusted as a good scientific theory.

This process of falsification, Popper argued, is a more sure way of determining the accuracy and value of a theory, because any theory can explain the world when it has the power to shape what we see through it. Good theories, though, make a prediction that we can test to see if the theory is false. Theories that don’t make predictions, then, are non-scientific. Theories that make predictions that turn out to be true represent good science for the time being but may be proven false in a future prediction. But theories that make predictions that turn out to be false are rejected as untrue.

Thus, evidence in favor of the theory doesn’t matter in determining its scientific value. All that matters is whether the theory contradicts itself by predicting things that are proven false, and if it makes no predictions that could be proven false, then it is a non-scientific theory.
In the case of learning styles, we can make very clear predictions of what should happen if learning styles exist, which allows the theory to pass Popper's falsifiability criterion, making it a scientific theory. We could, for instance, expect that if learning styles exist, then teaching students to their preferred styles should produce learning improvements. If this prediction doesn't pan out in rigorous studies, though, then, based on the requirement of falsification, we would have a strong justification to reject the theory, because it has been proven inaccurate no matter how much confirmatory evidence we might have collected along the way.

Learning Check

According to Popper, what is the central characteristic of "good" scientific theories?

- They can be proven true
- They can be proven false
- They can neither be proven true nor false
- They are found to be useful

Degenerative Science

The main critique for Popper's falsification criterion is that it sets "unrealistic standards for sound science" in many practical settings (Mercer, 2016, p. 1) and that he was making an argument for how science should be done in a purist form rather than how it is done in practice. Addressing the issue of falsification directly, Kuhn (1996) explained that "if any and every failure to fit [a falsification] were ground for theory rejection, all theories ought to be rejected at all times" (p. 146). Similarly, Mulkay and Gilbert (1981) explained that “Negative results ... may incline a scientist to abandon a hypothesis, but they will never require him to abandon it, on Popper's own admission. ... Thus the utter simplicity and clarity of Popper's logical point are lost as soon as he begins to take cognizance of some of the complexities of scientific practice and as soon as he makes the transition from his ideal scientific actor to real scientists engaged in research” (p. 391).

In other words, Popper idealistically gave a single negative result the power to topple an entire theory that might have mountains of confirmatory results, but in the real world, scientists balance falsifications and confirmations in a complex way to determine how long to responsibly stay with a theory that is starting to show some negative results.

Following a similar strand of reasoning, Lakatos (1974) argued that asking whether a theory is true or false is the wrong question. Rather, theories, like lenses, are tools for seeing novel things clearly and therefore can only either be progressive (empowering us to learn and see more) or degenerative (not empowering us to learn and see more). Falsification still plays a role in this classification but is seen more as a holistic or continual check on the overall health of a theory.

To explain, one example Lakatos gave was a comparison of two prominent theories from divergent fields: Marxism and Newtonian physics. As a once young Communist leader himself, Lakatos considered himself a disciple of Marx and zealously directed many Communist efforts in Hungary until he was forced to flee the country under Soviet threat. He took refuge in the United Kingdom, becoming a Cambridge professor and a prominent philosopher of science and mathematics. Revisiting Marxism in his later life, Lakatos (1974) explained:
Has … Marxism ever predicted a stunning novel fact successfully? Never! It has some famous unsuccessful predictions. It predicted the absolute impoverishment of the working class. It predicted that the first socialist revolution would take place in the industrially most developed society. It predicted that socialist societies would be free of revolutions. It predicted that there will be no conflict of interests between socialist countries. Thus the early predictions of Marxism were bold and stunning but they failed. Marxists explained all their failures: they explained the rising living standards of the working class by devising a theory of imperialism; they even explained why the first socialist revolution occurred in industrially backward Russia. They “explained” Berlin 1953, Budapest 1956, Prague 1968. They "explained" the Russian-Chinese conflict. But their auxiliary hypotheses were all cooked up after the event to protect Marxian theory from the facts. The Newtonian programme led to novel facts; the Marxian lagged behind the facts and has been running fast to catch up with them.

Lakatos considered the Marxism of his day to be degenerative science, because it consistently predicted novel facts that did not come about, and rather than rejecting the theory, adherents would devise ad hoc explanations for why the predictions failed, essentially propping up the theory with makeshift addenda.

Newtonian physics, on the other hand, made novel factual predictions that could not have been arrived at without the aid of the theory. And they were consistently confirmatory, making it a progressive theory for scientific work.

To apply this same principle to education research, it may be too purist of us to follow Popper’s falsification criterion too strictly, discounting a theory simply because it provided one false prediction. However, as theories are applied over time a pattern of falsification can emerge that should be troubling for researchers. For instance, if studies emerge showing that predictions made by learning styles are false, then we might explain away one or two false predictions on the basis of misapplication or extraneous factors (“auxiliary hypotheses” according to Lakatos), but if predictions continue to fail, then we should question the progressive value of adhering to a theory that is obviously degenerative in its predictive capabilities. This is especially true when alternative theories might exist that can more accurately and consistently explain results, such as multimodality, self-determination, and a host of others in the case of learning styles.

**Learning Check**

According to Lakatos, what is the central characteristic of degenerative scientific theories?

- a. They make a true prediction
- b. They make many true predictions
- c. They make a false prediction
- d. They make many false predictions

**Either/Or Fallacy**

Building off of this last point, just as the world is complex and messy, there might be a variety of theories that we could employ to explain anything that we see in it, meaning that there might actually
be many proposed theories that can explain any given phenomenon. For this reason, by focusing too heavily on a single theory and ignoring multiple alternatives, researchers can also quickly fall into an either/or (or false dichotomy) fallacy, where they assume that there are only two possible theoretical explanations for a given phenomenon.

This may sometimes be done as a persuasive tactic that education researchers use to bolster support for their own theoretical stances rather than an actual scientific error. In Kuhn's scenarios, there are always two competing theories — the dominant theory and the challenger — and the dominant theory will carry the day unless the challenger can show a capacity to outperform it in predicting and explaining the world. In education research, though, researchers are often hard-pressed to articulate what the dominant theory they are trying to respond to may be.

For this reason, you will rarely see critiques of formal theory in education research but will rather see veiled references to "the status quo," "one-size-fits-all," "lecture-based," "pencil-and-paper," or "industrial model," or researchers will use a nondescript catch-all term like "traditional" to serve as a strawman counterpoint to their proposition. This is often done to carve out a place for new theoretical approaches without seriously considering what the state of the field actually is, thereby ignoring theoretical plurality and alternative explanations.

To refer back to learning styles, if you showed me falsifying study results of the theory, I might respond incredulously with an either/or retort like "So, you think kids are all the same?!" Or, "You think we should just use one-size-fits-all curricula?!" Or, I might ignore the results and heroically proclaim that "I don't care what the so-called studies say; I'm going to differentiate for my students' diverse learning needs anyway." Yet, in doing so, I'm not showing a commitment to differentiation but merely a commitment to one, limited theory of differentiation, and my strategy for bringing up the alternatives I did would be to claim that any approach that ignores learning styles ignores the needs of students.

However, learning styles is not the only theory that empowers us to view students as unique individuals with unique needs. Given three diverse students — Juan (a first-generation, Christian, undocumented, English language learner from Guatemala), Suzy (a white, atheist, middle-class introvert), and Sofu (a physically active extrovert and child of Christian refugees from Nigeria) — it would be absurd to assume that these three students have the same learning needs or that they will respond identically to stories you read or videos you watch in class. But, it is also equally absurd to assume that the biggest differences between these students will be that one prefers pictures (visual), one prefers words (aural), and one prefers physical movement (kinesthetic). That is, learning styles theory is not flawed because it requires us to differentiate for student needs; rather, it is flawed, because it does not require us to differentiate enough in ways that matter for our students. It ignores socioeconomic, historical, language, and a variety of other factors that will influence a student's ability to learn.

Yet, whenever theory is brought up either in research or in more general contexts, specific theories are typically only compared against vague, Bogeyman-like alternatives. Perhaps this happens at times because we do not know what the alternatives actually are, but at other times it seems that such phantasmic comparisons are made for simple persuasive effect.
Barnum Effect

Another important and related problem with theory adoption borrows from a common psychological phenomenon known as the Barnum Effect. Named after the famous showman P. T. Barnum, the effect explains a technique charlatans, such as magicians, mediums, and crystal ball gazers, use to convince people that they have special powers and insightful knowledge about strangers' lives. They do this, in part, by purportedly telling strangers "secrets" that only they should know by actually making broad, generic statements about human nature and experience that could relate to anyone.

For instance, a psychic medium might tell a client that she "longs for freedom," "doesn't like to feel suffocated,” and "is focused on success," and the client might believe that the medium has special powers of insight because she keenly feels these statements to be true about herself. She would likely leave her session feeling energized and convinced that "yes, I am a woman who longs for freedom," not realizing that each of these statements is true for just about anyone and that she is being duped into building unjustifiable trust in the medium. In such situations, victims passionately believe the charlatan, because the statements are universally true, generic, and obvious, and because they align with the victims' biases and assumptions about the world. And once they believe that the charlatan has a gift of divination, they will gradually become less and less critical of subsequent statements.

Theories in education research can be structured in this same fortune-cookie-like way. Broad theories can be proposed that are generic and are impossible to refute, not because they are strong and accurate but because they are fluid, generic, and untestable. In other words, vagueness leads to a false perception of accuracy, and once we believe in the generalist theory, we begin to uncritically accept any addenda taught by adherents of the theory as inspired and also obviously true.

Learning styles is an interesting case of this, because individual scholars who propose the theory can be quite descriptive and precise in their formulations (making connections between specific styles and activity in particular areas of the brain), but among companies that provide commercial products to schools for testing and addressing learning styles, articulation tends to be broad and unfalsifiable (e.g., "kids are different and have different needs"). Once generalist theories are believed, though, all sorts of addenda can be appended without critical review, such as the adding of learning style upon learning style until people begin uncritically accepting that some students are inherently "naturalistic," some are "philosophical," some are "intrapersonal," ad infinitum, without any reasonable evidence other than that they have passionately bought into the base, generalist theory.

Proponents of many generalist theories can fall into this same trap, such as feminism, critical race theory, and Marxism, where adherents might believe in the obvious generalist theory (e.g., “that the sexes should be socially, economically, and politically equal”) and then might also uncritically believe more specific applications, manifestations, or addenda to the theory without reasonable confirmatory evidence or legitimate consideration of contradictions (e.g., “that girls’ outperformance of boys' in reading and writing must be a result of patriarchal oppression” or “that the prevalence of plant-based allergies is the result of human misogyny against female trees”).

In contrast, Kuhn explained that theories must be limited in "both scope and precision" in order to be useful (Kuhn, 1996, p. 23). Applying this to education, Burkhardt and Schoenfeld (2003) argued that "most of the theories that have been applied to education are quite broad. They lack what might be called 'engineering power'... [or] the specificity that helps to guide design, to take good ideas and make sure that they work in practice." (p. 10) For these reasons, "education lags far behind [other fields] in the range and reliability of its theories," because we have overestimated the strength of our
theories and allowed them to have too broad of a scope (p. 10). This leads to a precarious situation wherein we might uncritically believe any eventuality or formulation of a generalist theory simply because the base theory is so obviously true.

**Learning Check**

To avoid the Barnum Effect, theories in education research should be:

- a. Broad and generalist
- b. Focused and specific
- c. Useful and progressive
- d. Accurate and universal

**Irrationality**

And finally, we may also be influenced to adopt or reject a theory for purely irrational reasons, such as intuition, hunches, opportunism, and aesthetics. Feyerabend (1975) argued that rather than being an aberration, this is actually the historical norm of science, wherein revolutionary new theories have regularly been adopted and proliferated for no other reason than that they appealed to “valuable weaknesses of human thinking” (p. 126), requiring researchers to utilize "propaganda, emotion, *ad hoc* hypotheses, and appeal to prejudices of all kinds" to develop them and encourage their adoption (p. 119).

To illustrate this point, Feyerabend carefully chronicled the attempts of Galileo to convince the religio-scientific community of his day that the Copernican heliocentric theory of the universe was true and that geocentric theories were false. Based on the scientific evidence available at the time, there was not a legitimate scientific reason to reject geocentric theories, especially because existing geocentric theories (like the Tychonic system) were more accurate, complete, and non-contradictory than the Copernican theory (e.g., the case of stellar parallax). The main benefit that Copernicanism had over the Tychonic system at the time was that it was simpler and more elegant, which seemed to outweigh the need for evidence and accuracy in Galileo’s mind. Stellar parallax wasn’t detected for more than 200 years after Galileo’s death, and without that key piece of evidence, Copernicanism was less empirically accurate than existing geocentric alternatives. Yet, today, we celebrate Galileo’s commitment to Copernicanism not because it was the scientifically reasonable position to hold at the time, but because he stuck to simplicity and elegance (and ended up being right) even when the evidence at the time pointed the other way.

Some philosophers of science have built upon this point and gone on to question the potential truth-value of theories altogether, making irrationality in theory adoption not only a historical reality but also a necessity. In the realm of psychology, Greenwald, Pratkanis, Leippe, and Baumgardner (1986) explained that “no theory can be proven true by empirical data” and also that it is “impossible to prove one false” (p. 226). This suggests that the quality of a theory may be based primarily on its value to humans via irrational factors, rather than its accuracy or inherent truth value.

In the case of learning styles, this explains why the theory has proliferated despite contradicting evidence and inconsistencies. It appeals to Feyerabend’s “valuable weaknesses” or “noble prejudices” of humans by being intuitive and simple, which adopters seem to value more than research evidence.
Similar situations exist with various theories. A powerful example of this for me occurred once while attending an academic conference at Oxford University. I sat in a small audience while the presenter of a research study passionately shared his message and then, preempting any critiques or questions, flatly explained: “I know that you may argue with my theoretical stance, but I’ve lived it, I’ve experienced it, so it doesn’t matter what you think.” And he was right. I couldn't argue with him, not because his rationale was strong, but because he systematically rejected rationality as an expectation for theory adoption.

