Research in Mathematics Education

Series Editors: Jinfa Cai · James A. Middleton

James Hiebert · Jinfa Cai ·
Stephen Hwang · Anne K Morris ·
Charles Hohensee

Doing Research: A New Researcher's Guide

Springer

Research in Mathematics Education

Series Editors
Jinfa Cai, Newark, DE, USA
James A. Middleton, Tempe, AZ, UK

This series is designed to produce thematic volumes, allowing researchers to access numerous studies on a theme in a single, peer-reviewed source. Our intent for this series is to publish the latest research in the field in a timely fashion. This design is particularly geared toward highlighting the work of promising graduate students and junior faculty working in conjunction with senior scholars. The audience for this monograph series consists of those in the intersection between researchers and mathematics education leaders—people who need the highest quality research, methodological rigor, and potentially transformative implications ready at hand to help them make decisions regarding the improvement of teaching, learning, policy, and practice. With this vision, our mission of this book series is: (1) To support the sharing of critical research findings among members of the mathematics education community; (2) To support graduate students and junior faculty and induct them into the research community by pairing them with senior faculty in the production of the highest quality peer-reviewed research papers; and (3) To support the usefulness and widespread adoption of research-based innovation.

James Hiebert • Jinfa Cai • Stephen Hwang
Anne K Morris • Charles Hohensee

Doing Research: A New Researcher's Guide

Springer

James Hiebert
School of Education
University of Delaware
Newark, DE, USA

Jinfa Cai
Department of Mathematical Sciences
University of Delaware
Newark, DE, USA

Stephen Hwang
Department of Mathematical Sciences
University of Delaware
Newark, DE, USA

Anne K Morris
School of Education
University of Delaware
Newark, DE, USA

Charles Hohensee
School of Education
University of Delaware
Newark, DE, USA

ISSN 2570-4729 ISSN 2570-4737
Research in Mathematics Education
ISBN 978-3-031-19077-3 ISBN 978-3-031-19078-0 (eBook)
https://doi.org/10.1007/978-3-031-19078-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is a practical guide for doing research written for beginning researchers. You might be a doctoral student just learning how to conduct research. Or, you might be a new faculty member, having recently graduated from a doctoral program and looking forward to conducting research as part of your job. Or, you might be a postdoctoral fellow pursuing a new research agenda.

Although the target audience is beginning researchers, we believe more experienced researchers will find something useful as well. Indeed, we were surprised by how much *we* learned writing this book. As we were searching for ways to present the ideas, we realized that many of the concepts and procedures we describe are less about learning a set of rules and more about reasoning through a complicated but logical process. And getting better at reasoning is a lifelong journey. So, if you are a beginning researcher or someone more experienced, we hope you will learn something new and reflect on your own reasoning about the research process.

We are mathematics educators, so we use examples from mathematics teaching and learning, but we believe the practical guidelines we recommend apply to most content areas and domains of educational research. There is nothing special about mathematics with regard to the logic of the research process we describe.

The book can be used as a supplemental textbook in a graduate-level research methods course, as a primary textbook in a course on conducting research in (mathematics) education, and as a self-study guide for individuals or small groups of researchers who want to review and reconsider the key elements of the research process. The book includes exercises designed to engage you in working through the processes we describe. The narrative is punctuated with tips for researchers and recommendations for other sources to consult along the way.

We encourage you to work through the study exercises. You will learn more than you would otherwise. In the exercises, you will be asked to respond to specific questions designed to help you make progress in planning and conducting your study, and in writing your report of the study. For some questions, we will ask you to update your responses as you read succeeding chapters. You will be asked to try out specific aspects of conducting your study as well as writing a paper based on your study, and to write out examples to illustrate and sharpen the ideas presented in the text.

Many books on research methods are created as reference books; you can learn about a particular aspect of research by selecting and reading particular pages. This book is different. It is more like a story that develops as each idea builds on what came before. Individual chapters or sections will not make much sense unless the book is read as a whole, from the beginning to the end.

We treat conducting a research study and writing a research report as mutually beneficial processes. Consequently, the book includes suggestions for writing alongside doing research. Writing can be used to help you think through research issues, make more explicit your own thinking, and push your thinking forward. Writing also is used to communicate your research study to others. We address writing for both purposes.

One advantage we had writing this book came from our experience as a former editorial team for one of the leading journals in mathematics education—the *Journal for Research in Mathematics Education*. A major benefit of this experience is knowing what reviewers are likely to say. You can expect to receive the kind of feedback we examined regardless of which journal is reviewing your paper. Which features of papers do reviewers like and which features cause problems? At key points in the book, we cite information from reviewers to give you an inside look at what is expected if you want to get your work published.

We intend this book to be easy to read but challenging to complete. We believe the basic principles for conducting research are logical and quite easy to understand, but we also know that doing research can be difficult and even intimidating, especially in enormously complex fields like education. We hope you find this book useful in managing this complexity and helpful in guiding you through the challenging but exciting and rewarding work of educational research.

Newark, DE, USA

James Hiebert
Jinfa Cai
Stephen Hwang
Anne K. Morris
Charles Hohensee

Acknowledgments

We want to acknowledge the origins of the ideas we present. Some of the key ideas germinated during our discussions as we prepared four editorials for the *Journal for Research in Mathematics Education* (Issues 2–5, Vol. 50, 2019). As new associate editors, Michelle Cirillo and Steven Kramer participated in these discussions and helped compose the editorials. Victoria Robison served as the staff assistant for JRME and not only contributed to the analysis of reviews but also proofread all of the editorials. We want to credit their contributions. But they should not be blamed for the ways in which the ideas evolved as we drafted this book.

We would like to thank the beginning researchers who participated in two conferences, supported by the National Science Foundation, on doing research in mathematics education, one online and one in-person. During the conferences, we presented some of the key ideas contained in earlier drafts of this book. The insightful questions and comments of the participants during these conferences helped us refine the ideas in the final version.

We also want to thank several colleagues who graciously volunteered to read an earlier draft of this book and offered thoughtful comments and suggestions. We gratefully acknowledge their extraordinarily helpful advice and their unusual generosity. Our sincere thanks to Elena Bregier, Ronald Gallimore, Cynthia Langrall, and Benjamin Rott.

This work was supported by the National Science Foundation under Grant No. 1941494. Dr. Finbarr Sloane served as Program Director for this grant. Special thanks go to Dr. Sloane for his strong support in the process of preparing this book. However, the opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the National Science Foundation.

About the Authors

James Hiebert is the Robert J. Barkley Professor of Education at the University of Delaware, and a Fellow of the American Educational Research Association. He co-authored the books *The Teaching Gap, and Making Sense: Teaching and Learning Mathematics with Understanding*, and edited the books *Conceptual and Procedural Knowledge*: *The Case of Mathematics*, and *Teachers, Teaching, and Reform: Perspectives on Efforts to Improve Educational Outcomes*. He has served as PI or co-PI on 13 grants funded by the National Science Foundation or the U.S. Department of Education.

Jinfa Cai is the Kathleen and David Hollowell Professor of Mathematics Education at the University of Delaware, and a Fellow of the American Educational Research Association. He served as the Editor-in-Chief for the *Journal for Research in Mathematics Education* (2017-2020). He is interested in how students learn mathematics and pose/solve problems as well as how teachers can provide and create learning environments so that students can make sense of mathematics. He has served as Program Director at the National Science Foundation. He was the Editor of the 2017 *Compendium for Research in Mathematics Education* (published by NCTM).

Stephen Hwang is a researcher at the University of Delaware and is the Project Coordinator for the NSF-funded project, Supporting Teachers to Teach Mathematics Through Problem Posing. He served as the Assistant Editor for the *Journal for Research in Mathematics Education* (2017-2020). He is interested in mathematical problem posing and how teachers can learn to use problem posing to provide robust learning opportunities for their students. He is also interested in the processes of mathematical proof and justification and the development of mathematical habits of mind.

Anne K. Morris is a professor at the University of Delaware. She served as an Associate Editor for the *Journal for Research in Mathematics Education* (2017-2020). She is interested in how educators can store knowledge about teaching in instructional materials that are readily useable by teachers so that improvements accumulate over time. She is also interested in children's and preservice teachers' understanding of deductive argument and how to develop it through classroom instruction.

Charles Hohensee is an Associate Professor at the University of Delaware. He served as an Associate Editor for the *Journal for Research in Mathematics Education* (2017-2020). He is interested in backward transfer, which is how students' ways of reasoning about a previously-encountered mathematics concept changes when they engage in learning experiences about a new mathematics concept.

Contents

Prologue

Martha, Sam, Adrian, and Corin are (fictitious) beginning educational researchers you will meet in this book. Martha and Sam are doctoral students; Adrian and Corin are new assistant professors. We use them to illustrate the challenges common to all beginning researchers. Actually, these challenges are common to *all* educational researchers. They want to solve educational problems in order for students to, ultimately, have richer learning experiences. As you will see, they struggle with how to describe the problems so they can investigate possible solutions in a credible way.

Our purpose for writing this book is to help researchers like Martha, Sam, Adrian, and Corin conduct meaningful investigations of important educational problems. We draw on our experience as researchers to share what we wish we had known when we began conducting research. Consequently, what we describe is not based on what we have done but on what we have learned during our research careers.

Our overriding thesis is that educational research requires being as clear, explicit, precise, honest, and transparent as possible about what you want to know. What claims would you most like to make at the end of your study? What evidence will you need to make these claims? How could you design your study to give you the best chance of gathering such evidence, if it exists? Our thesis is that approaching research in this way requires lots of preparation: immersing yourself in the literature, talking extensively with colleagues, conducting pilot studies, and thinking deeply about your anticipated study for extended periods of time. Only with this kind of preparation can you clearly, explicitly, and precisely answer the preceding questions before you conduct your study. And answering these questions before you conduct your study greatly increases your chances of producing high-quality research.

We have learned there is major payoff to doing the hard intellectual work required to conduct research before beginning your study rather than after you finish. Instead of saying, “I’m interested in this issue so I will gather some data to see what’s going on,” we recommend saying, “I’m interested in this issue so I will read as much as I can about what others have found, I will talk to my colleagues about what I am thinking, and, if needed, I will gather a small amount of exploratory data so I can anticipate what I might find and explain why I think so.”

What we have just described is not a new or even unusual way of thinking about research. In fact, we see it as the most common, longstanding approach to conducting research. We believe the process of anticipating claims and checking if they are correct is part of *scientific inquiry*, a research process used across all disciplines.

A common way of talking about anticipating and checking claims is formulating, testing, and revising hypotheses. The process of formulating, testing, and revising hypotheses is at the heart of scientific inquiry. This enactment of scientific inquiry will guide our descriptions of research and our suggestions about how to conduct research. It is important to know, from the beginning, that we define "hypothesis" as a potential explanation for something based on what is currently known but not yet proven, or as a tentative explanation for reported observations that can guide further investigations. We do *not* mean a null statistical hypothesis encountered in graduate statistics courses.

We believe scientific inquiry, defined in this way, is equally useful for quantitative, qualitative, and mixed methods studies. No matter the kind of data collected or how they are analyzed, what matters is the goal of trying to solve an educational problem in a credible way. Scientific inquiry, defined this way, is also equally useful for a range of research designs, from descriptive, to correlational, to experimental. Descriptive case studies can benefit from scientific inquiry—formulating, testing, and revising hypotheses—just as can large-sample experimental studies.

Beginning researchers like Martha, Sam, Adrian, and Corin are faced not only with conducting credible studies but also with writing publishable reports of their studies. As the former editorial team for the *Journal for Research in Mathematics Education*, we aim to help with writing about, as well as conducting, research. In our view, scientific inquiry is as helpful for writing your research paper as for conducting your study.

Although the process of formulating, testing, and revising hypotheses is a common research approach across most disciplines, it has not always been applied systematically to conduct research in education. Because the ramifications of applying this process are not always part of the graduate school curriculum, many aspects might be new to readers. Consequently, we structured this book to build the concepts in a carefully sequenced way, from the ground up.

The first chapter starts at the beginning by reconsidering what counts as research and why we do it. We then examine the concept of a hypothesis and how it can be used to fortify the entire research process. In the third chapter, we tackle the knotty problem of theoretical frameworks and describe how formulating hypotheses can clarify what these are and how to create them. We then use the ideas developed in the first three chapters to frame the selection of methods as a task of crafting methods that provide the best test of the researcher's hypotheses. Finally, we take on the common "So what?" question all researchers face and consider how revising hypotheses can address this difficult but essential question.

Each chapter focuses on a major phase in the scientific inquiry process. The phases are tightly connected to form a coherent, integrated whole, so each chapter depends on what comes before and influences what comes after. In fact, we had some difficulty figuring out where one chapter should end and the next one should begin. So, this book will make sense only by working through each chapter, in turn. However, you will need to come back to earlier chapters as you read later chapters to refresh your memory of earlier ideas and to make connections among the ideas.

To reiterate, our purpose for writing this book is to help researchers, especially beginning researchers like Martha, Sam, Adrian, and Corin, conduct important and credible research and to publish this research both to benefit the educational community and to support their own professional careers. We hope you enjoy reading it as much as we enjoyed writing it.

Chapter 1
What Is Research, and Why Do People Do It?

Part I. What Is Research?

Have you ever studied something carefully because you wanted to know more about it? Maybe you wanted to know more about your grandmother's life when she was younger so you asked her to tell you stories from her childhood, or maybe you wanted to know more about a fertilizer you were about to use in your garden so you read the ingredients on the package and looked them up online. According to the dictionary definition, you were doing research.

Recall your high school assignments asking you to "research" a topic. The assignment likely included consulting a variety of sources that discussed the topic, perhaps including some "original" sources. Often, the teacher referred to your product as a "research paper."

Were you conducting research when you interviewed your grandmother or wrote high school papers reviewing a particular topic? Our view is that you were engaged in part of the research process, but only a small part. In this book, we reserve the word "research" for what it means in the scientific world, that is, for scientific research or, more pointedly, for *scientific inquiry*.

Exercise 1.1
Before you read any further, write a definition of what you think scientific inquiry is. Keep it short—Two to three sentences. You will periodically update this definition as you read this chapter and the remainder of the book.

This book is about scientific inquiry—what it is and how to do it. For starters, scientific inquiry is a process, a particular way of finding out about something that involves a number of phases. Each phase of the process constitutes one aspect of scientific inquiry. You are doing scientific inquiry as you engage in each phase, but

J. Hiebert et al., *Doing Research: A New Researcher's Guide*,
Research in Mathematics Education,
https://doi.org/10.1007/978-3-031-19078-0_1

you have not done scientific inquiry until you complete the full process. Each phase is necessary but not sufficient.

In this chapter, we set the stage by defining scientific inquiry—describing what it is and what it is not—and by discussing what it is good for and why people do it. The remaining chapters build directly on the ideas presented in this chapter.

A first thing to know is that scientific inquiry is not all or nothing. "Scientificness" is a continuum. Inquiries can be more scientific or less scientific. What makes an inquiry more scientific? You might be surprised there is no universally agreed upon answer to this question. None of the descriptors we know of are sufficient by themselves to define scientific inquiry. But all of them give you a way of thinking about some aspects of the process of scientific inquiry. Each one gives you different insights.

In this book, we reserve the word "research" for what it means in the scientific world, that is, for scientific research, or, more pointedly, for scientific inquiry.

Exercise 1.2
As you read about each descriptor below, think about what would make an inquiry more or less scientific. If you think a descriptor is important, use it to revise your definition of scientific inquiry.

Creating an Image of Scientific Inquiry

We will present three descriptors of scientific inquiry. Each provides a different perspective and emphasizes a different aspect of scientific inquiry. We will draw on all three descriptors to compose our definition of scientific inquiry.

Descriptor 1. Experience Carefully Planned in Advance

Sir Ronald Fisher, often called the father of modern statistical design, once referred to research as "experience carefully planned in advance" (1935, p. 8). He said that humans are always learning from experience, from interacting with the world around them. Usually, this learning is haphazard rather than the result of a deliberate process carried out over an extended period of time. Research, Fisher said, was learning from experience, but experience carefully planned in advance.

This phrase can be fully appreciated by looking at each word. The fact that scientific inquiry is based on *experience* means that it is based on interacting with the world. These interactions could be thought of as the *stuff* of scientific inquiry.

In addition, it is not just any experience that counts. The experience must be *carefully planned*. The interactions with the world must be conducted with an explicit, describable purpose, and steps must be taken to make the intended learning as likely as possible. This planning is an integral part of scientific inquiry; it is not just a preparation phase. It is one of the things that distinguishes scientific inquiry from many everyday learning experiences. Finally, these steps must be taken beforehand and the purpose of the inquiry must be articulated *in advance* of the experience. Clearly, scientific inquiry does not happen by accident, by just stumbling into something. Stumbling into something unexpected and interesting can happen while engaged in scientific inquiry, but learning does not depend on it and serendipity does not make the inquiry scientific.

Descriptor 2. Observing Something and Trying to Explain Why It Is the Way It Is

When we were writing this chapter and googled "scientific inquiry," the first entry was: "Scientific inquiry refers to the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work." The emphasis is on studying, or observing, and then *explaining*. This descriptor takes the image of scientific inquiry beyond carefully planned experience and includes explaining what was experienced.

According to the Merriam-Webster dictionary, "explain" means "(a) to make known, (b) to make plain or understandable, (c) to give the reason or cause of, and (d) to show the logical development or relations of" (Merriam-Webster, n.d.). We will use all these definitions. Taken together, they suggest that to explain an observation means to *understand it* by finding reasons (or causes) for why it is as it is. In this sense of scientific inquiry, the following are synonyms: explaining why, understanding why, and reasoning about causes and effects. Our image of scientific inquiry now includes planning, observing, and explaining why.

> *Our image of scientific inquiry now includes planning, observing, and explaining why.*

We need to add a final note about this descriptor. We have phrased it in a way that suggests "observing something" means you are observing something in real time—observing the way things are or the way things are changing. This is often true. But, observing could mean observing data that already have been collected, maybe by someone else making the original observations (e.g., secondary analysis of NAEP data or analysis of existing video recordings of classroom instruction). We will address secondary analyses more fully in Chap. 4. For now, what is important is that the process requires explaining why the data look like they do.

We must note that for us, the term “data” is not limited to numerical or quantitative data such as test scores. Data can also take many nonquantitative forms, including written survey responses, interview transcripts, journal entries, video recordings of students, teachers, and classrooms, text messages, and so forth.

“Data” is not limited to numerical or quantitative data such as test scores. Data can also take many nonquantitative forms, including written survey responses, interview transcripts, journal entries, video recordings of students, teachers, and classrooms, text messages, and so forth.

Exercise 1.3

(a) What are the implications of the statement that just “observing” is not enough to count as scientific inquiry? Does this mean that a detailed description of a phenomenon is not scientific inquiry?
(b) Find sources that define research in education that differ with our position, that say description alone, without explanation, counts as scientific research. Identify the precise points where the opinions differ. What are the best arguments for each of the positions? Which do you prefer? Why?

Descriptor 3. Updating Everyone’s Thinking in Response to More and Better Information

This descriptor focuses on a third aspect of scientific inquiry: updating and advancing the field’s understanding of phenomena that are investigated. This descriptor foregrounds a powerful characteristic of scientific inquiry: the *reliability* (or trustworthiness) of what is learned and the *ultimate inevitability* of this learning to advance human understanding of phenomena. Humans might choose *not* to learn from scientific inquiry, but history suggests that scientific inquiry always has the potential to advance understanding and that, eventually, humans take advantage of these new understandings.

Before exploring these bold claims a bit further, note that this descriptor uses “information” in the same way the previous two descriptors used “experience” and “observations.” These are the stuff of scientific inquiry and we will use them often, sometimes interchangeably. Frequently, we will use the term “data” to stand for all these terms.

An overriding goal of scientific inquiry is for everyone to learn from what one scientist does. Much of this book is about the methods you need to use so others have faith in what you report and can learn the same things you learned. This aspect of scientific inquiry has many implications.

One implication is that scientific inquiry is not a private practice. It is a public practice available for others to see and learn from. Notice how different this is from everyday learning. When you happen to learn something from your everyday experience, often only you gain from the experience. The fact that research is a public practice means it is also a social one. It is best conducted by interacting with others along the way: soliciting feedback at each phase, taking opportunities to present work-in-progress, and benefitting from the advice of others.

A second implication is that you, as the researcher, must be committed to sharing what you are doing and what you are learning in an open and transparent way. This allows all phases of your work to be scrutinized and critiqued. This is what gives your work credibility. The reliability or trustworthiness of your findings depends on your colleagues recognizing that you have used all appropriate methods to maximize the chances that your claims are justified by the data.

A third implication of viewing scientific inquiry as a collective enterprise is the reverse of the second—you must be committed to *receiving* comments from others. You must treat your colleagues as fair and honest critics even though it might sometimes feel otherwise. You must appreciate their job, which is to remain skeptical while scrutinizing what you have done in considerable detail. To provide the best help to you, they must remain skeptical about your conclusions (when, for example, the data are difficult for them to interpret) until you offer a convincing logical argument based on the information you share. A rather harsh but good-to-remember statement of the role of your friendly critics was voiced by Karl Popper, a well-known twentieth century philosopher of science: ". . . if you are interested in the problem which I tried to solve by my tentative assertion, you may help me by criticizing it as severely as you can" (Popper, 1968, p. 27).

A final implication of this third descriptor is that, as someone engaged in scientific inquiry, you have no choice but to update your thinking when the data support a different conclusion. This applies to your own data as well as to those of others. When data clearly point to a specific claim, even one that is quite different than you expected, you must reconsider your position. If the outcome is replicated multiple times, you need to adjust your thinking accordingly. Scientific inquiry does not let you pick and choose which data to believe; it *mandates* that everyone update their thinking when the data warrant an update.

Doing Scientific Inquiry

We define scientific inquiry in an operational sense—what does it mean to *do* scientific inquiry? What kind of process would satisfy all three descriptors: carefully planning an experience in advance; observing and trying to explain what you see; and, contributing to updating everyone's thinking about an important phenomenon?

We define scientific inquiry as formulating, testing, and revising hypotheses about phenomena of interest.

Of course, we are not the only ones who define it in this way. The definition for the scientific method posted by the editors of Britannica is: “a researcher develops a hypothesis, tests it through various means, and then modifies the hypothesis on the basis of the outcome of the tests and experiments” (Britannica, n.d.).

We define scientific inquiry as formulating, testing, and revising hypotheses about phenomena of interest.

Notice how defining scientific inquiry this way satisfies each of the descriptors. “Carefully *planning* an experience in advance” is exactly what happens when formulating a hypothesis about a phenomenon of interest and thinking about how to test it. “*Observing* a phenomenon” occurs when testing a hypothesis, and “*explaining*” what is found is required when revising a hypothesis based on the data. Finally, “updating everyone’s thinking” comes from comparing publicly the original with the revised hypothesis.

Doing scientific inquiry, as we have defined it, underscores the value of accumulating knowledge rather than generating random bits of knowledge. Formulating, testing, and revising hypotheses is an ongoing process, with each revised hypothesis begging for another test, whether by the same researcher or by new researchers. The editors of Britannica signaled this cyclic process by adding the following phrase to their definition of the scientific method: “The modified hypothesis is then retested, further modified, and tested again.” Scientific inquiry creates a process that encourages each study to build on the studies that have gone before. Through collective engagement in this process of building study on top of study, the scientific community works together to update its thinking.

Before exploring more fully the meaning of “formulating, testing, and revising hypotheses,” we need to acknowledge that this is not the only way researchers define research. Some researchers prefer a less formal definition, one that includes more serendipity, less planning, less explanation. You might have come across more open definitions such as “research is finding out about something.” We prefer the tighter hypothesis formulation, testing, and revision definition because we believe it provides a single, coherent map for conducting research that addresses many of the thorny problems educational researchers encounter. We believe it is the most useful orientation toward research and the most helpful to learn as a beginning researcher.

A final clarification of our definition is that it applies equally to qualitative and quantitative research. This is a familiar distinction in education that has generated much discussion. You might think our definition favors quantitative methods over qualitative methods because the language of hypothesis formulation and testing is often associated with quantitative methods. In fact, we do not favor one method over another. In Chap. 4, we will illustrate how our definition fits research using a range of quantitative and qualitative methods.

Exercise 1.4
Look for ways to extend what the field knows in an area that has already received attention by other researchers. Specifically, you can search for a program of research carried out by more experienced researchers that has some revised hypotheses that remain untested. Identify a revised hypothesis that you might like to test.

Unpacking the Terms Formulating, Testing, and Revising Hypotheses

To get a full sense of the definition of scientific inquiry we will use throughout this book, it is helpful to spend a little time with each of the key terms.

We first want to make clear that we use the term "hypothesis" as it is defined in most dictionaries and as it used in many scientific fields rather than as it is usually defined in educational statistics courses. By "hypothesis," we do *not* mean a null hypothesis that is accepted or rejected by statistical analysis. Rather, we use "hypothesis" in the sense conveyed by the following definitions: "An idea or explanation for something that is based on known facts but has not yet been proved" (Cambridge University Press, n.d.), and "An unproved theory, proposition, or supposition, tentatively accepted to explain certain facts and to provide a basis for further investigation or argument" (Agnes & Guralnik, 2008).

We distinguish two parts to "hypotheses." *Hypotheses consist of predictions and rationales*. Predictions are statements about what you expect to find when you inquire about something. Rationales are explanations for why you made the predictions you did, why you believe your predictions are correct. So, for us "formulating hypotheses" means making explicit predictions and developing rationales for the predictions.

"Testing hypotheses" means making observations that allow you to assess in what ways your predictions were correct and in what ways they were incorrect. In education research, it is rarely useful to think of your predictions as either right or wrong. Because of the complexity of most issues you will investigate, most predictions will be right in some ways and wrong in others.

By studying the observations you make (data you collect) to test your hypotheses, you can revise your hypotheses to better align with the observations. This means revising your predictions plus revising your rationales to justify your adjusted predictions. Even though you might not run another test, formulating revised hypotheses is an essential part of conducting a research study. Comparing your original and revised hypotheses informs everyone of what you learned by conducting your study. In addition, a revised hypothesis sets the stage for you or someone else to extend your study and accumulate more knowledge of the phenomenon.

We should note that not everyone makes a clear distinction between predictions and rationales as two aspects of hypotheses. In fact, common, non-scientific uses of the word "hypothesis" may limit it to only a prediction or only an explanation (or rationale). We choose to explicitly include both prediction and rationale in our definition of hypothesis, not because we assert this should be the universal definition, but because we want to foreground the importance of both parts acting in concert. Using "hypothesis" to represent both prediction and rationale could hide the two aspects, but we make them explicit because they provide different kinds of information. It is usually easier to make predictions than develop rationales because predictions can be guesses, hunches, or gut feelings about which you have little confidence. Developing a compelling rationale requires careful thought plus reading what other researchers have found plus talking with your colleagues. Often, while you are developing your rationale you will find good reasons to change your predictions. Developing good rationales is the engine that drives scientific inquiry. Rationales are essentially descriptions of how much you know about the phenomenon you are studying. Throughout this guide, we will elaborate on how developing good rationales drives scientific inquiry. For now, we simply note that it can sharpen your predictions and help you to interpret your data as you test your hypotheses.

We define a hypothesis to include both a prediction and a rationale. Both parts act in concert, and they provide different kinds of information. We discuss predictions in more detail in Chapter 2 and we detail how to build rationales in Chapter 3.

Hypotheses in education research take a variety of forms or types. This is because there are a variety of phenomena that can be investigated. Investigating educational phenomena is sometimes best done using qualitative methods, sometimes using quantitative methods, and most often using mixed methods (e.g., Hay, 2016; Weis et al. 2019a; Weisner, 2005). This means that, given our definition, hypotheses are equally applicable to qualitative and quantitative investigations.

Hypotheses take different forms when they are used to investigate different kinds of phenomena. Two very different activities in education could be labeled conducting experiments and descriptions. In an experiment, a hypothesis makes a prediction about anticipated changes, say the changes that occur when a treatment or intervention is applied. You might investigate how students' thinking changes during a particular kind of instruction.

A second type of hypothesis, relevant for descriptive research, makes a prediction about what you will find when you investigate and describe the nature of a situation. The goal is to understand a situation as it exists rather than to understand a change from one situation to another. In this case, your prediction is what you expect to observe. Your rationale is the set of reasons for making this prediction; it is your current explanation for why the situation will look like it does.

You will probably read, if you have not already, that some researchers say you do not need a prediction to conduct a descriptive study. We will discuss this point of view in Chap. 2. For now, we simply claim that scientific inquiry, as we have defined it, applies to all kinds of research studies. Descriptive studies, like others, not only benefit from formulating, testing, and revising hypotheses, but also *need* hypothesis formulating, testing, and revising.

One reason we define research as formulating, testing, and revising hypotheses is that if you think of research in this way you are less likely to go wrong. It is a useful guide for the entire process, as we will describe in detail in the chapters ahead. For example, as you build the rationale for your predictions, you are constructing the theoretical framework for your study (Chap. 3). As you work out the methods you will use to test your hypothesis, every decision you make will be based on asking, "Will this help me formulate or test or revise my hypothesis?" (Chap. 4). As you interpret the results of testing your predictions, you will compare them to what you predicted and examine the differences, focusing on how you must revise your hypotheses (Chap. 5). By anchoring the process to formulating, testing, and revising hypotheses, you will make smart decisions that yield a coherent and well-designed study.

Exercise 1.5
Compare the concept of formulating, testing, and revising hypotheses with the descriptions of scientific inquiry contained in *Scientific Research in Education* (NRC, 2002). How are they similar or different?

Exercise 1.6
Provide an example to illustrate and emphasize the differences between everyday learning/thinking and scientific inquiry.