Though such irrational approaches to theory seem to be anti-scientific (or, at best, non-scientific), they are nonetheless prevalent in education research. A first-generation college graduate might adopt critical race theory to study a K-12 school because it resonates with his experiences as a minoritized child, not because it is the most effective way for addressing the problem he is seeking to solve. An educational technology researcher might adopt a particular technology integration model because it is depicted with simple, intuitive, and colorful graphics, not because it is logically or pragmatically superior to alternatives. Or a behavioral scientist might adopt a granting agency's preferred theory of student behavior to improve chances for funding, not because it is accurate, helpful, or meaningful for understanding student behavior.

Clearly, as humans, education researchers are influenced and motivated by many irrational factors that will lead them to adopt some theories (like those that appeal to their [noble and ignoble] prejudices) while rejecting others (like those presented with a variety of typos), but as education researchers, we should consider to what extent we should allow irrationality to guide such decision making and to what extent we should strive to be more critical and rational.

**Good Theory**

Given all of these difficulties of theory adoption, education researchers might understandably lament "What are we to do?" Building off of Kuhn (1996), Rogers (2003), and others, I have previously outlined some considerations that may be helpful for education researchers broadly in identifying "good" theory (Kimmons, Graham, & West, 2020; Kimmons & Hall, 2016). In our current treatment, the most important of these seem to be the following: (1) clarity, (2) compatibility, (3) fruitfulness, (4) scope, and (5) ability to disprove/reject.

First, good education research theories should be clear, meaning that they are "simple and easy to understand conceptually and in practice" (Kimmons & Hall, 2016), thereby "eschewing explanations and constructs that invite confusion, misinterpretation, and 'hidden complexity'" (Kimmons, Graham, & West, 2020). The world is messy, but theories only allow us to make sense of it and to progress as a field if they themselves are clear and rigid. Clear theories can be tested against emerging evidence, reevaluated, and revised or rejected as needed. They allow for falsifiability and resist degenerative science and non-science by giving us concrete expectations.

Unclear, messy, esoteric, and amorphous theories, on the other hand, cannot be nailed down to specifics (often either in their form or in eventualities), meaning that they cannot be tested for accuracy. A theory that can be adapted to explain anything can be confirmed by anything and contradicted by nothing. Such procrustean obfuscation can be appealing to researchers who want their theories to survive no matter what the evidence actually says, but it is ultimately detrimental to the field because it allows theories to persist far beyond their shelf-life and prevents us ever from moving past the empirical ceiling necessarily established by the theory.
Second, a good theory is compatible with good practice (Kimmons & Hall, 2016). The whole purpose of having a theory is to inform practice, but not all practice is based on theory. This means that we may have a variety of good practices in education and research that are not based on theory, but if we are ever considering a theory to inform our practice, we should at least begin by making sure that the new theory doesn’t make our practice any worse. As Kuhn explained, theories are never adopted in a vacuum, but they are rather intentional shifts responding to emergent needs. Education theories come in all shapes and sizes; some are focused on design, others on understanding. Yet, in all cases, theories should always improve rather than merely replace existing ways of working. This means that there is always a comparison aspect of theory adoption in which we must ensure that by making any shift we are keeping any good practices and are only continually making our craft better.

Third, Kuhn (1977) proposed that good scientific theories should be fruitful, meaning that they should “produce new findings and discoveries,” thereby helping researchers “to reach solutions to research puzzles” and to solve problems (Ivani, 2019, p. 4). As stated earlier, even a theory that is not true (or cannot be proven true) may still be useful — Newton’s laws are not always true, but they are extremely useful. One way that this happens is by helping us to make predictions that come true. Another way is by helping us to make progress in other fields or areas of inquiry that are not central to the theory. Without such fruitfulness, we run the risk of falling into Lakatos’s depiction of degenerative science, where all of our work focuses on merely propping up the theory rather than learning from it.

Fourth, theories should be scoped properly to be useful. Broad, sprawling theories lack the engineered precision necessary to actually inform action, and they persist not because they are strong, by having a solid, finite backbone, but because they are weak, attempting to amorphously ooze into vast realms of human experience so that we can see them everywhere without actually learning anything.

And finally, adopting a theory should never be a terminal decision. That is, we should always operate on the assumption that any theory might be wrong or that there might come a time when it is no longer the most useful way of approaching our problems. We should be willing to reject any theory, given sufficient evidence and alternative ways of approaching our problems, rather than holding onto it with cult-like fanaticism. Theories are supposed to be our tools, not our masters. They serve us; we do not serve them. A fanatic might point to their unwillingness to reject a broken or unhelpful theory as a sign of heroic conviction, but in actuality, such an approach only stymies our collective progress toward addressing education problems that we all want to solve.

References


Psychological Science in the Public Interest, 9(3), 105-119.


1.4

Ethics

How should ethics and morality guide our work as education researchers?

Royce Kimmons

Suppose you are a veteran science teacher. After years of development, you've finally done it. You have created the best high school science education curriculum ever devised!

The curriculum takes three years to complete. You are now doing it with all of the students in your school and have found that the students have been performing much better on standardized tests than their peers did in previous years, and many more of them are being accepted to elite colleges.

Wanting to get funding to implement the program in other schools across the state, you approach a program officer for a major research organization. She suggests that your results are very promising but that in order to be competitive for their large grants, you will need to prove more reliably that the curriculum is having the effect that you are claiming and that it is not due to chance or some other factor.

"Can you try it out with some other schools," she asks?

"No," you reply, "not without more funding."

"Well," she responds, "it seems then that you'll need to do some form of a controlled trial at your current school, where you randomly assign half of the students to use the experimental curriculum and the other half to use the old curriculum. Then, you should be able to prove with greater certainty that it's the curriculum having the effect, and we can talk more optimistically about funding opportunities."

"But," you begin to carefully explain, "the curriculum is an entire package. It takes three years to complete. That means if we don't do it with half the students, they probably won't perform as well on standardized tests and won't get into good colleges at the same rate as their peers. I don't think I can do that to them."
She furrows her brow and nods understandably. "Yes, that would be a difficult choice. But if you want to really help more kids and make this thing big, then you're going to need to be a bit more scientific in how you do this. Some kids might fall through the cracks those first few years, but just think of all the kids you'll be able to help for years to come once you have the evidence you need!"

**Key Terms**

**Consequentialism**
An approach to ethics that holds that the morality of an action should be determined by its effects.

**Contractarianism**
A non-normative approach to ethics that holds that what is held to be right and good is merely determined by social contracts that are shared between people.

**Deontology**
An approach to ethics that holds that the morality of an action should be determined by its duty-bound adherence to particular laws or norms of behavior.

**Ethics**
The branch of knowledge that deals with rightness, goodness, and morality.

**Moral Relativism**
A non-normative approach to ethics that holds that what is right and good is only ever determined by references to individual or cultural norms or contexts.

**Normative Ethics**
Approaches to ethics that assume some level of universalizability of moral action (across cultures or contexts).

**Utilitarianism**
The consequentialist stance that moral behavior consists of doing what will have the greatest effect, typically in terms of doing the greatest good for the greatest number of people.

**Virtue Ethics**
An approach to ethics that holds that the morality of an action should be determined by its relationship to the moral agent's development or expression of fundamental virtues.

Now, pause. Ask yourself honestly: If you had to make this choice and there were no other options, what would you do? Would you jeopardize the futures of a few dozen students in order to potentially improve the lives of thousands? Or would you continue the curriculum with the few hundred knowing that doing so will likely prevent you from ever being able to have any impact beyond the walls of your school?

Though fictional, this story represents one of the many very real ethical dilemmas that education researchers must face. These dilemmas are shaped by competing values, competing beliefs about what is good, and competing needs of the self, other individuals, and society.

In this scenario, would it be good to continue providing the curriculum to all of your students? Of course. Would it also be good to do the research necessary to get funding to provide it to more students? Also, yes. But, if you cannot do both, then which good is greater, preeminent, or better? (I could further complicate the issue and point out that it also probably wouldn't be bad for you to be able to monetarily benefit from selling your curriculum at scale either, but we'll ignore that for now.)
Ethics, or the study of rightness, goodness, and morality, exists to help us determine which actions are right and wrong, under what circumstances, and why. Like everyone, researchers must grapple with the ethics of their decisions as individual people, but they must also grapple with the ethics of their decisions as researchers in a field where there are many opportunities to do good or harm, to act rightly or to act wrongly. In this chapter, I’ll provide an overview of the three dominant approaches to normative ethics, which will serve as a starting point for analyzing the morality of particular research behaviors and the enactment of research agendas through differing paradigms.

**Relativistic vs. Normative Ethics**

In the example above, we might argue that a science teacher’s withholding curriculum from a smaller group of students would be immoral, because intentionally withholding valuable educational experiences from any student is, as a rule, never acceptable. Or we could argue in the opposite direction that the potential benefit of helping more students by performing the experiment justifies any negative outcomes that the smaller group might experience. Or the teacher might claim that withholding the curriculum would constitute a betrayal of the teacher’s relationship to the students and their role as a teacher, while another might argue that the same action might be a betrayal of the teacher’s role as a researcher.

This is an example of a moral dilemma or a situation when the purported morality of an action is called into question and represents a conflict between opposing values or requirements (McConnell, 2018). In this case, the opposing requirements faced by the teacher would be requirements to a smaller group of current students vs. a larger group of potential students.

One might attempt to avoid the dilemma altogether by arguing for a stance of moral relativism or contractarianism, explaining that the moral choice for this teacher would be determined by their cultural context, individual beliefs, time, and place. Or in other words, there is no moral dilemma here, because whatever the teacher chose to do would simply be a matter of preference, cultural norms, habituation, etc. Such relativistic approaches to morality are differentiated from normative approaches to morality, which seek to universalize moral requirements in some manner.

Relativistic approaches to ethics are common in popular culture today and are often seen as “harbingers of tolerance, open-mindedness, and anti-authoritarianism” (Baghramian, 2015). At some level, non-normative approaches to morality are appealing, because they allow us to iconoclastically discount taboos and mundane behaviors that might be considered immoral in some cultures but not in others (such as wearing a hat indoors or extending one’s left hand to a stranger for a handshake). Yet, relativistic approaches to ethics are ultimately untenable for researchers, because the same relativistic argument that could be used to claim that there is nothing morally wrong with refusing to refer to an elder by their appropriate title could also be used to justify direct harm to individuals or to claim that Nazi war crimes committed in the name of research were merely reasonable expressions of the time and context.
History is replete with examples of researchers engaging in unethical and harmful activities:

- In the 1940s, Nazi scientists conducted thousands of experiments on prisoners in concentration camps, in which researchers intentionally harmed, tortured, and killed victims in various ways, including poisoning, blunt head trauma with hammers, mustard gas, freezing, jaundice, sterilization, burning, electroshock, body part transplantation, and many other methods.
- From 1932 to 1972, hundreds of economically disadvantaged African American men with syphilis in the Tuskegee Syphilis Study were provided with meals and other incentives so that researchers could study what syphilis did to their bodies. They were never offered penicillin treatment, even though it was accepted as the standard treatment by 1947, and they were not accurately told why the researchers were studying them.
- In the 1940s, inmates at Stateville Penitentiary were recruited for participation in a study on malaria and were paid $100. They were then infected with malaria via mosquito bites, assigned to treatment and control groups, and administered a variety of potential treatments, many of which were highly toxic. Some participants were intentionally not treated until their temperatures exceeded 108-degrees. They experienced jaundice and anemia as well as "severe headaches and eye pain, nausea, weakness, vertigo, vomiting, diarrhea, and Herpes simplex, which was very common and often severe" (Miller, 2013). At least one patient died.
- In 1939, researchers in the Monster Study experimented on war orphans to better understand the development and treatment of stuttering problems. As part of the study, they intentionally influenced many children who spoke fluently to develop stutters, which impacted their ongoing
ability to communicate and to be successful in their schoolwork.

- From 1963 to 1973, 67 male inmates in Washington and Oregon were offered $25 per session (along with suggestions of parole) to allow researchers to irradiate and biopsy their testicles. This allowed researchers to better understand how radiation might affect astronauts.

Such researcher actions are not merely relativistically immoral but are normatively reprehensible. We should not look at these examples and merely excuse them by saying that they "simply represent a different time" or culture. Rather, we need ethical measures and ways of reasoning that allow us to differentiate between moral and immoral behaviors across contexts in ways that relativistic morality cannot allow. Toward this end, ethicists approach moral dilemmas in a variety of normative ways, and as researchers, we must also be willing to approach ethics normatively to ensure that the morality of our work transcends our current time and culture.

Three Approaches to Normative Ethics

When faced with ethical dilemmas, we often find ourselves weighing multiple values, rights, requirements, or goods against one another, and determining which should be prioritized may not be clear. Ethicists propose that there are at least three aspects of the dilemma that we can focus on to guide our thinking: the action itself, the consequences of the action, or the nature and intentions of the person performing the action. By giving one of these aspects greater credence than the other two, we can create normative guidelines for what constitutes ethical behavior. Such guidelines will have far-reaching effects as they influence our moral reasoning across multiple situations, such as guiding us to always prioritize consequences over intentions.

By focusing more attention on one of these three aspects of a situation, we will subscribe to one of three common approaches to normative ethics: deontology, consequentialism, or virtue ethics. Ethicists continually disagree on which of these is the best approach, because each has its own benefits and disturbing consequences, but researchers will occasionally use each to make ethical claims about their work, and it is, therefore, our responsibility to better understand them and decide when they should be utilized.

Deontology

Deontology focuses on actions themselves and proposes that morality consists of adhering to universal laws of good action (Alexander, 2016). An example of this might be a parent teaching a child not to lie. "Even if it means you might get into trouble," a parent might say, "it is always better to tell the truth." By teaching this to a child, it is anticipated that the child will develop an intrinsic, duty-based motivation to tell the truth. If the child acts out of duty, then threats of punishment or promises of reward will not matter, as the child will tell the truth simply because it is "always the right thing to do."

A professional example of this would be the Hippocratic Oath (or variations that are still administered in some form to many graduating medical students in the U.S.) and the related injunction to "first, do no harm." Such oaths expect medical professionals to hold certain practices related to patient wellbeing as sacred and inviolate, and the morality of an action stipulated in the oath is interpreted solely upon rigid conformity to it in a generally black-and-white, legalistic manner.

Laws, policies, constitutions, and declarations of rights operate similarly. For instance, the United
Nation's Universal Declaration of Human Rights holds that "Everyone has the right to life, liberty and security of person." And based upon this moral dictum, governments, regimes, and individuals are judged as either moral or immoral by whether or not they follow it. Following the dictum is considered moral, while deviating from the dictum (no matter the reason, purpose, or context) is considered immoral.

Following the research abuses perpetrated by Nazis in World War II and the subsequent Nuremberg Trials, the Nuremberg Code was created and adopted as an international code of ethics governing research experiments. Some of the points of the code were clearly deontological in nature, such as requiring voluntary consent from research subjects and allowing subjects to withdraw from studies at any point, the duty-based principle underlying such points being that researchers have a duty to honor the self-determination and agency of participants. In other words, honoring participant self-determination and agency is simply the right thing for researchers to do.

In education, this translates into basic moral expectations of the researcher, in which we only study participants who have provided their informed consent, do not take advantage of vulnerable populations (such as young children and prisoners), inform participants of the nature of our work, do not attempt to coerce participation from our participants, do not share participant data without their consent, and allow participants to withdraw at will. Generally speaking, these are universal codes of conduct that we subscribe to simply because rigid conformity is the moral thing to do.

However, though many deontological approaches to ethics are commonplace, there are several obvious problems that can arise from this approach. First, by relying upon universal maxims for the right action, deontological ethics can seem too rigid in all situations and contexts. For instance, though honoring participant self-determination is paramount, are researchers required to be held to it if participants are attempting to self-harm? Or might there be another moral requirement placed upon the researcher to interfere out of concern for the subject’s wellbeing? In the case of the child learning to tell the truth, might there be situations when telling the truth is not the morally correct thing to do, such as when truth-telling will result in imminent harm to another person?

Second, since deontology defines morality as rigid adherence to maxims in a universal fashion, much harm can arise if those maxims are themselves harmful. Abraham Lincoln famously offered a deontological argument for following bad laws as follows:

When I so pressingly urge a strict observance of all the laws, let me not be understood as saying there are no bad laws.... But I do mean to say, that, although bad laws, if they exist, should be repealed as soon as possible, still while they continue in force, for the sake of example, they should be religiously observed.

In this regard, Lincoln argued for deontological rule-following but also recognized that any law might be wrong and in need of revision or abolition. In this view, then, rule-keeping can be seen in some ways to be more important than the rules themselves, which means that under a deontological view, a moral rule-keeper might be causing much harm but still be considered moral.

In a similar way, the faithful rule-keeper may sometimes actually be acting against the intent of the law by obeying the "letter of the law" while violating the "spirit of the law." For instance, as an instructor, I might establish and enforce assignment deadlines to ensure that my students are progressing through my course at a reasonable pace so that they will be able to successfully complete...
it, but if I enforce this rule too rigidly and do not account for students with special needs or unexpected trauma, then I would be allowing adherence to the rule to stand in the way of the purpose for the rule (i.e., student success).

Today, researchers do operate under several key duty-based ethical standards. As suggested earlier, guidelines on how to treat subjects, especially in terms of consent, privacy, self-determination, and safety are generally treated as vital, but not all are consistently treated as inviolate in every circumstance.

**Learning Check**

Which of the following statements are examples of deontological approaches to ethics?

a. "Thou shalt not kill." — The Holy Bible
b. "The end may justify the means as long as there is something that justifies the end." — Leon Trotsky
c. "Act only on that maxim through which you can at the same time will that it should become a universal law." — Immanuel Kant
d. "A knight must not complain of his wounds, though his bowels be dropping out." — Don Quixote (Man of La Mancha)
e. "Do what is right, let the consequence follow." — Latter-day Saint Hymn

**Consequentialism**

Consequentialism focuses on the consequences of actions and proposes that the morality of an action is determined by the desirability or harm of its results (Sinnott-Armstrong, 2019). To connect back to the truth-telling example above, a parent employing consequentialist ethics to teach a child about honesty might teach them that "you should normally tell the truth, because telling the truth makes people happy." In this scenario, truth-telling is not considered to be good on its own but only as a vehicle for making people happy. If, then, a situation ever arises when telling the truth would lead to heartache or sadness, then, in those cases, telling the truth would no longer be the morally right thing to do. This places all determinations of morality upon the effects of the actions, intended or not, and not the actions themselves.

As perhaps consequentialism’s most well-known and most-discussed formulation, utilitarianism holds that when judging between two potential actions, the moral choice will be the one that maximizes benefit, either in terms of numbers of people benefitted or the qualitative nature of the benefit (Driver, 2014). In other words, the goal of utilitarianism is utility or human happiness, and all actions are morally evaluated based on their influences upon happiness. Because of its popularity and development, I'll focus on utilitarianism as the prime example of consequentialism for moving forward.

Professional, social, and personal instances of utilitarianism are widespread. People will regularly structure their activities in ways that maximize their own happiness or that of their social group and will often even ignore moral maxims that they might generally hold to be true if they are perceived to interfere with this pursuit.

As a simple example, the original formulation of the Hippocratic Oath required physicians to not ever "use the knife," ostensibly because cutting a patient in any form inflicts harm on the patient’s body.
This is undoubtedly true, and anyone who has ever undergone surgery that involves incisions might still have the scars to prove it, but is there ever a time when surgery by incision is justified? Certainly. Take the case of a child who is born with a congenital heart defect. Surgeons have developed procedures to correct such defects, but doing so often requires breaking the infant's sternum, stopping their heart, transfusing blood from a donor, and months of recuperation. Why would anyone allow such a gruesome and dangerous procedure to be performed on an infant? The simple answer is that the net benefit to the child, in terms of years-lived and improved quality of life, justifies actions that might otherwise be unconscionable. That is, in this case, surgeons and parents believe that the consequence (improved life and happiness of the infant) justifies what would otherwise be an immoral act (intentionally harming the infant).

In fact, many of us would argue that if a parent did not allow reasonably-successful surgeries to be performed on a terminally-ill child then that parent would be guilty of negligence toward the child — meaning that failure to harm the child through surgery would be immoral.

Such reasoning is not only applied to doing harm to an individual to promote their overall wellbeing but is also commonly used to justify doing harm to an individual or small group in order to benefit society more broadly. We ask individuals to temporarily harm themselves by donating blood in order to save the lives of others. The very notion of drug testing in pharmaceuticals is based on this premise, where relatively small groups of people agree to take drugs that may have negative health side-effects in order to inform treatment for the larger population.

Many other points of the Nuremberg Code operate from this ethical stance, wherein researchers must "yield fruitful results for the good of society" that would be "unprocurable by other methods" and the "degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved." This means that risks to individuals and violations of general rules of conduct might be appropriate in situations where doing so is absolutely essential and of broad benefit to society.

Similar utilitarian reasoning is employed by education researchers to justify their studies and to occasionally violate maxims that otherwise would be followed. As the introductory dilemma to this chapter suggested, beneficial educational interventions might sometimes be denied to some students in order to prove their legitimacy for larger populations. Even in a single classroom, normal activities might be temporarily disrupted to try a new curricular approach, because it promises to have a net benefit to students in the long-run. And in situations where we want to study students' behaviors, we might temporarily not let them know that we are studying them because knowing would change their behaviors.

In all of these cases, note that a form of the term "temporary" was used, meaning that though education researchers might sometimes justify harming participants through their work, there remains an expectation that doing so will both (a) provide net benefits to society in terms of research outcomes and also, whenever possible, (b) correct any harm done to the individual, providing net benefits to them as well. So, if a student is placed in a control group and is temporarily denied access to a beneficial experimental education intervention, then it is expected that upon completion of the study the student will be allowed to benefit from the intervention as well. Though such reparative approaches to harming students are not always possible, the general principle is that education researchers will do everything possible to minimize risk of harm to participants and also do everything possible to correct any harm that occurs because of the study.
Though utilitarianism is common, there are at least two major problems that adherents of this approach must address. First, consequences are not always clearly discernible before acting. The child telling the truth might not know the effect it will have on the listener, and the surgeon might be asked to perform a surgery with a low likelihood of success. In these cases, if the consequence is increased harm (sadness for the listener or death for the patient), then the action was immoral. Yet, the entire purpose of ethics is to inform people on how they should act before they act, meaning that if the morality of an action cannot be determined until after the consequences are felt, then how are we to act? All morality, then, potentially becomes a guessing game of whether our actions will have the effects we intended, and if they don’t, then we are immoral as a matter of chance.

The second, and even more serious, difficulty with utilitarianism is that it can be used to rationalize the violation of fundamental rights and causing extreme harm to individuals as long as such atrocities are done in the name of promoting the greater good. Almost every despot, dictator, and perpetrator of genocide in history has claimed to act for the greater good. The entire Nazi propaganda machine was based on such premises: that in order for Germany and the majority of the population to thrive, a small minority of the population would need to be deprived of rights. Nazis justified horrible experiments on political prisoners upon the rationale that such suffering would benefit the German majority via better science, better medicine, and better soldiers. Utilitarian arguments can always be made to marginalize, harm, or destroy individuals and minorities as long as doing so provides a net benefit to society (e.g., the majority). Even genocide can be justified as a utilitarian moral good.

Today, researchers continue to use utilitarian reasoning to guide the morality of their actions, but, as these two problems highlight, utilitarianism must be tempered with certain inalienable maxims of right and wrong in order to guide actionable practice and to prevent the rationalization of atrocities.

**Learning Check**

Which of the following statements are examples of consequentialist approaches to ethics?

- a. "It is logical. The needs of the many outweigh the needs of the few." — Spock (Star Trek)
- b. "Ask not what your country can do for you but what you can do for your country." — John F. Kennedy
- c. "The means we employ must be as pure as the ends we seek." — Martin Luther King, Jr.
- d. "Society's needs come before the individual's needs." — Adolf Hitler

**Virtue Ethics**

Virtue ethics focuses on the development and expression of virtues themselves as the fundamental indicator of moral action. Whereas a consequentialist might say that virtues are good insofar as their development leads to social benefits and deontologists might say that fulfilling one's duty to universal laws is the only guiding virtue, virtue ethicists "will resist the attempt to define virtues in terms of some other concept that is taken to be more fundamental," such as happiness or duty (Hursthouse, 2017).

Virtues are excellent traits or dispositions that are considered to be worthy of cultivation in the moral actor. In the case of the child being taught not to lie, a parent might explain "you should be an honest person, and honest people do not lie." Becoming an honest person, then, is the reason that the child should not lie, without reference to universal maxims or consequences of dishonesty. Developing the
virtue (in this case, the disposition of honesty) is the measure of morality. An honest person, in this way, will still generally tell the truth but will do so with consideration for complexities of others' feelings, the ramifications of the truth, an abhorrence of dishonesty, and a consideration of other important virtues (e.g., love and concern for the other). Whether or not a person is truly honest (or has developed the virtue of honesty) will be manifested in their overall behavior, not a single act, and will take into account their intentions and reasoning.

Though developing virtue is considered in this approach to be the fundamental goal of ethics, it is tempered with an understanding that practical wisdom is necessary in all things. Just as we might say that "so-and-so is honest to a fault," we can simultaneously value honesty but also recognize that its enactment requires tact, thoughtfulness, and balance. An honest child might behave differently than an honest adult not because of a different natural propensity to honesty but simply because the adult has learned to wisely enact honesty in any situation.

Though originally postulated by Aristotle over 2,000 years ago, virtue ethics remains popular and is heavily debated. In recent years, new approaches to ethical reasoning have been proposed that have marked similarities to virtue ethics, such as care ethics (Sander-Staudt, 2020) and feminist ethics (Norlock, 2019). In these approaches, love, caring, and relationship-building are considered essential virtues, to be understood and appreciated in non-reductive ways. In the education literature as well, being caring (Laine, Bauer, Johnson, Kroeger, Troup, & Meyer, 2010), student-centered, and humanizing (Salazar, Lowenstein, & Brill, 2010) are typically treated as essential virtues for teachers to cultivate. Importantly, such caring dispositions are not seen as "fixed personality traits" but, rather, "are commitments and habits of thought and action that grow as the teacher learns, acts, and reflects" (Diez & Murrell, 2010, p. 14), meaning that in education we expect professionals to develop specific virtues in relation to their practice.

It seems reasonable that this should be applied to education researchers as well and that caring for our students and our research participants is an essential virtue that should be cultivated by every teacher and researcher. However, the obvious difficulty with applying virtue ethics to professional decision-making is that it requires us to determine, define, and agree upon what those core virtues are. Thus, we might agree that the virtue of caring is essential to education research, or we might not. And even if we do agree that caring is essential, we might not agree on what this actually means or how it is wisely enacted in every case.

Early in my career, I was discussing a curricular decision with a senior colleague, wherein I was pushing him to consider a small group of students in his class whose needs were not being met. He interpreted the gist of my comments as an unspoken question “Don’t you care about your students?” He then responded curtly: “I don’t care about my students individually. I only care about them collectively.”

Though not prevalent, I have occasionally heard different variations on this same theme many times since then from diverse education professionals who interpret the virtue of caring as an abstract concept that is applied to groups in a disinterested, generalist manner rather than to individuals in a focused, concrete manner. I see this as a gross misunderstanding of the virtue of caring because caring for a group implies caring for the individuals within it. So, if you do not care for the individuals, then you can never care for the group. And just as I would not want my child to have a teacher that failed to cultivate the virtue of caring toward them, I would not want to participate in a study with a researcher who did not cultivate the virtue of caring toward me. Thus, even if we agree on virtues, what any virtue actually means in practice may be contested.
To further explore this issue, part of the reason that virtue-based decision-making requires wisdom is that virtues can lead to contradictory conclusions. A simple case arises from the virtues of justice and mercy. Most would agree that both justice and mercy are virtues, but whenever they are applied in specific cases they are almost always at odds with one another (Kidder, 2009). If a student plagiarized a paper but was later penitent, then a just instructor might fail them while a merciful instructor might give them another chance. Since the actions are contradictory in this situation, which virtue should the instructor enact, and does it depend on the context (e.g., the student's understanding of what plagiarism is and whether they were properly taught)? Again, in my professional practice I have grappled with colleagues in these sorts of cases, and it can be very difficult for our shared wisdom to see eye-to-eye.