Learning from Doing Scientific Inquiry

We noted earlier that a measure of what you have learned by conducting a research study is found in the differences between your original hypothesis and your revised hypothesis based on the data you collected to test your hypothesis. We will elaborate this statement in later chapters, but we preview our argument here.

Even before collecting data, scientific inquiry requires cycles of making a prediction, developing a rationale, refining your predictions, reading and studying more to strengthen your rationale, refining your predictions again, and so forth. And, even if you have run through several such cycles, you still will likely find that when you test your prediction you will be partly right and partly wrong. The results will support some parts of your predictions but not others, or the results will "kind of" support your predictions. A critical part of scientific inquiry is making sense of your results by interpreting them against your predictions. Carefully describing what aspects of your data supported your predictions, what aspects did not, and what data fell outside of any predictions is not an easy task, but you cannot learn from your study without doing this analysis.

Even before collecting data, scientific inquiry requires cycles of making a prediction, developing a rationale, refining your predictions, reading and studying more to strengthen your rationale, refining your predictions again, and so forth.

Analyzing the matches and mismatches between your predictions and your data allows you to formulate different rationales that would have accounted for more of the data. The best revised rationale is the one that accounts for the most data. Once you have revised your rationales, you can think about the predictions they best justify or explain. It is by comparing your original rationales to your new rationales that you can sort out what you learned from your study.

Suppose your study was an experiment. Maybe you were investigating the effects of a new instructional intervention on students' learning. Your original rationale was your explanation for why the intervention would change the learning outcomes in a particular way. Your revised rationale explained why the changes that you observed occurred like they did and why your revised predictions are better. Maybe your original rationale focused on the potential of the activities if they were implemented in ideal ways and your revised rationale included the factors that are likely to affect how teachers implement them. By comparing the before and after rationales, you are describing what you learned—what you can explain now that you could not before. Another way of saying this is that you are describing how much more you understand now than before you conducted your study.

Revised predictions based on carefully planned and collected data usually exhibit some of the following features compared with the originals: more precision, more completeness, and broader scope. Revised rationales have more explanatory power and become more complete, more aligned with the new predictions, sharper, and overall more convincing.

Part II. Why Do Educators Do Research?

Doing scientific inquiry is a lot of work. Each phase of the process takes time, and you will often cycle back to improve earlier phases as you engage in later phases. Because of the significant effort required, you should make sure your study is worth it. So, from the beginning, you should think about the purpose of your study. Why do you want to do it? And, because research is a social practice, you should also think about whether the results of your study are likely to be important and significant to the education community.

If you are doing research in the way we have described—as scientific inquiry—then one purpose of your study is to *understand*, not just to describe or evaluate or report. As we noted earlier, when you formulate hypotheses, you are developing rationales that explain why things might be like they are. In our view, trying to understand and explain is what separates research from other kinds of activities, like evaluating or describing.

One reason understanding is so important is that it allows researchers to see how or why something works like it does. When you see how something works, you are better able to predict how it might work in other contexts, under other conditions. And, because conditions, or contextual factors, matter a lot in education, gaining insights into applying your findings to other contexts increases the contributions of your work and its importance to the broader education community.

Consequently, the purposes of research studies in education often include the more specific aim of identifying and understanding the conditions under which the phenomena being studied work like the observations suggest. A classic example of this kind of study in mathematics education was reported by William Brownell and Harold Moser in 1949. They were trying to establish which method of subtracting whole numbers could be taught most effectively—the regrouping method or the equal additions method. However, they realized that effectiveness might depend on the conditions under which the methods were taught—"meaningfully" versus "mechanically." So, they designed a study that crossed the two instructional approaches with the two different methods (regrouping and equal additions). Among other results, they found that these conditions did matter. The regrouping method was more effective under the meaningful condition than the mechanical condition, but the same was not true for the equal additions algorithm.

What do education researchers want to understand? In our view, the ultimate goal of education is to offer all students the best possible learning opportunities. So, we believe the *ultimate* purpose of scientific inquiry in education is to develop understanding that supports the improvement of learning opportunities for all students. We say "ultimate" because there are lots of issues that must be understood to improve learning opportunities for all students. Hypotheses about many aspects of education are connected, ultimately, to students' learning. For example, formulating and testing a hypothesis that preservice teachers need to engage in particular kinds of activities in their coursework in order to teach particular topics well is, ultimately, connected to improving students' learning opportunities. So is hypothesizing that school districts often devote relatively few resources to instructional leadership training or hypothesizing that positioning mathematics as a tool students can use to combat social injustice can help students see the relevance of mathematics to their lives.

We do not exclude the importance of research on educational issues more removed from improving students' learning opportunities, but we do think the argument for their importance will be more difficult to make. If there is no way to imagine a connection between your hypothesis and improving learning opportunities for students, even a distant connection, we recommend you reconsider whether it is an important hypothesis within the education community.

Notice that we said the ultimate goal of education is to offer *all* students the best possible learning opportunities. For too long, educators have been satisfied with a goal of offering rich learning opportunities for lots of students, sometimes even for just the majority of students, but not necessarily for all students. Evaluations of success often are based on outcomes that show high averages. In other words, if many students have learned something, or even a smaller number have learned a lot, educators may have been satisfied. The problem is that there is usually a pattern in the groups of students who receive lower quality opportunities—students of color and students who live in poor areas, urban and rural. This is not acceptable. Consequently, we emphasize the premise that the purpose of education research is to offer rich learning opportunities to *all* students.

One way to make sure you will be able to convince others of the importance of your study is to consider investigating some aspect of teachers' shared instructional problems. Historically, researchers in education have set their own research agendas, regardless of the problems teachers are facing in schools. It is increasingly recognized that teachers have had trouble applying to their own classrooms what researchers find. To address this problem, a researcher could partner with a teacher—better yet, a small group of teachers—and talk with them about instructional problems they all share. These discussions can create a rich pool of problems researchers can consider. If researchers pursued one of these problems (preferably alongside teachers), the connection to improving learning opportunities for all students could be direct and immediate. "Grounding a research question in instructional problems that are experienced across multiple teachers' classrooms helps to ensure that the answer to the question will be of sufficient scope to be relevant and significant beyond the local context" (Cai et al., 2019b, p. 115).

As a beginning researcher, determining the relevance and importance of a research problem is especially challenging. We recommend talking with advisors, other experienced researchers, and peers to test the educational importance of possible research problems and topics of study. You will also learn much more about the issue of research importance when you read Chap. 5.

Exercise 1.7

Identify a problem in education that is closely connected to improving learning opportunities and a problem that has a less close connection. For each problem, write a brief argument (like a logical sequence of if-then statements) that connects the problem to all students' learning opportunities.

Part III. Conducting Research as a Practice of Failing Productively

Scientific inquiry involves formulating hypotheses about phenomena that are not fully understood—by you or anyone else. Even if you are able to inform your hypotheses with lots of knowledge that has already been accumulated, you are likely to find that your prediction is not entirely accurate. This is normal. Remember, scientific inquiry is a process of constantly updating your thinking. More and better information means revising your thinking, again, and again, and again. Because you never fully understand a complicated phenomenon and your hypotheses never produce completely accurate predictions, it is easy to believe you are somehow failing.

The trick is to fail upward, to fail to predict accurately in ways that inform your next hypothesis so you can make a better prediction. Some of the best-known researchers in education have been open and honest about the many times their predictions were wrong and, based on the results of their studies and those of others, they continuously updated their thinking and changed their hypotheses.

A striking example of publicly revising (actually reversing) hypotheses due to incorrect predictions is found in the work of Lee J. Cronbach, one of the most distinguished educational psychologists of the twentieth century. In 1955, Cronbach delivered his presidential address to the American Psychological Association. Titling it "Two Disciplines of Scientific Psychology," Cronbach proposed a rapprochement between two research approaches—correlational studies that focused on individual differences and experimental studies that focused on instructional treatments controlling for individual differences. (We will examine different research approaches in Chap. 4). If these approaches could be brought together, reasoned Cronbach (1957), researchers could find interactions between individual characteristics and treatments (aptitude-treatment interactions or ATIs), fitting the best treatments to different individuals.

In 1975, after years of research by many researchers looking for ATIs, Cronbach acknowledged the evidence for simple, useful ATIs had not been found. Even when trying to find interactions between a few variables that could provide instructional guidance, the analysis, said Cronbach, creates "a hall of mirrors that extends to infinity, tormenting even the boldest investigators and defeating even ambitious designs" (Cronbach, 1975, p. 119).

As he was reflecting back on his work, Cronbach (1986) recommended moving away from documenting instructional effects through statistical inference (an approach he had championed for much of his career) and toward approaches that probe the reasons for these effects, approaches that provide a "full account of events in a time, place, and context" (Cronbach, 1986, p. 104). This is a remarkable change in hypotheses, a change based on data and made fully transparent. Cronbach understood the value of failing productively.

Closer to home, in a less dramatic example, one of us began a line of scientific inquiry into how to prepare elementary preservice teachers to teach early algebra. Teaching early algebra meant engaging elementary students in early forms of algebraic reasoning. Such reasoning should help them transition from arithmetic to algebra. To begin this line of inquiry, a set of activities for preservice teachers were developed. Even though the activities were based on well-supported hypotheses, they largely failed to engage preservice teachers as predicted because of unanticipated challenges the preservice teachers faced. To capitalize on this failure, follow-up studies were conducted, first to better understand elementary preservice teachers' challenges with preparing to teach early algebra, and then to better support preservice teachers in navigating these challenges. In this example, the initial failure was a necessary step in the researchers' scientific inquiry and furthered the researchers' understanding of this issue.

We present another example of failing productively in Chap. 2. That example emerges from recounting the history of a well-known research program in mathematics education.

Making mistakes is an inherent part of doing scientific research. Conducting a study is rarely a smooth path from beginning to end. We recommend that you keep the following things in mind as you begin a career of conducting research in education.

First, do not get discouraged when you make mistakes; do not fall into the trap of feeling like you are not capable of doing research because you make too many errors.

Second, learn from your mistakes. Do not ignore your mistakes or treat them as errors that you simply need to forget and move past. Mistakes are rich sites for learning—in research just as in other fields of study.

Third, by reflecting on your mistakes, you can learn to make better mistakes, mistakes that inform you about a productive next step. You will not be able to eliminate your mistakes, but you can set a goal of making better and better mistakes.

Exercise 1.8
How does scientific inquiry differ from everyday learning in giving you the tools to fail upward? You may find helpful perspectives on this question in other resources on science and scientific inquiry (e.g., *Failure: Why Science is So Successful* by Firestein, 2015).

Exercise 1.9
Use what you have learned in this chapter to write a new definition of scientific inquiry. Compare this definition with the one you wrote before reading this chapter. If you are reading this book as part of a course, compare your definition with your colleagues' definitions. Develop a consensus definition with everyone in the course.

Part IV. Preview of Chap. 2

Now that you have a good idea of what research is, at least of what *we* believe research is, the next step is to think about how to actually begin doing research. This means how to begin formulating, testing, and revising hypotheses. As for all phases of scientific inquiry, there are lots of things to think about. Because it is critical to start well, we devote Chap. 2 to getting started with formulating hypotheses.

Chapter 2
How Do You Formulate (Important) Hypotheses?

Part I. Getting Started

We want to begin by addressing a question you might have had as you read the title of this chapter. You are likely to hear, or read in other sources, that the research process begins by asking *research questions*. For reasons we gave in Chap. 1, and more we will describe in this and later chapters, we emphasize formulating, testing, and revising hypotheses. However, it is important to know that asking and answering research questions involve many of the same activities, so we are not describing a completely different process.

We acknowledge that many researchers do not actually begin by formulating hypotheses. In other words, researchers rarely get a researchable idea by writing out a well-formulated hypothesis. Instead, their initial ideas for what they study come from a variety of sources. Then, *after* they have the idea for a study, they do lots of background reading and thinking and talking before they are ready to formulate a hypothesis. So, for readers who are at the very beginning and do not yet have an idea for a study, let's back up. Where do research ideas come from?

There are no formulas or algorithms that spawn a researchable idea. But as you begin the process, you can ask yourself some questions. Your answers to these questions can help you move forward.

1. What are you curious about? What are you passionate about? What have you wondered about as an educator? These are questions that look inward, questions about yourself.
2. What do you think are the most pressing educational problems? Which problems are you in the best position to address? What change(s) do you think would help all students learn more productively? These are questions that look outward, questions about phenomena you have observed.
3. What are the main areas of research in the field? What are the big questions that are being asked? These are questions about the general landscape of the field.

J. Hiebert et al., *Doing Research: A New Researcher's Guide*,
Research in Mathematics Education,
https://doi.org/10.1007/978-3-031-19078-0_2

4. What have you read about in the research literature that caught your attention? What have you read that prompted you to think about extending the profession's knowledge about this? What have you read that made you ask, "I wonder why this is true?" These are questions about how you can build on what is known in the field.
5. What are some research questions or testable hypotheses that have been identified by other researchers for future research? This, too, is a question about how you can build on what is known in the field. Taking up such questions or hypotheses can help by providing some existing scaffolding that others have constructed.
6. What research is being done by your immediate colleagues or your advisor that is of interest to you? These are questions about topics for which you will likely receive local support.

Exercise 2.1
Brainstorm some answers for each set of questions. Record them. Then step back and look at the places of intersection. Did you have similar answers across several questions? Write out, as clearly as you can, the topic that captures your primary interest, at least at this point. We will give you a chance to update your responses as you study this book.

Part II. Paths from a General Interest to an Informed Hypothesis

There are many different paths you might take from conceiving an idea for a study, maybe even a vague idea, to formulating a prediction that leads to an informed hypothesis that can be tested. We will explore some of the paths we recommend.

We will assume you have completed Exercise 2.1 in Part I and have some written answers to the six questions that preceded it as well as a statement that describes your topic of interest. This very first statement could take several different forms: a description of a problem you want to study, a question you want to address, or a hypothesis you want to test. We recommend that you begin with one of these three forms, the one that makes most sense to you. There is an advantage to using all three and flexibly choosing the one that is most meaningful at the time and for a particular study. You can then move from one to the other as you think more about your research study and you develop your initial idea. To get a sense of how the process might unfold, consider the following alternative paths.

Beginning with a Prediction If You Have One

Sometimes, when you notice an educational problem or have a question about an educational situation or phenomenon, you quickly have an idea that might help solve the problem or answer the question. Here are three examples.

You are a teacher, and you noticed a problem with the way the textbook presented two related concepts in two consecutive lessons. Almost as soon as you noticed the problem, it occurred to you that the two lessons could be taught more effectively in the reverse order. You *predicted* better outcomes if the order was reversed, and you even had a preliminary rationale for why this would be true.

You are a graduate student and you read that students often misunderstand a particular aspect of graphing linear functions. You *predicted* that, by listening to small groups of students working together, you could hear new details that would help you understand this misconception.

You are a curriculum supervisor and you observed sixth-grade classrooms where students were learning about decimal fractions. After talking with several experienced teachers, you *predicted* that beginning with percentages might be a good way to introduce students to decimal fractions.

We begin with the path of making predictions because we see the other two paths as leading into this one at some point in the process (see Fig. 2.1). Starting with this path does not mean you did not sense a problem you wanted to solve or a question you wanted to answer.

Notice that your predictions can come from a variety of sources—your own experience, reading, and talking with colleagues. Most likely, as you write out your predictions you also think about the educational problem for which your prediction is a potential solution. Writing a clear description of the problem will be useful as you proceed. Notice also that it is easy to change each of your predictions into a question. When you formulate a prediction, you are actually answering a question, even though the question might be implicit. Making that implicit question explicit can generate a first draft of the research question that accompanies your prediction.

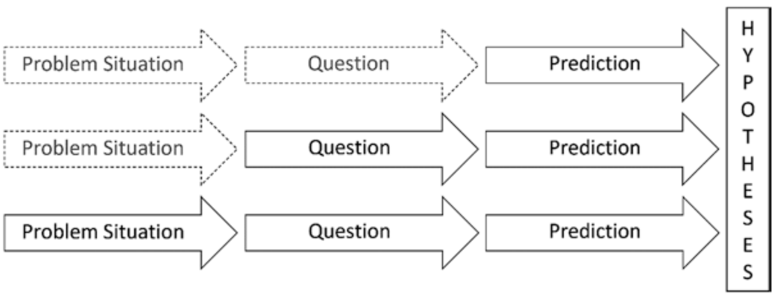

Fig. 2.1 Three Pathways to Formulating Informed Hypotheses

For example, suppose you are the curriculum supervisor who predicts that teaching percentages first would be a good way to introduce decimal fractions. In an obvious shift in form, you could ask, "In what ways would teaching percentages benefit students' initial learning of decimal fractions?"

The difference between a question and a prediction is that a question simply asks what you will find whereas a prediction also says what you expect to find.

There are advantages to starting with the prediction form if you can make an educated guess about what you will find. Making a prediction forces you to think *now* about several things you will need to think about at some point anyway. It is better to think about them earlier rather than later. If you state your prediction clearly and explicitly, you can begin to ask yourself three questions about your prediction: Why do I expect to observe what I am predicting? Why did I make that prediction? (These two questions essentially ask what your rationale is for your prediction.) And, how can I test to see if it's right? This is where the benefits of making predictions begin.

Asking yourself why you predicted what you did, and then asking yourself why you answered the first "why" question as you did, can be a powerful chain of thought that lays the groundwork for an increasingly accurate prediction and an increasingly well-reasoned rationale. For example, suppose you are the curriculum supervisor above who predicted that beginning by teaching percentages would be a good way to introduce students to decimal fractions. Why did you make this prediction? Maybe because students are familiar with percentages in everyday life so they could use what they know to anchor their thinking about hundredths. Why would that be helpful? Because if students could connect hundredths in percentage form with hundredths in decimal fraction form, they could bring their meaning of percentages into decimal fractions. But how would that help? If students understood that a decimal fraction like 0.35 meant 35 of 100, then they could use their understanding of hundredths to explore the meaning of tenths, thousandths, and so on. Why would that be useful? By continuing to ask yourself why you gave the previous answer, you can begin building your rationale and, as you build your rationale, you will find yourself revisiting your prediction, often making it more precise and explicit. If you were the curriculum supervisor and continued the reasoning in the previous sentences, you might elaborate your prediction by specifying the way in which percentages should be taught in order to have a positive effect on particular aspects of students' understanding of decimal fractions.

Developing a Rationale for Your Predictions

Keeping your initial predictions in mind, you can read what others already know about the phenomenon. Your reading can now become targeted with a clear purpose.

HELPFUL TIP

You can search for chapters or literature reviews related to your research topic in recent research handbooks and compendia or in journals. Reading these will help inform your predictions and provide helpful reference lists of other sources.

By reading and talking with colleagues, you can develop more complete reasons for your predictions. It is likely that you will also decide to revise your predictions based on what you learn from your reading. As you develop sound reasons for your predictions, you are creating your rationales, and your predictions together with your rationales become your hypotheses. The more you learn about what is already known about your research topic, the more refined will be your predictions and the clearer and more complete your rationales. We will use the term *more informed hypotheses* to describe this evolution of your hypotheses.

As you develop sound reasons for your predictions, you are creating your rationales, and your predictions together with your rationales become your hypotheses.

Developing more informed hypotheses is a good thing because it means: (1) you understand the reasons for your predictions; (2) you will be able to imagine how you can test your hypotheses; (3) you can more easily convince your colleagues that they are important hypotheses—they are hypotheses worth testing; and (4) at the end of your study, you will be able to more easily interpret the results of your test and to revise your hypotheses to demonstrate what you have learned by conducting the study.

Imagining Testing Your Hypotheses

Because we have tied together predictions and rationales to constitute hypotheses, testing hypotheses means testing predictions and rationales. Testing predictions means comparing empirical observations, or findings, with the predictions. Testing rationales means using these comparisons to evaluate the adequacy or soundness of the rationales.

Imagining how you might test your hypotheses does not mean working out the details for exactly how you would test them. Rather, it means thinking ahead about how you could do this. Recall the descriptor of scientific inquiry: "experience carefully planned in advance" (Fisher, 1935). Asking whether predictions are testable and whether rationales can be evaluated is simply planning in advance.

You might read that testing hypotheses means simply assessing whether predictions are correct or incorrect. In our view, it is more useful to think of testing as a means of gathering enough information to compare your findings with your predictions, revise your rationales, and propose more accurate predictions. So, asking yourself whether hypotheses can be tested means asking whether information could be collected to assess the accuracy of your predictions and whether the information will show you how to revise your rationales to sharpen your predictions.

Cycles of Building Rationales and Planning to Test Your Predictions

> Scientific reasoning is a dialogue between the possible and the actual, an interplay between hypotheses and the logical expectations they give rise to: there is a restless to-and-fro motion of thought, the formulation and rectification of hypotheses (Medawar, 1982, p.72).

As you ask yourself about how you could test your predictions, you will inevitably revise your rationales and sharpen your predictions. Your hypotheses will become more informed, more targeted, and more explicit. They will make clearer to you and others what, exactly, you plan to study.

When will you know that your hypotheses are clear and precise enough? Because of the way we define hypotheses, this question asks about both rationales and predictions. If a rationale you are building lets you make a number of quite different predictions that are equally plausible rather than a single, primary prediction, then your hypothesis needs further refinement by building a more complete and precise rationale. Also, if you cannot briefly describe to your colleagues a believable way to test your prediction, then you need to phrase it more clearly and precisely.

Each time you strengthen your rationales, you might need to adjust your predictions. And, each time you clarify your predictions, you might need to adjust your rationales. The cycle of going back and forth to keep your predictions and rationales tightly aligned has many payoffs down the road. Every decision you make from this point on will be in the interests of providing a transparent and convincing test of your hypotheses and explaining how the results of your test dictate specific revisions to your hypotheses. As you make these decisions (described in the succeeding chapters), you will probably return to clarify your hypotheses even further. But, you will be in a much better position, at each point, if you begin with well-informed hypotheses.

Beginning by Asking Questions to Clarify Your Interests

Instead of starting with predictions, a second path you might take devotes more time at the beginning to asking questions as you zero in on what you want to study. Some researchers suggest you start this way (e.g., Gournelos et al., 2019). Specifically, with this second path, the first statement you write to express your research interest would be a question. For example, you might ask, “Why do ninth-grade students change the way they think about linear equations after studying quadratic equations?” or “How do first graders solve simple arithmetic problems before they have been taught to add and subtract?”

The first phrasing of your question might be quite general or vague. As you think about your question and what you *really* want to know, you are likely to ask follow-up questions. These questions will almost always be more specific than your first question. The questions will also express more clearly what you want to know. So, the question “How do first graders solve simple arithmetic problems before they have been taught to add and subtract” might evolve into “Before first graders have been taught to solve arithmetic problems, what strategies do they use to solve arithmetic problems with sums and products below 20?” As you read and learn about what others already know about your questions, you will continually revise your questions toward clearer and more explicit and more precise versions that zero in on what you really want to know. The question above might become, “Before they are taught to solve arithmetic problems, what strategies do beginning first graders use to solve arithmetic problems with sums and products below 20 if they are read story problems and given physical counters to help them keep track of the quantities?”

Imagining Answers to Your Questions

If you monitor your own thinking as you ask questions, you are likely to begin forming some guesses about answers, even to the early versions of the questions. What do students learn about quadratic functions that influences changes in their proportional reasoning when dealing with linear functions? It could be that if you analyze the moments during instruction on quadratic equations that are extensions of the proportional reasoning involved in solving linear equations, there are times when students receive further experience reasoning proportionally. You might predict that these are the experiences that have a “backward transfer” effect (Hohensee, 2014).

These initial guesses about answers to your questions are your first predictions. The first predicted answers are likely to be hunches or fuzzy, vague guesses. This simply means you do not know very much yet about the question you are asking. Your first predictions, no matter how unfocused or tentative, represent the most you know at the time about the question you are asking. They help you gauge where you are in your thinking.

Shifting to the Hypothesis Formulation and Testing Path

Research questions can play an important role in the research process. They provide a succinct way of capturing your research interests and communicating them to others. When colleagues want to know about your work, they will often ask "What are your research questions?" It is good to have a ready answer.

However, research questions have limitations. They do not capture the three images of scientific inquiry presented in Chap. 1. Due, in part, to this less expansive depiction of the process, research questions do not take you very far. They do not provide a guide that leads you through the phases of conducting a study.

Consequently, when you can imagine an answer to your research question, we recommend that you move onto the hypothesis formulation and testing path. Imagining an answer to your question means you can make plausible predictions. You can now begin clarifying the reasons for your predictions and transform your early predictions into hypotheses (predictions along with rationales). We recommend you do this as soon as you have guesses about the answers to your questions because formulating, testing, and revising hypotheses offers a tool that puts you squarely on the path of scientific inquiry. It is a tool that can guide you through the *entire* process of conducting a research study.

This does not mean you are finished asking questions. Predictions are often created as answers to questions. So, we encourage you to continue asking questions to clarify what you want to know. But your target shifts from *only* asking questions to also proposing predictions for the answers and developing reasons the answers will be accurate predictions. It is by predicting answers, and explaining why you made those predictions, that you become engaged in scientific inquiry.

Cycles of Refining Questions and Predicting Answers

An example might provide a sense of how this process plays out. Suppose you are reading about Vygotsky's (1987) *zone of proximal development* (ZPD), and you realize this concept might help you understand why your high school students had trouble learning exponential functions. Maybe they were outside this zone when you tried to teach exponential functions. In order to recognize students who would benefit from instruction, you might ask, "How can I identify students who are within the ZPD around exponential functions?" What would you predict? Maybe students in this ZPD are those who already had knowledge of related functions. You could write out some reasons for this prediction, like "students who understand linear and quadratic functions are more likely to extend their knowledge to exponential functions." But what kind of data would you need to test this? What would count as "understanding"? Are linear and quadratic the functions you should assess? Even if they are, how could you tell whether students who scored well on tests of linear and quadratic functions were within the ZPD of exponential functions? How, in the end, would you measure what it means to be in this ZPD? So, asking a series of reasonable questions raised some red flags about the way your initial question was phrased, and you decide to revise it.

You set the stage for revising your question by defining ZPD as the zone within which students can solve an exponential function problem by making only one additional conceptual connection between what they already know and exponential functions. Your revised question is, “Based on students’ knowledge of linear and quadratic functions, which students are within the ZPD of exponential functions?” This time you know what kind of data you need: the number of conceptual connections students need to bridge from their knowledge of related functions to exponential functions. How can you collect these data? Would you need to see into the minds of the students? Or, are there ways to test the number of conceptual connections someone makes to move from one topic to another? Do methods exist for gathering these data? You decide this is not realistic, so you now have a choice: revise the question further or move your research in a different direction.

Notice that we do not use the term *research question* for all these early versions of questions that begin clarifying for yourself what you want to study. These early versions are too vague and general to be called research questions. *In this book, we save the term research question for a question that comes near the end of the work and captures exactly what you want to study*. By the time you are ready to specify a research question, you will be thinking about your study in terms of hypotheses and tests. When your hypotheses are in final form and include clear predictions about what you will find, it will be easy to state the research questions that accompany your predictions.

To reiterate one of the key points of this chapter: hypotheses carry much more information than research questions. Using our definition, hypotheses include predictions about what the answer might be to the question plus reasons for why you think so. Unlike research questions, hypotheses capture all three images of scientific inquiry presented in Chap. 1 (planning, observing and explaining, and revising one’s thinking). *Your hypotheses represent the most you know, at the moment, about your research topic.* The same cannot be said for research questions.

Beginning with a Research Problem

When you wrote answers to the six questions at the end of Part I of this chapter, you might have identified a research interest by stating it as a problem. This is the third path you might take to begin your research. Perhaps your description of your problem might look something like this: “When I tried to teach my middle school students by presenting them with a challenging problem without showing them how to solve similar problems, they didn’t exert much effort trying to find a solution but instead waited for me to show them how to solve the problem.” You do not have a specific question in mind, and you do not have an idea for why the problem exists, so you do not have a prediction about how to solve it. Writing a statement of this problem as clearly as possible could be the first step in your research journey.

As you think more about this problem, it will feel natural to ask questions about it. For example, why did some students show more initiative than others? What could I have done to get them started? How could I have encouraged the students to keep trying without giving away the solution? You are now on the path of asking questions—not research questions yet, but questions that are helping you focus your interest.

As you continue to think about these questions, reflect on your own experience, and read what others know about this problem, you will likely develop some guesses about the answers to the questions. They might be somewhat vague answers, and you might not have lots of confidence they are correct, but they are guesses that you can turn into predictions. Now you are on the hypothesis-formulation-and-testing path. This means you are on the path of asking yourself why you believe the predictions are correct, developing rationales for the predictions, asking what kinds of empirical observations would test your predictions, and refining your rationales and predictions as you read the literature and talk with colleagues.

A simple diagram that summarizes the three paths we have described is shown in Fig. 2.1. Each row of arrows represents one pathway for formulating an informed hypothesis. The dotted arrows in the first two rows represent parts of the pathways that a researcher may have implicitly travelled through already (without an intent to form a prediction) but that ultimately inform the researcher's development of a question or prediction.

Part III. One Researcher's Experience Launching a Scientific Inquiry

Martha was in her third year of her doctoral program and beginning to identify a topic for her dissertation. Based on (a) her experience as a high school mathematics teacher and a curriculum supervisor, (b) the reading she has done to this point, and (c) her conversations with her colleagues, she has developed an interest in what kinds of professional development experiences (let's call them learning opportunities [LOs] for teachers) are most effective. Where does she go from here?

Exercise 2.2
Before you continue reading, please write down some suggestions for Martha about where she should start.

A natural thing for Martha to do at this point is to ask herself some additional questions, questions that specify further what she wants to learn: What kinds of LOs do most teachers experience? How do these experiences change teachers' practices and beliefs? Are some LOs more effective than others? What makes them more effective?

To focus her questions and decide what she really wants to know, she continues reading but now targets her reading toward everything she can find that suggests possible answers to these questions. She also talks with her colleagues to get more ideas about possible answers to these or related questions. Over several weeks or months, she finds herself being drawn to questions about what makes LOs effective, especially for helping teachers teach more conceptually. She zeroes in on the question, "What makes LOs for teachers effective for improving their teaching for conceptual understanding?"

This question is more focused than her first questions, but it is still too general for Martha to define a research study. How does she know it is too general? She uses two criteria. First, she notices that the predictions she makes about the answers to the question are all over the place; they are not constrained by the reasons she has assembled for her predictions. One prediction is that LOs are more effective when they help teachers learn content. Martha makes this guess because previous research suggests that effective LOs for teachers include attention to content. But this rationale allows lots of different predictions. For example, LOs are more effective when they focus on the content teachers will teach; LOs are more effective when they focus on content beyond what teachers will teach so teachers see how their instruction fits with what their students will encounter later; and LOs are more effective when they are tailored to the level of content knowledge participants have when they begin the LOs. The rationale she can provide at this point does not point to a particular prediction.

A second measure Martha uses to decide her question is too general is that the predictions she can make regarding the answers seem very difficult to test. How could she test, for example, whether LOs should focus on content beyond what teachers will teach? What does "content beyond what teachers teach" mean? How could you tell whether teachers use their new knowledge of later content to inform their teaching?

Before anticipating what Martha's next question might be, it is important to pause and recognize how *predicting* the answers to her questions moved Martha into a new phase in the research process. As she makes predictions, works out the reasons for them, and imagines how she might test them, she is immersed in scientific inquiry. This intellectual work is the main engine that drives the research process. Also notice that revisions in the questions asked, the predictions made, and the rationales built represent the *updated thinking* (Chap. 1) that occurs as Martha continues to define her study.

Based on all these considerations and her continued reading, Martha revises the question again. The question now reads, "Do LOs that engage middle school mathematics teachers in studying mathematics content help teachers teach this same content with more of a conceptual emphasis?" Although she feels like the question is more specific, she realizes that the answer to the question is either "yes" or "no." This, by itself, is a red flag. Answers of "yes" or "no" would not contribute much to understanding the relationships between these LOs for teachers and changes in their teaching. Recall from Chap. 1 that *understanding* how things work, *explaining why* things work, is the goal of scientific inquiry.

Martha continues by trying to understand why she believes the answer is “yes.” When she tries to write out reasons for predicting “yes,” she realizes that her prediction depends on a variety of factors. If teachers already have deep knowledge of the content, the LOs might not affect them as much as other teachers. If the LOs do not help teachers develop their own conceptual understanding, they are not likely to change their teaching. By trying to build the rationale for her prediction—thus formulating a hypothesis—Martha realizes that the question still is not precise and clear enough.

Martha uses what she learned when developing the rationale and rephrases the question as follows: “*Under what conditions* do LOs that engage middle school mathematics teachers in studying mathematics content help teachers teach this same content with more of a conceptual emphasis?” Through several additional cycles of thinking through the rationale for her predictions and how she might test them, Martha specifies her question even further: “Under what conditions do middle school teachers who lack conceptual knowledge of linear functions benefit from LOs that engage them in conceptual learning of linear functions as assessed by changes in their teaching toward a more conceptual emphasis on linear functions?”

Each version of Martha’s question has become more specific. This has occurred as she has (a) identified a starting condition for the teachers—they lack conceptual knowledge of linear functions, (b) specified the mathematics content as linear functions, and (c) included a condition or purpose of the LO—it is aimed at conceptual learning.

Because of the way Martha’s question is now phrased, her predictions will require thinking about the conditions that could influence what teachers learn from the LOs and how this learning could affect their teaching. She might predict that if teachers engaged in LOs that extended over multiple sessions, they would develop deeper understanding which would, in turn, prompt changes in their teaching. Or she might predict that if the LOs included examples of how their conceptual learning could translate into different instructional activities for their students, teachers would be more likely to change their teaching. Reasons for these predictions would likely come from research about the effects of professional development on teachers’ practice.

As Martha thinks about testing her predictions, she realizes it will probably be easier to measure the conditions under which teachers are learning than the changes in the conceptual emphasis in their instruction. She makes a note to continue searching the literature for ways to measure the “conceptualness” of teaching.

As she refines her predictions and expresses her reasons for the predictions, she formulates a hypothesis (in this case several hypotheses) that will guide her research. As she makes predictions and develops the rationales for these predictions, she will probably continue revising her question. She might decide, for example, that she is not interested in studying the condition of different numbers of LO sessions and so decides to remove this condition from consideration by including in her question something like “. . . over five 2-hour sessions . . .”

At this point, Martha has developed a research question, articulated a number of predictions, and developed rationales for them. Her current question is: “Under what conditions do middle school teachers who lack conceptual knowledge of linear functions benefit from five 2-hour LO sessions that engage them in conceptual learning of linear functions as assessed by changes in their teaching toward a more conceptual emphasis on linear functions?" Her hypothesis is:

- **Prediction:** Participating teachers will show changes in their teaching with a greater emphasis on conceptual understanding, with larger changes on linear function topics directly addressed in the LOs than on other topics.
- **Brief Description of Rationale:** (1) Past research has shown correlations between teachers' specific mathematics knowledge of a topic and the quality of their teaching of that topic. This does not mean an increase in knowledge *causes* higher quality teaching but it allows for that possibility. (2) Transfer is usually difficult for teachers, but the examples developed during the LO sessions will help them use what they learned to teach for conceptual understanding. This is because the examples developed during the LO sessions are much like those that will be used by the teachers. So larger changes will be found when teachers are teaching the linear function topics addressed in the LOs.

Notice it is more straightforward to imagine how Martha could test this prediction because it is more precise than previous predictions. Notice also that by asking how to test a particular prediction, Martha will be faced with a decision about whether testing this prediction will tell her something she wants to learn. If not, she can return to the research question and consider how to specify it further and, perhaps, constrain further the conditions that could affect the data.

As Martha formulates her hypotheses and goes through multiple cycles of refining her question(s), articulating her predictions, and developing her rationales, she is constantly building the *theoretical framework* for her study. Because the theoretical framework is the topic for Chap. 3, we will pause here and pick up Martha's story in the next chapter. Spoiler alert: Martha's experience contains some surprising twists and turns.

Before leaving Martha, however, we point out two aspects of the process in which she has been engaged. First, it can be useful to think about the process as identifying (1) the *variables* targeted in her predictions, (2) the *mechanisms* she believes explain the relationships among the variables, and (3) the *definitions* of all the terms that are special to her educational problem. By variables, we mean things that can be measured and, when measured, can take on different values. In Martha's case, the variables are the conceptualness of teaching and the content topics addressed in the LOs. The mechanisms are cognitive processes that enable teachers to see the relevance of what they learn in PD to their own teaching and that enable the transfer of learning from one setting to another. Definitions are the precise descriptions of how the important ideas relevant to the research are conceptualized. In Martha's case, definitions must be provided for terms like conceptual understanding, linear functions, LOs, each of the topics related to linear functions, instructional setting, and knowledge transfer.

A second aspect of the process is a practice that Martha acquired as part of her graduate program, a practice that can go unnoticed. Martha writes out, in full sentences, her thinking as she wrestles with her research question, her predictions of the answers, and the rationales for her predictions. Writing is a tool for organizing thinking and we recommend you use it throughout the scientific inquiry process. We say more about this at the end of the chapter.

Here are the questions Martha wrote as she developed a clearer sense of what question she wanted to answer and what answer she predicted. The list shows the increasing refinement that occurred as she continued to read, think, talk, and write.

Early questions: What kinds of LOs do most teachers experience? How do these experiences change teachers' practices and beliefs? Are some LOs more effective than others? What makes them more effective?

First focused question: What makes LOs for teachers effective for improving their teaching for conceptual understanding?

Question after trying to predict the answer and imagining how to test the prediction: Do LOs that engage middle school mathematics teachers in studying mathematics content help teachers teach this same content with more of a conceptual emphasis?

Question after developing an initial rationale for her prediction: *Under what conditions* do LOs that engage middle school mathematics teachers in studying mathematics content help teachers teach this same content with more of a conceptual emphasis?

Question after developing a more precise prediction and richer rationale: Under what conditions do middle school teachers who lack conceptual knowledge of linear functions benefit from five 2-hour LO sessions that engage them in conceptual learning of linear functions as assessed by changes in their teaching toward a more conceptual emphasis on linear functions?

Part IV. An Illustrative Dialogue

The story of Martha described the major steps she took to refine her thinking. However, there is a lot of work that went on behind the scenes that wasn't part of the story. For example, Martha had conversations with fellow students and professors that sharpened her thinking. What do these conversations look like? Because they are such an important part of the inquiry process, it will be helpful to "listen in" on the kinds of conversations that students might have with their advisors.

Here is a dialogue between a beginning student, Sam (S), and their advisor, Dr. Avery (A). They are meeting to discuss data Sam collected for a course project. The dialogue below is happening very early on in Sam's conceptualization of the study, prior even to systematic reading of the literature.

S: Thanks for meeting with me today. As you know, I was able to collect some data for a course project a few weeks ago, but I'm having trouble analyzing the data, so I need your help. Let me try to explain the problem. As you know, I wanted to understand what middle-school teachers do to promote girls' achievement in a mathematics class. I conducted four observations in each of three teachers' classrooms. I also interviewed each teacher once about the four lessons I observed, and I interviewed two girls from each of the teachers' classes. Obviously, I have a ton of data. But when I look at all these data, I don't really know what I learned about my topic. When I was observing the teachers, I thought I might have observed some ways the teachers were promoting girls' achievement, but then I wasn't sure how to interpret my data. I didn't know if the things I was observing were actually promoting girls' achievement.

A: What were some of your observations?

S: Well, in a couple of my classroom observations, teachers called on girls to give an answer, even when the girls didn't have their hands up. I thought that this might be a way that teachers were promoting the girls' achievement. But then the girls didn't say anything about that when I interviewed them and also the teachers didn't do it in every class. So, it's hard to know what effect, if any, this might have had on their learning or their motivation to learn. I didn't want to ask the girls during the interview specifically about the teacher calling on them, and without the girls bringing it up themselves, I didn't know if it had any effect.

A: Well, why didn't you want to ask the girls about being called on?

S: Because I wanted to leave it as open as possible; I didn't want to influence what they were going to say. I didn't want to put words in their mouths. I wanted to know what *they* thought the teacher was doing that promoted their mathematical achievement and so I only asked the girls general questions, like "Do you think the teacher does things to promote girls' mathematical achievement?" and "Can you describe specific experiences you have had that you believe do and do not promote your mathematical achievement?"

A: So then, how did they answer those general questions?

S: Well, with very general answers, such as that the teacher knows their names, offers review sessions, grades their homework fairly, gives them opportunities to earn extra credit, lets them ask questions, and always answers their questions. Nothing specific that helps me know what teaching actions specifically target girls' mathematics achievement.

A: OK. Any ideas about what you might do next?

S: Well, I remember that when I was planning this data collection for my course, you suggested I might want to be more targeted and specific about what I was looking for. I can see now that more targeted questions would have made my data more interpretable in terms of connecting teaching actions to the mathematical achievement of girls. But I just didn't want to influence what the girls would say.

A: Yes, I remember when you were planning your course project, you wanted to keep it open. You didn't want to miss out on discovering something new and interesting. What do you think now about this issue?

S: Well, I still don't want to put words in their mouths. I want to know what *they* think. But I see that if I ask really open questions, I have no guarantee they will talk about what I want them to talk about. I guess I still like the idea of an open study, but I see that it's a risky approach. Leaving the questions too open meant I didn't constrain their responses and there were too many ways they could interpret and answer the questions. And there are too many ways I could interpret their responses.

******

By this point in the dialogue, Sam has realized that open data (i.e., data not testing a specific prediction) is difficult to interpret. In the next part, Dr. Avery explains why collecting open data was not helping Sam achieve goals for her study that had motivated collecting open data in the first place.

******

A: Yes, I totally agree. Even for an experienced researcher, it can be difficult to make sense of this kind of open, messy data. However, if you design a study with a more specific focus, you can create questions for participants that are more targeted because you will be interested in their answers to these specific questions. Let's reflect back on your data collection. What can you learn from it for the future?

S: When I think about it now, I realize that I didn't think about the distinction between all the different constructs at play in my study, and I didn't choose which one I was focusing on. One construct was the teaching moves that teachers think could be promoting achievement. Another is what teachers deliberately do to promote girls' mathematics achievement, if anything. Another was the teaching moves that actually do support girls' mathematics achievement. Another was what teachers were doing that supported *girls'* mathematics achievement versus the mathematics achievement of all students. Another was students' perception of what their teacher was doing to promote girls' mathematics achievement. I now see that any one of these constructs could have been the focus of a study and that I didn't really decide which of these was the focus of my course project prior to collecting data.

A: So, since you told me that the topic of this course project is probably what you'll eventually want to study for your dissertation, which of these constructs are you most interested in?

S: I think I'm more interested in the teacher moves that teachers deliberately do to promote girls' achievement. But I'm still worried about asking teachers directly and getting too specific about what they do because I don't want to bias what they will say. And I chose qualitative methods and an exploratory design because I thought it would allow for a more open approach, an approach that helps me see what's going on and that doesn't bias or predetermine the results.

A: Well, it seems to me you are conflating three issues. One issue is how to conduct an unbiased study. Another issue is how specific to make your study. And the third issue is whether or not to choose an exploratory or qualitative study design. Those three issues are not the same. For example, designing a study that's more open or more exploratory is not how researchers make studies fair and unbiased. In fact, it would be quite easy to create an *open* study that is biased. For example, you could ask very open questions and then interpret the responses in a way that unintentionally, and even unknowingly, aligns with what you were hoping the findings would say. Actually, you could argue that by adding more specificity and narrowing your focus, you're creating constraints that prevent bias. The same goes for an exploratory or qualitative study; they can be biased or unbiased. So, let's talk about what is meant by getting more specific. Within your new focus on what teachers deliberately do, there are many things that would be interesting to look at, such as teacher moves that address math anxiety, moves that allow girls to answer questions more frequently, moves that are specifically fitted to student thinking about specific mathematical content, and so on. What are one or two things that are most interesting to you? One way to answer this question is by thinking back to where your interest in this topic began.

******

In the preceding part of the dialogue, Dr. Avery explained how the goals Sam had for their study were not being met with open data. In the next part, Sam begins to articulate a prediction, which Sam and Dr. Avery then sharpen.

******

S: Actually, I became interested in this topic because of an experience I had in college when I was in a class of mostly girls. During whole class discussions, we were supposed to critically evaluate each other's mathematical thinking, but we were too polite to do that. Instead, we just praised each other's work. But it was so different in our small groups. It seemed easier to critique each other's thinking and to push each other to better solutions in small groups. I began wondering how to get girls to be more critical of each other's thinking in a whole class discussion in order to push *everyone's* thinking.

A: Okay, this is great information. Why not use this idea to zoom-in on a more manageable and interpretable study? You could look specifically at how teachers support girls in critically evaluating each other's thinking during whole class discussions. That would be a much more targeted and specific topic. Do you have predictions about what teachers could do in that situation, keeping in mind that you are looking specifically at girls' mathematical achievement, not students in general?

S: Well, what I noticed was that small groups provided more social and emotional support for girls, whereas the whole class discussion did not provide that same support. The girls felt more comfortable critiquing each other's thinking in small groups. **So, I guess I predict that when the social and emotional supports that are present in small groups are extended to the whole class discussion, girls would be more willing to evaluate each other's mathematical thinking critically during whole class discussion**. I guess ultimately, I'd like to know how the whole class discussion could be used to enhance, rather than undermine, the social and emotional support that is present in the small groups.

A: Okay, then where would you start? Would you start with a study of what the teachers say they will do during whole class discussion and then observe if that happens during whole class discussion?

S: But part of my prediction also involves the small groups. So, I'd also like to include small groups in my study if possible. If I focus on whole groups, I won't be exploring what I am interested in. My interest is broader than just the whole class discussion.

A: That makes sense, but there are many different things you could look at as part of your prediction, more than you can do in one study. For instance, if your prediction is that **when the social and emotional supports that are present in small groups are extended to whole class discussions, girls would be more willing to evaluate each other's mathematical thinking critically during whole class discussions**, then you could ask the following questions: What are the social and emotional supports that are present in small groups?; In which small groups do they exist?; Is it groups that are made up only of girls?; Does every small group do this, and for groups that do this, when do these supports get created?; What kinds of small group activities that teachers ask them to work on are associated with these supports?; Do the same social and emotional supports that apply to small groups even apply to whole group discussion?

S: All your questions make me realize that my prediction about extending social and emotional supports to whole class discussions first requires me to have a better understanding of the social and emotional supports that exist in small groups. In fact, I first need to find out whether those supports commonly exist in small groups or is that just my experience working in small groups. So, I think I will first have to figure out what small groups do to support each other and then, in a later study, I could ask a teacher to implement those supports during whole class discussions and find out how you can do that. Yeah, now I'm seeing that.

******

The previous part of the dialogue illustrates how continuing to ask questions about one's initial prediction is a good way to make it more and more precise (and researchable). In the next part, we see how developing a precise prediction has the added benefit of setting the researcher up for future studies.

******

A: Yes, I agree that for your first study, you should probably look at small groups. In other words, you should focus on only a part of your prediction for now, namely the part that says **there are social and emotional supports in small groups that support girls in critiquing each other's thinking**. That begins to sharpen the focus of your prediction, but you'll want to continue to refine it. For example, right now, the question that this prediction leads to is a question with a yes or no answer, but what you've said so far suggests to me that you are looking for more than that.

S: Yes, I want to know more than just whether there are supports. I'd like to know what kinds. That's why I wanted to do a qualitative study.

A: Okay, this aligns more with my thinking about research as being prediction driven. It's about collecting data that would help you revise your existing predictions into better ones. What I mean is that you would focus on collecting data that would allow you to refine your prediction, make it more nuanced, and go beyond what is already known. Does that make sense, and if so, what would that look like for your prediction?

S: Oh yes, I like that. I guess that would mean that, based on the data I collect for this next study, I could develop a more refined prediction that, for example, more specifically identifies and differentiates between different kinds of social and emotional supports that are present in small groups, or maybe that identifies the kinds of small groups that they occur in, or that predicts when and how frequently or infrequently they occur, or about the features of the small group tasks in which they occur, etc. I now realize that, although I chose qualitative research to make my study be more open, really the reason qualitative research fits my purposes is because it will allow me to explore fine-grained aspects of social and emotional supports that may exist for girls in small groups.

A: Yes, exactly! And then, based on the data you collect, you can include in your revised prediction those new fine-grained aspects. Furthermore, you will have a story to tell about your study in your written report, namely the story about your evolving prediction. In other words, your written report can largely tell how you filled out and refined your prediction as you learned more from carrying out the study. And even though you might not use them right away, you are also going to be able to develop new predictions that you would not have even thought of about social and emotional supports in small groups and your aim of extending them to whole-class discussions, had you not done this study. That will set you up to follow up on those new predictions in future studies. For example, you might have more refined ideas after you collect the data about the goals for critiquing student thinking in small groups versus the goals for critiquing student thinking during whole class discussion. You might even begin to think that some of the social and emotional supports you observe are not even replicable or even applicable to or appropriate for whole-class discussions, because the supports play different roles in different contexts. So, to summarize what I'm saying, what you look at in this study, even though it will be very focused, sets you up for a research program that will allow you to more fully investigate your broader interest in this topic, where each new study builds on your prior body of work. That's why it is so important to be explicit about the best place to start this research, so that you can build on it.

S: I see what you are saying. We started this conversation talking about my course project data. What I think I should have done was figure out explicitly what I needed to learn with that study with the intention of then taking what I learned and using it as the basis for the next study. I didn't do that, and so I didn't collect data that pushed forward my thinking in ways that would guide my next study. It would be as if I was starting over with my next study.

******

Sam and Dr. Avery have just explored how specifying a prediction reveals additional complexities that could become fodder for developing a systematic research program. Next, we watch Sam beginning to recognize the level of specificity required for a prediction to be testable.

******

A: One thing that would have really helped would have been if you had had a specific prediction going into your data collection for your course project.

S: Well, I didn't really have much of an explicit prediction in mind when I designed my methods.

A: Think back, you must have had some kind of prediction, even if it was implicit.

S: Well, yes, I guess I was predicting that teachers would enact moves that supported girls' mathematical achievement. And I observed classrooms to identify those teacher moves, I interviewed teachers to ask them about the moves I observed, and I interviewed students to see if they mentioned those moves as promoting their mathematical achievement. The goal of my course project was to identify teacher moves that support girls' mathematical achievement. And my specific research question was: What teacher moves support girls' mathematical achievement?

A: So, really you were asking the teacher and students to show and tell you what those moves are and the effects of those moves, as a result putting the onus on your participants to provide the answers to your research question for you. I have an idea, let's try a thought experiment. You come up with data collection methods for testing the prediction that **there are social and emotional supports in small groups that support girls in critiquing each other's thinking** that still puts the onus on the participants. And then I'll see if I can think of data collection methods that would not put the onus on the participants.

S: Hmm, well. .. I guess I could simply interview girls who participated in small groups and ask them "are there social and emotional supports that you use in small groups that support your group in critiquing each other's thinking and if so, what are they?" In that case, I would be putting the onus on them to be aware of the social dynamics of small groups and to have thought about these constructs as much as I have. Okay now can you continue the thought experiment? What might the data collection methods look like if I didn't put the onus on the participants?

A: First, I would pick a setting in which it was only girls at this point to reduce the number of variables. Then, personally I would want to observe a lot of groups of girls interacting in groups around tasks. I would be looking for instances when the conversation about students' ideas was shut down and instances when the conversation about students' ideas involved critiquing of ideas and building on each other's thinking. I would also look at what happened just before and during those instances, such as: did the student continue to talk after their thinking was critiqued, did other students do anything to encourage the student to build on their own thinking (i.e., constructive criticism) or how did they support or shut down continued participation. In fact, now that I think about it, "critiquing each other's thinking" can be defined in a number of different ways. I could mean just commenting on someone's thinking, judging correctness and incorrectness, constructive criticism that moves the thinking forward, etc. If you put the onus on the participants to answer your research question, you are stuck with their definition, and they won't have thought about this very much, if at all.

S: I think that what you are also saying is that my definitions would affect my data collection. If I think that critiquing each other's thinking means that the group moves their thinking forward toward more valid and complete mathematical solutions, then I'm going to focus on different moves than if I define it another way, such as just making a comment on each other's thinking and making each other feel comfortable enough to keep participating. In fact, am I going to look at individual instances of critiquing or look at entire sequences in which the critiquing leads to a goal? This seems like a unit of analysis question, and I would need to develop a more nuanced prediction that would make explicit what that unit of analysis is.

A: I agree, your definition of "critiquing each other's thinking" could entirely change what you are predicting. One prediction could be based on defining critiquing as a one-shot event in which someone makes one comment on another person's thinking. In this case the prediction would be that **there are social and emotional supports in small groups that support girls in making an evaluative comment on another student's thinking.** Another prediction could be based on defining critiquing as a back-and-forth process in which the thinking gets built on and refined. In that case, the prediction would be something like that **there are social and emotional supports in small groups that support girls in critiquing each other's thinking in ways that do not shut down the conversation but that lead to sustained conver-**

sations that move each other toward more valid and complete solutions.

S: Well, I think I am more interested in the second prediction because it is more compatible with my long-term interests, which are that I'm interested in extending small group supports to whole class discussions. The second prediction is more appropriate for eventually looking at girls in whole class discussion. During whole class discussion, the teacher tries to get a sustained conversation going that moves the students' thinking forward. So, if I learn about small group supports that **lead to sustained conversations that move each other toward more valid and complete solutions**, those supports might transfer to whole class discussions.

******

In the previous part of the dialogue, Dr. Avery and Sam showed how narrowing down a prediction to one that is testable requires making numerous important decisions, including how to define the constructs referred to in the prediction. In the final part of the dialogue, Dr. Avery and Sam begin to outline the reading Sam will have to do to develop a rationale for the specific prediction.

******

A: Do you see how your prediction and definitions are getting more and more specific? You now need to read extensively to further refine your prediction.

S: Well, I should probably read about micro dynamics of small group interactions, anything about interactions in small groups, and what is already known about small group interactions that support sustained conversations that move students' thinking toward more valid and complete solutions. I guess I could also look at research on whole-class discussion methods that support sustained conversations that move the class to more mathematically valid and complete solutions, because it might give me ideas for what to look for in the small groups. I might also need to focus on research about how learners develop understandings about a particular subject matter so that I know what "more valid and complete solutions" look like. I also need to read about social and emotional supports but focus on how they support students cognitively, rather than in other ways.

A: Sounds good, let's get together after you have processed some of this literature and we can talk about refining your prediction based on what you read and also the methods that will best suit testing that prediction.

S: Great! Thanks for meeting with me. I feel like I have a much better set of tools that push my own thinking forward and allow me to target something specific that will lead to more interpretable data.

Part V. Is It Always Possible to Formulate Hypotheses?

In Chap. 1, we noted you are likely to read that research does not require formulating hypotheses. Some sources describe doing research without making predictions and developing rationales for these predictions. Some researchers say you cannot always make predictions—you do not know enough about the situation. In fact, some argue for the value of *not* making predictions (e.g., Glaser & Holton, 2004; Merton, 1968; Nemirovsky, 2011). These are important points of view, so we will devote this section to discussing them.

Can You Always Predict What You Will Find?

One reason some researchers say you do not need to make predictions is that it can be difficult to imagine what you will find. This argument comes up most often for descriptive studies. Suppose you want to describe the nature of a situation you do not know much about. Can you still make a prediction about what you will find? We believe that, although you do not know exactly what you will find, you probably have a hunch or, at a minimum, a very fuzzy idea. It would be unusual to ask a question about a situation you want to know about without at least a fuzzy inkling of what you might find. The original question just would not occur to you. We acknowledge you might have only a vague idea of what you will find and you might not have much confidence in your prediction. However, we expect if you monitor your own thinking you will discover you have developed a suspicion along the way, regardless how vague the suspicion might be. Through the cyclic process we discussed above, that suspicion or hunch gradually evolves and turns into a prediction.

The Benefits of Making Predictions Even When They Are Wrong: An Example from the 1970s

One of us was a graduate student at the University of Wisconsin in the late 1970s, assigned as a research assistant to a project that was investigating young children's thinking about simple arithmetic. A new curriculum was being written, and the developers wanted to know how to introduce the earliest concepts and skills to kindergarten and first-grade children. The directors of the project did not know what to expect because, at the time, there was little research on five- and six-year-olds' pre-instruction strategies for adding and subtracting.

After consulting what literature was available, talking with teachers, analyzing the nature of different types of addition and subtraction problems, and debating with each other, the research team formulated some hypotheses about children's performance. Following the usual assumptions at the time and recognizing the new curriculum would introduce the concepts, the researchers predicted that, before instruction, most children would not be able to solve the problems. Based on the rationale that some young children did not yet recognize the simple form for written problems (e.g., 5 + 3 = ___), the researchers predicted that the best chance for success would be to read problems as stories (e.g., Jesse had 5 apples and then found 3 more. How many does she have now?). They reasoned that, even though children would have difficulty on all the problems, some story problems would be easier because the semantic structure is easier to follow. For example, they predicted the above story about adding 3 apples to 5 would be easier than a problem like, "Jesse had some apples in the refrigerator. She put in 2 more and now has 6. How many were in the refrigerator at the beginning?" Based on the rationale that children would need to count to solve the problems and that it can be difficult to keep track of the numbers, they predicted children would be more successful if they were given counters. Finally, accepting the common reasoning that larger numbers are more difficult than smaller numbers, they predicted children would be more successful if all the numbers in a problem were below 10.

Although these predictions were not very precise and the rationales were not strongly convincing, these hypotheses prompted the researchers to design the study to test their predictions. This meant they would collect data by presenting a variety of problems under a variety of conditions. Because the goal was to describe children's thinking, problems were presented to students in individual interviews. Problems with different semantic structures were included, counters were available for some problems but not others, and some problems had sums to 9 whereas others had sums to 20 or more.

The punchline of this story is that gathering data under these conditions, prompted by the predictions, made all the difference in what the researchers learned. Contrary to predictions, children could solve addition and subtraction problems before instruction. Counters were important because almost all the solution strategies were based on counting which meant that memory was an issue because many strategies require counting in two ways simultaneously. For example, subtracting 4 from 7 was usually solved by counting down from 7 while counting up from 1 to 4 to keep track of counting down. Because children acted out the stories with their counters, the semantic structure of the story was also important. Stories that were easier to read and write were also easier to solve.

To make a very long story very short, other researchers were, at about the same time, reporting similar results about children's pre-instruction arithmetic capabilities. A clear pattern emerged regarding the relative difficulty of different problem types (semantic structures) and the strategies children used to solve each type. As the data were replicated, the researchers recognized that kindergarten and first-grade teachers could make good use of this information when they introduced simple arithmetic. This is how *Cognitively Guided Instruction* (CGI) was born (Carpenter et al., 1989; Fennema et al., 1996).

To reiterate, the point of this example is that the study conducted to describe children's thinking would have looked quite different if the researchers had made no predictions. They would have had no reason to choose the particular problems and present them under different conditions. The fact that some of the predictions were completely wrong is not the point. The predictions created the conditions under which the predictions were tested which, in turn, created learning opportunities for the researchers that would not have existed without the predictions. The lesson is that even research that aims to simply describe a phenomenon can benefit from hypotheses. As signaled in Chap. 1, this also serves as another example of "failing productively."

Suggestions for What to Do When You Do Not Have Predictions

There likely are exceptions to our claim about being able to make a prediction about what you will find. For example, there could be rare cases where researchers truly have no idea what they will find and can come up with no predictions and even no hunches. And, no research has been reported on related phenomena that would offer some guidance. If you find yourself in this position, we suggest one of three approaches: revise your question, conduct a pilot study, or choose another question.