Today, virtue ethics is applied in a variety of professional practices where key dispositions are identified as being paramount, even though specific enactment of those dispositions may be left uncodified to allow for wise application to each context or situation. According to the Nuremberg Code, researchers should exercise "careful judgment," "good faith," risk-avoidance, and humanitarianism, all of which seem to be dispositional or virtue-based in nature. In addition, some common dispositions expected of education researchers might include being caring, empathetic, just, equitable, committed, competent, respectful, and contemplative, just to name a few.

**Learning Check**

Which of the following statements are examples of virtue ethics?

a. "By doing just acts the just man is produced, by doing temperate acts, the temperate man; without acting well no one can become good." — Aristotle

b. "Be ye therefore perfect, even as your Father which is in heaven is perfect." — The Holy Bible

c. "Success is not final, failure is not fatal: it is the courage to continue that counts." — Winston Churchill

d. "Human happiness and moral duty are inseparably connected." — George Washington

**Confronting Ethical Dilemmas**

From this brief introduction to ethics, it is not surprising that there is so much disagreement about what constitutes moral action and that people often struggle to determine what they should do in a given circumstance. I will now close this chapter with some practical guidance on how to move forward as a researcher and circle back to the original dilemma presented at the beginning, providing a solution for whether the science teacher should perform the study or not.

The reason that some of the choices we face are dilemmas is that it is often difficult to parse out where we can compromise and where we cannot. For instance, if we believe in two conflicting maxims (such as "it's wrong to lie" and "it's wrong to hurt someone's feelings"), then how do we know which to follow?

Furthermore, the role of human agency and intentionality in ethics is even more difficult to understand with certainty, because sometimes we believe that an action is good even if it has bad results as long as the intent was honorable (e.g., Good Samaritan clauses) and at other times we believe that intentionality doesn't matter for determining the morality of an action at all (e.g., a man who sexually harasses or abuses a woman but feels there was nothing inappropriate about his
To help us through this, here's a procedure that employs all three approaches to ethics in a way that attempts to capitalize on their strengths and to defuse their weaknesses. It is specifically designed to help you determine whether or not (and how) to conduct a research study, and for your convenience, Table 1 can also be used as an organizing tool for documenting your reasoning as you work through the dilemma.

To solve the dilemma, you should first reflectively identify the core virtues that you're seeking after in your personal and professional life by filling in the blank of “I am (virtue).” This allows you to clarify the central tenets and guiding principles that shape your vision for who you are and who you want to be. We do this first so that we can ensure that these virtues are foremost in our thinking and to ensure that any course of action we take will not violate our central moral purpose. As you do this, you do not need to list absolutely every virtue that you strive for or think is good but only those that relate directly to the problem at hand. In the case of the science teacher, she might identify her core pertinent virtues as being caring (she cares about her students’ wellbeing), equitable (she doesn’t want to disadvantage one student to another), helpful (she wants to benefit as many students as possible), career-focused (she wants to improve her career), and competent (she wants to do good research to produce the results that the grantor is asking for). Place these core virtues in Column 1 of the table.

Second, you should identify any requirements that are placed upon you either by general maxims you have adopted (which may come directly from your core virtues), laws that you must abide by, or expectations that are placed upon you by an outsider: Do this by completing the statement “I must (requirement).” Again, you should focus here only on the requirements that pertain directly to the matter at hand and do not need to include truisms that might matter to you in a different circumstance. In this case, the science teacher recognizes that she is required by the grantor to conduct studies using control groups, and she also has a requirement emanating from her core virtues dictating that she will not disadvantage students in the long-term. Place these requirements in Column 2.

These requirements, along with the core virtues in Column 1, represent our uncompromisables, meaning that violating any of them would constitute either a moral violation or a technical impossibility in the given case. In this situation, there is nothing inherently moral about conducting studies using control groups, but it is, nonetheless, a requirement placed upon the teacher by the grantor and must, therefore, be deontologically observed. As uncompromisables, any conflicts that exist within these columns constitute the heart of the dilemma, and we must eventually find a way to move them out.

Once you have drafted your uncompromisables, then you are ready to engage in the utilitarian process of asking what the potential benefits of the proposed action would be. Do this by completing the statement “Wouldn’t it be great if (potential benefit),” and place your answers in Column 3. These are your goals or the best-case results of the proposed action.

The last two columns mirror Columns 1 and 2 but represent Secondary Virtues (instead of Core) and Guidelines (instead of Requirements). These take the same grammatical form as the Core Virtues and Requirements but are of lower priority and are therefore compromisable in comparison to the uncompromisables. Add any virtues or expectations that you considered for Columns 1 and 2 that you ended up not including because they weren’t absolutely core or necessary. In the science teacher's
case, she recognized that being helpful to the research community and benefiting more students through her work would both be good things to do but that they shouldn't be prioritized at the same level as her other entries.

**Table 1**

*Example table for solving the science teacher's dilemma*

<table>
<thead>
<tr>
<th>Core Virtues</th>
<th>Requirements</th>
<th>Potential Benefits</th>
<th>Secondary Virtues</th>
<th>Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;I am...&quot;</td>
<td>&quot;I must...&quot;</td>
<td>&quot;It would be great if...&quot;</td>
<td>&quot;It's generally good to be...&quot;</td>
<td>&quot;Whenever possible, I should...&quot;</td>
</tr>
<tr>
<td>Caring</td>
<td>Not disadvantage students in the long-term 1</td>
<td>We had more solid evidence that the curriculum was working</td>
<td>Helpful</td>
<td>Benefit more students through my work</td>
</tr>
<tr>
<td>Equitable</td>
<td>Gain informed consent/assent from participants</td>
<td>More students could be impacted by the curriculum</td>
<td>Committed</td>
<td>Not disadvantage students in the short-term</td>
</tr>
<tr>
<td></td>
<td>Allow participants to withdraw 2</td>
<td></td>
<td>Career-Focused</td>
<td>Advance my career</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Competent</td>
<td>Only conduct studies using long-term control groups 1 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Uncompromisables</th>
<th>Potentialities</th>
<th>Compromisables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

1 2 These items are in conflict and cannot remain together in the Requirements column.

As you engage in this process, feel free to move items back and forth between Columns 1 and 4 and between Columns 2 and 5. As you do this, you are engaging in a process of discerning what is essential to the morality of the issue in question and what is not. By seeing the requirements lain out like this, you can quickly see if any conflicts exist, and if they do, then you must resolve the conflict by moving one of the items to the Compromisables category.

In the case of the science teacher, she determines that some of her initial Uncompromisables are actually Compromisables. Being career-focused, for instance, and seeking to advance her career are treated as secondary when compared to the other items in Columns 1 and 2 and are therefore moved to Columns 4 and 5.

The central dilemma that eventually becomes clear to the teacher is a conflict between being competent (and only conducting studies using long-term control groups) and being caring and equitable (and not disadvantaging students in the long-term). After weighing these against each other, she concludes that being equitable and caring is more important than being competent and moves the latter to Column 4.

This effectively resolves the dilemma and makes clear that though it would be good for the teacher to...
do the study and follow the guidelines in Column 5, it would be wrong for her to do it in a way that violated any items in Column 2. She, therefore, concludes that she will not do the study unless the grantor changes the "long-term control group" requirement to a guideline.

The conclusion that the science teacher arrives at, then, is that choosing to conduct the study in its present form would have been immoral and that though it is unfortunate that she will not be able to conduct the study, she has nonetheless behaved morally by choosing to abstain.

The purpose of this chapter has been to lay out normative approaches to ethics and to introduce the problems of moral relativism when applied to research. In future chapters, I will explain specific moral requirements that are placed upon researchers, how ethical research considerations are arbitrated by institutional review boards, and how specific ethical problems should be approached across multiple research paradigms.

References


Paradigms

What are research paradigms, and why are they important in education research?

Imagine you are invited to work as part of a research team. You show up for your first day and discover, unbelievably, that you are joined by many famous historical figures. Whisked away from their own times, these thinkers sit down with you around a massive table, and you are tasked with working together to learn about the universe.

How do you imagine this meeting unfolding? Would Marie Curie and Carl Jung agree on fundamental assumptions about the nature of the universe? Would Galileo and Harriet Beecher Stowe agree on what should count as evidence? Would Malcolm X and Hannah Arendt agree on what the end goal should be of any work you are doing together? Would Mohandas Gandhi and Jeanne d'Arc agree on the means for achieving those goals? And ultimately would you be able to work together to do anything meaningful?
Key Terms

Axiology
The branch of knowledge that deals with value, mattering, and importance.

Epistemology
The branch of knowledge that deals with knowing, such as what is known, how it is known, and how it is proven.

Ethics
The branch of knowledge that deals with rightness, goodness, and morality.

Incommensurability
The assumption that differing paradigms or worldviews are not readily compatible with or reducible to one another.

Method
The specific procedure, action, or steps taken when doing research, such as distributing a survey, conducting an interview, or statistically analyzing numeric data.

Methodology
The technical guidelines followed when performing methods, such as how a survey instrument should be constructed, how an interview should be conducted, or how data should be validly analyzed.

Ontology
The branch of knowledge that deals with being and reality.

Paradigm
A model or pattern we follow when conducting research, including both our surface-level methods as well as our deeper, often hidden, beliefs, attitudes, expectations, and values.

Pluralism
The assumption that multiple views, approaches, or stances can be contextually legitimate.

Differing Worldviews

The optimist in me hopes that a room of such brilliant people could find sufficient consensus to work together toward some valuable goal, but the realist in me recognizes that the underlying worldviews of these people would make any type of joint endeavor difficult if not altogether impossible.

Why? Why might bright, capable people with common goals struggle to work together?

To answer this question, we need to recognize that the way people approach problems and even what they identify to be problems is based on certain fundamental beliefs, habits, and dispositions. It encompasses how they see themselves, how they see other people, and how they see the world.

For instance, when faced with a particular instance of human suffering, Mother Teresa might respond by rushing to the aid of the person in an attempt to alleviate their immediate needs. She would do this operating from a metaphysical belief that good and evil currently exist in the world as a result of divine design (or at least divine allowance) and also a related ethical belief that, as God’s children, other people should be loved and cared for and that it is her duty to do so.

In contrast, when faced with the same scenario, Karl Marx might respond by rushing to critique and correct the social and economic system in which the human suffering originated. He does this
operating from a belief that the welfare of groups of people has historically been dictated by the social and economic institutions surrounding them and that human suffering can only be overcome by correcting the systems that have allowed the wealthy and powerful to take advantage of the masses. His guiding ethical belief, then, is that all people should be equally empowered and liberated and that such liberation is a group or community affair. Marx himself famously explained this difference of perspective as follows: "When I give food to the poor, they call me a saint. When I ask why they are poor, they call me a communist."

Thus, even though Mother Teresa and Karl Marx might have similar aims to reduce human suffering, the assumptions that they make about suffering and about their roles in addressing it might be a source of friction. Karl Marx might view Mother Teresa's approaches as quaint or ultimately futile because she ignores larger social institutions and only works to address specific instances or symptoms of suffering rather than its root causes, which he believes reside in the inequitable distribution of wealth, capital, and production.

Mother Teresa, on the other hand, might equally view Karl Marx's approaches as dangerous or ultimately futile, because he ignores individuals and only works to address historical symptoms of a deeper problem, which she believes resides in the selfishness and pride of the individual human heart.

Such differences are not trivial and are not limited to the realm of political action. For a more academic spin, we might in this imagined scenario ask any of our historical cadre of researchers how we could learn about the nature of human suffering, and they would have very different answers.

- Karl Marx and Michel Foucault would likely suggest a critical, historical study of human economic and political institutions.
- Malcolm X would likely adapt this same approach in a manner that focused squarely upon racialized suffering perpetrated by white supremacy and European colonialism, while Simone de Beauvoir might similarly adapt this approach to focus squarely on the effects of gendered violence perpetrated by the patriarchal status quo.
- Sir Isaac Newton and Galileo would likely suggest conducting experiments or observations to test what physical stimuli influence responses of suffering (or its alleviation) in many subjects.
- Mother Teresa and Mohandas Gandhi would likely suggest a prayerful, meditative, self-reflective study of sacred scripture, coupled with selfless service to reduce suffering around them.
- Sigmund Freud and Carl Jung would likely suggest the psychoanalysis of people or groups.
- Friedrich Nietzsche and Bertrand Russell would likely suggest a rational grappling with what suffering is, whether its existence is a necessity, and what this means for the rational mind.

This list is an oversimplification, but it serves the point of illustrating that how these people saw the world and themselves in it would have direct impacts on how they would go about studying and interpreting it.

Even among thinkers in similar time periods and similar fields, this holds true, and even if we look squarely at the hard sciences, any historical examination of leading thinkers would fail “to reveal any gold standard research method in common, in virtue of which they all could be held to be doing science” (Phillips, 2006).

As education researchers, the situations we operate in may often feel similar to this imagined
scenario, where a wide diversity of people might come together for achieving a broad, common goal, such as improving education, understanding learners, or training teachers, only to discover that our beliefs, attitudes, and methods for achieving these aims might be very different and seemingly incompatible.

This is because in education research we are typically dealing with applied or surface-level issues, problems, and questions—questions of practice, innovation, improvement, and adaptation—rather than deeper issues of existence, meaning, knowing, valuing, and being. Yet, these surface education problems that we are trying to address are nonetheless heavily influenced by deeper issues that we are not, and how we respond to these deeper issues and are shaped by them is often hidden to others and even to ourselves.

**Deep Research Paradigms**

If we think of our approach to doing education research as an iceberg (Figure 1), then the visible methods we use (e.g., interviews, surveys, randomized controlled trials) may belie deep, fundamental assumptions we are making about the universe and our relationship to it, such as how our methods should technically be employed (methodology), how morality should guide those methods (ethics), why our research and the questions we ask matter (axiology), what we are able to learn by using these methods (epistemology), and how this whole endeavor connects to reality at a fundamental
level (ontology).

Our paradigm, or the model or pattern we follow when conducting research (cf. Feyerabend, 1975; Kuhn, 1996), includes both our surface-level methods as well as the deeper, often hidden, beliefs, attitudes, expectations, and values that guide us.

**Learning Check**

Which of the following would NOT be part of a research paradigm?

- a. Method
- b. Methodology
- c. Ethics
- d. [ All of these are part of the paradigm. ]

To test this, conduct a simple experiment. Ask ten people “What is the purpose of K-12 education?” and record their answers. Odds are that you will get a wide variety of responses, including statements like “to prepare kids for jobs,” “to teach kids how to live well,” “to help kids self-actualize,” “to prepare kids for college,” “to cultivate a love of learning,” or “to sort kids for the capitalist economic machine.”