Because there are many advantages to making predictions explicit and then writing out the reasons for these predictions, one approach is to adjust your question just enough to allow you to make a prediction. Perhaps you can build on descriptions that other researchers have provided for related situations and consider how you can extend this work. Building on previous descriptions will enable you to make predictions about the situation you want to describe.

A second approach is to conduct a small pilot study or, better, a series of small pilot studies to develop some preliminary ideas of what you might find. If you can identify a small sample of participants who are similar to those in your study, you can try out at least some of your research plans to help make and refine your predictions. As we detail later, you can also use pilot studies to check whether key aspects of your methods (e.g., tasks, interview questions, data collection methods) work as you expect.

A third approach is to return to your list of interests and choose one that has been studied previously. Sometimes this is the wisest choice. It is very difficult for beginning researchers to conduct research in brand-new areas where no hunches or predictions are possible. In addition, the contributions of this research can be limited. Recall the earlier story about one of us "failing productively" by completing a dissertation in a somewhat new area. If, after an exhaustive search, you find that no one has investigated the phenomenon in which you are interested or even related phenomena, it can be best to move in a different direction. You will read recommendations in other sources to find a "gap" in the research and develop a study to "fill the gap." This can be helpful advice if the gap is very small. However, if the gap is large, too large to predict what you might find, the study will present severe challenges. It will be more productive to extend work that has already been done than to launch into an entirely new area.

Should You Always Try to Predict What You Will Find?

In short, our answer to the question in the heading is "yes." But this calls for further explanation.

Suppose you want to observe a second-grade classroom in order to investigate how students talk about adding and subtracting whole numbers. You might think, "I don't want to bias my thinking; I want to be completely open to what I see in the classroom." Sam shared a similar point of view at the beginning of the dialogue: "I wanted to leave it as open as possible; I didn't want to influence what they were going to say." Some researchers say that beginning your research study by making predictions is inappropriate precisely because it will bias your observations and results. The argument is that by bringing a set of preconceptions, you will confirm what you expected to find and be blind to other observations and outcomes. The following quote illustrates this view: "The first step in gaining theoretical sensitivity is to enter the research setting with as few predetermined ideas as possible—especially logically deducted, a priori hypotheses. In this posture, the analyst is able to remain sensitive to the data by being able to record events and detect happenings without first having them filtered through and squared with pre-existing hypotheses and biases" (Glaser, 1978, pp. 2–3).

We take a different point of view. In fact, we believe there are several compelling reasons for making your predictions explicit.

Making Your Predictions Explicit Increases Your Chances of Productive Observations

Because your predictions are an extension of what is already known, they prepare you to identify more nuanced relationships that can advance our understanding of a phenomenon. For example, rather than simply noticing, in a general sense, that students talking about addition and subtraction leads them to better understandings, you might, based on your prediction, make the specific observation that talking about addition and subtraction in a particular way helps students to think more deeply about a particular concept related to addition and subtraction. Going into a study without predictions can bring less sensitivity rather than more to the study of a phenomenon. Drawing on knowledge about related phenomena by reading the literature and conducting pilot studies allows you to be much more sensitive and your observations to be more productive.

Making Your Predictions Explicit Allows You to Guard Against Biases

Some genres and methods of educational research are, in fact, rooted in philosophical traditions (e.g., Husserl, 1929/1973) that explicitly call for researchers to temporarily "bracket" or set aside existing theory as well as their prior knowledge and experience to better enter into the experience of the participants in the research. However, this does not mean ignoring one's own knowledge and experience or turning a blind eye to what has been learned by others. Much more than the simplistic image of emptying one's mind of preconceptions and implicit biases (arguably an impossible feat to begin with), the goal is to be as reflective as possible about one's prior knowledge and conceptions and as transparent as possible about how they may guide observations and shape interpretations (Levitt et al., 2018).

We believe it is better to be honest about the predictions you are almost sure to have because then you can deliberately plan to minimize the chances they will influence what you find and how you interpret your results. For starters, it is important to recognize that acknowledging you have some guesses about what you will find does not make them more influential. Because you are likely to have them anyway, we recommend being explicit about what they are. It is easier to deal with biases that are explicit than those that lurk in the background and are not acknowledged.

What do we mean by "deal with biases"? Some journals require you to include a statement about your "positionality" with respect to the participants in your study and the observations you are making to gather data. Formulating clear hypotheses is, in our view, a direct response to this request. The reasons for your predictions are your explicit statements about your positionality. Often there are methodological strategies you can use to protect the study from undue influences of bias. In other words, making your vague predictions explicit can help you design your study so you minimize the bias of your findings.

Making Your Predictions Explicit Can Help You See What You Did Not Predict

Making your predictions explicit does not need to blind you to what is different than expected. It does not need to force you to see only what you want to see. Instead, it can actually increase your sensitivity to noticing features of the situation that are surprising, features you did *not* predict. Results can stand out when you did not expect to see them.

In contrast, not bringing your biases to consciousness might subtly shift your attention away from these unexpected results in ways that you are not aware of. This path can lead to claiming no biases and no unexpected findings without being conscious of them. You cannot observe everything, and some things inevitably will be overlooked. If you have predicted what you will see, you can design your study so that the unexpected results become more salient rather than less.

Returning to the example of observing a second-grade classroom, we note that the field already knows a great deal about how students talk about addition and subtraction. Being cognizant of what others have observed allows you to enter the classroom with some clear predictions about what will happen. The rationales for these predictions are based on all the related knowledge you have before stepping into the classroom, and the predictions and rationales help you to better deal with what you see. This is partly because you are likely to be surprised by the things you did not anticipate. There is almost always something that will surprise you because your predictions will almost always be incomplete or too general. This sensitivity to the unanticipated—the sense of surprise that sparks your curiosity—is an indication of your openness to the phenomenon you are studying.

Making Your Predictions Explicit Allows You to Plan in Advance

Recall from Chap. 1 the descriptor of scientific inquiry: "Experience carefully planned in advance." If you make no predictions about what might happen, it is very difficult, if not impossible, to plan your study in advance. Again, you cannot observe everything, so you must make decisions about what you will observe. What kind of data will you plan to collect? Why would you collect these data instead of others? If you have no idea what to expect, on what basis will you make these consequential decisions? Even if your predictions are vague and your rationales for the predictions are a bit shaky, at least they provide a direction for your plan. They allow you to explain why you are planning *this* study and collecting *these* data. They allow you to "carefully plan in advance."

Making Your Predictions Explicit Allows You to Put Your Rationales in Harm's Way

Rationales are developed to justify the predictions. Rationales represent your best reasoning about the research problem you are studying. How can you tell whether your reasoning is sound? You can try it out with colleagues. However, the best way to test it is to put it in "harm's way" (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003 p. 10). And the best approach to putting your reasoning in harm's way is to test the predictions it generates. Regardless if you are conducting a qualitative or quantitative study, rationales can be improved only if they generate testable predictions. This is possible only if predictions are explicit and precise. As we described earlier, rationales are evaluated for their soundness and refined in light of the specific differences between predictions and empirical observations.

Making Your Predictions Explicit Forces You to Organize and Extend Your (and the Field's) Thinking

By writing out your predictions (even hunches or fuzzy guesses) and by reflecting on why you have these predictions and making these reasons explicit for yourself, you are advancing your thinking about the questions you really want to answer. This means you are making progress toward formulating your research questions and your final hypotheses. Making more progress in your own thinking before you conduct your study increases the chances your study will be of higher quality and will be exactly the study you intended. Making predictions, developing rationales, and imagining tests are tools you can use to push your thinking forward before you even collect data.

Suppose you wonder how preservice teachers in your university's teacher preparation program will solve particular kinds of math problems. You are interested in this question because you have noticed several PSTs solve them in unexpected ways. As you ask the question you want to answer, you make predictions about what you expect to see. When you reflect on why you made these predictions, you realize that some PSTs might use particular solution strategies because they were taught to use some of them in an earlier course, and they might believe you expect them to solve the problems in these ways. By being explicit about why you are making particular predictions, you realize that you might be answering a different question than you intend ("How much do PSTs remember from previous courses?" or even "To what extent do PSTs believe different instructors have similar expectations?"). Now you can either change your question or change the design of your study (i.e., the sample of students you will use) or both. You are advancing your thinking by being explicit about your predictions and why you are making them.

The Costs of Not Making Predictions

Avoiding making predictions, for whatever reason, comes with significant costs. It prevents you from learning very much about your research topic. It would require *not* reading related research, *not* talking with your colleagues, and *not* conducting pilot studies because, if you do, you are likely to find a prediction creeping into your thinking. Not doing these things would forego the benefits of advancing your thinking before you collect data. It would amount to conducting the study with as little forethought as possible.

Part VI. How Do You Formulate *Important* Hypotheses?

We provided a partial answer in Chap. 1 to the question of a hypothesis' importance when we encouraged considering the ultimate goal to which a study's findings might contribute. You might want to reread Part III of Chap. 1 where we offered our opinions about the purposes of doing research. We also recommend reading the March 2019 editorial in the *Journal for Research in Mathematics Education* (Cai et al., 2019b) in which we address what constitutes important educational research.

As we argued in Chap. 1 and in the March 2019 editorial, a worthy ultimate goal for educational research is to improve the learning opportunities for *all* students.

However, arguments can be made for other ultimate goals as well. To gauge the importance of your hypotheses, think about how clearly you can connect them to a goal the educational community considers important. In addition, given the descriptors of scientific inquiry proposed in Chap. 1, think about how testing your hypotheses will help you (and the community) *understand* what you are studying. Will you have a better explanation for the phenomenon after your study than before?

HELPFUL TIP

One potentially useful way to start finding an important area of mathematics education in which to conduct research is to consult with teachers about a problem of practice that affects their students' learning opportunities. If you can connect that problem to research that helps you develop a prediction, you may have a promising candidate for a good research problem.

Although we address the question of importance again, and in more detail, in Chap. 5, it is useful to know here that you can determine the significance or importance of your hypotheses when you formulate them. The importance need not depend on the data you collect or the results you report. The importance can come from the fact that, based on the results of your study, you will be able to offer revised hypotheses that help the field better understand an important issue. In large part, it is these revised hypotheses rather than the data that determine a study's importance.

A critical caveat to this discussion is that few hypotheses are self-evidently important. They are important only if you make the case for their importance. Even if you follow closely the guidelines we suggest for formulating an important hypothesis, you must develop an argument that convinces others. This argument will be presented in the research paper you write.

Few hypotheses are self-evidently important. They are important only if you make the case for their importance.

Consider Martha's hypothesis presented earlier. When we left Martha, she predicted that "Participating teachers will show changes in their teaching with a greater emphasis on conceptual understanding with larger changes on linear function topics directly addressed in the LOs than on other topics." For researchers and educators not intimately familiar with this area of research, it is not apparent why someone should spend a year or more conducting a dissertation to test this prediction. Her rationale, summarized earlier, begins to describe why this could be an important hypothesis. But it is by writing a clear argument that explains her rationale to readers that she will convince them of its importance.

How Martha fills in her rationale so she can create a clear written argument for its importance is taken up in Chap. 3. As we indicated, Martha's work in this regard led her to make some interesting decisions, in part due to her own assessment of what was important.

Part VII. Beginning to Write the Research Paper for Your Study

It is common to think that researchers conduct a study and then, after the data are collected and analyzed, begin writing the paper about the study. We recommend an alternative, especially for beginning researchers. We believe it is better to write drafts of the paper at the same time you are planning and conducting your study. The paper will gradually evolve as you work through successive phases of the scientific inquiry process. Consequently, we will call this paper your *evolving research paper*.

We believe it is better to write drafts of the paper at the same time you are planning and conducting your study.

You will use your evolving research paper to communicate your study, but you can also use writing as a tool for thinking and organizing your thinking while planning and conducting the study. Used as a tool for thinking, you can write drafts of your ideas to check on the clarity of your thinking, and then you can step back and reflect on how to clarify it further. Be sure to avoid jargon and general terms that are not well defined. Ask yourself whether someone not in your field, maybe a sibling, a parent, or a friend, would be able to understand what you mean. You are likely to write multiple drafts with lots of scribbling, crossing out, and revising.

Used as a tool for communicating, writing the best version of what you know before moving to the next phase will help you record your decisions and the reasons for them before you forget important details. This best-version-for-now paper also provides the basis for your thinking about the next phase of your scientific inquiry.

At this point in the process, you will be writing your (research) questions, the answers you predict, and the rationales for your predictions. The predictions you make should be direct answers to your research questions and should flow logically from (or be directly supported by) the rationales you present. In addition, you will have a written statement of the study's purpose or, said another way, an argument for the importance of the hypotheses you will be testing. It is in the early sections of your paper that you will convince your audience about the importance of your hypotheses.

In our experience, presenting research questions is a more common form of stating the goal of a research study than presenting well-formulated hypotheses. Authors sometimes present a hypothesis, often as a simple prediction of what they might find. The hypothesis is then forgotten and not used to guide the analysis or interpretations of the findings. In other words, authors seldom use hypotheses to do the kind of work we describe. This means that many research articles you read will not treat hypotheses as we suggest. We believe these are missed opportunities to present research in a more compelling and informative way. We intend to provide enough guidance in the remaining chapters for you to feel comfortable organizing your evolving research paper around formulating, testing, and revising hypotheses.

While we were editing one of the leading research journals in mathematics education (*JRME*), we conducted a study of reviewers' critiques of papers submitted to the journal. Two of the five most common concerns were: (1) the research questions were unclear, and (2) the answers to the questions did not make a substantial contribution to the field. These are likely to be major concerns for the reviewers of all research journals. We hope the knowledge and skills you have acquired working through this chapter will allow you to write the opening to your evolving research paper in a way that addresses these concerns. Much of the chapter should help make your research questions clear, and the prior section on formulating "important hypotheses" will help you convey the contribution of your study.

Exercise 2.3

Look back at your answers to the sets of questions before part II of this chapter.

(a) Think about how you would argue for the importance of your current interest.
(b) Write your interest in the form of (1) a research problem, (2) a research question, and (3) a prediction with the beginnings of a rationale. You will update these as you read the remaining chapters.

Part VIII. The Heart of Scientific Inquiry

In this chapter, we have described the process of formulating hypotheses. This process is at the heart of scientific inquiry. It is where doing research begins. Conducting research always involves formulating, testing, and revising hypotheses. This is true regardless of your research questions and whether you are using qualitative, quantitative, or mixed methods. Without engaging in this process in a deliberate, intense, relentless way, your study will reveal less than it could. By engaging in this process, you are maximizing what you, and others, can learn from conducting your study.

In the next chapter, we build on the ideas we have developed in the first two chapters to describe the purpose and nature of *theoretical frameworks*. The term theoretical framework, along with closely related terms like conceptual framework, can be somewhat mysterious for beginning researchers and can seem like a requirement for writing a paper rather than an aid for conducting research. We will show how theoretical frameworks grow from formulating hypotheses—from developing rationales for the predicted answers to your research questions. We will propose some practical suggestions for building theoretical frameworks and show how useful they can be. In addition, we will continue Martha's story from the point at which we paused earlier—developing her theoretical framework.

Chapter 3
Building and Using Theoretical Frameworks

Part I. What Are Theoretical Frameworks?

As the name implies, a theoretical framework is a type of theory. We will define it as the *custom-made theory* that focuses specifically on the hypotheses you want to test and the research questions you want to answer. It is custom-made for your study because it explains why your predictions are plausible. It does no more and no less. Building directly on Chap. 2, as you develop more complete rationales for your predictions, you are actually building a theory to support your predictions. Our goal in this chapter is for you to become comfortable with what theoretical frameworks are, with how they relate to the general concept of theory, with what role they play in scientific inquiry, and with why and how to create one for your study.

As you build a more complete rationale for your predictions, you are actually building a theory to support your predictions.

As you read this chapter, it will be helpful to remember that our definitions of terms in this book, such as theoretical framework, are based on our view of scientific inquiry as formulating, testing, and revising hypotheses. We define theoretical framework in ways that continue the coherent story we lay out across all phases of scientific inquiry and all the chapters this book. You are likely to find descriptions of theoretical frameworks in other sources that differ in some ways from our description. In addition, you are likely to see other terms that we would include as synonyms for theoretical framework, including conceptual framework. We suggest that when you encounter these special terms, make sure you understand how the authors are defining them.

J. Hiebert et al., *Doing Research: A New Researcher's Guide*,
Research in Mathematics Education,
https://doi.org/10.1007/978-3-031-19078-0_3

HELPFUL TIP

Although we treat terms like "theoretical framework" and "conceptual framework" as synonyms, some authors use these terms to mean different things. In this book, we encourage you to focus less on the exact terminology and more on what should be in a theoretical framework.

Definitions of Theories

We begin by stepping back and considering how theoretical frameworks fit within the concept of theory, as usually defined. There are many definitions of theory; you can find a huge number simply by googling "theory." Educational researchers and theorists often propose their own definitions but many of these are quite similar. Praetorius and Charalambous (2022) reviewed a number of definitions to set the stage for examining theories of teaching. Here are a few, beginning with a dictionary definition:

- Lexico.com Dictionary (Oxford University Press, 2021): "A supposition or a system of ideas intended to explain something, especially one based on general principles independent of the thing to be explained."
- Biddle and Anderson (1986): "By scientific theory we mean the system of concepts and propositions that is used to represent, think about, and predict observable events. Within a mature science that theory is also explanatory and formalized. It does not represent ultimate 'truth,' however; indeed, it will be superseded by other theories presently. Instead, it represents the best explanation we have, at present, for those events we have so far observed" (p. 241).
- Kerlinger (1964): "A theory is a set of interrelated constructs (concepts), definitions and propositions which presents a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting phenomena" (p. 11).
- Colquitt and Zapata-Phelan (2007): The authors say that theories allow researchers to *understand* and *predict* outcomes of interest, *describe* and *explain* a process or sequence of events, *raise consciousness* about a specific set of concepts as well as *prevent* scholars from "being dazzled by the complexity of the empirical world by providing a linguistic tool for organizing it" (p. 1281).

For our purposes, it is important to notice two things that most definitions of theories share: They are descriptions of a connected set of facts and concepts, and they are created to predict and/or explain observed events. You can connect these ideas to Chaps. 1 and 2 by noticing that the language for the descriptors of scientific inquiry we suggested in Chap. 1 are reflected in the definitions of theories. In particular, notice in the definitions two of the descriptors: "Observing something and trying to explain why it is the way it is" and "Updating everyone's thinking in response to more and better information." Notice also in the definitions the emphasis on the elements of a theory similar to the elements of a rationale described in Chap. 2: definitions, variables, and mechanisms that explain relationships.

Exercise 3.1
Before you continue reading, in your own words, write down a definition for "theoretical framework."

Theoretical Frameworks Are Local Theories

There are strong similarities between building theories and doing scientific inquiry (formulating, testing, and revising hypotheses). In both cases, the researcher (or theorist) develops explanations for phenomena of interest. Building theories involves describing the concepts and conjectures that predict and later explain the events, and specifying the predictions by identifying the variables that will be measured. Doing scientific inquiry involves many of the same activities: formulating predictions for answers to questions about the research problem and building rationales to explain why the predictions are appropriate and reasonable.

As you move through the cycles described in Chap. 2—cycles of asking questions, making predictions, writing out the reasons for these predictions, imagining how you would test the predictions, reading more about what scholars know and have hypothesized, revising your predictions (and maybe your questions), and so on—your theoretical rationales will become both more complete and more precise. They will become more complete as you find new arguments and new data in the literature and through talking with others, and they will become sharper as you remove parts of the rationales that originally seemed relevant but now create mostly distractions and noise. They will become increasingly customized local theories that support your predictions.

In the end, your framework should be as clean and frugal as possible without missing arguments or data that are directly relevant. In the language of mathematics, you should use an idea *if and only if* it makes your framework stronger, more convincing. On the one hand, including more than you need becomes a distraction and can confuse both you, as you try to conceptualize and conduct your research, and others, as they read your reports of your research. On the other hand, including less than you need means your rationale is not yet as convincing as it could be.

The set of rationales, blended together, constitute a precisely targeted custom-made theory that supports *your* predictions. Custom designing your rationales for your specific predictions means you probably will be drawing ideas from lots of sources and combining them in new ways. You are likely to end up with a unique local theory, a theoretical framework that has not been proposed in exactly the same way before.

A common misconception among beginning researchers is that they should borrow a theoretical framework from somewhere else, especially from well-known scholars who have theories named after them or well-known general theories of learning or teaching. You are likely to use ideas from these theories (e.g., Vygotsky's theory of learning, Maslow's theory of motivation, constructivist theories of learning), but you will combine specific ideas from multiple sources to create your own framework. When someone asks, "What theoretical framework are you using?" you would not say, "A Vygotskian framework." Rather, you would say something like, "I created my framework by combining ideas from different sources so it explains why I am making these predictions."

Your set of custom-designed rationales for your predictions is your theoretical framework.

You should think of your theoretical framework as a potential contribution to the field, all on its own. Although it is unique to your study, there are elements of your framework that other researchers could draw from to construct theoretical frameworks for their studies, just as you drew from others' frameworks. In rare cases, other researchers could use your framework as is. This might happen if they want to replicate your study or extend it in very specific ways. Usually, however, researchers borrow parts of frameworks or modify them in ways that better fit their own studies. And, just as you are doing with your own theoretical framework, those researchers will need to justify why borrowing or modifying parts of your framework will help them explain the predictions they are making.

Considering your theoretical framework as a contribution to the field means you should treat it as a central part of scientific inquiry, not just as a required step that must be completed before moving to the next phase. To be useful, the theoretical framework should be constructed as a critical part of conceptualizing and carrying out the research (Cai et al., 2019c). This also means you should write out your framework as you are developing it. This will be a key part of your evolving research paper. Because your framework will be adjusted multiple times, your written document will go through many drafts.

If you are a graduate student, do not think of the potential audience for your written framework as only your advisor and committee members. Rather, consider your audience to be the larger community of education researchers. You will need to be sure all the key terms are defined and each part of your argument is clear, even to those who are not familiar with your study. This is one place where writing out your framework can benefit your study—it is easy to assume key terms are clear, but then you find out they are not so clear, even to you, when trying to communicate them. Failing to notice this lack of clarity can create lots of problems down the road.

HELPFUL TIP

When you write up your theoretical framework, consider your audience to be the larger community of education researchers. Define all of the key terms and make sure each part of your argument is clear.

Exercise 3.2
Researchers have used a number of different metaphors to describe theoretical frameworks. Maxwell (2005) referred to a theoretical framework as a "coat closet" that provides "places to 'hang' data, showing their relationship to other data," although he cautioned that "a theory that neatly organizes some data will leave other data disheveled and lying on the floor, with no place to put them" (p. 49). Lester (2005) referred to a framework as a "scaffold" (p. 458), and others have called it a "blueprint" (Grant & Osanloo, 2014). Eisenhart (1991) described the framework as a "skeletal structure of justification" (p. 209). Spangler and Williams (2019) drew an analogy to the role that a house frame provides in preventing the house from collapsing in on itself. What aspects of a theoretical framework does each of these metaphors capture? What aspects does each fail to capture? Which metaphor do you find best fits your definition of a theoretical framework? Why? Can you think of another metaphor to describe a theoretical framework?

Part II. Why Do You Need Theoretical Frameworks?

Theoretical frameworks do lots of work for you. They have four primary purposes. They ensure (1) you have sound reasons to expect your predictions will be accurate, (2) you will craft appropriate methods to test your predictions, (3) you can interpret appropriately what you find, and (4) your interpretations will contribute to the accumulation of a knowledge base that can improve education. How do they do this?

Supporting Your Predictions

In previous chapters and earlier in this chapter, we described how theoretical frameworks are built *along* with your predictions. In fact, the rationales you develop for convincing others (and yourself) that your predictions are accurate are used to refine your predictions, and vice versa. So, it is not surprising that your refined framework provides a rationale that is fully aligned with your predictions. In fact, you could think of your theoretical framework as your best explanation, at any given moment during scientific inquiry, for why you will find what you think you will find.

Throughout this book, we are using "explanation" in a broad sense. As we noted earlier, an explanation for why your predictions are accurate includes all the concepts and definitions about mechanisms (Kerlinger's, 1964 definition of "theory") that help you describe why you think the predictions you are making are the best predictions possible. The explanation also identifies and describes all the variables that make up your predictions, variables that will be measured to test your predictions.

Crafting Appropriate Methods

Critical decisions you make to test your hypotheses form the methods for your scientific inquiry. As we have noted, imagining how you will test your hypotheses helps you decide whether the empirical observations you make can be compared with your predictions or whether you need to revise the methods (or your predictions). Remember, the theoretical framework is the coherent argument built from the rationales you develop as part of each hypothesis you formulate. Because each rationale explains why you make that prediction, it contains helpful cues for which methods would provide the fairest and most complete test of that prediction. In fact, your theoretical framework provides a logic against which you can check every aspect of the methods you imagine using.

You might find it helpful to ask yourself two questions as you think about which methods are best aligned with your theoretical framework. One is, "After reading my theoretical framework, will anyone be surprised by the methods I use?" If so, you should look back at your framework and make sure the predictions are clear and the rationales include all the reasons for your predictions. Your framework should telegraph the methods that make the most sense. The other question is, "Are there some predictions for which I can't imagine appropriate methods?" If so, we recommend you return to your hypotheses—to your predictions and rationales (theoretical framework)—to make sure the predictions are phrased as precisely as possible and your framework is fully developed. In most cases, this will help you imagine methods that could be used. If not, you might need to revise your hypotheses.

Exercise 3.3

Kerlinger (1964) stated, "A theory is a set of interrelated constructs (concepts), definitions and propositions which presents a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting phenomena" (p. 11). What role do definitions play in a theoretical framework and how do they help in crafting appropriate methods?

Exercise 3.4
Sarah is in the beginning stages of developing a study. Her initial prediction is: There is a relationship between pedagogical content knowledge and ambitious teaching. She realizes that in order to craft appropriate measures, she needs to develop definitions of these constructs. Sarah's original definitions are: Pedagogical content knowledge is knowledge about subject matter that is relevant to teaching. Ambitious teaching is teaching that is responsive to students' thinking and develops a deep knowledge of content. Sarah recognizes that her prediction and her definitions are too broad and too general to work with. She wants to refine the definitions so they can guide the refinement of her prediction and the design of the study. Develop definitions of these two constructs that have clearer implications for the design and that would help Sarah to refine her prediction. (tip: Sarah may need to reduce the scope of her prediction by choosing to focus only on one aspect of pedagogical content knowledge and one aspect of ambitious teaching. Then, she can more precisely define those aspects.)

Guiding Interpretations of the Data

By providing rationales for your predictions, your theoretical framework explains why you think your predictions will be accurate. In education, researchers almost always find that if they make *specific* predictions (which they should), the predictions are not entirely accurate. This is a consequence of the fact that theoretical frameworks are never complete. Recall the definition of theories from Biddle and Anderson (1986): A theory "does not represent ultimate 'truth,' however; indeed, it will be superseded by other theories presently. Instead, it represents the best explanation we have, at present, for those events we have so far observed" (p. 241). If you have created your best developed and clearly stated theoretical framework that explains why you expected certain results, you can focus your interpretation on the ways in which your theoretical framework should be revised.

Focusing on realigning your theoretical framework with the data you collected produces the richest interpretation of your results. And it prevents you from making one of the most common errors of beginning researchers (and veteran researchers, as well): claiming that your results say more than they really do. Without this anchor to ground your interpretation of the data, it is easy to overgeneralize and make claims that go beyond the evidence.

In one of the definitions of theory presented earlier, Colquitt and Zapata-Phelan (2007) say that theories prevent scholars from "being dazzled by the complexity of the empirical world" (p. 1281). Theoretical frameworks keep researchers grounded by setting parameters within which the empirical world can be interpreted.

Exercise 3.5
Find two published articles that explicitly present theoretical frameworks (not all articles do). Where do you see evidence of the researchers using their theoretical frameworks to inform, shape, and connect other parts of their articles?

Showing the Contribution of Your Study

Theoretical frameworks contain the arguments that define the contribution of research studies. They do this in two ways, by showing how your study extends what is known and by setting the parameters for your contribution.

Showing How Your Study Extends What Is Known

Because your theoretical framework is built from what is already known or has been proposed, it situates your study in work that has occurred before. A clearly written framework shows readers how your study will take advantage of what is known to extend it further. It reveals what is new about what you are studying. The predictions that are generated from your framework are predictions that have never been made in quite the same way. They predict you will find something that has not been found previously in exactly this way. Your theoretical framework allows others to see the contributions that your study is likely to make even before the study has been conducted.

Setting the Parameters for Your Contribution

Earlier we noted that theoretical frameworks keep researchers grounded by setting parameters within which they should interpret their data. They do this by providing an initial explanation for why researchers expect to find particular results. The explanation is custom-built for each study. This means it uniquely explains the expected results. The results will almost surely turn out somewhat differently than predicted. Interpreting the data includes revising the initial explanation. So, you will end up with two versions of your theoretical framework, one that explains what you expected to find plus a second, updated framework that explains what you actually found.

The two frameworks—the initial version and the updated version—define the parameters of your study's contribution. The difference between the two frameworks is what can be learned from your study. The first framework is a claim about what is known before you conduct your study about the phenomenon you are studying; the updated framework is a claim about how what is known has changed based on your results. It is the new aspects of the updated framework that capture the important contribution of your work.

Here is a brief example. Suppose you study the errors fourth graders make after receiving ordinary instruction on adding and subtracting decimal fractions. Based on empirical findings from past research, on theories of student learning, and on your own experience, you develop a rationale which predicts that a common error on "ragged" addition problems will be adding the wrong numerals. One of the reasons for this prediction is that students are likely to ignore the values of the digit positions and "line up" the numerals as they do with whole numbers. For instance, if they are asked to add 53.2 + .16, they are likely to answer either 5.48 or 54.8.

When you conduct your study, you present problems, handwritten, in both horizontal and vertical form. The horizontal form presents the numbers using the format shown above. The vertical form places one numeral over the other but not carefully aligned:

$$\begin{array}{l} \quad 53.2 \\ +\quad .16 \\ \hline \end{array}$$

You find the predicted error occurs, but only for problems written in vertical form. To interpret these data, you look back at your theoretical framework and realize that students might ignore the value of the digits if the format reminded them of the way they lined up digits for whole number addition but might consider the value of the digits if they are forced to align the digits themselves, either by rewriting the problem or by just adding in their heads. A measure of what you (and others) learned from this study is the change in possible explanations (your theoretical frameworks). This does not mean your updated theoretical framework is "correct" or will make perfectly accurate predictions next time. But, it does mean that you are very likely moving toward more accurate predictions *and* toward a deeper understanding of how students think about adding decimal fractions.

Anchoring the Coherence of Your Study (and Your Evolving Research Paper)

Your theoretical framework serves as the anchor or center point around which all other aspects of your study should be aligned. This does not mean it is created first or that all other aspects are changed to align with the framework after it is created. The framework also changes as other aspects are considered. However, it is useful to always check alignment by beginning with the framework and asking whether other aspects are aligned and, if not, adjusting one or the other. This process of checking alignment is equally true when writing your evolving research paper as when planning and conducting your study.

Part III. How Do You Construct a Theoretical Framework for Your Study?

How do you start the process? Because constructing a theoretical framework is a natural extension of constructing rationales for your predictions, you already started as soon as you began formulating hypotheses: making predictions for what you will find and writing down reasons for why you are making these predictions. In Chap. 2, we talked about beginning this process. In this section, we will explore how you can continue building out your rationales into a full-fledged theoretical framework.

Building a Theoretical Framework in Phases

Building your framework will occur in phases and proceed through cycles of clarifying your questions, making more precise and explicit your predictions, articulating reasons for making *these* predictions, and imagining ways of testing the predictions. The major source for ideas that will shape the framework is the research literature. That said, conversations with colleagues and other experts can help clarify your predictions and the rationales you develop to justify the predictions.

As you read relevant literature, you can ask: What have researchers found that help me predict what I will find? How have they explained their findings, and how might those explanations help me develop reasons for my predictions? Are there new ways to interpret past results so they better inform my predictions? Are there ways to look across previous results (and claims) and see new patterns that I can use to refine my predictions and enrich my rationales? How can theories that have credibility in the research community help me understand what I might find and help me explain why this is the case? As we have said, this process will go back and forth between clarifying your predictions, adjusting your rationales, reading, clarifying more, adjusting, reading, and so on.

One Researcher's Experience Constructing a Theoretical Framework: The Continuing Case of Martha

In Chap. 2, we followed Martha, a doctoral student in mathematics education, as she was working out the topic for her study, asking questions she wanted to answer, predicting the answers, and developing rationales for these predictions. Our story concluded with a research question, a sample set of predictions, and some reasons for Martha's predictions. The question was: "Under what conditions do middle school teachers who lack conceptual knowledge of linear functions benefit from five 2-hour learning opportunity (LO) sessions that engage them in conceptual learning of linear functions as assessed by changes in their teaching toward a more conceptual emphasis of linear functions?" Her predictions focused on particular conditions that would affect the outcomes in particular ways. She was beginning to build rationales for these predictions by returning to the literature and identifying previous research and theory that were relevant. We continue the story here.

Imagine Martha continuing to read as she develops her theoretical framework—the rationales for her predictions. She tweaks some of her predictions based on what other researchers have already found. As she continues reading, she comes across some related literature on learning opportunities for teachers. A number of articles describe the potential of another form of LOs that might help teachers teach mathematics more conceptually—analyzing videos of mathematics lessons.

The data suggested that teachers can improve their teaching by analyzing videos of other teachers' lessons as well as their own. However, the results were mixed so researchers did not seem to know exactly what makes the difference. Martha also read that teachers who *already* can analyze videos of lessons and spontaneously describe the mathematics that students are struggling with and offer useful suggestions for how to improve learning opportunities for students teach toward more conceptual learning goals, and their students learn more (Kersting et al., 2010, 2012). These findings caught Martha's attention because it is unusual to find correlates with conceptual teaching *and* better achievement. What is not known, realized Martha, is whether teachers who learn to analyze videos in this way, through specially designed LOs, would look like the teachers who already could analyze them. Would teachers who *learned* to analyze videos teach more conceptually?

It occurred to Martha she could bring these lines of research together by extending what is known along both lines. Recall our earlier suggestion of looking across the literature and noticing new patterns that can inform your work. Martha thought about studying how, exactly, these two skills are related: analyzing videos in particular ways and teaching conceptually. Would the relationships reported in the literature hold up for teachers who *learn* to describe the mathematics students are struggling with and make useful suggestions for improving students' LOs?

Martha was now conflicted. She was well on her way to developing a testable hypothesis about the effects of learning about linear functions, but she was really intrigued by the work on analyzing videos of teaching. In addition, she saw several advantages of switching to this new topic:

- The research question could be formulated quite easily. It would be something like: "What are the relationships between learning to analyze videos of mathematics teaching in particular ways (specified from prior research) and teaching for conceptual understanding?"
- She could imagine predicting the answers to this question based directly on previous research. She would predict connections between particular kinds of analy-

sis skills and levels of conceptual teaching of mathematics in ways that employed these skills.

- The level of conceptual teaching, a challenging construct to define with her previous topic (the effects of professional development on the teaching of linear functions), was already defined in the work on analyzing videos of mathematics teaching, so that would solve a big problem. The definition foregrounded particular sets of behaviors and skills such as identifying key learning moments in a lesson and focusing on students' thinking about the key mathematical idea during these moments. In other words, Martha saw ways to adapt a definition that had already been used and tested.
- The issue of transfer—another challenging issue in her original hypothesis—was addressed more directly in this setting because the learning environment—analyzing videos of classroom teaching—is quite close to the classroom environment in which participants' conceptual teaching would be observed.
- Finally, the nature of learning opportunities, an aspect of her original idea she still needed to work through, had been explored in previous studies on this new topic, and connections were found between studying videos and changes in teaching.

Given all these advantages, Martha decided to change her topic and her research question. We applaud this decision for two major reasons. First, Martha's interest grew as she explored this new topic. She became excited about conducting a study that might answer the research question she posed. It is always good to be passionate about what you study. Second, Martha was more likely to contribute important new insights if she could extend what is already known rather than explore a new area. Exploring something quite new requires lots of effort defining terms, creating measures, making new predictions, developing reasons for the predictions, and so on. Sometimes, exploring a new area has payoffs. But, as a beginning researcher, we suggest you take advantage of work that has already been done and extend it in creative ways.

Although Martha's idea of extending previous work came with real advantages, she still faced a number of challenges. A first, major challenge was to decide whether she could build a rationale that would predict learning to analyze videos *caused* more conceptual teaching. Or, could she only build a rationale that would predict that there was a relationship between changes in analyzing videos and level of conceptual teaching? Perhaps a cause-effect relationship existed but in the opposite direction: If teachers learned to teach more conceptually, their analysis of teaching videos would improve. Although most of the literature described learning to analyze videos as the potential cause of teaching conceptually, Martha did not believe there was sufficient evidence to build a rationale for this prediction. Instead, she decided to first determine if a relationship existed and, if so, to *understand* the relationship. Then, if warranted, she could develop and test a hypothesis of causation in a future study. In fact, the direction of the causation might become clearer when she understood the relationship more clearly.

A second major challenge was whether to study the relationship as it existed or as one (or both) of the constructs was changing. Past research had explored the relationship as it existed, without inducing changes in either analyzing videos or teaching conceptually. So, Martha decided she could learn more about the relationship if one of the constructs was changing in a planned way. Because researchers had argued that teachers' analysis of video could be changed with appropriate LOs, and because changing teachers' teaching practices has resisted simple interventions, Martha decided to study the relationship as she facilitated changes in teachers' analysis of videos. This would require gathering data on the relationship at more than one point in time.

Even after resolving these thorny issues, Martha faced many additional challenges. Should she predict a closer relationship between learning to analyze video and teaching for conceptual understanding before teachers began learning to analyze videos or after? Perhaps the relationship increases over time because conceptual teaching often changes slowly. Should she predict a closer relationship if the content of the videos teachers analyzed was the same as the content they would be teaching? Should she predict the relationship will be similar across pairs of similar topics? Should she predict that some analysis skills will show closer relationships to levels of conceptual teaching than others? These questions and others occurred to Martha as she was formulating her predictions, developing justifications for her predictions, and considering how she would test the predictions.

Based on her reading and discussions with colleagues, Martha phrased her initial predictions as follows:

1. There will be a significant positive correlation between teachers' performance on analysis of videos and the extent to which they create conceptual learning opportunities for their students both before and after proposed learning experiences.
2. The relationship will be stronger:
   (a) Before the proposed opportunities to learn to analyze videos of teaching;
   (b) When the videos and the instruction are about similar mathematical topics; and,
   (c) When the videos analyzed display conceptual misunderstandings among students.
3. Of the video analysis skills that will be assessed, the two that will show the strongest relationship are spontaneously describing (1) the mathematics that students are struggling with and (2) useful suggestions for how to improve the conceptual learning opportunities for students.

Martha's rationales for these predictions—her theoretical framework—evolved along with her predictions. We will not detail the framework here, but we will note that the rationale for the first prediction was based on findings from past research. In particular, the prediction is generated by reasoning that if there has been no special intervention, the tendency to analyze videos in particular ways and to teach conceptually develop together. This might explain Kersting's findings described earlier. The second and third predictions were based on the literature on teachers' learning, especially their learning from analyzing videos of teaching.

Before leaving Martha at this point in her journey, we want to make an important point about the change she made to her research topic. Changes like this occur quite often as researchers do the hard intellectual work of developing testable hypotheses that guide research studies. When this happens to you, it can feel like you have lost ground. You might feel like you wasted your time on the original topic. In Chap. 1, we described inevitable "failure" when engaged in scientific inquiry. Failure is often associated with realizing the data you collected do not come close to supporting your predictions. But a common kind of failure occurs when researchers realize the direction they have been pursuing should change before they collect data. This happened in Martha's case because she came across a topic that was more intriguing to her and because it helped solve some problems she was facing with the previous topic. This is an example of "failing productively" (see Chap. 1). Martha did not succeed in pursuing her original idea, but while she was recognizing the problems, she was also seeing new possibilities.

Constantly Improving Your Framework

We will use Martha's experience to be more specific about the back-and-forth process in which you will engage as you flesh out your framework. We mentioned earlier your review of the literature as a major source of ideas and evidence that will affect your framework.

Reviewing Published Empirical Evidence

One of the best sources for helping you specify your predictions are studies that have been conducted on related topics. The closer to your topic, the more helpful will be the evidence for anticipating what you will find. Many beginning researchers worry they will locate a study just like the one they are planning. This (almost) never happens. Your study will be different in some ways, and a study that is very similar to yours can be extraordinarily helpful in specifying your predictions. Be excited instead of terrified when you come across a study with a title similar to yours.

Try to locate all the published research that has been conducted on your topic. What does "on your topic" mean? How widely should you cast your net? There are no rules here; you will need to use your professional judgment. However, here is a general guide: If the study does not help you clarify your predictions, change your confidence in them, or strengthen your rationale, then it falls outside your net.

In addition to helping specify your predictions, prior research studies can be a goldmine for developing and strengthening your theoretical framework. How did researchers justify their predictions or explain why they found what they did? How can you use these ideas to support (or change) your own predictions?

By reading research on similar topics, you might also imagine ways of testing your predictions. Maybe you learn of ways you could design your study, measures you could use to collect data, or strategies you could use to analyze your data. As you find helpful ideas, you will want to keep track of where you found these ideas so you can cite the appropriate sources as you write drafts of your evolving research paper.

Examining Theories

You will read a wide range of theories that provide insights into why things might work like they do. When the phenomena addressed by the theory are similar to those you will study, the associated theories can help you think through your own predictions and why you are making them. Returning to Martha's situation, she could benefit from reading theories on adult learning, especially teacher learning, on transferring knowledge from one setting to another, on professional development for teachers, on the role of videos in learning, on the knowledge needed to teach conceptually, and so on.

Focusing on Variables and Mechanisms

As you review the literature and search for evidence and ideas that could strengthen your predictions and rationales, it is useful to keep your eyes on two components: the variables you will attend to and the mechanisms that might explain the relationships between the variables. Predictions could be considered statements about expected behaviors of the variables. The theoretical framework could be thought of as a description of all the variables that will be deliberately attended to plus the mechanisms conjectured to account for these relationships.

In Martha's case, the most obvious variables are the responses teachers give to questions about their analysis of the videos and the features observed in their teaching practices. The mechanism of primary interest is the (mental and social) process that transforms the skills, knowledge, and attention involved in analyzing videos into particular kinds of teaching practices—or vice versa. The definition of conceptual teaching she adopted from previous studies gave her a clue about the mechanisms—about how and why learning to analyze videos might affect classroom teaching. The definition included attending to key learning moments in a lesson and tracking students' thinking during these moments. Martha predicted that if teachers learned to attend to these aspects of teaching when viewing videos, they might attend to them when planning and implementing their own teaching.

As Martha reviewed the literature, she identified a number of variables that might affect the likelihood and extent of this translation. Here are some examples: how well teachers understand the mathematics in the videos and the mathematics they will teach; the nature of the videos themselves; the number of opportunities teachers have to analyze videos and the ways in which these opportunities are structured;

teachers' analysis of videos and their teaching practices before the learning opportunities begin; and how much time they have to apply what they learn to their own teaching.

Martha identified these additional variables because she learned they might have a direct influence on the mechanisms that could explain the relationship between analyzing videos and teaching. Some variables might support these mechanisms, and some might interfere. Martha's task at this point in her work is to identify and describe all the variables that could play a meaningful role in the outcome of her study. This means to identify each variable for which it is possible to establish a clear and direct connection between the variable and the relationship she planned to investigate. Using the outcome of this task, Martha then needs to update her description of the mechanisms that could account for the relationships she expects to see and review her predictions and theoretical framework with these variables and mechanisms in mind.

Exercise 3.6
Review the predictions that Martha made and identify the variables that play a role in these predictions. Even though you might not be immersed in this literature, think about the alignment between the variables included in the predictions and those that could impact the relationships in which Martha is interested. Are there other missing variables that should be included in her predictions?

How Do You Know When You Have Finished Building Your Theoretical Framework?

The question of when your theoretical framework is finished could be answered in several ways. First, it is never really finished. As you continue to write your evolving research paper, you will continue strengthening your framework. You might even refine the framework as you write the final draft of your paper, *after* you have collected and analyzed your data. Furthermore, if you do follow-up studies, you will continue to build your framework.

A second answer is that you should invest the time and effort to build a theoretical framework that is as finished as possible at each point in the research process. As you write each draft of your evolving research paper, you should feel as if you have the strongest, most robust rationale you can have for your current predictions. In other words, you should feel that with each succeeding draft you have finished building your framework, even though you are quite sure you have not.

A third answer addresses a common, related question: "How do I know when I have included enough ideas and borrowed from enough sources? Would including another idea or citing another source be useful?" The answer is that you should include *only* those ideas that contribute to building a stronger framework. When you wonder whether you should include another idea or reference, ask yourself whether doing so would make your framework stronger in all the ways we described earlier.

Exercise 3.7
In 2–3 pages (single spaced), write out the plan for your study. The plan should include your research questions, your predictions of the answers, your rationale for the predictions (i.e., your theoretical framework), and your imagined plan for testing the predictions. Be as explicit and precise as you can. Be sure you have identified the critical variables and described the mechanism(s) that could explain the phenomena, the relationships, and/or the changes you predict. Look back to see if the logic connecting the parts is obvious. Ask yourself whether the tests you plan are what anyone familiar with your framework would expect (i.e., there should be no surprises).

Part IV. Refining a Theoretical Framework: A Scholarly Dialogue

As we noted above, conversations with colleagues and other experts can help you refine your theoretical framework by clarifying your predictions and digging into the details of the rationales you develop to justify those predictions. This is as true for experienced researchers as it is for beginning researchers. The dialogue below is an example of how two colleagues, Adrian (A) and Corin (C), work together to gradually formulate a testable hypothesis. Some of their conversation will look familiar as they refine their prediction through multiple steps of discussion:

- Narrowing the focus of their prediction.
- Making their prediction more testable.
- Being more specific about what they want to study.
- Engaging their prediction in cycles of refinements.
- Determining the appropriate level/grainsize of their prediction (zoom in, zoom out).
- Adding more predictions.
- Thinking about underlying mechanisms (i.e., what explains the relationships between their variables).
- Putting their predictions on a continuum (going from black and white to grey).

In addition, they construct their theoretical framework to match their hypotheses through multiple steps:

- Defining and rationalizing their variables.
- Re-evaluating their initial rationales in response to changes in their initial predictions.
- Asking themselves "why" questions about predictions and rationales.

- Finding empirical evidence and theory that better supports their evolving predictions.
- Keeping in mind what they are going to be measuring.
- Making sure their rationales support each link in their chain of reasoning.
- Identifying underlying mechanisms.
- Making sure that statements are included in their rationale if and only if they directly support their predictions and are essential to the argument.

They begin with the following hypothesis:

- **Prediction:** Students will exhibit more persistence in mathematical course taking in high school if they work in groups.
- **Brief Description of Rationale:** When people work in groups, they feel more competent and learn better (Cohen & Lotan, 2014; Jansen, 2012). When people feel more competent, they persist in additional mathematical course taking (Bandura & Schunk, 1981; Dweck, 1986).

A: So, do we think this hypothesis is testable?

C: Well actually, who these students are is probably something we need to be more specific about.

A: Good point, and also, since Algebra 2 is the bridge to additional course taking (i.e., the first course students don't have to take), perhaps we should target Algebra 2. How about if we change our prediction to the following: *Algebra 2 students will exhibit more mathematical persistence in mathematical course taking in high school if they work in groups in Algebra 2.*

C: Okay, but another problem is that it would take a long time to collect data that would inform a prediction about the courses students take, and over that amount of time I'm not sure we could even tell if groupwork was responsible. What if we limited our prediction to: *Algebra 2 students will exhibit more mathematical persistence in Algebra 2 if they work in groups.*

A: Good idea! But when we talk about persistence, do we mean students don't quit, or that they don't drop the course, or *productively* struggle during class, or turn in their homework, or is it something else we mean? To me, what would be testable about mathematical persistence would be persistence at the problem level, such as when students get stuck on a problem, but they don't give up.

C: I agree. So, let's predict the following: Algebra 2 students will exhibit more mathematical persistence in Algebra 2 when they get stuck on problems if they work in groups. That's something I think we could test.

A: Yes, but I think we need to be even more specific about what we mean by mathematical persistence when students get stuck on problems.

C: Hmm, what if we focused specifically on mathematical persistence that involves staying engaged in trying to solve a problem for the duration of a problem-solving session or until the problem gets solved? But that also makes me wonder if we want to be focusing on persistence at the individual level or at the group level?

A: Umm, I think we should focus on persistence at the individual level, because that's more consistent with our original interest in persistence in course taking, which is about individual students, not about groups.

C: Okay, that makes sense. So then how about this for a prediction: If Algebra 2 students work in groups, they will be more likely to stay engaged in trying to solve problems for the duration of a problem-solving session or until they solve the problem.

******

To this point in the dialogue, Adrian and Corin are developing a theoretical framework by sharpening what they mean by their prediction and making sure their prediction is testable. In the next part, they return to their original idea to make sure they have not strayed too far by making their prediction more precise. The dialogue illustrates how making predictions should support the goal of understanding the relationship between variables and the mechanisms for change.

******

A: Yes, I'm liking the way this prediction is evolving. However, I also feel like our prediction is now so focused that we've lost a bit of our initial idea of competence and learning, which is what we were initially interested in. Could we do something to bring those ideas back? Perhaps we could create more predictions to get at more of those ideas?

C: Great idea! Okay, so to help us see what we are missing now, let's look back at the initial links in our chain of reasoning. We initially said that Working in Groups leads to Feeling Competent & Learning Better leads to Persistence in Math Course Taking. But our chain of reasoning has changed. I think it's more like this: Working in Groups on Problems leads to Staying Engaged in Problem Solving leads to Greater Sense of Competence and Learning Better leads to More Persistence in Course Taking.

A: Okay, so if that's the case, it looks like our new prediction just tests the first link in this chain, the link between Working in Groups on Problems and Staying Engaged in Problem Solving. It looks like there are three other potential predictions we could make; we could make a prediction about the relationship between Staying Engaged in Problem Solving and having a Greater Sense of Competence, between Staying Engaged in Problem Solving and Learning Better, and between having a Greater Sense of Competence/Learning Better and More Persistence in Course Taking.

C: Clearly that's too many predictions for us to tackle in one study and actually I am aware of several studies that already address the third prediction. So, we can use those studies as part of our rationale and don't need to study that link.

A: I agree. Let's just add one prediction, one about the link between Staying Engaged and Sense of Competence. In our initial prediction, we just had a vague connection between Working in Groups and Sense of Competence. But in our new prediction, we were more specific that working in groups helps students stay engaged until the end of a problem-solving session. So, I guess we could say for a second prediction then that *When Algebra 2 students stay engaged in problem solving until the end of a problem-solving session, they develop a greater sense of competence.*

C: Okay so we will have two predictions to examine with our study: Prediction 1 is: If Algebra 2 students work in groups, they will be more likely to stay engaged in trying to solve problems for the duration of a problem-solving session or until they solve the problem. This prediction deals with the first link in our chain of reasoning. And then Prediction 2 is: If Algebra 2 students try to solve problems for the duration of a problem-solving session or until they solve the problem, they will be more likely to develop a sense of competence. Oh, as soon as I finished stating that prediction, the thought just came to me, "sense of competence about what?"

A: How about if we focused on sense of competence in being able to solve similar problems in the future? Actually, maybe that's too limited. Maybe we should expand our prediction a bit more so we include a sense of competence that's at least somewhat closer to more course taking? Something like sense of competence that involves feeling capable of understanding future Algebra 2 concepts. That's at least bigger than sense of competence at solving similar problems. If students feel they're capable of understanding future Algebra 2 concepts, then they will probably be more likely to persist in course taking too.

C: Okay, that makes sense. So, then our Prediction 2 could be: If Algebra 2 students try to solve problems for the duration of a problem-solving session or until they solve the problem, they will be more likely to feel they will be capable of understanding future Algebra 2 concepts.

A: Oh, I just had an additional idea! What if we changed the two predictions one more time to allow for more or less of the variables? For example, Prediction 1 could be: *The more Algebra 2 students work in groups, the more likely they will stay engaged in trying to solve problems for the duration of a problem-solving session or until they solve the problem.*

C: Yes, great. So, that would mean Prediction 2 could be: The more Algebra 2 students try to solve problems for the duration of a problem-solving session or until they solve the problem, the more likely they will feel they are capable of understanding future Algebra 2 concepts.

A: So, I think we're happy with our predictions for now, but I think we need to work on our rationales for those predictions because they no longer apply very well.

C: Okay, to recap, our original chain of reasoning was Working in Groups leads to Feeling Competent & Learning Better leads to Persistence in Math Course Taking. Our initial rationales were the following: For the link between working in groups and feeling competent, we based that link on Cohen and Lotan's (2014) book on Designing Groupwork, in which they explain why and how all students can feel competent through their engagement in groupwork. We also based this link on that 2012 Jansen study that found that groupwork helped students enact their competence in math. Then, for the link between competence and persistence, we based that link on the Bandura and Schunk (1981) study and on the work by Carol Dweck (1986) that show that children who feel more competent in arithmetic, tend to persist more.

******

Corin and Adrian have looked back at their initial research idea. In doing so, they illustrated how developing a theoretical framework involves developing and refining a chain of reasoning. They continue by working on developing rationales for their predictions.

******

A: Okay, so let's think if any of our previous rationales still work. How about Elizabeth Cohen's work? I still think her work applies because it shows that groupwork can affect engagement. But now that I think about it, another part of her work indicates that groupwork needs particular norms in order to be effective. So maybe we should tighten up our predictions to focus just on groupwork that has particular norms?

C. But, on the other hand, what about Jo Boaler's (1998) "Open and Closed Mathematics" article? In that study, students at the Phoenix Park School did not have much structure, and in spite of that, groupwork worked quite well for those students, better than individual work did for students at the Amber Hill School who had highly structured instruction.

A. That's a good point. So maybe we should leave our predictions about groupwork as is (i.e., not focus on particular norms). Also, the ideas in the Boaler article would be good to add to our theoretical framework because it deals with secondary students, which aligns better with the ages of the Algebra 2 students we are planning on studying.

C: Okay, so we're adding the ideas in the Boaler article. I also think we need to find literature that specifies the kind of engagement we want to focus on. Looking at the engagement literature would sharpen our thinking about the engagement we are most interested in. We should consider Brigid Barron's (2003) study, "When Smart Groups Fail." In her study, students produced better products if they engaged with each other and with the content. But that makes me think that we are mostly just focused on the latter, namely on how individuals engage with the content.

A: I agree we're focused on individuals' engagement with the content. Come to think of it, the fact that we're focused on how individuals engage with content rather than how groups engage further justifies why we're not looking at groupwork norms. But let me ask a question we need to answer. Why are we focusing on how individuals engage with content? It's not just a preference. It's because we think individual engagement with content is related to feeling capable. So, our decision to focus on individual engagement aligns with our predictions. And even though we're not including Barron's work in our framework, considering her work helped sharpen our thinking about what we're focusing on.

C: You know, we are kind of in a weird space because we're focusing on individual engagement with content at the same time as we are predicting that groupwork leads to more engagement. In other words, we are and aren't taking a social perspective. But what this reminds me of is how, from the perspective of the theory of constructivism, even though individuals have to make sense of things for themselves, social interactions are what drives sense making. In fact, here's a quote from von Glasersfeld (1995): "Piaget has stressed many times that the most frequent cause of accommodation is the interaction" (p. 66). So, I think we can use constructivism as a theoretical justification for predicting that the social activity of groupwork is what is related to individual engagement with content.

A: Interesting! Yes, makes sense. When you were describing that, I had another insight from constructivism. You know how when someone experiences a perturbation, it also creates a need in them to resolve the perturbation, right? So maybe perturbations are the mechanism explaining why groupwork leads to more individual engagement with content. Groupwork potentially generates perturbations, meaning the person engages more to try to resolve those perturbations.

C: Okay, now that we have brought in the idea of perturbations as potentially being the mechanism that drives how working in groups leads to staying more engaged, perhaps we need to reconsider what we will be measuring in our study. Will it be perturbations, or will it be staying engaged that we should be measuring?

A: I think what we are saying is that the need to resolve perturbations is part of the underlying mechanism, but measuring the need to resolve perturbations would be difficult if not impossible. So, instead, I think we should focus on measuring the variable *staying engaged*, a variable we can measure. And then if we find that more working in groups leads to more staying engaged, that also gives us more evidence that our theoretical framework with perturbations as a mechanism is viable. In other words, mechanisms are part of our framework and by testing our prediction, we are testing our theoretical framework (i.e., our rationales) too.

******

This final part of the dialogue illustrates that the rationale for a study continues to develop as the predictions continue to be refined and testability continues to be considered. In other words, the development of the predictions and rationale (i.e., the theoretical framework) should be iterative and ongoing.

******

Through their discussion, Adrian and Corin have refined both their predictions and their rationales. In the process, the key ideas they have drawn on contributed to their rationales and thus to constructing their theoretical framework.

Part V. Distinctions Between Rationales, Theoretical Frameworks, and Literature Reviews

We have introduced a number of terms that play critical roles in the scientific inquiry process. Because they refer to related and sometimes overlapping ideas, keeping straight their meanings and uses can be challenging. It might be helpful to revisit each of them briefly to describe how they are similar to, and different from, each other.

To distinguish between rationales, theoretical frameworks, and literature reviews, it is useful to consider the roles they play as you plan and conduct a study compared to the roles they play when you write the report of your study.

Thinking Through a Study

The chronology of the thinking process often moves through many cycles of identifying a research problem or asking a question, and then reading the literature to learn more about the problem, and then refining and narrowing the scope of a question that would add to or extend what is known, and then predicting (guessing) an answer to the question and asking yourself why you predicted this answer and writing a first draft of your rationale, and then reading the literature to improve your rationale, and then realizing you can refine the question further along with specifying a clearer and more targeted prediction, and then reading the literature to further improve your rationale, and then realizing you can refine the question further along with a clearer and more targeted prediction, and so on.

The primary activity that generates more specific and clearer hypotheses is searching and *reviewing literature*. You can return to the literature as often as you need to build your *rationales*. As your rationales develop, they morph into your *theoretical framework*. The theoretical framework is a coherent argument that threads together the individual rationales and explains why your predictions are the best predictions the field can make at this time.

If you have one research question and one prediction you will have one rationale. In this case, your rationale is essentially the same as your theoretical framework. If you have more than one research question, you will have multiple predictions and multiple rationales. As you develop rationales for each prediction, you might find lots of overlap. Maybe the literatures you read to refine each prediction and develop each rationale overlap, and maybe the arguments you piece together include many of the same elements. Your theoretical framework emerges from weaving the rationales together into one coherent argument. Although this process is more complicated than the thinking process for one prediction, it is more common. If you find few connections among the rationales for each prediction, we recommend stepping back and asking whether you are conducting more than one study. It might make more sense to sort the questions into two or more studies because the rationales for the predicted answers are drawing from different literatures.