Whenever I have done this experiment (even with educators working on the same project or at the same institution), I rarely receive answers that are repeated multiple times, which means that even among those engaged in the “work of education” or surface-level problems of how to improve teaching and learning, there is little deep consensus about what the ultimate purpose of that work is and what it actually entails.

This leads to obvious problems and misunderstandings. After all, how can we work together on curriculum (e.g., liberal arts vs. vocational training) if we cannot agree on where we should prepare our kids to be once they graduate (e.g., workforce vs. college)? How can we work together on improving standardized methods if we have not agreed on what learning should be standard to all students? And how can we work together on improving behavioral and disciplinary mechanisms if we do not agree on how we want our children to behave?

Paradigms are deep and complicated and help us understand why two people that might be engaged in solving the same problem (such as Mother Theresa and Karl Marx or Mohandas Gandhi and Jeanne d’Arc) can operate so differently from one another.

For instance, you may have heard of the “quantitative vs. qualitative” distinction in education research, where quantitative researchers utilize statistical methods to determine predictive or correlational relationships between numerically-represented data variables and qualitative researchers utilize naturalistic methods to tease apart the complexity or essential nature of phenomena being studied.

An example of this might be a quantitative researcher attempting to construct a hierarchical linear model to determine how a child not having access to lunch at school (interacting with other variables, such as family socioeconomic status, parental educational attainment, and so forth) impacts their performance on a standardized test.
In contrast, a qualitative researcher might attempt to understand what minoritized children’s experiences are like when they are expected to perform on standardized tests without their basic physiological and social needs being met.

In this case, both the quantitative and the qualitative researcher are loosely working in the realm of student lunch and academic achievement, but what they do, how they do it, and why they do it will differ greatly, because of the hidden parts of their paradigms. Some of these differences may be visible on the surface, but others won’t.

Thus, the difference between "quantitative" and "qualitative" education researchers cannot merely be that some folks like numbers and others don’t. Differences go much deeper, and focusing only on the method may make us think that differences are greater than they are or, alternatively, that differences don’t exist when methods simply look similar.

**Exploring Paradigms**

In this section of the book, we will explore various paradigms for approaching education research. We will not treat "qualitative" and "quantitative" as paradigms, because these terms deal only with surface-level methods or methodologies, but will instead attempt to identify some of the rich paradigmatic differences that have historically been present in education research, paying attention to deep differences and assumptions that researchers are making about the world, knowing, and being.

There currently is no consensus on what constitutes a paradigm in education research, how many paradigms there are, and what their names should be. Nonetheless, I will proceed in subsequent chapters by identifying a number of deep paradigms that are either currently common in education research or that have historically influenced education researchers. These will include the following:

- Positivism / Post-Positivism
- Interpretivism
- Critical Theory
- Feminism
- Postmodernism
- Pragmatism and Design

**Learning Check**

Which of the following would be an example of a research paradigm (in the deep sense)?

a. Axiology  
b. Interviews  
c. Quantitative  
d. Positivism

Like discussing politics, exploring and navigating such diverse paradigms can often be difficult for learners, because their stark diversity can lead to deep-seated discomfort, cognitive dissonance, and disagreement. I submit, however, that such discomfort is healthy and instructive as long as we allow ourselves to be guided by two key principles: pluralism and incommensurability (cf. Kimmons &
Pluralism merely means that we recognize that there might be multiple, legitimate ways of viewing the world or of solving problems (Kimmons & Johnstun, 2019). The behavioral scientist experimenting on rats in a sterile laboratory is doing very different work than the phenomenologist uncovering the experiences of students in disadvantaged schools. The two researchers are asking different questions, but both are important, and we would be reasonably concerned if the phenomenologist began treating students in disadvantaged schools like experimental rats or if the behavioral scientist began asking rats deep questions about their experiences. Phillips (2006) similarly explains that “It seems incontrovertible that Newton, Einstein, Galileo, Harvey, Pasteur, Darwin, Bohr, Boyle, Rutherford, and the Curies all were doing science, yet according to [some stipulative definitions of research], none of them was.” Many types of work are legitimate and necessary. So, we don’t need to engage in a process of deciding which is the one, right way of doing education research but should rather seek to recognize the benefits and limits of different paradigms for understanding the world and for solving problems.

Similarly, incommensurability merely means that we recognize that different paradigms are just that: different (Oberheim, 2018; Yanchar & Williams, 2006). They don’t need to be reconcilable to one another. We don’t need to understand interpretivism through a positivist lens or feminism through a pragmatist lens. Rather, we should seek to understand each paradigm on its own terms, using its own language, and avoiding the temptation to believe that different paradigms are merely different wrappers on the same box of truth. Paradigms do not need to work together, only people do, and in order for people to work together, we have to at times acknowledge and honor the depths of differences that separate us. So, interpretivism is not just a “poor person’s positivism,” and feminism is not just interpretivism applied to women as a group. Rather, each paradigm is distinct, rich, and deep and should be understood as its own unique entity.

References


2.1

Inferential Statistics

Phillip Isaac Pfleger

Everyone makes inferences, general statements drawn from specific evidences or experiences, as they learn about and act in the world around them. Inferential statistics are powerful tools for making inference that rely on frequencies and probabilities. Consequently, an understanding of inferential statistics can improve one’s ability to make decisions, form predictions, and conduct research. It can also protect one from the misused and misinterpreted statistics that are all too common occurrences.

This chapter is not meant to teach all statistical principles or to convince the skeptic of the value of quality statistical inference. Instead it is meant to provide a brief taste of inferential statistics, just enough to help the reader decide whether or not to pursue more information on the topic. Three general topics will be covered in the chapter: (1) the importance of a representative sample, (2) the types of questions that can be answered by statistics, and (3) the most common branch of statistical analysis, which is called Null Hypothesis Significance Testing (NHST).

Sampling

We make inferences when we do not have access to the whole picture. For example, a candy company may want to be certain of the quality of their candies, so they taste a few. It is ludicrous to expect the company to taste all of their candies, because they would no longer have anything to sell. However, when they say that a whole batch is good or bad based on a sample, they are wading into uncertain territory. The same is true in inferential statistics. The process of inferential statistics has been labeled, “decision making under uncertainty” (Panik, 2012, p. 2). To reduce uncertainty it is necessary for the sample to represent the population (the whole batch of candies in this case). If the sample is not representative, then the inferences drawn about the population would be incorrect.

Theoretically, the best way to get a representative sample is called simple random sampling (SRS). Simple random sampling means that every person in the population, or every candy in the batch, has an equal chance of being selected. In practice this is often difficult or impossible. Researchers cannot force people to participate in their studies, so they are automatically limited to those who are interested in the study in the first place. With many other limitations preventing a truly random sample, many other options become necessary. These quasi-experimental designs tend to be complicated, leading some researchers to gather whatever sample is convenient. However, convenience sampling is not a good practice, and it greatly increases the chance of a non-representative sample, which invalidates the generalizability of the research. Instead, the aspiring researcher should familiarize himself or herself with the more complex quasi-experimental designs.
Statistical Questions

Foundational to the design of the experiment or study is the selection of the research question. The selection of the question leads naturally into the selection of an analysis and therefore requirements on the data that can and should be gathered.

Many different types of analyses are available, and each one lends itself to a different type of question or set of questions. A regression, for example, will tell you how strong the relationship is between one variable of interest and another. It will also tell you if one variable predicts the other and helps you make predictive models. A simple t-test will tell how probable it is for one group to be different from another. While each test may answer different questions, it is important to consider that all statistical analyses share one limitation in particular. Inferential statistics can only answer questions of how many, how much, and how often.

This limit on the types of questions a researcher can ask comes, because inferential statistics rely on frequencies and probabilities to make inferences. Consequently, only certain types of data may be used: nominal, ordinal, interval, or ratio (Panik, 2012, p. 4).

Nominal data consists only of a classification into groups, such as male or female, or control group or experimental group. Ordinal data is also categorical in nature but includes an order placed on the data. For example, first and second place in a race tell us nothing about the relationship between the two runners other than the fact that the first place runner came before the second.

Interval data and ratio data are very similar to each other and are often grouped together under the terms numerical or quantitative data. Interval data are like temperature in degrees Celsius. They are numbers that have meaning, but the zero is not an absolute zero. In the case of degrees Celsius, a zero does not mean a complete lack of temperature. It just means the point where water freezes. The temperature scale of Kelvins is different. Zero on that scale means absolutely no heat, making this scale a ratio scale.

Ratio data is often, but not always, the ideal data for an analysis. However the best way to determine what type of data to gather goes back to the research question. The research question will not only help you decide if statistics will help you, but it will also help you decide what type of data you should gather.

Null Hypothesis Significance Testing

Most people who have read an academic article have been exposed to something called a p-value. The p-value is fundamental to the most common statistical practice today, Null Hypothesis Significance Testing (NHST). NHST involves estimating the probability that the average of your sample is different from some other expected value (the null hypothesis). This probability estimate is the p-value. For example, if a researcher was investigating whether or not two groups were different, the null hypothesis would be “the difference between group A and group B is zero.” If the difference between the groups was 3.7, and the p-value was .03, then there would be a 3% chance that the difference in our sample was 3.7 if the true difference was zero.

For the novice statistician this can seem like a bit of a black box. When examined fully, however, it is not too hard to understand. The whole process involves giving the null hypothesis a score based on
how many standard deviations away from the sample mean it is. The p-value is calculated from this
score, and if the p-value is below a preset value (usually .05), then we say that it is “significant.”

Airline Example

To better clarify the process associated with many statistical inferences, consider the data in Table 1
(R Core Team, 2016).

Table 1. Airline Passenger Data

<table>
<thead>
<tr>
<th></th>
<th>1959</th>
<th>1960</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>360</td>
<td>342</td>
<td>391</td>
<td>49</td>
</tr>
<tr>
<td>406</td>
<td>406</td>
<td>419</td>
<td>13</td>
</tr>
<tr>
<td>396</td>
<td>396</td>
<td>461</td>
<td>65</td>
</tr>
<tr>
<td>420</td>
<td>420</td>
<td>472</td>
<td>52</td>
</tr>
<tr>
<td>472</td>
<td>472</td>
<td>535</td>
<td>63</td>
</tr>
<tr>
<td>548</td>
<td>548</td>
<td>622</td>
<td>74</td>
</tr>
<tr>
<td>559</td>
<td>559</td>
<td>606</td>
<td>47</td>
</tr>
<tr>
<td>463</td>
<td>463</td>
<td>508</td>
<td>45</td>
</tr>
<tr>
<td>407</td>
<td>407</td>
<td>461</td>
<td>54</td>
</tr>
<tr>
<td>362</td>
<td>362</td>
<td>390</td>
<td>28</td>
</tr>
<tr>
<td>405</td>
<td>405</td>
<td>432</td>
<td>27</td>
</tr>
</tbody>
</table>

This is the number of passengers that flew each month on a certain airline in 1959 and 1960, as well
as the differences between the two. A researcher may want to know if there was a difference in
passengers between the two years. This researcher would first need to clarify the null and alternative
hypotheses and set the alpha level (the level our p-value has to be before we will believe the
conclusions).

H0: The average of 1959 = the average of 1960. (i.e the difference = 0)

Ha: The average of 1959 ≠ the average of 1960. (i.e the difference ≠ 0)

α = 0.05

In other words, the researcher is assuming the two are the same but will have enough evidence to
support that they are different if the p-value is less than 0.05.

The differences between the two groups is found in column three. Our first step is to find the mean of
this column by adding all of the values and then dividing by the number of data points we added
together. This gives us a mean of 47.83. This sample mean is a point estimate, or an approximation,
of the true difference. We know this data follows a certain pattern (Figure 1), called a normal
distribution.
Consequently, we know that 68% of the data is within one standard deviation, and 95% is within two standard deviations. A standard deviation is a measure of uncertainty. It is the average distance between the data points and the sample mean. We calculate the standard deviation using this formula (Moses, 1986, p. 50):

\[ SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \]

The standard deviation for the airline data is 17.58. A test statistic is obtained using the following formula (Vaughan, 2013, p. 47):
The test statistic for the airline question is 9.425. The p-value is the probability of getting a test-statistic as extreme or more extreme than the one you got, given the null hypothesis is true (Brase & Brase, 2016, p. 425). In other words, it is the probability of getting 47.83 as the average distance if the true average difference was 0. With a p-value of 0.000013, which is less than the .05 standard the researcher set at the start, there is enough evidence to reject the null hypothesis. Thus, the researcher concludes that there is a difference between the two groups. It is important to note that conclusions based on p-values alone lead to an incorrect answer 5% of the time. Consequently, it is good practice to interpret p-values in the context of other inferential statistics, such as effect sizes and confidence intervals. This approach is neither perfect, nor the only approach available, it is simply the most common.

Conclusion

Inferential statistics are an extension of the natural human tendency toward inference. They are powerful tools that can help answer questions such as how much, how many, or how often. An understanding of the process of statistics can help us be better consumers of research, prevent us from being misled by invalid or misinterpreted statistics, and give us another tool in the search for knowledge.

References


Keywords: inference, inferential statistics, null hypothesis significance testing
Most research studies in education require some form of sampling. Because you can’t always study everyone or everything, sampling means that you only study part of a larger group and (hopefully) are still able to draw meaningful conclusions.

For instance, if you wanted to study university students over the age of 50, you might randomly survey a few hundred people by phone and then draw conclusions about other people within that demographic group more generally. In such a study, you would be concerned about two different groups of people: the larger population you are trying to draw conclusions about (designated by a capital N) and the smaller sample you are actually studying (designated by a lower-case n).

How we go about sampling is very important and will be dependent upon a variety of factors, such as the questions we’re trying to answer, what we’re studying, and the context. Because goals and needs vary, there is not a single, correct way to sample, but there are appropriate (and inappropriate) ways to sample depending on our study.

In this chapter, I’ll explain four basic types of sampling — convenience, purposeful, random, and census — along with guidance on specific approaches for each type, as well as examples and cautions. I will then conclude by providing some considerations that should guide you in selecting appropriate approaches in your own research.
Key Terms

Census
A study that includes every member of the population (removing the need for sampling, generalizability, etc.).