Writing the Evolving Research Paper

We recommend that you write drafts of the research report as you think through your study and make decisions about how to proceed. Although your thinking will be fluid and evolving, we recommend that you follow the conventions of academic writing as you write drafts. For example, we recommend that you structure the paper using the five typical major sections of a journal article: introduction, theoretical framework, methods, results, and discussion. Each of these sections will go through multiple drafts as you plan your study, collect the data, analyze the data, and interpret the results.

In the introduction, you will present the research problem you are studying. This includes describing the problem, explaining why it is significant, defining the special terms you use, and often presenting the research questions you will address along with the answers you predict. Sometimes the questions and predictions are part of the next section—the theoretical framework.

In the theoretical framework, you will present your best arguments for expecting the predicted answers to the research questions. You will *not* trace the many cycles in which you engaged to get to the best versions of your arguments but rather present the latest and best version. The report of a study does not describe the chronology of the back-and-forth messiness always involved in thinking through all aspects of the study. What you learned from reviewing the literature will be an integral part of your arguments. In other words, the review of research will be included in the presentation of your theoretical framework rather than in a separate section.

The report of a study does not describe the chronology of the back-and-forth messiness always involved in thinking through all aspects of the study.

The literature you choose to include to present your theoretical framework is not all the literature you reviewed for conducting your study. Rather, the literature cited in your paper should be the literature that contributed to building your theoretical framework, and *only* that literature. In other words, the theoretical framework places the boundaries on what you should review in the paper.

Beginning researchers are often tempted to review much of what they read. Researchers put lots of time into reading, and leaving lots of it out when writing the paper can make all that reading feel like a waste of time. It is not a waste of time; it is always part of the research process. But, reviewing more than you need in the paper becomes a distraction and diverts the reader from the main points.

The literature cited in your paper should be the literature that contributed to building your theoretical framework, and only *that literature.*

What should you do if the editor of the journal requires, or recommends, a section titled "review of research"? We recommend you create a somewhat more elaborated review for this section and then show exactly how you used the literature to build your rationale in the theoretical framework section.

Reviewers notice when the theoretical framework and the literature reviewed do not provide sufficient justification for the research questions (or the hypotheses). We found that about 13% of *JRME* reviews noted an especially important gap—the research questions in a paper were not sufficiently motivated. We expect the same would be true for other research journals. Reviewers also note when manuscripts either do not have an explicit theoretical framework or when they seem to be juggling more than one theoretical framework.

Part VI. Moving to Methods

A significant benefit of building rich and precise theoretical frameworks is the guidance they provide for selecting and creating the methods you will use to test your hypotheses. The next phase in the process of scientific inquiry is crafting your methods: choosing your research design, selecting your sample, developing your measures, deciding on your data analysis strategies, and so on. In Chap. 4, we discuss how you can do this in ways that keep your story coherent.

Chapter 4
Crafting the Methods to Test Hypotheses

Part I. What Does It Mean to Test Your Hypotheses?

From the beginning, we have talked about formulating and testing hypotheses. We will briefly review relevant points from the first three chapters and then consider some additional issues you will encounter as you craft the methods you will use to test your hypotheses.

In Chap. 1, we proposed a distinction between hypotheses and predictions. Predictions are guesses you make about answers to your research questions; hypotheses are the predictions plus the reasons, or rationales, for your predictions. We tied together predictions and rationales as constituent parts of hypotheses because it is beneficial to keep them connected throughout the process of scientific inquiry. When we talk about testing *hypotheses*, we mean gathering information (data) to see how close your predictions were to being correct *and then* assessing the soundness of your rationales. So, testing hypotheses is really a two-step process: (1) comparing predictions with empirical observations or data, and (2) assessing the soundness of the rationales that justified these predictions.

In Chap. 2, we suggested that making predictions and explaining why you made them should happen at the same time. Along with your first guesses about the answers to your research questions, you should write out your explanations for why you think the answers will be accurate. This will be a back-and-forth process because you are likely to revise your predictions as you think through the reasons you are making them. In addition, we suggested asking how you could test your predictions. This often leads to additional revisions in your predictions.

We also noted that, because education is filled with complexities, answers to substantive questions can seldom be predicted with complete accuracy. Consequently, testing predictions does not mean deciding whether or not they were correct but rather how you can revise them to improve their correctness. In addition, testing predictions means reexamining your rationales to improve the soundness of your

J. Hiebert et al., *Doing Research: A New Researcher's Guide*,
Research in Mathematics Education,
https://doi.org/10.1007/978-3-031-19078-0_4

reasoning. In other words, testing predictions involves gathering the kind of information that guides revisions to your hypotheses.

As a final reminder from Chap. 2, we asked you to imagine *how* you could test your hypotheses. This involves anticipating what information (data) would best show how accurate your predictions were and would inform revisions to your rationales. Imagining the best ways to test hypotheses is essential for moving through the early cycles of scientific inquiry. In this chapter, we extend the process by crafting the actual methods you will use to test your hypotheses.

In Chap. 3, you considered further the multiple cycles of asking questions, articulating your predictions, developing your rationales, imagining testing your predictions and rationales, adjusting your rationales, revising your predictions, and so on. You learned that a significant consequence of repeating this cycle many times is the increasingly clear, justifiable, and complete rationales that turn into the theoretical framework for your study. This comes, in large part, from the clear descriptions of the variables you will attend to and the mechanisms you conjecture are at work. The theoretical framework allows you to imagine with greater confidence, and in more detail, the kind of data you will need to test your hypotheses and how you could collect them.

In this chapter, we will examine many of the issues you must consider as you choose and adapt methods to fit your study. By “methods,” we mean the entire set of procedures you will use, including the basic design of the study, measures for collecting data, and analytic approaches. As in previous chapters, we will focus on issues that are critical for conducting scientific inquiry but often are not sufficiently discussed in more standard methods textbooks. We will also cite sources where you can find more information. For example, the Institute of Education Sciences and the National Science Foundation (2013) jointly developed guidelines for researchers about the different methods that can be used for different types of research. These guidelines are meant to inform researchers who seek funding from these agencies.

Exercise 4.1

Choose a published empirical study that includes clearly stated research questions, explicit hypotheses (predictions about the answers to the research questions plus the rationales for the predictions), and the methods used. Identify the variables studied and describe the mechanisms, embedded in the hypotheses and conjectured to create the predicted answers. Analyze the appropriateness of the methods used to answer the research questions (i.e., test the predictions). Notes: (1) you might have trouble finding a clear statement of the hypotheses; if so, imagine what the researchers had in mind; and (2) although we have not discussed all of the information you might need to complete this exercise in detail, writing out your response in as much detail as possible will prepare you to make sense of this chapter.

Part II. What Are the Best Methods for Your Study?

The best methods for your study are the procedures that give you the richest information about how near your predictions were to the actual findings and how they could be adjusted to be more accurate. Said another way, choose the methods that provide the clearest answers to your research questions. There are many decisions you will need to make about which methods to use, and it is likely that, at a detailed level, there are different combinations of decisions that would be equally effective. So, we will not assume there is a single *best* combination. Rather, from this point on we will talk about *appropriate* methods.

Choose the methods that provide the clearest answers to your research questions.

Most research questions in education are too complicated to be fully answered by conducting only one study using one set of methods. Different methods offer different perspectives and reveal different aspects of educational phenomena. "Science becomes more certain in its progression if it has the benefits of a wide array of methods and information. Science is not improved by subtracting but by adding methods" (Sechrest et al., 1993, p. 230). You will need to craft one set of methods for your study but be aware that, in the future, other researchers could use another set of methods to test similar hypotheses and report somewhat different findings that would lead to further revisions of the hypotheses. The methods you craft should be aligned with your theoretical framework, as noted earlier, but there are likely to be other sets of methods that are aligned as well.

A useful organizational scheme for crafting your methods divides the process into three phases: choosing the design of your study, developing the measures and procedures for gathering the data, and choosing methods to analyze the data (in order to compare your findings to your predictions). We will not repeat most of what you can find in textbooks on research methods. Rather, we will focus on the issues within each phase of crafting your methods that are often difficult for beginning researchers. In addition, we will identify areas that manuscript reviewers for *JRME* often say are inadequately developed or described. Reviewers' concerns are based on what they read, so the problems they identify could be with the study itself or the way it is reported. We will deal first with issues of conducting the study and then talk about related issues with communicating your study to others.

Choosing the Design for Your Study

One of the first decisions you need to make is what design you will use. By design we mean the overall strategy you choose to integrate the different components of the study in a coherent and logical way. The design offers guidelines for the sampling procedure, the development of measures, the collection of data, and the analysis of data. Depending on the textbook you consult, there are different classification schemes that identify different designs. One common scheme is to distinguish between experimental, correlational, and descriptive research.

In our view, each design is tailored to *explain* different features of phenomena. Experiments, as we define them, are tailored to explain changes in phenomena. Correlations are tailored to explain relationships between two or more phenomena. And descriptions are tailored to explain phenomena as they exist. We unpack these ideas in the following discussions.

Experiment

In education, most experiments take the form of intervention studies. They are conducted to test the effects of an intervention designed to change something (e.g., students' achievement). If you choose an experimental design, your research questions probably ask whether an intervention will improve certain outcomes. For example: "Will professional development that engages teachers in analyzing videos of teaching help them teach more conceptually? If so, under what conditions does this occur?" There are several good sources to read about designing experiments in education research (e.g., Cook et al., 2002; Gall et al., 2007; Kelly & Lesh, 2000). We will focus our attention on several specific issues.

Causation Many experiments aim to determine if something *causes* something else. This is another way of saying the aim is to produce change in something and explain why the change occurred. In education, experiments often try to explain whether and why an intervention is "effective," or whether and why intervention A is more effective than intervention B. Effective usually means the treatment causes or explains the outcomes of interest. If your investigation is situated in an actual classroom or another authentic educational setting, it is usually difficult to claim causal effects. There are many reasons for this, most tied to the complicated nature of educational settings. You should consider the following three issues when designing an experiment.

First, in education, the strict requirements for an experimental design are rarely met. For example, usually students, teachers, schools, and so forth, cannot be randomly assigned to receive one or the other of the interventions that are being compared. In addition, it is almost impossible to double-blind education experiments (that is, to ensure that the participants do not know which treatment they are receiving and that the researchers do not know which participants are receiving which treatment—like in medical drug trials). These design constraints limit your ability to claim causal effects of an intervention because they make it difficult to explain the reasons for the changes. Consequently, many studies that are called experiments are better labeled “quasi-experiments.” See Campbell et al. (1963) and Gopalan et al. (2020) for more details.

Second, even when you are aware of these constraints and consider your study a quasi-experiment, it is still tempting to make causal claims not supported by your findings. Suppose you are testing your prediction that a specially designed five-lesson unit will help students understand adding and subtracting fractions with unlike denominators. Suppose you are fortunate enough to randomly assign many classrooms to your intervention and an equal number to a common textbook unit. Suppose students in the experimental classrooms perform significantly better on a valid measure of understanding fraction addition and subtraction. Can you claim your treatment caused the better outcomes?

Before making this basic causal claim, you should ask yourself, “What, exactly, was the treatment? To what do I attribute better performance?” When implemented in actual, real classrooms, your intervention will have included many (interacting) elements, some of which you might not even be aware of. That is, in practice, the “treatment” may no longer be defined precisely enough to make a strong claim about the effects of the treatment you planned. And, because each classroom operates under different conditions (e.g., different groups of students, different expectations), the aspects of the intervention that really mattered in each classroom might not be apparent. An average effect over all classrooms may mask aspects of the intervention that matter in some classrooms but not others.

Despite the challenges outlined above with making causal claims, it remains important for education researchers to pursue a greater understanding of the causes behind effects. As the National Research Council (2002) says: “An area of research that, for example, does not advance beyond the descriptive phase toward more precise scientific investigation of causal effects and mechanisms for a long period of time is clearly not contributing as much to knowledge as one that builds on prior work and moves toward more complete understanding of the causal structure” (NRC, 2002, p. 101).

Many of the problems with developing convincing explanations for changes and making causal claims have become more visible as researchers find it difficult to replicate findings (Makel & Plucker, 2014; Open Science Collaboration, 2015). And if findings cannot be replicated, it is impossible to accumulate knowledge—a hallmark of scientific inquiry (Campbell, 1961). Even when efforts are made to implement a particular intervention in another setting with as much fidelity as possible, the findings usually look different. The real challenge is to identify the conditions under which the intervention works as it did.

This leads to a third issue. Be sure to consider the nature of data that will best help you establish connections between interventions and outcomes. Quantitative data often are the data of choice because analyses can be applied to detect the probability the outcomes occurred as a consequence of the intervention. This information is important, but it does not, by itself, *explain* why the connections exist. Along with Maxwell (2004), we recommend that qualitative data also play a role in establishing causation. Qualitative data can provide insights into the mechanisms that are responsible for the connections between interventions and outcomes. Identifying mechanisms that explain changes in outcomes is key to making causal claims. Whereas quantitative data are helpful in showing whether an intervention could have caused particular outcomes, qualitative data can explain how or why this could have occurred.

Beyond Causation Do the challenges of using experiments mean experimental designs should be avoided? No. There are a number of considerations that can make experimental designs informative. Remember that the overriding purpose of research is to *understand* what you are studying. We equate this to *explaining* why what you found might look like it does (see Chaps. 1 and 2). Experiments that simply compare one treatment with another or with "business as usual" do not help you understand what you are studying because the data do not help you explain why the differences occurred. They do not help you refine your predictions and revise your rationales. However, experiments do not need to be conducted simply to determine the winner of two treatments.

If you are conducting an experiment to increase the accuracy of your predictions and the adequacy of your rationales, your research questions will almost certainly ask about the conditions under which your predicted outcomes will occur. Your predictions will likely focus on the ways in which the outcomes are significantly different from before the intervention to after the intervention, and on how the intervention plus the conditions *might* explain or have caused these changes. Your experiment will be designed to test the effects of these conditions on the outcomes. Testing conditions is a direct way of trying to understand the reasons for the outcomes, to explain why you found what you did. In fact, understanding under what conditions an intervention works or does not work is the essence of scientific inquiry that follows an experimental design.

By providing as much detail as you can in the hypotheses, by making your predictions as precise as possible, you can set boundaries on how and to what you will generalize your findings. Making hypotheses precise often requires including the conditions under which you believe the intervention might work best, the conditions under which your predictions will be true.

Another way of saying this is that you should subject your hypotheses to *severe tests*. The more precise your predictions, the more severe your tests. Consider a meteorologist predicting, a month in advance, that it will rain in the State of Delaware in April. This is not a precise hypothesis, so the test is not severe. No one would be surprised if the prediction was true. Suppose she predicts it will rain in the city of Newark, Delaware, during the second week in April. The hypothesis is more precise, the test is more severe, and her colleagues will be a bit more interested in her rationale (why she made the prediction). Now suppose she predicts it will rain on the University of Delaware campus on April 16. This is a very precise prediction, the test would be considered very severe, and lots of people will be interested in understanding her rationale (even before April 16).

In education, making precise predictions about the conditions under which a classroom intervention might cause changes in particular learning outcomes and subjecting your predictions to severe tests often requires gathering lots of data at high levels of detail or small grain sizes. Graham Nuthall (2004, 2005) provides a useful analysis of the challenges involved in designing a study with the grain size of data he believes is essential. Your study will probably not be as ambitious as that described by Nuthall (2005), but the lesson is to think carefully about the grain size of data you need to test your (precise) predictions.

Additional Considerations Although you can find details about experimental designs in several sources, some issues might not be emphasized in these sources even though they deserve attention.

First, if you are comparing the changes that occurred during your intervention to the changes that occurred during a control condition, your interpretation of the effectiveness of your intervention is only as useful as the quality of the control condition. That is, if the control condition is not expected to produce much change, and if your analyses are designed primarily to show statistical differences in outcomes, then your claim about the better effects of your intervention is not very interesting or educationally important.

Second, the significance in the size of the changes from before to after the intervention are usually reported using values that describe the probability the changes would have occurred by chance (statistical significance). But these values are affected by factors other than the size of the change, such as the size of the sample. Recently, journals have started encouraging or requiring researchers to report the size of the changes in more meaningful ways, both in terms of what the statistical result really means and in terms of the educational importance of the changes. “Effect size” is often used for these purposes. See Bakker et al. (2019) for further discussion of effect size and related issues.

Third, you should consider what “better performance” means when you compare interventions. Did all the students in the experimental classrooms outperform their peers in the control classrooms, or was the better average performance due to some students performing much better to make up for some students performing worse? Do you want to claim the intervention was effective when some students found it less effective than the control condition?

Fourth, you need to consider how fully you can describe the nature of the intervention. Because you want to explain changes in outcomes by referencing aspects of the intervention, you need to describe the intervention in enough detail to provide meaningful explanations. Describing the intervention means describing how it was implemented, not how it was planned. The degree to which the intervention was implemented as planned is sometimes referred to as fidelity of implementation (O’Donnell, 2008). Fidelity of implementation is especially critical when an intervention is implemented by multiple teachers in different contexts.

Based on our experience as an editorial team, there are a few additional considerations you should keep in mind. These considerations concern inadequacies that were often commented on by reviewers, so they are about the research paper and not always about the study itself. But many of them can be traced back to decisions the authors made about their research methods.

- Sample is not big enough to conduct the analyses presented. If you are planning to use quantitative methods, we strongly recommend conducting a statistical power analysis. This is a method of determining if your sample is large enough to detect the anticipated effects of an intervention.
- Measures used do not appear to assess what the authors claim they assess.
- Methods (including coding rubrics) are not described in enough detail. (A good rule of thumb for "enough" is that readers should be able to replicate the study if they wish.)

Methods are different from those expected based on the theoretical framework presented in the paper.

Special Experimental Designs Three designs that fit under the general category of experiments are specially crafted to examine the possible reasons for changes observed before and after an intervention. Sometimes, these designs are used to explore the conditions under which changes occur before conducting a larger study. These designs are defined somewhat differently by different researchers. Our goal is to introduce the designs but not to settle the differences in the definitions.

Because these designs include features that fall outside the conventional experiment, researchers face some unique challenges both conducting and reporting these studies. One such feature is the repeated implementation of an intervention, with each implementation containing small revisions based on the previous outcomes, in order to improve the intervention *during* the study. There are no agreed upon practices for reporting these studies. Should every trial and every small change in outcomes and subsequent interventions be reported? Should all the revised versions of the hypotheses that guided the next trial be reported? Keep these challenges in mind as you consider the following designs.

Teaching Experiments During the 1980s, mathematics educators began focusing closely on how students changed their thinking during instruction (Cobb & Steffe, 1983; Steffe & Thompson, 2000). The aim was to describe these changes in considerable detail and to explain how the instructional activities prompted them. Teaching experiments were developed as a design to follow changes in students' thinking as they received small, well-defined episodes of teaching. In some cases, mapping *developmental* changes in student thinking was of primary interest; instruction was simply used to induce and accelerate these changes.

Most teaching experiments can be described as a sequence of teaching episodes designed for testing hypotheses about how students learn and reason. A premium is placed on getting to know students well, so the number of students is usually small, and the teacher is the researcher. Predictions are made before each episode about how students' (often *each* student's) thinking will change based on the features of the teaching activity. Data are gathered at a small grain size to test the predictions and revise the hypotheses for the next episode. Until they gain the insights they intend, researchers often continue the following cycles of activities: teaching to test hypotheses, collecting data, analyzing data to compare with predictions, revising predictions and rationales, teaching to test the revised hypotheses, and so on.

Design-Based Research Following the introduction of teaching experiments, the concept was elaborated and expanded into an approach called design-based research (Akker et al., 2006; Cobb et al., 2017; Collins, 1992; Design-Based Research Collaborative, 2003; Puntambekar, 2018). There are many forms of this research design but most of them are tailored to developing topic-specific instructional theories that can be shared with teachers and educational designers.

Like teaching experiments, design-based research consists of continuous cycles of formulating hypotheses that connect instructional activities with changes in learning, designing the learning environment to test the hypotheses, implementing instruction, gathering and analyzing data on changes in learning, and revising the hypotheses. The grain size of data matches the needs of teachers to make day-to-day instructional decisions. Often, this research is carried out through researcher–teacher partnerships, with researchers focused on developing theories (systematic explanations for changes in students' learning) and teachers focused on implementing and testing theories. In addition, unlike many teaching experiments, design-based research has the design of instructional products as one of its goals.

These designs initially aimed to develop full explanations or theories of the learning processes through which students developed understanding for a topic complemented with theories of instructional activities that support such processes. The design was quickly expanded to study learning situations of all kinds, including, for example, teacher professional development (Gravemeijer & van Eerde, 2009).

Other forms of design-based research have also emerged, each with the same basic principles but with different emphases. For example, "Design-Based Implementation Research" (Fishman & Penuel, 2018) focuses on improving the implementation of promising instructional approaches for meeting the needs of diverse students in diverse classrooms. Researcher–teacher partnerships produce adaptations that are scalable and sustainable through cycles of formulating, testing, and revising hypotheses.

Continuous Improvement Research An approach to research that shares features with design-based research but focuses more directly on improving professional practices is often called either continuous improvement, improvement science, or implementation science. This approach has shown considerable promise outside of education in fields such as medicine and industry and could be adapted to educational settings (Bryk et al., 2015; Morris & Hiebert, 2011). A special issue of the *American Psychologist* in 2020 explored the possibilities of implementation science to address the challenge posed in its first sentence, "Reducing the gap between science and practice is the great challenge of our time" (Stirman & Beidas, 2020, p. 1033).

The cycles of formulating, testing, and revising hypotheses in the continuous improvement model are characterized by four features (Morris & Hiebert, 2011).

First, the research problems are drawn from practice because the aim is to improve these practices. Second, the outcome is a concrete product that holds the knowledge gained from the research. For example, an annotated lesson plan could serve as a product of research directed toward improving instructional practice of a particular concept or skill. Third, the interventions test a series of small changes to the product, each built on the previous version, by collecting just enough data to tell whether the change was an improvement. Finally, the research process involves the users as well as the researchers. If the goal is to improve practice, practitioners must be an integral part of the process.

Shared Goals of Useful Education Experiments All experimental designs that we recommend have two things in common. One is they try to change something and then study the possible mechanisms for the change and the conditions under which the change occurred. Experimental designs that study the reasons and conditions for a change offer greater understanding of the phenomena they are studying. The noted sociologist Kurt Lewin said, “If you want truly to understand something, try to change it” (quoted in Tolman et al., 1996, p. 31). Recall that understanding phenomena was one of the basic descriptors of scientific inquiry we introduced in Chap. 1.

In our view, a second feature of useful experiments in education is that they formulate, test, and revise hypotheses at a grain size that matches the needs of educators to make decisions that improve the learning opportunities for all students. Often, research questions that motivate useful experiments address instructional problems that teachers face in their classrooms. We will return to these two features in Chap. 5.

HELPFUL TIP

If you are conducting an experiment, consider beginning with a small experiment and planning follow-up experiments (after your dissertation) that gradually increase in size and scope. Many resear chers find it helpful to work out the conceptual issues while conducting a small study to increase the chances that a larger, more expensive study will be worth the time and resources.

Correlation

Correlational designs investigate and explain the relationship between two or more variables. Researchers who use this design might ask questions like Martha’s: “What is the relationship between how well teachers analyze videos of teaching and how conceptually they teach?”

Notice the difference between this research question and the earlier one posed for an experimental design ("Will professional development that engages teachers in analyzing videos of teaching help them teach more conceptually? If so, under what conditions does this occur?"). In the experimental case, researchers hypothesized that analyzing videos of teaching would *cause* more conceptual teaching; in the correlational case they are acknowledging they are not ready to make this prediction. However, they believe there *is* a sufficiently strong rationale (theoretical framework) to predict a relationship between the two. In other words, although predicting that one event causes another cannot be justified, a rationale can be developed for predicting a relationship between the events.

Correlations in Education Are Rarely Simple When two or more events appear related, the explanation might be quite complicated. It might be that one event causes another, but there are many more possibilities. Recall Martha's research question: "What are the relationships between learning to analyze videos of teaching in particular ways (specified from prior research) and teaching for conceptual understanding?" Her research question fits a correlational design because she could not develop a clear rationale explaining why one event (learning to analyze videos) should cause changes in another (changes in teaching conceptually).

Martha could imagine three reasons for a relationship: (1) an underlying factor could be responsible for both events varying together (maybe developing more pedagogical content knowledge is the underlying factor that enables teachers to both analyze videos more insightfully and teach more conceptually); (2) there could be a causal relation but in the reverse direction (maybe teachers who already teach quite conceptually build on students' thinking, which then helps them analyze videos of teaching in particular ways); or (3) analyzing videos well could lead to more conceptual teaching but through a complicated path (maybe analyzing video helps focus teachers' attention on key learning moments during a lesson which, in turn, helps them plan lessons with these moments in mind which, in turn, shifts their emphasis to engaging students in these moments which, in turn, results in more conceptual instruction).

Simple correlational designs involve investigating and explaining relationships between just two variables. But simple correlations can get complicated quickly. Researchers might, for example, hypothesize the relationship exists only under particular conditions—when other factors are controlled. In these situations, researchers often remove the effect of these variables and investigate the "partial correlations" between the two variables of primary interest. Many sophisticated statistical techniques have been developed for investigating more complicated relationships between multiple variables (e.g., exploratory and confirmatory factor analysis, Gorsuch, 2014).

Correlational Designs We Recommend The correlational designs we recommend are those that involve collecting data to test your predictions about the extent of the relationship between two (or more) variables and assess how well your rationales (theoretical framework) explain why these relationships exist. By predicting the extent of the relationships and formulating rationales for the degree of the relationships, the findings will help you adjust your predictions and revise your rationales.

Because correlations often involve multiple variables, your rationales might have proposed which variables are most important for, or best explain, the relationship. The findings could help you revise your thinking about the roles of different variables in determining the observed relationship.

For example, analyzing videos insightfully could be unpacked into separate variables, such as the nature of the video, the aspects of the video that could be attended to, and the knowledge needed to comment on each aspect. Teaching conceptually could also be unpacked into many individual variables. To explain or *understand* the predicted relationship, you would need to study which variables are most responsible for the relationship.

Some researchers suggest that correlational designs precede experimental designs (Sloane, 2008). The logic is that correlational research can document that relationships exist and can reveal the key variables. This information can enable the development of rationales for why changes in one construct or variable might cause changes in another construct or variable.

HELPFUL TIP

Our previous tip was to plan and conduct a small experimental study before a large one. Conducting a correlation study can serve a similar purpose—To work out the conceptual issues so you know what variables are critical and you have clear conjectures about the mechanisms that could account for the relationships among the variables.

Description

In some ways, descriptions are the most basic design. They are tailored to describe a phenomenon and then explain why it exists as it does. If the research questions ask about the status of a situation or about the nature of a phenomenon and there is no interest, at the moment, in trying to change something or to relate one thing with another, then a descriptive design is appropriate. For example, researchers might be interested in describing the ways in which teachers analyze video clips of classroom instruction or in describing the nature of conceptual teaching in a particular school district.

In this type of research, researchers would predict what they expect to find, and rationales would explain why these findings are expected. As an example, consider the case above of researchers describing the ways teachers analyze video clips of classroom instruction. If Martha had access to such a description and an explanation for why teachers analyzed videos in this way, she could have used this information to formulate her hypotheses regarding the relationship between analysis of videos and conceptual teaching (see Chap. 3). Based on the literature describing what teachers notice when observing classroom instruction (e.g., Sherin et al., 2001) and on the researchers' experience working with teachers to explain why they notice particular features, researchers might predict that many teachers will focus more on specific pedagogical skills of the teacher, such as classroom management and organization, and less on the nature of the content being discussed and the strategies students use to solve problems. If these predictions are partially confirmed, the predictions and their rationales would support the rationale for Martha's hypothesis of a growing relationship between analyzing videos and conceptual teaching as teachers move from focusing on pedagogical skills to focusing on the way in which students are interacting with the content.

In some research programs, descriptive studies logically precede correlation studies (Sloane, 2008). Until researchers know they can describe, say, conceptual teaching, there is no point in asking how such teaching relates to other variables (e.g., analyzing videos of teaching) or how to improve the level of conceptual teaching.

As with other designs, there are several types of descriptive studies. We encourage you to read more about the details of each (in, e.g., Miles et al., 2014; de Freitas et al., 2017).

Case Study A case study is usually defined as the in-depth study of a particular instance or of a single unit or case. The instance must be identifiable with clear boundaries and must be sufficiently meaningful to warrant detailed observation, data collection, and analysis. At the outset, you need to describe what the case is a case of. The goal is to understand the case—how it works, what it means, why it looks like it does—within the context in which it functions. To describe conceptual teaching more fully, for example, researchers might investigate a case of one teacher teaching several lessons conceptually.

Some researchers use a case study to show something *exists*. For example, suppose a researcher notices that students change the way they think about two-dimensional geometric figures after studying three-dimensional objects. The researcher might propose a concept of backward transfer (Hohensee, 2014) and design a case study with a small group of students and a targeted set of instructional activities to study this phenomenon in detail. The goal is to determine whether this effect exists and to explain its existence by identifying some of the conditions under which it occurs. Notice that this example also could be considered a "teaching experiment." There are overlaps between some designs and boundaries between them are not always clear.

Ethnography The term "ethnography" often is used to name a variety of research approaches that provide detailed and comprehensive accounts of educational phenomena. The approaches include participant observation, fieldwork, and even case studies. For a useful example, see Weisner et al. (2001). See the following for further descriptions of ethnographic research from various perspectives (Atkinson et al., 2007; Denzin & Lincoln, 2017).

Survey Survey designs are used to gather information from groups of participants, often large groups that fit specific criteria (e.g., fourth-grade teachers in Delaware), to learn about their characteristics, opinions, attitudes, and so on. Usually, surveys are conducted by administering a questionnaire, either written or oral. The responses to the questions form the data for the study. See Wolf et al. (2016) for more complete descriptions of survey methodology.

Like for previous designs, we recommend that each of these designs be used to test predictions about what will be found and assess the soundness of the rationales for these predictions. In all these settings, the goal remains to understand and explain what you are studying.

HELPFUL

TIP

If your goal is to gain insights into why participants in your study are responding in particular ways, surveys will probably not be the best design. Surveys usually rely on written responses gathered at one point in time. If your questions or tasks are not phrased exactly right or if a participant misinterprets the item, the data might not be helpful, and you will not have a chance to follow-up.

Developing Measures and Procedures for Gathering Data

This a critical phase of crafting your methods because your study is only as good as the quality of the data you gather. And, the quality of data is determined by the measures you use. "Measures" means tests, questionnaires, observation instruments, and anything else that generates data. The research methods textbooks and other resources we cited above include lots of detail about this phase. However, we will note a few issues that journal reviewers often raise and that we have found are problematic for beginning researchers.

Craft Measures That Produce Data at an Appropriate Grain Size

A critical step in the scientific inquiry process is comparing the results you find with those you predicted based on your rationales. Thinking ahead about this part of the process (see Chap. 3) helps you see that, for this comparison to be useful for revising your hypotheses, the predictions you make must be at the same level of detail, or grain size, as the results. If your predictions are at too general of a level, you will not be able to make this comparison in a meaningful way. After making predictions, you must craft measures that generate data at the same grain size as your predictions.

To illustrate, we return to Martha, the doctoral student investigating "What are the relationships between learning to analyze videos of teaching in particular ways (specified from prior research) and teaching for conceptual understanding?" In Chap. 3, one of Martha's predictions was: "Of the video analysis skills that will be assessed, the two that will show the strongest relationship are spontaneously describing (1) the mathematics that students are struggling with and (2) useful suggestions for how to improve the conceptual learning opportunities for students." To test this prediction, Martha will need to craft measures that assess separately different kinds of responses when analyzing the videos. Notice that in her case, the predictions are precise enough to specify the nature and grain size of the data that must be collected (i.e., the measures must yield information on the teachers' spontaneous descriptions of the mathematics that students are struggling with plus their suggestions for how to improve conceptual learning opportunities for students).

Develop Your Own Measures or Borrow from Others?

When crafting the measures for gathering data, weigh carefully the benefits and costs of designing your own measures versus using measures designed and already used by other researchers.

The benefits of developing your own measures come mostly from targeting your measures to assess exactly what you need so you can test your predictions. Sometimes, creating your own measures is critical for the success of your study.

Weigh carefully the benefits and costs of designing your own measures versus using measures designed and already used by other researchers.

However, there also are costs to consider. One is convincing others that your measures are both reliable and valid. In general, reliability of a measure refers to how consistently it will yield the same outcomes; validity means how accurately the measure assesses what you say you are measuring (see Gournelos et al., 2019). Establishing reliability and validity for new measures can be challenging and expensive in terms of time and resources.

A second cost of creating your own measures is not being able to compare your data to those of other researchers who have studied similar phenomenon. Knowledge accumulates as researchers build on the work of others and extend and refine hypotheses. This is partially enabled by comparing results across different studies that have addressed similar research questions. When you formulate hypotheses that extend previous research, it is often natural (and even obvious) to borrow measures that were used in previous studies. Consider Martha's predictions described in Chap. 3, one of which is presented above. Because the prediction builds directly on previous work, testing the predictions would almost require Martha to use the same measures used previously.

If you find it necessary to design your own measures, you should ask yourself whether you are reaching too far beyond previous work. Maybe you could tie your work more closely to past research by tweaking your research questions and hypotheses so existing, validated measures are what you need to test your predictions. In other words, use the time when you are crafting measures as a chance to ask whether you are extending previous research in the most productive way. If you decide to keep your original research questions and design new measures, we recommend considering a combination of previously validated measures and your own custom-made measures.

Whichever approach you choose, be sure to describe your measures in enough detail that others can use them if they are studying related phenomenon or if they would like to replicate your study. Also, if you use measures developed by others be sure to credit them.

Using Data that Already Exist

Most educational researchers collect their own data as part of the study. We have written the previous sections assuming this is the case. Is it possible to conduct an important study using data that have been collected by someone else? Yes. But we suggest you consider the following issues if you are planning a study using an existing set of data.

First, we recommend that your study begin with a hypothesis or research question, just like for a study in which you collect your own data. A common warning about choosing research methods is that you should *not* choose a method (e.g., hierarchical linear modeling) and then look for a research question. Your hypotheses, or research questions, should drive everything else. Similarly for choosing data to analyze. The data should be chosen because they are the best data to test your hypothesis, not because they exist.

Of course, you might be familiar with a data set and wonder what it would tell you about a particular research problem. Even in this case, however, you should formulate a hypothesis that is important on its own merits. It is easy to tell whether this is true by sharing your hypothesis with colleagues who are not aware of the existing data set and asking them to comment on the value of testing the hypothesis. Would a tested and revised hypothesis make a contribution to the field?

> *You should* not *choose a method and then look for a research question. Your hypotheses, or research questions, should drive everything else.*

A second issue to consider when using existing data is the alignment of the methods used to collect the data and your theoretical framework. Although you didn't choose the methods, you need to be familiar with the methods that were used and be able to justify the appropriateness of the methods, just as you would with methods you craft. Justifying the appropriateness of methods is another way of saying you need to convince others you are using the best data possible to test your hypotheses. As you read the remaining sections of this chapter, think about what you would need to do if you use existing data. Could you satisfy the same expectations as researchers who are collecting their own data?

Exercise 4.2
There are several large data sets that are available to researchers for secondary analyses, including data from the National Assessment of Educational Progress (NAEP), the Programme for International Student Assessment (PISA), and the Trends in International Mathematics and Science Study (TIMSS). Locate a published empirical study that uses an existing data set and clearly states explicit hypotheses or research questions. How do the authors justify their use of the existing data set to address their hypotheses or research questions? What advantages do you think the authors gained by choosing to use existing data? What constraints do you think that choice placed on them?

Choosing Methods to Analyze Data and Compare with Predictions

As with the first two phases of crafting your methods, there are a number of sources that describe issues to think about when putting together your data analysis strategies (e.g., de Freitas et al., 2017; Sloane & Wilkins, 2017). Beyond what you will read in these sources, or to emphasize some things you might read, we identify a few issues that you should attend to with extra care.

Create Coding Rubrics

Frequently, research in education involves collecting data in the form of interview responses by participants (students, teachers, teacher educators, etc.) or written responses to tasks, problems, or questionnaires, as well as in other forms that researchers must interpret before conducting analyses. This interpretation process is often referred to as *coding* data, and coding requires developing a rubric that describes, in detail, how the responses will be coded.

There are two main reasons to create a rubric. First, you must code responses that have the same meaning in the same way. This is sometimes called *intracoder* *reliability*: an individual coder is coding similar responses consistently. Second, you must communicate to readers and other researchers exactly how you coded the responses. This helps them interpret your data and make their own decisions about whether your claims are warranted. Recall from Chap. 1 an implication of the third descriptor of scientific inquiry which pointed to the public nature of research: "It is a public practice that occurs in the open and is available for others to see and learn from."

As you code, you will almost always realize that the initial definitions you created for your codes are insufficient to make borderline judgments, and you will need to revise and elaborate the coding rubric. For example, you might decide to split a code into several codes because you realize that the responses you were coding as similar are not as similar as you initially thought. Or you might decide to combine codes that at first seemed to describe different kinds of responses but you now realize are too hard to distinguish reliably. This process helps you clarify for yourself exactly what your codes mean and what the data are telling you.

Determine Intercoder Reliability

In addition to ensuring that you are coding consistently with yourself, you must make sure others would code the same way *if* they followed your rubric. Determining *intercoder reliability* involves training someone else to use your rubric to code the same responses and then comparing codes for agreement. There are several ways to calculate intercoder reliability (see, e.g., Stemler, 2004).

There are two main reasons to determine intercoder reliability. First, it is important to convince readers that the rubric holds all the information you used to code the responses. It is easy to use lots of implicit knowledge to code responses, especially if you are familiar with the data (e.g., if you conducted the interviews). Using implicit knowledge to code responses hides from others why you are coding responses as you are. This creates bias that interferes with the principles of scientific inquiry (being open and transparent). Establishing acceptable levels of intercoder reliability shows others that the knowledge made explicit in the rubric is all that was needed to code the responses.

A second reason to determine intercoder reliability is that doing so improves the completeness and specificity of the definitions for the codes. As you compare your coding with that of another coder, you will realize that your definitions were not as clear as you thought. You can learn what needs to be added or revised so the definition is clearer; sometimes this includes examples to help clarify the boundary between one code and another. As you reach sufficient levels of agreement, your rubric will reach its final version. This is the version that you will likely include as an appendix in a written report of your study. It tells the reader what each code means.

Beyond the Three Phases

We have discussed three phases of crafting methods (choosing the design of your study, developing the measures and procedures you need to gather the data, and selecting the analysis procedures to compare your findings with your predictions). There are some issues that cut across all three phases. You will read about some of these in the sources we suggested, but several could benefit from special attention.

Quantitative and Qualitative Data

For some time, educators have debated the value of quantitative versus qualitative data (Hart et al., 2008). As the labels suggest, quantitative data refers to data that can be expressed with numbers (frequencies, amounts, etc.). Most of the common statistical analyses require quantitative data. Qualitative data are not automatically transformed into numbers. Coding of qualitative data, as described above, can produce numbers (e.g., frequencies) but the data themselves are often words—written or spoken. Corresponding to these two forms of data, some types of research are referred to as quantitative research and some types as qualitative. As an easy reference point, experimental and correlational designs often foreground quantitative data and descriptive designs often foreground qualitative data. We recommend keeping several things in mind when reading about these two types of research.

First, it is best *not* to begin developing a study by saying you want to do a quantitative study or a qualitative study. We recommend, as we did earlier, that you begin with questions or hypotheses that are of most interest and then decide whether the methods that will best test your predictions require collecting quantitative or qualitative data.

Second, many hypotheses in education are best examined using both kinds of data. You are not limited to using one or the other. Often, studies that use both are referred to as mixed methods studies. Our guess is that if you are investigating an important hypothesis, your study could take advantage of, and benefit from, mixed methods (Hay, 2016; Weis et al. 2019a). As we noted earlier, different methods offer different perspectives so multiple methods are more likely to tell a more complete story (Sechrest et al., 1993). Some useful resources for reading about quantitative, qualitative, *and* mixed methods are Miles et al. (2014); de Freitas et al. (2017); Weis et al. (2019b); Small (2011); and Sloane and Wilkins (2017).

Defining a Unit of Analysis

The unit of analysis in your study is the "who" or the "what" that you are analyzing and want to make claims about. There are several ways in which this term is used. Your unit of analysis could be an individual student, a group of students, an individual task, a classroom, and so forth. It is important to understand that, in these cases, your unit of analysis might not be the same as your unit of observation. For example, you might gather data about individual students (unit of observation) but then compare the averages among groups of students, say in classrooms or schools (unit of analysis).

Unit of analysis can also refer to what is coded when you analyze qualitative data. For example, when analyzing the transcript of an interview or a classroom lesson, you might want to break up the transcript into segments that focus on different topics, into turns that each speaker takes, into sentences or utterances, or into other chunks. Again, the unit of analysis might not be the same as your unit of observation (the unit in which your findings are presented).

We recommend keeping two things in mind when you consider the unit of analysis. First, it is not uncommon to use more than one unit of analysis in a study. For example, when conducting a textbook analysis, you might use "page" as a unit of analysis (i.e., you treat each page as a single, separate object to examine), and you might also use "instructional task" as a unit of analysis (i.e., you treat each instructional task as a single object to examine, whether it takes up less than one page or many pages). Second, when the data collected have a nested nature (e.g., students nested in classrooms nested in schools), it is necessary to determine what is the most appropriate unit of analysis. Readers can refer to Sloane and Wilkins (2017) for a more detailed discussion of such analyses.

Ensuring Your Methods Are Fair to All Students

Regardless of which methods you use, remember they need to help you fulfill the purpose of your study. Suppose, as we suggested in earlier chapters, the purpose furthers the goal of understanding how educators can improve the learning opportunities for *all* students. It is worth thinking, separately, about whether the methods you are using are fully inclusive and are not (unintentionally) leading you to draw conclusions that systematically ignore groups of students with specific characteristics—race, ethnicity, gender, sexual orientation, and special education needs.

For example, if you want to investigate the correlation between students' participation in class and their sense of efficacy for the subject, you need to include students at different levels of achievement, with different demographics, with different entry efficacy levels, and so on. Your hypotheses should be perfectly clear about which variables that might influence this correlation are being included in your design. This issue is also directly related to our concern about generalizability: it would be inappropriate to generalize to populations or conditions that you have not accounted for in your study.

Researchers in education and psychology have also considered methodological approaches to ensure that research does not unfairly marginalize groups of students. For example, researchers have made use of back translation to ensure the translation equivalency of measures when a study involves students using different languages. Jonson and Geisinger (2022) and Zieky (2013) discuss ways to help ensure the fairness of educational assessments.

Part III. Crafting the Most Appropriate Methods

With the background we developed in Part III, we can now consider how to craft the methods you will use. In Chap. 3, we discussed how the theoretical framework you create does lots of work for you: (1) it helps you refine your predictions and backs them up with sound reasons or explanations; (2) it provides the parameters within which you craft your methods by providing clear rationales for some methods but not others; (3) it ensures that you can interpret your results appropriately by comparing them with your predictions; and, (4) it describes how your results connect with the prior research you used to build the rationales for your hypotheses. In this part of Chap. 4, we will explore the ways in which your theoretical framework guides, and even determines, the methods you craft for your study.

In Chap. 3, we described a cyclical process that produced the theoretical framework: asking questions, articulating predictions, developing rationales, imagining testing predictions, revising questions, adjusting rationales, revising predictions, and so on, and so on. We now extend this process beyond imagining how you could test your predictions.

The best way to craft appropriate methods that you will use is to try them out. Instead of only imagining how you could test your predictions, the cyclical process we described in Chap. 3 will be extended to trying out the methods you think you will use. This means trying out the measures you plan to use, the coding rubric (if you are coding data), the ways in which you will collect data, and how you will analyze data. By "try out" we mean a range of activities.

Write Out Your Methods

The first way you should try out your methods is by writing them out for yourself (*actually writing them out*) and then asking yourself two main questions. First, do the reasons or rationales in the theoretical framework point to using these specific measures, this coding rubric, and so forth? In other words, would anyone who reads your theoretical framework be the least bit surprised that you plan to use *these* methods? They should not be. In fact, you would expect anyone who read your theoretical framework to choose from the same set of reasonable, appropriate methods. If you plan to use methods for reasons other than those you find in your theoretical framework (perhaps because the framework is silent about this part of your study) or if you are using methods that are different from what would be expected, you probably need to either revise your framework (maybe to fill in some gaps or revise the arguments you make) or change your methods.

A second question to ask yourself after you have written a description of your methods is: "Can I imagine using these methods to generate data I could compare with my predictions?" Are the grain sizes similar? Can you plan how you will compare the data with the predictions? If you are unsure about this, you should consider changing your predictions (and your hypotheses and theoretical rationales) or changing your methods.

As described in Chap. 3, your writing will serve two purposes. It will help you think through and reflect on your methods, trying them out in your head. And it will also constitute another part of your evolving research paper that you create while you are designing, conducting, and then documenting your research study. Writing is a powerful tool for thinking as well as the most common form of communicating your work to others. So, the writing you do here is not just scratch work that you will discard. It should be a draft for what will become your final research paper. Treat it seriously. That said, it is still just a draft; do not take it *so* seriously that you find yourself stuck and unable to put words to paper because you are not certain what you are writing is good enough.

Ask Others

The second way you can try out your methods is to solicit feedback and advice from other people. Scientific inquiry is not only an individual process but a social process as well (recall again the third descriptor of scientific inquiry in Chap. 1). Doing good scientific inquiry requires the assistance of others. It is impossible to see everything you will need to think about by yourself; you need to present your ideas and get feedback from others. Here are several things to try.

First, if you are a doctoral student, describe your planned methods to your advisor. That is probably already your go-to strategy. If you are a beginning professor, you can seek advice from former and current colleagues.

Second, try out your ideas by making a more formal presentation to an audience of friendly critics (e.g., colleagues). Perhaps you can invite colleagues to a special "seminar" in which you present your study (without the results). Ask for suggestions, maybe about specific issues you are struggling with and about any aspects of your study that could be clarified and even revised. You do not need to have the details of your methods worked out before showing your preliminary plans to your colleagues. If your research questions and initial predictions are clear, getting feedback on your preliminary plans (design, measures, and data analysis) can be very helpful and can prevent wasting time on things you will end up needing to change. We recommend getting feedback earlier rather than later and getting feedback in multiple settings multiple times.

Finally, regardless of your current professional situation, we encourage you to join, or create, a *community of learners* who interact regularly. Such communities are not only intellectually stimulating but socially supportive.

Exercise 4.3
Ask a few colleagues to spend 45–60 min with you. Present your study as you have imagined it to this point (20 min): Research questions, predictions about the answers, rationales for your predictions (i.e., your theoretical framework), and methods you will use to test your predictions (design, measures, data collection, and data analysis to check your predictions). Ask for their feedback (especially about the methods you will use, but also about any aspect of the planned study). Presenting all this information is challenging but is good practice for thinking about the most critical pieces of your plan and your reasons for them. Use the feedback to revise your plan.

Conduct Pilot Studies

The value of conducting small, repeated, pilot studies cannot be overstated. It is hugely undervalued in most discussions of crafting methods for research studies. Conducting pilot studies is well worth the time and effort. It is probably the best way to try out the methods you think will work.

Conducting pilot studies is probably the best way to try out the methods you think will work.

Pilot studies can be quite small, both in terms of time spent and number of participants. You can keep pilot studies small by using a very small sample of participants or a small sample of your measures. The sample of participants can be participants who are easy to find. Just try to select a small sample that represents the larger sample you plan to use. Then, see if the data you collect are like those you expected and if these data will test your predictions in the way you hoped. If not, you are likely to find that your methods are not aligned well enough with your theoretical framework. Even one pilot study can be very useful and save you tons of time; several follow-up pilots are even better because you can check whether your revisions solved the problem. Do not think of pilot studies as speed bumps that slow your progress but rather as course corrections that help you stay aimed squarely at your goal and save you time in the long run.

Small pilot studies can be conducted for various purposes. Here are a few.

Help Specify Your Predictions

Pilot studies can help you specify your predictions. Sometimes it might be difficult to anticipate the answers to your research questions. Rather than conducting a complete study with little idea of what will happen, it is much more productive to do some preliminary work to help you formulate predictions. If you conduct your study without doing this, you are likely to realize too late that your study could have been much more informative if you used a different sample of participants, if you asked different or additional questions during your interviews, if you used different measures (or tasks) to gather the data, if your data looked different so you could have used different analyses, and so forth.

In our view, this is an especially important use of pilot studies because it is our response to the argument we rebutted earlier that asserted research can be productive even if researchers have no idea what to expect and cannot make testable predictions. Throughout this book, we have argued that scientific inquiry requires predictions and rationales, regardless how weak or uncertain. We have claimed that, if the research is worth doing, it is possible and productive to make predictions. It is hard for us to imagine conducting research that builds on past work yet having no idea what to expect. If a researcher is charting new territory, then pilot studies are essential. Conducting one or more small pilot studies will provide some initial guesses and should trigger some ideas for why these guesses will be correct. As we noted earlier, however, we do not recommend beginning researchers chart completely new territory.

Improve Your Predictions

Even if you have some predictions, conducting a pilot study or two will tell you whether you are close. The more accurate you are with your predictions for the main study, the more precisely you can revise your predictions after the study and formulate very good explanations for why these new predictions should be accurate.

Refine Your Measures

Pilot studies can be very useful for making sure your measures will produce the kinds of data you need. For example, if your study includes participants who are asked to complete tasks of various kinds, you need to make sure the tasks generate the information you need.

Suppose you ask whether second graders improve their understanding of place value after an instructional intervention. You need to use tasks that help you interpret how well they understand place value before and after the intervention. You might ask two second graders and two third graders to complete your tasks to see if they generate the expected variation in performance and whether this variation can be tied to inferred levels of understanding. Also, ask a few colleagues to interpret the responses and check if they match with your interpretations.

Suppose you want to know whether middle school teachers interact differently with boys and girls about the most challenging problems during math class. Find a lesson or two in the curriculum that includes challenging problems and sit in on these lessons in several teachers' classrooms. Test whether your observation instrument captures the differences that you think you notice.

Test Your Analytic Procedures

You can use small pilot studies to check if your data analysis procedures will work. This can be extremely useful if your procedures are more than simple quantitative comparisons such as *t* tests. Suppose you will conduct interviews with teachers and code their responses for particular features or patterns. Conducting two or three interviews and coding them can tell you quickly whether your coding rubric will work. Even more important, coding the interviews will tell you whether the interview questions are the right ones or whether they need to be revised to produce the data you need.

Other Purposes of Pilot Studies

In addition to the purposes we identified above, pilot studies can tell you whether the sample you identified will give you the information you need, whether your measures can be administered in the time you allocated, and whether other details of your data collection and analysis plans work as you hope. In summary, pilot studies allow you to rehearse your methods so you can be sure they will provide a strong test of your predictions.

After you conduct a pilot study, make the revisions needed to the framework or to the methods to ensure you will gather more informative data. Be sure to update your evolving research paper to reflect these changes. Each draft of this paper should be the draft which matches your current reasoning and decisions regarding your study.

> *Pilot studies allow you to rehearse your methods so you can be sure they will provide a strong test of your predictions.*

Part IV. Writing Your Evolving Research Paper and Revisiting Alignment

We continue here to elaborate our recommendation that you compose drafts of your evolving research paper as you make decisions along the way. It is worth describing several advantages in writing the paper and planning the study in parallel.

Advantages of Writing Your Research Paper While Planning Your Study

One of the major challenges researchers face as they plan and conduct research studies is aligning all parts of the study with a visible and tight logic tying all the parts together. You will find that as you make decisions about your study and write about these decisions, you are faced with this alignment challenge in both settings. Working out the alignment in one setting will help in the other. They reinforce each other. For example, as you write a record of your decisions while you plan your study, you might notice a gap in your logic. You can then fill in the gap, both in the paper and in the plans for the study.

As we have argued, writing is a useful tool for thinking. Writing out your questions and your predictions of the answers helps you decide if the questions are the ones you really want to ask and if your predictions are testable; writing out your rationales for your predictions helps you decide if you have sound reasons for your predictions, and if your theoretical framework is complete and convincing; writing out your theoretical rationales also helps you decide which methods will provide a strong test of your predictions.

Your evolving research paper will become the paper you will use to communicate your study to others. Writing drafts as you make decisions about how to conduct your study and why to conduct it as you did will prevent you from needing to reconstruct the logic you used as you planned each successive phase of your study. In addition, composing the paper as you go ensures that you consider the logic connecting each step to the next one. One of the major complaints reviewers are likely to have is that there is a lack of alignment. By following the processes we have described, you have no choice but to find, in the end, that all parts of the study are connected by an obvious logic.

We noted in Chap. 3 that writing your evolving research paper along with planning and conducting your study does not mean creating a chronology of all the decisions you made along the way. At each point in the process, you should step back and think about how to describe your work in the easiest-to-follow and clearest way for the reader. Usually, readers want to know only about your final decisions and, in many cases, your reasons for making these decisions.

Journal Reviewers' Common Concerns

The concerns of reviewers provide useful guides for where you need to be especially careful to conduct a well-argued and well-designed study and to write a coherent paper reporting the study. As the editorial team for *JRME*, we found that one of the most frequent concerns raised by reviewers was that the research questions were not well connected to other parts of the paper. Of all manuscripts sent out for review, nearly 30% of the reviewers expressed concern that the paper was not coherent because parts of the paper were not connected back to the research questions. This could mean, for example, reviewers were not clear why or how the methods crafted for the study were appropriate to test the hypotheses or to answer the questions. The lack of clear connections could be due to either choices made planning and implementing the study or writing the research paper. Sometimes the connections exist but have been left implicit in the research report or even in the conceptualization of the study. Conceptualizing a study and writing the research report require making all the connections explicit. As noted above, these disconnects are less likely if you are composing the evolving research paper simultaneously with planning and implementing the study.

A further concern raised by many reviewers speaks to alignment and coherence: One or more of the research questions were not answered fully by the study. Although we will deal with this concern further in the next chapter, we believe it is relevant for the choice of methods because if you do not ensure that the methods are appropriate to answer your research questions (i.e., to test your hypotheses), it is likely they will not generate the data you need to answer your questions. In contrast, if you have aligned all parts of your study, you are likely to collect the data you need to answer your questions (i.e., to test and revise your hypotheses).

In summary, there are many reasons to compose your evolving research paper along with planning and conducting your study. As we have noted several times, your paper will not be a chronology of all the back-and-forth cycles you used to refine aspects of your study as you moved to the next phase, but it will be a faithful description of the ultimate decisions you made and your reasons for making them. Consequently, your evolving research paper will gradually build as you describe the following parts and explain the logic connecting them: (1) the purpose of your study, (2) your theoretical framework (i.e., the rationales for your predictions woven into a coherent argument), (3) your research questions plus predictions of the answers (generated directly from your theoretical rationales), (4) the methods you used to test your predictions, (5) the presentation of results, and (6) your interpretation of results (i.e., comparison of predicted results with the results reported plus proposed revisions to hypotheses). We will continue the story by addressing parts 5 and 6 in Chap. 5.

Chapter 5
Significance of a Study: Revisiting the "So What" Question

Part I. Setting the Groundwork

One of the most common questions asked of researchers is "So what?" What difference does your study make? Why are the findings important? The "so what" question is one of the most basic questions, often perceived by novice researchers as the most difficult question to answer. Indeed, addressing the "so what" question continues to challenge even experienced researchers. It is not always easy to articulate a convincing argument for the importance of your work. It can be especially difficult to describe its importance *without* falling into the trap of making claims that reach beyond the data.

That this issue is a challenge for researchers is illustrated by our analysis of reviewer comments for *JRME*. About one-third of the reviews for manuscripts that were ultimately rejected included concerns about the importance of the study. Said another way, reviewers felt the "So what?" question had not been answered. To paraphrase one journal reviewer, "The manuscript left me unsure of what the contribution of this work to the field's knowledge is, and therefore I doubt its significance." We expect this is a frequent concern of reviewers for all research journals.

Our goal in this chapter is to help you navigate the pressing demands of journal reviewers, editors, and readers for demonstrating the importance of your work while staying within the bounds of acceptable claims based on your results. We will begin by reviewing what we have said about these issues in previous chapters. We will then clarify one of the confusing aspects of developing appropriate arguments—the absence of consensus definitions of key terms such as significance, contributions, and implications. Based on the definitions we propose, we will examine the critical role of alignment for realizing the potential significance of your study. Because the importance of your study is communicated through your evolving research paper, we will fold suggestions for writing your paper into the discussion of creating and executing your study.

J. Hiebert et al., *Doing Research: A New Researcher's Guide*,
Research in Mathematics Education,
https://doi.org/10.1007/978-3-031-19078-0_5

A confusing aspect of developing appropriate arguments is the absence of consensus definitions of key terms such as significance, contributions, and implications.

We laid the groundwork in Chap. 1 for what we consider to be important research in education:

> In our view, the ultimate goal of education is to offer all students the best possible learning opportunities. So, we believe the *ultimate* purpose of scientific inquiry in education is to support the improvement of learning opportunities for all students…. If there is no way to imagine a connection to improving learning opportunities for students, even a distant connection, we recommend you reconsider whether it is an important hypothesis within the education community.

Of course, you might prefer another "ultimate purpose" for research in education. That's fine. The critical point is that the argument for the importance of the hypotheses you are testing should be connected to the value of a long-term goal you can describe. As long as most of the educational community agrees with this goal, and you can show how testing your hypotheses will move the field forward to achieving this goal, you will have developed a convincing argument for the importance of your work.

In Chap. 2, we argued the importance of your hypotheses can and should be established before you collect data. Your theoretical framework should carry the weight of your argument because it should describe how your hypotheses will extend what is already known. Your methods should then show that you will test your hypotheses in an appropriate way—in a way that will allow you to detect how the results did, and did not, confirm the hypotheses. This will, in turn, allow you to formulate revised hypotheses. We described establishing the importance of your study by saying, "The importance can come from the fact that, based on the results, you will be able to offer revised hypotheses that help the field better understand an issue relevant for improving all students' learning opportunities."

The ideas from Chaps. 1, 2, and 3 go a long way toward setting the parameters for what counts as an important study and how its importance can be determined. Chapter 4 focused on ensuring that the importance of a study can be realized. The next section fills in the details by proposing definitions for the most common terms used to claim importance: significance, contributions, and implications.

You might notice that we do not have a chapter dedicated to discussing the presentation of the findings—that is, a "results" chapter. We do not mean to imply that presenting results is trivial. However, we believe that if you follow our recommendations for writing your evolving research paper, presenting the results will be quite straightforward. The key is to present your results so they can be most easily compared with your predictions. This means, among other things, organizing your presentation of results according to your earlier presentation of hypotheses.

Part II. Clarifying Importance by Revisiting the Definitions of Key Terms

What does it mean to say your findings are significant? Statistical significance is clear. There are widely accepted standards for determining the statistical significance of findings. But what about educational significance? Is this the same as claiming that your study makes an important contribution? Or, that your study has important implications? Different researchers might answer these questions in different ways. When key terms like these are overused, their definitions gradually broaden or shift, and they can lose their meaning. That is unfortunate, because it creates confusion about how to develop claims for the importance of a study.