Convenience Sampling
An approach to sampling, common in design and professional practices, that chooses subjects that are accessible to the researcher, such as testing with a colleague, interviewing a spouse, etc.

Population
The group facing the problem researchers are trying to solve or to whom researchers will generalize their results (e.g., K-12 students), represented by the uppercase "N" variable.

Purposeful Sampling
An approach to sampling, common in qualitative research, that chooses subjects that will provide insight for answering the study questions.

Random Sampling
An approach to sampling, common in quantitative research, that chooses subjects randomly from a target group or population.

Sample
The actual group being studied, represented by the lowercase "n" variable.

Convenience Sampling

Convenience sampling consists of studying those who are close to us or who are easy to study. A web developer might put an in-progress website in front of other web developers for feedback, a graduate student might survey other students in a class to get their opinions on educational reform, or a teacher might ask other teachers about their ideas for lesson improvement. In all of these cases, the primary factor guiding who is sampled is the ease at which data can be collected. Because it makes data collection easy, convenience sampling is common in user testing, design settings, and other situations that utilize non-traditional or less-formal research methodologies, such as action, design, or guerrilla research.

Of all the sampling methods, convenience sampling tends to be the easiest to do, which is why it is commonly used in teaching and design scenarios, but it is also the least accurate and most
problematic. It is great for getting quick data and guidance on topics that might deal with universal
human experiences, but its results are not generalizable to the population and also do not allow you
to acknowledge varieties in the population.

In the web developer example above, putting an educational site in front of a few colleagues might
provide you with useful, instant guidance on how to improve some aspects of the design, but because
your sample subjects will tend to be like you, it will often fail to provide you with guidance to meet
the needs of all potential users (e.g., you will not know how a person with a vision disability might
struggle with the site unless one of your close colleagues has a vision disability). Thus, results derived
from convenience sampling will typically not meet rigor expectations of formal research, but they can
be helpful for the initial shaping of research questions as well as iterative designs of instruments,
tools, and products.

**Purposeful Sampling**

Purposeful sampling consists of selecting subjects specifically out of a population that the researcher
believes will help them to most meaningfully and accurately answer the research question. A
hallmark of qualitative methodologies, purposeful sampling takes at least four common forms, which I
will now explain: informant, extreme case and intensity, quota, and snowball (see Table 1).

**Table 1**

*Four Common Non-Random-Sampling Strategies with Examples*

Example numbers are used for illustration only and do not accurately reflect necessary comparative
sample sizes needed for each approach as this would be dependent upon the methodology, research
question, and other considerations. For appropriate sample sizes, consult other studies that use
similar methodologies to your own.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Strategy Example</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informant</td>
<td>Interview students representing minoritized ethnic groups.</td>
<td>10 students from the target ethnic group</td>
</tr>
<tr>
<td>Extreme Case and Intensity</td>
<td>Interview students representing both majority and minoritized groups.</td>
<td>5 students from the minoritized group + 5 students from the majority group = 10 students</td>
</tr>
<tr>
<td>Quota</td>
<td>Interview the same number of students from each target ethnic group.</td>
<td>2 students x 5 different ethnic groups = 10 students</td>
</tr>
<tr>
<td>Snowball</td>
<td>Interview a student representing a target ethnic group, then ask them to help you to identify additional students in the target group, and so forth.</td>
<td>1 student from the target group + 2 of that student’s friends + 7 of their friends = 10 students</td>
</tr>
</tbody>
</table>

**Informant**
Informant sampling selects subjects based upon their expertise or ability to provide insight for answering the research question. They are intentionally selected from the population for their experience with the topic under study and their ability to help the researcher understand it. For example, if we wanted to understand women’s experiences of being harassed in academia, we might interview women who have had experiences of being harassed. Doing so would not allow us to generalize their results to all women or to say how prevalent harassment is in academia, but it would allow us to understand the nuances of harassment, what it means to be harassed, what the effects of harassment might be, and how to identify it in practice.

The strength of informant sampling is that it allows us to develop a deep understanding of our topic from those best equipped to help us understand it. Yet, any results we derive from such an approach will not be reliably attributable to the whole population and will be dependent upon the criteria we use for selecting our informants. If in the case above we only select ethnic majority women to study, for instance, then we will not understand how ethnic or racial differences among women influence the forms that harassment takes, and we also will not understand whether and how men might face harassment in similar contexts. This is not a weakness of the sampling approach so much as it is an intentional delimitation, allowing the researcher to hone in on the issue or sub-population they desire to study the most.

**Extreme Case and Intensity**

Extreme case sampling selects subjects from opposite ends or contradictory sides of the phenomenon being studied to give a sense for the breadth of the topic and divergence in experiences, opinions, or characteristics. This requires the researcher to have a sense for what the extremities might be prior to selection but allows them to show a range in their sampling. For example, if you wanted to study
how political affiliations of teachers influence their pedagogy, then you might sample teachers from both the far-right and the far-left of the political spectrum. This would allow you to effectively see how drastically experiences and effects diverge but would give you limited insight into more moderate cases of political divergence, which would likely represent the majority of the population.

As a more moderate approach, intensity sampling also tries to account for a diversity of subjects in the sample, but it is less interested in the extremity of the cases and is more focused on identifying richness in sampled subjects. To apply intensity sampling to the previous example, we might still select subjects representing the major political parties but will focus on those who will help us to best understand the phenomenon in question. For instance, interviewing an extreme case of a politically anarchistic mathematics teacher might not be very helpful for understanding the phenomenon in question, because there might not be much overlap in the teacher's political stances and how they teach fractions. However, interviewing intensity samples of politically active social studies teachers would likely be more informative even though subjects might be more politically moderate simply because their subject areas would allow the researcher to explore a richer relationship between each teacher's politics and how they teach their related subject area.

Quota sampling selects fixed or equivalent numbers of subjects from predetermined groups to ensure that the study accounts for important differences between the groups. This requires the researcher to make a priori assumptions about what types of differences between subjects are important, and the researcher then will treat each group of subjects like a separate population pool. For example, if you wanted to understand teacher attitudes toward hiring practices and to see how experiences might differ based on the teacher’s gender, you might choose to interview five recently-hired women and five recently-hired men to see commonalities and differences between their accounts. Similarly, if you were studying the experiences of racially minoritized students in low-SES schools, you might choose to interview four students from each major racial group represented in your school’s demographic data (e.g., four Latinx students, four Black students, and four Asian American students). In such approaches to sampling, the overall representation of the subgroup in the larger population is irrelevant (e.g., the Asian American student population of the school might be much smaller than the Black student population), but the goal would be to make sure that you have sampled sufficiently from each group to be able to draw meaningful conclusions.

Quota sampling is useful if researchers have a priori categories of analysis that matter to their research questions and if researchers are able to identify subjects based on these categories. By
conducting parallel analyses of groups of subjects, researchers can construct group-based narratives for comparison and ensure that diverse perspectives (in terms of the grouping criterion) are explored in a focused manner. If, for instance, in the hiring example above, researchers only sampled one woman and nine men, then women’s hiring experiences likely would not be well-represented in the findings, but by setting a quota of five women, researchers can help ensure that women’s experiences will be treated equally to men’s.

If, however, our *a priori* categories are messy, poorly defined, or unnecessary to our research questions, then quota sampling is less useful. In the example of racially minoritized students, for instance, sampling four Asian American students might not yield very helpful results once we realize that the experiences of students from Korea, China, Japan, and various Pacific islands might be quite distinct or that the experiences of Black students from poor families might be very different from those of Black students from more affluent families. Thus, the key to quota sampling is to ensure that your *a priori* categories are important, accurate, and clear.

**Snowball**

Snowball sampling selects a small number of initial subjects and then gradually grows the sample via additional sampling through the subjects. This approach is especially useful if researchers have difficulty identifying or recruiting subjects in the target population (due to ignorance, lack of trust, or outsider status). For instance, if you were studying homeschool parents, you might not be able to identify many subjects at the outset of the study due to lack of knowledge, privacy, etc. So, you might begin by only identifying one parent, interviewing them, and then, as part of your interview protocol, asking them if they know of other parents who homeschool as well.

This approach allows the researcher to utilize subjects’ knowledge and social networks as sampling tools and is very useful for gaining entry to insular, invisible, or poorly understood populations. However, because it relies so heavily upon the knowledge and networks of individual subjects, it may have difficulty providing representation of the whole population. Following the homeschoolers example above, the subsequent parents that we sample will be dependent upon the initial parent, which would mean that the researchers might only ever be able to identify subjects that are like the original. If there are different networks of homeschool parents (e.g., religious homeschool communities vs. military families), then this sampling strategy might only allow entry to one of the networks and not others, and if the networks exhibit different values, behaviors, or norms from one another, then results will not account for the totality of homeschool parent experiences. Thus, as researchers traverse these networks, they should be aware of the potential for insularity in selection.
and attempt to identify alternative networks that might also exist in the population.

**Random Sampling**

Random sampling consists of selecting a sufficient number of subjects at random from a population to ensure that results are indicative of the population at large. If your research goal is to generalize your results from your sample to an entire population, then you must utilize a random sampling strategy. Failure to randomize your sample would mean that your study might be biased toward particular sub-groups in the population (e.g., those that are near you or are like you).

For instance, if I wanted to know how students feel about their school lunch programs, the easiest thing to do would be to interview my own kids or my neighbors’ kids. But would this tell me anything about school lunch programs more broadly?

Similarly, if I wanted to understand teacher beliefs about technology integration in my state, and I proposed to survey teachers via email, which teachers I sent the survey to (and which teachers responded) would be very important. If I only sent the survey to teachers in one subject area, geographic location, grade level, school socioeconomic classification, etc., then my results would not reflect the beliefs of teachers in my state broadly. And even if I sent the survey to every teacher in the state and only 10% responded, might there be reasons that the other 90% did not respond that would influence their beliefs about technology integration (such as not having a computer to complete the survey with)?

The goal of randomization is to allow you to study a relatively small proportion of the population and get results that can reasonably be said to represent the population more broadly. This makes research studies possible that would otherwise be infeasible due to exorbitant cost or other limitations (e.g., there’s no way that you could ever interview every child about how they feel about school lunches), but for randomization to work for achieving generalization, it must be done in a manner that is appropriate for your research question, context, and population.

Education researchers employ at least four common random-sampling strategies for doing this well. These include simple, stratified, proportional, and cluster sampling (see Table 2), which I will now explain in more detail.

**Table 2**

*Four Common Random-Sampling Strategies with Examples*

Example numbers are used for illustration only and do not accurately reflect necessary comparative sample sizes needed for each approach as this would be dependent upon population sizes and between- and within-group variances.
### Table 1: Sampling Strategies

<table>
<thead>
<tr>
<th>Approach</th>
<th>Strategy Example</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simple</strong></td>
<td>Survey random teachers from a state-level list of teachers.</td>
<td>1,000 teachers</td>
</tr>
<tr>
<td></td>
<td>Survey teachers from each type of school.</td>
<td>334 teachers from traditional public schools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>333 teachers from private schools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>333 teachers from charter schools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>= 1,000 teachers</td>
</tr>
<tr>
<td><strong>Stratified</strong></td>
<td>Survey a representative number of teachers from each type of school based on the teacher population of each type of school.</td>
<td>800 teachers from traditional public schools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100 teachers from private schools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100 teachers from charter schools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>= 1,000 teachers</td>
</tr>
<tr>
<td><strong>Proportional</strong></td>
<td>Survey all teachers from randomly-selected schools.</td>
<td>300 teachers from school A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300 teachers from school B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400 teachers from school C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>= 1,000 teachers</td>
</tr>
<tr>
<td><strong>Cluster</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Simple

Simple random sampling selects subjects from the target population without consideration for sub-groups, categories, or differences between the subjects. This allows researchers to show that their results are representative of the population broadly while giving them limited power to answer more nuanced questions about sub-groups.

In the teacher email survey example, you might acquire a list of teacher emails from your state department of education and email a random selection of 1,000 teachers from the list. This would give you a sense for how teachers in your state broadly think about technology integration, but suppose that you wanted to then see if teacher beliefs varied based on whether they worked for a charter vs. a traditional public school. If your random sample only provided results from 10 teachers representing charter schools, then you probably would not have enough data to meaningfully compare their results to the group at large.

As this case illustrates, simple randomization is great for fast generalizability, but it will often not provide sufficient data for more nuanced analysis especially if our questions deal with sub-groups that are relatively small or that might be less likely to show up in the data. Charter schools represent a small proportion of schools generally, and so results for their teachers will not be well-represented in the dataset. Similarly, results from teachers representing racial or ethnic minoritized groups would also likely be limited for comparison, because such a high proportion of the overall sample will...
represent the majority group. And teachers with limited access to technology would likely be ignored altogether because the sampling procedure did not take into account the role that email access might play in ensuring that those teachers' responses were actually counted.

If such considerations are not pertinent to the research question and context, however, then simple randomization might be appropriate and tends to be the easiest method for collecting generalizable data quickly and efficiently.

**Stratified**

Stratified sampling selects subjects from different groups, levels, or strata in the population so that results will account for important variation in the population. Like quota sampling above, this requires *a priori* identification of the strata that differentiate subjects from one another and for researchers to then sample an appropriate number of subjects from each stratum. This allows for the sample to reflect the diversity of the larger population and prevents findings for smaller groups in the population to be ignored due to lack of representation.

For instance, if you wanted to study high school drop-out rates and believed that rates might differ based on the type of high school (e.g., charter vs. private vs. general public), then doing a simple random sample would probably result in general public high schools comprising most of your dataset. If charter schools only represented 1% of the schools in your state, then they would only be expected to represent 1% of your dataset, which might make it difficult to draw robust statistical conclusions about charter schools. To combat this, you might intentionally sample in such a way that 33.3% of your dataset represented each type of school. This would allow you to compare results between groups.

Stratified sampling is preferable to simple random sampling when you want to be able to compare results between *a priori* groups. To do it properly, though, you need to ensure that your sample size for each stratum is sufficient to make generalizable claims, and doing so can be a double-edged sword in that it allows you to draw conclusions about relatively small segments of the population, but you will then need to be careful of how you interpret results, because stratum-based results (especially for strata with smaller populations) may not generalize to the entire population.

**Proportional**
Like stratified sampling, probability-proportional-to-size (or PPS) sampling (also known as proportional stratified random sampling) selects subjects from different groups, levels, or strata in the population but does so at a rate proportional to their overall representation in the population. This allows for the sample to reflect key characteristics of the population at large that could have been lost through simple random sampling and also allows results to be comparable between strata and also generalizable to the entire population.

For instance, if you wanted to survey teachers on their experiences with home-life balance, you could reasonably assume that gender would be an important factor that would influence results, due to different social expectations that are placed on women vs. men. This means that if the representation of women vs. men in your sample did not match that of your population, then your results would be less accurate. Similarly, if you wanted to study student achievement on a standardized test across a state, you could reasonably assume that results would be influenced by the type of school students attend (traditional public vs. private vs. public charter). So, if your state's student population has a breakdown of 80%/10%/10% for public/private/charter, then you would want to ensure that your sample has a similar breakdown to be valid. This would mean sampling from all three types of schools until you could be confident that the sample accurately reflects the population at large.

PPS sampling is superior to simple random sampling when key \textit{a priori} strata are identifiable at the outset because it can help to ease sampling errors that are always possible with randomization (e.g., accidentally overrepresenting men or private schools in a dataset).

**Cluster**

Cluster sampling selects subjects from organic groups or clusters of the population for analysis. This
allows researchers to focus their efforts on one or more clusters (like specific schools) and generalize their results to the entire population. For cluster sampling to work, the clusters that researchers study should be similar to the larger population both in terms of matching its characteristics and having internal variance or diversity within the cluster. For instance, if you want to study the home lives of students in a state, you might go door-to-door in a few random neighborhoods that are representative of the state at large. This requires researchers to first identify clusters that might be appropriate for analysis and then to randomly select which cluster(s) to study.

Most applications of cluster sampling are geographic in nature when researchers are doing work on-the-ground and cannot feasibly travel to every school, city, or state in their population. Instead, they will identify schools, cities, or states that are representative of the larger population and then randomly sample from this list to reduce travel time and other difficulties inherent in shifting between contexts.

The major benefit of cluster sampling is feasibility because it allows researchers to study subjects as organic clusters in finite settings. However, for results to be valid, clusters must be representative of the population, and results are further problematized if clusters vary in size and internal diversity or variance from one another.

**Census**

And finally, a census is a study in which all members of a population are treated as subjects. It is not a sampling strategy per se but is rather any study where the population (N) and the sample (n) are identical. Census studies are not very common in education research mainly because it's often difficult to study large groups (e.g., all K-12 teachers in a country), but in situations where populations are relatively small, a census study might be appropriate.

For instance, if you are studying a database of 100 lesson plans and want to be able to determine what percent of the lesson plans in the database have particular characteristics, like learning objectives or standards alignment, then your population would be relatively small (N = 100). Even assuming that you aren't trying to generalize to another set of lesson plans, sampling a large enough number to generalize to the database would require you to analyze about 92 lesson plans to achieve a confidence interval of +/-3% at the 95% confidence level (or n = 92% of the population). But in a situation like this, if you are already going to analyze 92% of the population, then why not just go ahead and analyze the remaining 8%?

When it is possible to do a census study, doing so provides clear benefits to other sampling procedures, because you can avoid sampling errors and can rely upon basic descriptive statistics (like
Considerations

Before choosing a sampling method, you should consider at least two aspects of your study that will determine appropriateness. These include (a) the goal of your research and (b) the contextual feasibility of your study and questions. I’ll now explain each of these in more depth.

Research Goal

First, research is done for a variety of purposes. Sometimes we do research to determine norms or trends in a large population, such as whether U.S. legislator mandates for high-stakes testing negatively impact student wellbeing. At other times, we do research to gain a deeper understanding of a phenomenon, such as what it is like to be an undocumented immigrant student in a K-12 school while learning English as a second language. And at other times, we do research to improve a product or intervention, such as iteratively testing a reading intervention program for fourth-grade students. In each of these scenarios, our goal for doing research is different, and our sampling method should reflect this intended goal.

If iterative improvement is our goal, then we will often rely upon sampling methods that are fast and easy or that allow us to answer targeted, finite questions for improvement, thereby favoring convenience or purposeful sampling approaches. If our goal is to understand a phenomenon in-depth (and to remove noise from the larger population), then we will use approaches that are more focused or purposeful. And if our goal is to generalize or to get a broad understanding, then we will use approaches that randomize sampling or that allow us to study every member of the population via a census.

To drive this point home a bit more, many researchers initially believe that randomization is always valuable for research, but randomization is only a tool for attempting to achieve generalization from the sample to the population. It is only appropriate if generalization is the goal. Conversely, if convenience or purposeful sampling is utilized in a study, then generalization can never be achieved. Thus, researchers need to be sure that their method of sampling aligns with their research goals, and the decision of how to sample must be driven by that goal.

Contextual Feasibility

And second, the context of our study and our questions matter in determining sampling methods. Some topics of study might be simpler to observe, measure, or identify, such as a family’s income level, while other topics might be more complex and ephemeral, such as a child’s emotional wellbeing. If the topic of your study is simpler and easier to observe in a single subject, then you can rely upon approaches that allow for greater numbers of subjects and randomization, but if your topic is more complex, then you will need to use approaches that are more purposeful.

Albert Einstein hinted at this when he famously quipped that “Not everything that counts can be counted, and not everything that can be counted counts.” As an example, let’s say that we work for a state that has called for schools to submit grant proposals. If we want to study whether the proposals meet specific formatting requirements, we could scan each document with relative ease to see if the
proper headings, font size, and margins were used; or we could write a simple program to do this for us. If, however, we were tasked with evaluating the proposals on scholarly merit or likelihood of success, then this would require much deeper reading of the proposals and would rely upon much more expertise and interpretation on our part. In this scenario, a single researcher could feasibly study formatting issues of a sample of 1,000 proposals with relative ease, but determining each proposal's implementation success likelihood or evidentiary merit would take much more time and effort.

This complexity is also influenced by the field's prior work and knowledge regarding our research topic. If we are studying a topic that is brand new and upon which little previous work exists in terms of theoretical constructs, measurements, tools of observation, and so forth, then we will often need to rely upon more purposeful and focused approaches that allow us to go deep to unpack nuances. If, however, our topic has been studied deeply already and others have provided us with lenses and tools for measuring and observing, then we can use sampling approaches that take a broader view, such as randomization, because we will have a clearer idea of what we are looking for and how to go about finding it.

This issue of feasibility is further complicated in the education research setting because education research is a social science that deals with intentional, agentic humans with rich, complex life histories, which means that the complexity of our study context will also be informed by how we view our subjects and what we expect from them in the course of the study. If we are merely studying external, observable characteristics of our subjects, then larger samples and randomization will be more possible, but if we are expecting to unpack or reveal deep experiences, beliefs, or interpretations, such as trying to understand traumatic experiences among young children, then we will need to be much more purposeful and directed in how we select and interact with our subjects and come to view them not merely as subjects but as participants, informants, or collaborators of the research.

Thus, if we are studying simple topics, we can more feasibly rely upon large sample sizes and randomization to make generalizable determinations, but if we are studying more complex topics, then we may need to reduce sample sizes and focus our efforts on purposefully targeting specific subjects for deeper analysis.
2.3

Qualitative Rigor

How do I conduct qualitative research in a rigorous manner?

David Dwayne Williams & Royce Kimmons

Imagine you are reading the abstract of a study, and the researcher claims something like “our interviews showed that games can be great for teaching collaboration skills.”

What questions would you have for the researcher?

Would you believe them and trust them?

Or what would you need to know about how the study was conducted, who was interviewed, how the interviews were analyzed, and so forth before you would be willing to agree that “yes, indeed, games are great for teaching collaboration skills?”

Because qualitative inquiry is messy, relies on relatively few participant accounts, and often requires heavy interpretation on the part of the researcher, qualitative researchers need to be careful both in how they go about doing their work and how they report on it to ensure that their process is reasonable and their results are believable.
Key Terms

Confirmability
An expectation that results should be supported by participants, other researchers, and existing literature.

Credibility
An expectation that study results should be believable to critical readers, approved by participants, and otherwise true or accurate.

Dependability
An expectation that methods, logic, and reasoning guiding a study should be clear, stable, and consistent.

Discipline
An expectation that the researcher is thoughtful and methodical, following key standards and norms that are generally accepted by other researchers who use similar methodologies.

Generalizability
The ability to take the results of a study focusing on a sample of a population and to apply them to the overall population.

Rigor
An expectation (common in all research methodologies) that the researcher is being thorough, responsible, reasonable, and accurate.

Transferability
The ability of a reader to apply or transfer the results of a study to their own situation or context.

Trustworthiness
An expectation in qualitative methodologies that the researcher should provide enough explanation, transparency, and evidence that their results can be confidently believed.

Earlier we established that research is both systematic and auditable. Embedded in this is the notion that our practices and the results of our efforts should be rigorous, meaning that we strictly adhere to specific criteria of quality. In qualitative methodologies, this rigor typically means two things: (a) that we are disciplined (Cronbach & Suppes, 1969), meaning that we follow certain key standards established by other researchers, and (b) that we are trustworthy (Guba & Lincoln, 1989; Lincoln & Guba, 1985), meaning that whoever reads our work should be able to feel informed and confident in both what we did and what we concluded. For these reasons, qualitative researchers have developed a variety of standards to help ensure both discipline and trustworthiness, which we will proceed in this chapter to explain in more detail.

The rationale for trustworthiness as the central objective of these standards is centered on the desire most people have for truth. Qualitative researchers agree that most claims people make are based on their subjective constructions of reality. A major objective in sharing our findings from inquiry thus becomes the persuasion of others that our constructions of reality are of value and should also be considered in their constructions. Whether or not these claims are true in any ultimate sense can only be tested over time through many different experiences in a variety of contexts (cf., intersubjectivity), but for any given study, the objective is one of persuasiveness — providing evidence that is compelling enough that audiences are willing to listen to and consider the claims made. In other words, the more the researcher can do to make the inquiry trustworthy, the more likely it is that readers will be persuaded to read on.
The standards presented by Lincoln and Guba (1985) and by Guba and Lincoln (1989) provide an excellent core of standards for a beginning. They suggested four types of standards be used to ensure trustworthiness: credibility, transferability, dependability, and confirmability. They also recommended several techniques for conducting studies so that they meet these standards. Although no single study is likely to adhere to all of these standards, which we will discuss below, as we meet or address more standards we make our work more believable and increasingly influential to people who will read it.

**Learning Check**

Which of the following are rigor requirements in qualitative research?

a. Validity  
b. Confirmability  
c. Trustworthiness  
d. Dependability

**Credibility**

Credibility is the standard by which a qualitative study is expected to be believable to critical readers and to be approved by the persons who provided the information gathered during the study. Lincoln and Guba recommended several techniques researchers may use to enhance the credibility of their research, including prolonged engagement, persistent observation, triangulation, peer debriefing, negative case analysis, progressive subjectivity checks, and member checking, which we will now explain.

**Prolonged Engagement**

Prolonged engagement is a technique whereby researchers immerse themselves in the site or context of the study long enough to build trust with the participants and for the researcher to experience the breadth of variation and to overcome distortions due to their presence (cf., Hawthorne Effect). This may mean an entire year or longer for some studies or as little as a month for others, depending on the size of the study and the level of depth needed for the researcher to become part of a community and understand what is happening. There is no set amount of time a qualitative inquiry should last, but the proper length can be estimated by the researcher once they have spent some time in the site.

For example, if a researcher wanted to understand the phenomenon of Texas high school football, this would require being present at least through a full season of the sport and may also require presence in the pre-season and the off-season, whereas a researcher who simply showed up for a championship game would have little understanding of the nuances, histories, difficulties, perplexities, and larger context of what they were witnessing. If a researcher can be present in a setting long enough to see the range of things to be expected in such a site (e.g., not just the championship game), then the results produced will be more credible.

**Persistent Observation**

Persistent observation is a technique that ensures depth of experience and understanding in addition to the broad scope encouraged through prolonged engagement. To be persistent, the researcher
must explore details of the phenomenon under study to a deep enough level that they can decide what is important and what is irrelevant and focus on the most relevant aspects.

For instance, if a researcher wanted to understand the impacts of homework on marginalized students, they might begin by talking to students at school but would likely need to observe them after school or at home as well. Issues that such students might face, such as working jobs to support their families, providing childcare to working parents, not having access to a quiet study location, and so forth, would not be readily apparent if the researcher didn't leave the confines of the school or the 9 am to 3 pm hours of the established school day.

Failure to engage in persistent observation would mean that the researcher would learn very little detail about particular aspects of the phenomenon under study. Even if the researcher engaged in prolonged engagement (i.e., sufficient time), this alone would not mean that they had explored the phenomenon in sufficient depth, persistently learning more about the phenomenon from a variety of angles and in a variety of ways. Without such persistence, results would be limited in scope and less credible.

**Triangulation**

Triangulation is the verification of findings through (a) referring to multiple sources of information (including literature), (b) using multiple methods of data collection, and often (c) conducting observations with multiple researchers. If a conclusion is based on one person's report, given during one interview to only one interviewer, then it will be less credible than if several people confirmed the finding at different points in time, during multiple interviews, through various unstructured observations, in response to queries from several independent researchers, and in the review of literature.

For instance, if a researcher wanted to understand racially-motivated bullying in schools, a single interview with a bullied child might provide rich data about that child's experience, but readers might be left wondering how pervasive such experiences are and how race played a role in the child being targeted. If, however, the researcher can interview multiple students who have had similar experiences, connect this to other datasets (such as instances of bullying in school behavioral referrals), connect this to previous studies or theoretical literature, and find similar results from other researchers in other contexts, then the results will be more credible to a reader. Although all three forms of triangulation are not required for every conclusion the researcher makes, credibility is increased as more and more triangulation occurs.

**Peer Debriefing**

Peer debriefing is a technique whereby a researcher meets with a disinterested peer so that the peer can question the researcher's methods, emerging conclusions, and biases. A disinterested peer might include anyone who is willing to ask probing questions and who is not a participant or researcher in the setting where the study is being conducted.

For instance, a researcher studying the experiences of immigrant students in rural communities might periodically meet with a colleague who studies early childhood literacy. This peer would then challenge the assumptions that the researcher is starting to make about the context or results and would encourage them to both think critically about what they are observing and to figure out how to make results understandable to someone who is outside the context.
This technique is meant to keep the researcher honest by having someone else independently point out the implications of what they are doing. If a researcher can provide evidence of having engaged in peer debriefing and can show the reader how the report was modified through the influence of the peer, credibility is improved.

**Negative Case Analysis**

Negative case analysis is an analytic procedure that is meant to refine conclusions until they account for all possible cases, without any exceptions. The process involves developing hypotheses based on extensive fieldwork and then searching for cases or instances within the site under study that may contradict the conclusions proposed by the hypotheses.

For instance, if a researcher begins to conclude that poverty is having a serious, negative impact on student achievement in a community, then they should seek to find any negative cases in which a child from an impoverished family is actually excelling academically. This will help them to better understand the interaction between poverty and achievement and what some individuals or families must do to mitigate it.

If no contradictory cases are found after extensive searching, then the hypotheses are considered more credible because no evidence has been found to negate them. If such evidence is found, however, then the hypotheses are modified to account for the new data associated with the negative cases. This process continues until the hypotheses have been modified to account for all negative cases and no new negative cases can be found.

If a researcher completes such an extensive process, the resulting qualitative inquiry report is considered very credible. Though single studies sometimes fail to account for negative cases due to limited exposure or familiarity with the topic, as researchers engage in series of studies over time on the same topic, it is expected that they will eventually grapple with negative cases in a robust way.

**Progressive Subjectivity Checking**

Progressive subjectivity checking is a technique whereby a researcher archives their changing expectations and assumptions for a study. These expectations might include *a priori* and emerging constructions or interpretations of what is being learned or what is going on.

For instance, if a female academic is studying gender disparities in higher education, then she might bring to light her own experiences of inequity, memo how her participants' experiences reflect, reinforce, or challenge her own assumptions and experiences, and acknowledge how her thinking on the topic is changing as she progresses through the study.

In all qualitative research, the researcher is responsible for revealing their biases and preferences in reports, field notes, and the audit trail both initially and over time, and as Guba and Lincoln (1989) explain "if the [researcher] 'finds' only what he or she expected to find, initially, or seems to become 'stuck' or 'frozen' on some intermediate construction [interpretation], credibility suffers" (p. 238).

**Emic Perspective**

Emic or the folk perspective of participants is the insider view of how participants see and understand themselves from the inside-out, and researchers improve credibility by showing that they
are able to understand and communicate about the phenomenon being studied as an insider.

For instance, any researcher studying young children would not just need to be able to interpret behaviors through etic theoretical or psychological lenses (e.g., phobias, apathy, intelligence) but would also need to show that they understand the children as they understand themselves, often using their same words (e.g., "love," "hate," "smart," "dumb," "mad").

That is, it should be clear to readers that the researcher discovered something of the viewpoints held by the people they studied and can see them as they see themselves. If only the researcher's outsider perspective is present, then the study will lack one of the most critical characteristics of a qualitative study: the type of understanding that can only come from empathy.

This also helps ensure that the researcher is not pigeonholed by a priori assumptions that might blind them to new discoveries. If the researcher's original hypotheses are simply confirmed, then qualitative inquiry probably is not the appropriate approach to use, but by discovering emic perspectives, researchers can add richness to existing understandings of phenomena and make their results more credible.

**Learning Check**

Which of the following would be an example of researchers employing emic perspectives?

- a. Observing a kindergarten classroom through a one-way mirror
- b. Explaining teenagers' moral reasoning in their own words
- c. Utilizing a standardized test to determine intelligence
- d. Analyzing usage data for a web-based learning app

**Member Checking**

Member checking is a technique whereby a researcher provides the data record, interpretations, and/or reports for review by the participants who provided the data - the natives. This validates that the represented emic perspective is accurate and is one of the most important techniques for ensuring credibility.

For instance, if a researcher interviewed parents about their reasons for enrolling their students in charter schools, the researcher might provide each parent with a transcript of the interview as well as the researcher's summary and key takeaways of what was said.

This allows the participant to either tell the researcher "Yes, you got it!" or "No, I actually meant something else." If they agree that their perspectives have been adequately represented and that the conclusions reached in the report are accurate to them, then the reader will be more convinced that the qualitative inquiry itself is credible.

However, because member checking might require participants to read, understand, and provide feedback on data and results, researchers may need to employ some creativity and empathy in how they go about doing member checks with diverse participants. Young children might not be able to read, second language learners may have difficulty understanding reports, and the lay public may not understand technical terms. In such situations, the researcher might find alternative ways to share what they are concluding in understandable ways, perhaps simplifying reports or reading segments...
to a participant and then relying on oral feedback and reactions.

**Transferability**

Because qualitative studies are not designed to be generalizable like quantitative studies, their results should never be framed as universal truths or as conclusions that are true in all contexts and settings. Attempting to generalize is a common mistake in qualitative research and can signal to your readers that you are not aware of the limits of your own methods and do not recognize the actual complexity of the phenomena you are studying. To avoid generalization, be sure that you are using softening words, like “may,” “suggests,” “perhaps,” and so forth, and that you are aware that what you see in your research setting may not always be true in other research settings (or even in the same setting at a different time).

Yet, given sufficient detail, qualitative studies can provide insight into what is happening in new contexts that you, as the researcher, may not be aware of. Transferability is the standard by which qualitative study results are expected to be able to be transferred or applied to new, novel contexts.

In other words, the qualitative researcher should consider whether their findings, which were discovered in one situated context, can apply to other contexts or settings as well (such as where the reader is working). Whether findings can be transferred or not is an empirical question, which cannot be answered by the researcher alone, because the reader’s context must be compared to the research context to identify similarities. The more similar, the more likely it is that the findings will be transferable. Thus, readers must be the ones to determine whether the qualitative inquiry is transferable, not the researcher.

The researcher is expected to facilitate transferability, however, by providing clear descriptions of the time and context in which results and conclusions are developed, providing thick descriptions of the phenomena under study, and providing as much explanation about the context in which the study took place as possible. In short, more details give readers more power to discern which results might transfer to their contexts and which might not, and the rigorous qualitative researcher provides readers with sufficient detail to determine for themselves whether study results will transfer to their unique contexts.

**Learning Check**

How is transferability different from generalizability?

- a. They are synonyms or have the same meaning.
- b. Transferability is contextual, whereas generalizability is universal (to a population).
- c. Transferability requires significance testing, whereas generalizability requires proper sampling.
- d. Transferability is the first step toward generalizability and can become generalizability if done properly and often enough.

**Dependability**

Dependability is the standard by which the logic, reasoning, methods, and results are expected to be
stable or consistent over time. To check the dependability of a qualitative study, one looks to see if the researcher has been careless or made mistakes in conceptualizing the study, collecting the data, interpreting the findings and reporting results. The logic used for selecting people and events to observe, interview, and include in the study should be clearly presented. The more consistent the researcher has been in this research process, the more dependable are the results.

For instance, a study that was attempting to understand African American students' experiences in an inner-city school but then shifted to interviewing white students, rural students, etc. would have deviated from the established reasoning and methods proposed in the study. Such deviations often occur out of convenience to the researcher (e.g., a target population is no longer available for study), but they represent a serious threat to dependability.

A major technique for assessing dependability is a dependability audit in which an independent auditor reviews the activities of the researcher (as recorded in an audit trail in field notes, archives, and reports) to see how well the techniques for meeting the credibility and transferability standards have been followed. If the researcher does not maintain any kind of audit trail, then the dependability cannot be assessed, thereby diminishing it along with overall trustworthiness.

**Confirmability**

Confirmability is the standard by which a qualitative study is expected to be supported by informants (participants) who are involved in the study and by events that are independent of the researcher. Reference to literature and findings by other authors that confirm the researcher's interpretations can strengthen confirmability of the study in addition to information and interpretations by people other than the researcher from within the inquiry site itself.

For instance, if a researcher studied a few women's experiences in computer science and found that they felt empowered and treated equally to their male counterparts, but many external examples of harassment and mistreatment of women were arising both in the literature and mainstream news, then this would lead the reader to wonder whether the researcher's findings actually represented the real experiences of women in computer science or were merely an anomaly or a misinterpretation. In this case, it seems possible that a researcher was misinterpreting the experiences of women they were interviewing or simply didn't interview enough women to understand the issue fully.

This does not mean that qualitative research results must always agree with all other sources of information, but it does mean that there should be ways to confirm research results, either by reviewing data sources (such as transcripts), repeating the study in different contexts, or comparing results to other evidence.

To do this, a confirmability audit can be conducted at the same time as the dependability audit, as the auditor asks if the data and interpretations made by the researcher are supported by material in the audit trail, are internally coherent, and represent more than "figments of the researcher's imagination" (Guba & Lincoln, 1989, p. 243). If such an audit attests to the confirmability of the study, it is more likely to be accepted by readers.

**Other Criteria**

In addition to the standards discussed above, several other important considerations are suggested in
the literature, including meaningfulness, appropriateness, natural conditions, ethical treatment, and audit trails.

**Meaningfulness**

Meaningfulness is the expectation that a study will address a worthwhile problem or issue, and if it doesn't, then it is not worth doing. This holds true for all research, not just qualitative inquiry. There should be a rationale providing justification for the time, money, and other resources devoted to the study. Deciding whether a problem is meaningful or not is a subjective determination, but the researcher can provide evidence and logic to support his or her decision, which will allow the reader to make an informed decision as to whether the study merits attention.

** Appropriateness**

Appropriateness is the expectation that a study's methods align with its intended goals. Not all research is or should be qualitative. If the needs call for it and the researcher can justify the application of a qualitative approach, then qualitative methods can be reasonably used. If the goal or need is something else, such as generalizability or design of an intervention, then qualitative methods alone are likely not appropriate.

The danger here is that researchers might approach qualitative research with a means-oriented mindset, wherein they apply qualitative methodologies inappropriately to problems that they are not equipped to solve, thereby overstepping the limits of the paradigm and failing the incommensurability test.

Studies should, therefore, provide a rationale to readers that both the goal of the research study is meaningful (i.e., meaningful ends) and also that qualitative methods are the right or best way of achieving the intended goal (i.e., appropriate means).

**Natural Conditions**

The expectation of natural conditions should be conducted under the most natural conditions possible. Manipulation of the participants through random assignment, submission to unnatural measurement instruments, or exposure to unnatural treatments should be avoided. The researcher should be as unobtrusive as possible so participants are acting essentially as they would if the researcher were simply another participant in the setting and not also conducting inquiry.

**Ethical Treatment**

Though all research should follow ethical guidelines, qualitative research places an especially high emphasis on valuing participant self-determination and social and psychological wellbeing. This may mean that practices sometimes used in other research projects, such as deception, may not be appropriate in qualitative settings and indeed may reduce the legitimacy of the qualitative approach.

This generally means that participants should be given the opportunity to react to the data record and have their disagreements with the researcher's interpretations taken seriously. Participants should also be given anonymity in any reports, and there should be no indications that participants were treated with disrespect or cruelty.
Audit Trail

An audit trail is simply the records kept of how a qualitative study was conducted. The audit trail should include all field notes and any other records kept of what the researcher does, sees, hears, thinks, etc. These notes describe the researcher's evolving relationship to what they are observing and what is being learned, and they also describe the researcher's thoughts about how to proceed with the study, sampling decisions, ethical concerns, and so on.

Each researcher is free to create a unique audit trail that fits the study being conducted, and the audit trail may be used as a reference throughout the study to review what has been done and to consider alternative plans, in addition to serving as part of the dependability and confirmability audits described above. Often, audit trails and field notes are the same, and if field notes are kept current and are easily accessible, no extra audit trail may be necessary (although some people like to keep a separate file for audit trail documentation).

To help an auditor, many researchers create a brief chronological index to their study. They list choices they made each day of the study, actions they engaged in, and some of their thoughts about how the study is going at each stage. The auditor can then go from this listing to the field notes, audio and video recordings, and other files associated with the inquiry to reconstruct how the study was conducted, to understand how conclusions were reached, and to make the dependability and confirmability judgments described earlier.

Signaling Rigor

In addition to following these standards, rigor must also be effectively communicated to the reader in order to serve its purpose. For this reason, reports should be well written to include description, analysis, and synthesis (cf., Wolcott, 1994), as well as to reveal the biases and assumptions of the researchers involved. Attempts to share what the researcher is learning should be communicated clearly. Descriptions should develop a sense of "being there" for the reader. Analyses should be logically presented. The audience for the report should be identified, and the report should address the concerns of that audience. The grammar and use of language should be of the highest quality.

Although necessary balances between description, analysis, and synthesis will vary depending on the length of the report and the purposes of the inquiry, readers need to have some raw description of scenes from the research site to use in judging the conclusions that are reached and to make their own conclusions independently. They also should see some synthesis of results by the researcher, in which all contradictions in findings are analyzed and/or resolved. Although there are paradoxes in the world, a report that presents conflicting pieces of evidence without discussing them and trying to discern their nature (i.e., whether it is a true paradox or whether one side of the issue is erroneous) needs to be improved.

Relevant characteristics of the researcher should also be clearly revealed so that the reader can understand the context from which the study emerged more completely. This may be done either explicitly in an appendix, in the foreword, or in the body of the text. Or it may be done implicitly in the text as the researcher describes his or her methods, decisions, reasons for doing the study, and so on.

Also, as researchers employ any of the techniques described above, they should explicitly state so in the methods sections of their reports and only mention techniques that they intentionally employed.
This should consist of more than a list but should include necessary descriptions, such as how member checking was conducted, who the peer debriefers were, what some examples of negative cases were, and so forth. By doing this, qualitative researchers can better establish the rigor of their work to discerning and critical readers and also legitimize their processes and results as being worthy of consideration.

References


3

Learning Theories
Appendix A

Supplements
Glossary
References
Book Authors
Dr. Royce Kimmons is an Associate Professor of Instructional Psychology and Technology at Brigham Young University where he studies digital participation divides specifically in the realms of social media, open education, and classroom technology use. More information about his work may be found at http://roycekimmons.com, and you may also dialogue with him on Twitter @roycekimmons.
https://edtechbooks.org/education_research

**CC BY**: This work is released under a CC BY license, which means that you are free to do with it as you please as long as you properly attribute it.