By clarifying the definitions, we hope to clarify what is required to claim that a study is *significant*, that it makes a *contribution*, and that it has important *implications*. Not everyone defines the terms as we do. Our definitions are probably a bit narrower or more targeted than those you may encounter elsewhere. Depending on where you want to publish your study, you may want to adapt your use of these terms to match more closely the expectations of a particular journal. But the way we define and address these terms is not antithetical to common uses. And we believe ridding the terms of unnecessary overlap allows us to discriminate among different key concepts with respect to claims for the importance of research studies. It is not necessary to define the terms exactly as we have, but it is critical that the ideas embedded in our definitions be distinguished and that all of them be taken into account when examining the importance of a study.

We will use the following definitions:

- *Significance:* The importance of the problem, questions, and/or hypotheses for improving the learning opportunities for all students (you can substitute a different long-term goal if its value is widely shared). Significance can be determined *before* data are gathered. *Significance is an attribute of the research problem, not the research findings*.
- *Contributions*: The value of the findings for revising the hypotheses, making clear what has been learned, what is now better understood.
- *Implications*: Deductions about what can be concluded from the findings that are not already included in "contributions." The most common deductions in educational research are for improving educational practice. Deductions for research practice that are not already defined as contributions are often suggestions about research methods that are especially useful or methods to avoid.

Significance

The significance of a study is built by formulating research questions and hypotheses you connect through a careful argument to a long-term goal of widely shared value (e.g., improving learning opportunities for all students). Significance applies both to the domain in which your study is located and to your individual study. The significance of the domain is established by choosing a goal of widely shared value and then identifying a domain you can show is connected to achieving the goal. For example, if the goal to which your study contributes is improving the learning opportunities for all students, your study might aim to understand more fully how things work in a domain such as teaching for conceptual understanding, or preparing teachers to attend to all students, or designing curricula to support all learners, or connecting learning opportunities to particular learning outcomes.

The significance of your individual study is something you *build*; it is not predetermined or self-evident. Significance of a study is established by making a case for it, not by simply choosing hypotheses everyone already thinks are important. Although you might believe the significance of your study is obvious, readers will need to be convinced.

Significance can be determined before data are gathered. Significance is an attribute of the research problem, not the research findings.

Significance is something you develop in your evolving research paper. The theoretical framework you present connects your study to what has been investigated previously. Your argument for significance of the domain comes from the significance of the line of research of which your study is a part. The significance of your study is developed by showing, through the presentation of your framework, how your study advances this line of research. This means the lion’s share of your answer to the “So what?” question will be developed as part of your theoretical framework.

Although defining significance as located in your paper prior to presenting results is not a definition universally shared among educational researchers, it is becoming an increasingly common view. In fact, there is movement toward evaluating the significance of a study based only on the first sections of a research paper—the sections prior to the results (Makel et al., 2021).

In addition to addressing the “So what?” question, your theoretical framework can address another common concern often voiced by readers: “What is so interesting? I could have predicted those results.” Predictions do not need to be surprising to be interesting and significant. The significance comes from the rationales that show how the predictions extend what is currently known. It is irrelevant how many researchers could have made the predictions. What makes a study significant is that the theoretical framework and the predictions make clear how the study will increase the field’s understanding toward achieving a goal of shared value.

What makes a study significant is that the theoretical framework and the predictions make clear how the study will increase the field's understanding toward achieving a goal of shared value.

An important consequence of interpreting significance as a carefully developed argument for the importance of your research study within a larger domain is that it reveals the advantage of conducting a series of connected studies rather than single, disconnected studies. Building the significance of a research study requires time and effort. Once you have established significance for a particular study, you can build on this same argument for related studies. This saves time, allows you to continue to refine your argument across studies, and increases the likelihood your studies will contribute to the field.

Contributions

As we have noted, in fields as complicated as education, it is unlikely that your predictions will be entirely accurate. If the problem you are investigating is significant, the hypotheses will be formulated in such a way that they extend a line of research to understand more deeply phenomena related to students' learning opportunities or another goal of shared value. Often, this means investigating the conditions under which phenomena occur. This gets complicated very quickly, so the data you gather will likely differ from your predictions in a variety of ways. The contributions your study makes will depend on how you interpret these results in light of the original hypotheses.

A study's contribution lies in the value of its findings for revising the hypotheses, making clear what has been learned, what is now better understood.

Contributions Emerge from Revisions to your Hypotheses

We view interpreting results as a process of comparing the data with the predictions and then examining the way in which hypotheses should be revised to more fully account for the results. Revising will almost always be warranted because, as we noted, predictions are unlikely to be entirely accurate. For example, if researchers expect Outcome A to occur under specified conditions but find that it does not occur to the extent predicted or actually does occur but without all the conditions, they must ask what changes to the hypotheses are needed to predict more accurately the conditions under which Outcome A occurred. Are there, for example, essential conditions that were not anticipated and that should be included in the revised hypotheses?

Consider an example from a recently published study (Wang et al., 2021). A team of researchers investigated the following research question: "How are students' perceptions of their parents' expectations related to students' mathematics-related beliefs and their perceived mathematics achievement?" The researchers predicted that students' perceptions of their parents' expectations would be highly related to students' mathematics-related beliefs and their perceived mathematics achievement. The rationale was based largely on prior research that had consistently found parents' general educational expectations to be highly correlated with students' achievement.

The findings showed that Chinese high school students' perceptions of their parents' educational expectations were positively related to these students' mathematics-related beliefs. In other words, students who believed their parents expected them to attain higher levels of education had more desirable mathematics-related beliefs.

However, students' perceptions of their parents' expectations about *mathematics* achievement were not related to students' mathematics-related beliefs in the same way as the more general parental educational expectations. Students who reported that their parents had *no* specific expectations possessed *more desirable* mathematics-related beliefs than all other subgroups. In addition, these students tended to perceive their mathematics achievement rank in their class to be higher on average than students who reported that their parents expressed some level of expectation for mathematics achievement.

Because this finding was not predicted, the researchers revised the original hypothesis. Their new prediction was that students who believe their parents have no specific mathematics achievement expectations possess more positive mathematics-related beliefs and higher perceived mathematics achievement than students who believe their parents do have specific expectations. They developed a revised rationale that drew on research on parental pressure and mathematics anxiety, positing that parents' specific mathematics achievement expectations might increase their children's sense of pressure and anxiety, thus fostering less positive mathematics-related beliefs. The team then conducted a follow-up study. Their findings aligned more closely with the new predictions and affirmed the better explanatory power of the revised rationale. The contributions of the study are found in this increased explanatory power—in the new understandings of this phenomenon contained in the revisions to the rationale.

Interpreting findings in order to revise hypotheses is not a straightforward task. Usually, the rationales blend multiple constructs or variables and predict multiple outcomes, with different outcomes connected to different research questions and addressed by different sets of data. Nevertheless, the contributions of your study depend on specifying the differences between your original hypotheses and your revised hypotheses. What can you explain now that you could not explain before?

We believe that revising hypotheses is an optimal response to any question of contributions because a researcher's initial hypotheses plus the revisions suggested by the data are the most productive way to tie a study into the larger chain of research of which it is a part. Revised hypotheses represent growth in knowledge. Building on other researchers' revised hypotheses and revising them further by more explicitly and precisely describing the conditions that are expected to influence the outcomes in the next study accumulates knowledge in a form that can be recorded, shared, built upon, and improved.

The significance of your study is presented in the opening sections of your evolving research paper whereas the contributions are presented in the final section, after the results. In fact, the central focus in this "Discussion" section should be a specification of the contributions (note, though, that this guidance may not fully align with the requirements of some journals).

Contributions Answer the Question of Generalizability

A common and often contentious, confusing issue that can befuddle novice and experienced researchers alike is the generalizability of results. All researchers prefer to believe the results they report apply to more than the sample of participants in their study. How important would a study be if the results applied only to, say, two fourth-grade classrooms in one school, or to the exact same tasks used as measures? How do you decide to which larger population (of students or tasks) your results could generalize? How can you state your claims so they are precisely those justified by the data?

To illustrate the challenge faced by researchers in answering these questions, we return to the *JRME* reviewers. We found that 30% of the reviews expressed concerns about the match between the results and the claims. For manuscripts that ultimately received a decision of Reject, the majority of reviewers said the authors had overreached—the claims were not supported by the data. In other words, authors generalized their claims beyond those that could be justified.

The Connection Between Contributions and Generalizability In our view, claims about contributions can be examined productively alongside considerations of generalizability. To make the case for this view, we need to back up a bit. Recall that the purpose of research is to understand a phenomenon. To understand a phenomenon, you need to determine the conditions under which it occurs. Consequently, productive hypotheses specify the conditions under which the predictions hold and explain why and how these conditions make a difference. And the conditions set the parameters on generalizability. They identify when, where, and for whom the effect or situation will occur. So, hypotheses describe the extent of expected generalizability, and revised hypotheses that contain the contributions recalibrate generalizability and offer new predictions within these parameters.

An Example That Illustrates the Connection In Chap. 4, we introduced an example with a research question asking whether second graders improve their understanding of place value after a specially designed instructional intervention. We suggested asking a few second and third graders to complete your tasks to see if they generated the expected variation in performance. Suppose you completed this pilot study and now have satisfactory tasks. What conditions might influence the effect of the intervention? After careful study, you developed rationales that supported three conditions: the entry level of students' understanding, the way in which the intervention is implemented, and the classroom norms that set expectations for students' participation.

Suppose your original hypotheses predicted the desired effect of the intervention *only if* the students possessed an understanding of several concepts on which place value is built, *only if* the intervention was implemented with fidelity to the detailed instructional guidelines, and *only if* classroom norms encouraged students to participate in small-group work and whole-class discussions. Your claims of generalizability will apply to second-grade settings with these characteristics.

Now suppose you designed the study so the intervention occurred in five second-grade classrooms that agreed to participate. The pre-intervention assessment showed all students with the minimal level of entry understanding. The same well-trained teacher was employed to teach the intervention in all five classrooms, none of which included her own students. And you learned from prior observations and reports of the classroom teachers that three of the classrooms operated with the desired classroom norms, but two did not. Because of these conditions, your study is now designed to test one of your hypotheses—the desired effect will occur *only if* classroom norms encouraged students to participate in small-group work and whole-class discussions. This is the only condition that will vary; the other two (prior level of understanding and fidelity of implementation) are the same across classrooms so you will not learn how these affect the results.

Suppose the classrooms performed equally well on the post-intervention assessments. You expected lower performance in the two classrooms with less student participation, so you need to revise your hypotheses. The challenge is to explain the higher-than-expected performance of these students. Because you were interested in understanding the effects of this condition, you observed several lessons in all the classrooms during the intervention. You can now use this information to explain why the intervention worked equally well in classrooms with different norms.

Your revised hypothesis captures this part of your study's contribution. You can now say more about the ways in which the intervention can help students improve their understanding of place value because you have different information about the role of classroom norms. This, in turn, allows you to specify more precisely the nature and extent of the generalizability of your findings. You now can generalize your findings to classrooms with different norms. However, because you did not learn more about the impact of students' entry level understandings or of different kinds of implementation, the generalizability along these dimensions remains as limited as before.

This example is simplified. In many studies, the findings will be more complicated, and more conditions will likely be identified, some of which were anticipated and some of which emerged while conducting the study and analyzing the data. Nevertheless, the point is that generalizability should be tied to the conditions that are expected to affect the results. Further, unanticipated conditions almost always appear, so generalizations should be conservative and made with caution and humility. They are likely to change after testing the new predictions.

Contributions Are Assured When Hypotheses Are Significant and Methods Are Appropriate and Aligned

We have argued that the contributions of your study are produced by the revised hypotheses you can formulate based on your results. Will these revisions always represent contributions to the field? What if the revisions are minor? What if your results do not inform revisions to your hypotheses?

We will answer these questions briefly now and then develop them further in Part IV of this chapter. The answer to the primary question is "yes," your revisions will always be a contribution to the field *if* (1) your hypotheses are significant and (2) you crafted appropriate methods to test the hypotheses. This is true even if your revisions are minor or if your data are not as informative as you expected. However, this is true *only if* you meet the two conditions in the earlier sentence. The first condition (significant hypotheses) can be satisfied by following the suggestions in the earlier section on significance. The second condition (appropriate methods) is addressed further in Part III in this chapter.

Implications

Before examining the role of methods in connecting significance with important contributions, we elaborate briefly our definition of "implications." We reserve implications for the conclusions you can logically deduce from your findings that are not already presented as contributions. This means that, like contributions, implications are presented in the Discussion section of your research paper.

Many educational researchers present two types of implications: implications for future research and implications for practice. Although we are aware of this common usage, we believe our definition of "contributions" cover these implications. Clarifying why we call these "contributions" will explain why we largely reserve the word "implications" for recommendations regarding methods.

Implications for Future Research

Implications for future research often include (1) recommendations for empirical studies that would extend the findings of this study, (2) inferences about the usefulness of theoretical constructs, and (3) conclusions about the advisability of using particular kinds of methods. Given our earlier definitions, we prefer to label the first two types of implications as contributions.

Consider recommendations for empirical studies. After analyzing the data and presenting the results, we have suggested you compare the results with those predicted, revise the rationales for the original predictions to account for the results, and make new predictions based on the revised rationales. It is precisely these new predictions that can form the basis for recommending future research. Testing these new predictions is what would most productively extend this line of research. It can sometimes sound as if researchers are recommending future studies based on hunches about what research might yield useful findings. But researchers can do better than this. It would be more productive to base recommendations on a careful analysis of how the predictions of the original study could be sharpened and improved.

Now consider inferences about the usefulness of theoretical constructs. Our argument for labeling these inferences as contributions is similar. Rationales for predictions are where the relevant theoretical constructs are located. Revisions to these rationales based on the differences between the results and the predictions reveal the theoretical constructs that were affirmed to support accurate predictions and those that must be revised. In our view, usefulness is determined through this revision process.

Implications that do not come under our meaning of contributions are in the third type of implications, namely the appropriateness of methods for generating rich contributions. These kinds of implications are produced by your evaluation of your methods: research design, sampling procedures, tasks, data collection procedures, and data analyses. Although not always included in the discussion of findings, we believe it would be helpful for researchers to identify particular methods that were useful for conducting their study and those that should be modified or avoided. We believe these are appropriately called implications.

Implications for Practice

If the purpose of research is to better understand how to improve learning opportunities for all students, then it is appropriate to consider whether implications for improving educational practice can be drawn from the results of a study. How are these implications formulated? This is an important question because, in our view, these claims often come across as an afterthought, "Oh, I need to add some implications for practice." But here is the sobering reality facing researchers: By any measure, the history of educational research shows that identifying these implications has had little positive effect on practice.

Perhaps the most challenging task for researchers who attempt to draw implications for practice is to interpret their findings for appropriate settings. A researcher who studied the instructional intervention for second graders on place value and found that average performance in the intervention classrooms improved more than in the textbook classrooms might be tempted to draw implications for practice. What should the researcher say? That second-grade teachers should adopt the intervention? Such an implication would be an overreach because, as we noted earlier, the findings cannot be generalized to all second-grade classrooms. Moreover, an improvement in average performance does not mean the intervention was better for all students.

The challenge is to identify the conditions under which the intervention would improve the learning opportunities for all students. Some of these conditions will be identified as the theoretical framework is built because the predictions need to account for these conditions. But some will be unforeseen, and some that are identified will not be informed by the findings. Recall that, in the study described earlier, a condition of entry level of understanding was hypothesized but the design of the study did not allow the researcher to draw any conclusions about its effect.

What can researchers say about implications for practice given the complexities involved in generalizing findings to other settings? We offer two recommendations. First, because it is difficult to specify all the conditions under which a phenomenon occurs, it is rarely appropriate to *prescribe* an educational practice. Researchers cannot anticipate the conditions under which individual teachers operate, conditions that often require adaptation of a suggested practice rather than implementation of a practice as prescribed.

Our second recommendation comes from returning to the purpose for educational research—to understand more fully how to improve learning opportunities for all students (or to achieve another goal of widely shared value). As we have described, understanding comes primarily from building and reevaluating rationales for your predictions. If you reach a new understanding related to improving learning opportunities, an understanding that could have practical implications, we recommend you share this understanding as an implication for practice.

For example, suppose the researcher who found better average performance of second graders after the intervention on place value had also studied several conditions under which performance improved. And suppose the researcher found that most students who did not improve their performance misunderstood a concept that appeared early in the intervention (e.g., the multiplicative relationship between positional values of a numeral). An implication for practice the researcher might share would be to describe the potential importance of understanding this concept early in the sequence of activities if teachers try out this intervention.

If you use our definitions, these implications for practice would be presented as contributions because they emerge directly from reevaluating and revising your rationales. We believe it is appropriate to use “Contributions” as the heading for this section in the Discussion section of your research paper. However, if editors prefer “Implications” we recommend following their suggestion.

We want to be clear that the terms you use for the different ways your study is important is not critical. We chose to define the terms significance, contributions, and implications in very specific and not universally shared ways to distinguish all the meanings of importance you should consider. Some of these can be established through your theoretical framework, some by the revisions of your hypotheses, and some by reflecting on the value of particular methods. The important thing, from our point of view, is that the ideas we defined for each of these terms are distinguished and recognized as specific ways of determining the importance of your study.

Part III. The Role of Methods in Determining Contributions

We have argued that every part of the study (and of the evolving research paper) should be aligned. All parts should be connected through a coherent chain of reasoning. In this chapter, we argue that the chain of reasoning is not complete until the methods are presented and the results are interpreted and discussed. The methods of the study create a bridge that connects the introductory material (research questions, theoretical framework, literature review, hypotheses) with the results and interpretations.

The role that methods play in scientific inquiry is to ensure that your hypotheses will be tested appropriately so the significance of your study will yield its potential contributions. To do this, the methods must do more than follow the standard guidelines and be technically correct (see Chap. 4). They must also fit with the surrounding parts of the study. We call this coherence.

The role that methods play in scientific inquiry is to ensure that your hypotheses will be tested appropriately so the significance of your study will yield its potential contributions.

Coherence Across the Phases of Scientific Inquiry

Coherence means the parts of a whole are fully aligned. When doing scientific inquiry, the early parts or phases should motivate the later phases. The methods you use should be motivated or explained by the earlier phases (e.g., research questions, theoretical framework, hypotheses). Your methods, in turn, should produce results that can be interpreted by comparing them with your predictions. Methods are aligned with earlier phases when you can use the rationales contained in your hypotheses to decide what kinds of data are needed to test your predictions, how

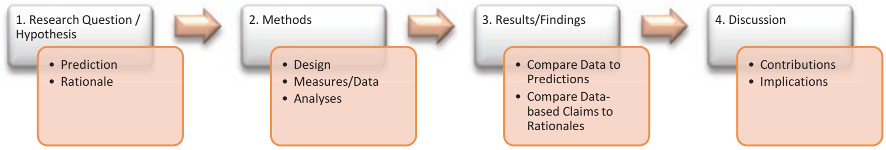

Fig. 5.1 The Chain of Coherence That Runs Through All Parts of a Research Study

best to gather these kinds of data, and what analyses should be performed (see Chap. 4 and Cai et al., 2019a).

For a visual representation of this coherence, see Fig. 5.1. Each box identifies an aspect of scientific inquiry. Hypotheses (shown in Box 1) include the rationales and predictions. Because the rationales encompass the theoretical framework and the literature review, Box 1 establishes the significance of the study. Box 2 represents the methods, which we defined in Chap. 4 as the entire set of procedures you will use, including the basic design, measures for collecting data, and analytic approaches. In Fig. 5.1, the hypothesis in Box 1 points you to the methods you will use. That is, you will choose methods that provide data for analyses that will generate results or findings (Box 3) that allow you to make comparisons against your predictions. Based on those comparisons, you will revise your hypotheses and derive the contributions and implications of your study (Box 4).

We intend Fig. 5.1 to carry several messages. One is that coherence of a study and the associated research paper require all aspects of the study to flow from one into the other. Each set of prior entries must motivate and justify the next one. For example, the data and analyses you intend to gather and use in Box 2 (Methods) must be those that are motivated and explained by the research question and hypothesis (prediction and rationale) in Box 1.

A second message in the figure is that coherence includes Box 4, "Discussion." Aligned with the first three boxes, the fourth box flows from these boxes but is also constrained by them. The contributions and implications authors describe in the Discussion section of the paper cannot go beyond what is allowed by the original hypotheses and the revisions to these hypotheses indicated by the findings.

HELPFUL TIP

For each hypothesis (and thus each research question) in your study, you should be able to trace an entire chain of coherence. In a complex study with multiple hypotheses (and thus multiple research questions), it can be extremely helpful to diagram these connections (or make a table of them) so that you can explicitly link each research question to the data collected for that question, to the analyses that will be conducted on those data to address that question, to the results obtained for that question, and finally to the contributions related to that question. A diagram or table of these links can help you to maintain coherence both while conducting your study and while writing your research paper.

Methods Enable Significance to Yield Contributions

We begin this section by identifying a third message conveyed in Fig. 5.1. The methods of the study, represented by Box 2, provide a bridge that connects the significance of the study (Box 1) with the contributions of the study (Box 4). The results (Box 3) indicate the nature of the contributions by determining the revisions to the original hypotheses.

In our view, the connecting role played by the methods is often underappreciated. Crafting appropriate methods aligned with the significance of the study, on one hand, and the interpretations, on the other, can determine whether a study is judged to make a contribution.

If the hypotheses are established as significant, and if appropriate methods are used to test the predictions, the study will make important contributions even if the data are not statistically significant. We can say this another way. When researchers establish the significance of the hypotheses (i.e., convince readers they are of interest to the field) and use methods that provide a sound test of these hypotheses, the data they present will be of interest regardless of how they turn out. This is why Makel et al. (2021) endorse a review process for publication that emphasizes the significance of the study as presented in the first sections of a research paper.

Treating the methods as connecting the introductory arguments to the interpretations of data prevent researchers from making a common mistake: When writing the research paper, some researchers lose track of the research questions and/or the predictions. In other words, results are presented but are not interpreted as answers to the research questions or compared with the predictions. It is as if the introductory material of the paper begins one story, and the interpretations of results ends a different story. Lack of alignment makes it impossible to tell one coherent story.

A final point is that the alignment of a study cannot be evaluated and appreciated if the methods are not fully described. Methods must be described clearly and completely in the research paper so readers can see how they flow from the earlier phases of the study and how they yield the data presented. We suggested in Chap. 4 a rule of thumb for deciding whether the methods have been fully described: "Readers should be able to replicate the study if they wish."

Part IV. Special Considerations that Affect a Study's Contributions

We conclude Chap. 5 by addressing two additional issues that can affect how researchers interpret the results and make claims about the contributions of a study. Usually, researchers deal with these issues in the Discussion section of their research paper, but we believe it is useful to consider them as you plan and conduct your study. The issues can be posed as questions: How should I treat the limitations of my study? How should I deal with findings that are completely unexpected?

Limitations of a Study

We can identify two kinds of limitations: (1) limitations that constrain your ability to interpret your results because of unfortunate choices you made, and (2) limitations that constrain your ability to generalize your results because of missing variables you could not fit into the scope of your study or did not anticipate. We recommend different ways of dealing with these.

Limitations Due to Unfortunate Choices

All researchers make unfortunate choices. These are mistakes that could have been prevented. Often, they are choices in how a study was designed and/or executed. Maybe the sample did not have the characteristics assumed, or a task did not assess what was expected, or the intervention was not implemented as planned. Although many unfortunate choices can be prevented by thinking through the consequences of every decision or by conducting a well-designed pilot study or two, some will occur anyway. How should you deal with them?

The consequence of unfortunate choices is that the data do not test the hypotheses as precisely or completely as hoped. When this happens, the data must be interpreted with these constraints in mind. Almost always, this limits the researcher to making fewer or narrower claims than desired about differences and similarities between the results and the predictions. Usually this means conclusions about the ways in which the rationales must be revised require extra qualifications. In other words, claims about contributions of the study must be made with extra caution.

Research papers frequently include a subsection in the Discussion called "Limitations of the Study." Researchers often use this subsection to identify the study's limitations by describing the unfortunate choices, but they do not always spell out how these limitations should affect the contributions of the paper. Sometimes, it appears that researchers are simply checking off a requirement to identify the limitations by saying something like "The results should be interpreted with caution." But this does not help readers understand exactly what cautions should be applied and it does not hold researchers accountable for the limitations.

We recommend something different. We suggest you do the hard work of figuring out how the data should be interpreted *in light of the limitations* and share these details with the readers. You might do this when the results are presented or when you interpret them. Rather than presenting your claims about the contributions of the study and then saying readers should interpret these with "caution" because of the study's limitations, we suggest presenting only those interpretations and claims of contributions that can be made with the limitations in mind.

We suggest you do the hard work of figuring out how the data should be interpreted in light of the limitations and share these details with the readers. Rather than presenting your claims about the contributions of the study and then saying readers should interpret these with caution, present only those claims that can be made with the limitations in mind.

One way to think about the constraints you will likely need to impose on your interpretations is in terms of generalizability. Recall that earlier in this chapter, we described the close relationship between contributions and generalizability. When generalizability is restricted, so are contributions.

Limitations Due to Missing Variables

Because of the complexity of problems, questions, and hypotheses explored in educational research, researchers are unlikely to anticipate in their studies all the variables that affect the data and results. In addition, tradeoffs often must be made. Researchers cannot study everything at once, so decisions must be made about which variables to study carefully and which to either control or ignore.

In the earlier example of studying whether second graders improve their understanding of place value after a specially designed instructional intervention, the researcher identified three variables that were expected to influence the effect of the intervention: students’ entry level of understanding, implementation of the intervention, and norms of the classrooms in which the intervention was implemented. The researcher decided to control the implementation variable by hiring one experienced teacher to implement the intervention in all the classrooms. This meant the variable of individual teacher differences was not included in the study and the researcher could not generalize to classrooms with these differences.

Some researchers might see controlling the implementation of the intervention as a limitation. We do not. As a factor that is not allowed to vary, it constrains the generalizations a researcher can make, but we believe these kinds of controlled variables are better treated as opportunities for future research. Perhaps the researcher’s observations in the classroom provided information that could be used to make some predictions about which elements of the intervention are essential and which are optional—about which aspects of the intervention must be implemented as written and which can vary with different teachers. When revising the rationales to show what was learned in this study, the researcher could include rationales for new, tentative predictions about the effects of the intervention in classrooms where implementation differed in specified ways. These predictions create a genuine contribution of the study. If you use our definitions, these new predictions, often presented under "implications for future research," would be presented as "contributions."

Notice that if you follow our advice, you would *not* need to include a separate section in the Discussion of your paper labeled "Limitations." We acknowledge, however, that some journal editors recommend such a subsection. In this case, we suggest you include this subsection along with treating the two different kinds of limitations as we recommend. You can do both.

Dealing with Unexpected Findings

Researchers are often faced with unexpected and perhaps surprising results, even when they have developed a convincing theoretical framework, posed research questions tightly connected to this framework, presented predictions about expected outcomes, and selected methods that appropriately test these predictions. Indeed, the unexpected findings can be the most interesting and valuable products of the study. They can range from mildly surprising to "Wow. I didn't expect that." How should researchers treat such findings? Our answer is based on two principles.

The first principle is that the value of research does not lie in whether the predictions are completely accurate but in helping the field learn more about the explanatory power of theoretical frameworks. That is, the value lies in the increased understanding of phenomena generated by examining the ability of theoretical frameworks (or rationales) to predict outcomes and explain results. The second principle, a corollary to the first, is to treat unexpected findings in a way that is most educative for the reader.

Based on our arguments to this point, you could guess we will say there will always be unexpected findings. Predicted answers to significant research questions in education will rarely, if ever, be entirely accurate. So, you can count on dealing with unexpected findings.

Consistent with the two principles above, your goal should be to use unexpected findings to understand more fully the phenomenon under investigation. We recommend one of three different paths. The choice of which path to take depends on what you decide after reflecting again on the decisions you made at each phase of the study.

The first path is appropriate when researchers reexamine their theoretical framework in light of the unexpected findings and decide that it is still a compelling framework based on previous work. They reason that readers are likely to have been convinced by this framework and would likely have made similar predictions. In this case, we believe that it is educative for researchers to (a) summarize their initial framework, (b) present the findings and distinguish those that were aligned with the predictions from those that were not, and (c) explain why the theoretical framework was inadequate and propose changes to the framework that would have created more alignment with the unexpected findings.

Revisions to initial hypotheses are especially useful if they include explanations for why a researcher might have been wrong (and researchers who ask significant questions in domains as complex as education are almost always wrong in some way). Depending on the ways in which the revised framework differs from the original, the authors have two options. If the revised framework is an expansion of the original, it would be appropriate for the authors to propose directions for future research that would extend this study. Alternatively, if the revised framework is still largely within the scope of the original study and consists of revisions to the original hypotheses, the revisions could guide a second study to check the adequacy of the revisions. This second study could be conducted by the same researchers (perhaps before the final manuscript is written and presented as two parts of the same report) or it could be proposed in the Discussion as a specific study that could be conducted by other researchers.

The second path is appropriate when researchers reexamine their theoretical framework in light of the unexpected findings and recognize serious flaws in the framework. The flaws could result from a number of factors, including defining elements of the framework in too general a way to formulate well-grounded hypotheses, failing to include a variable, or not accounting carefully enough for the previous work in this domain, both theoretical and empirical. In many of these cases, readers would not be well served by reading a poorly developed framework and then learning that the framework, which had not been convincing, did not accurately predict the results. Before scrapping the study and starting over, we suggest stepping back and reexamining the framework. Is it possible to develop a more coherent, complete, and convincing framework? Would this framework predict the results more accurately? If the findings remain unexpected based on the predictions generated by this revised, more compelling framework, then the first path applies.

It is likely that the new framework will better predict the findings. After all, the researchers now know the findings they will report. However, it is unlikely that the framework will accurately predict all the findings. This is because the framework is not built around the findings of this study of which authors are now aware (but have not yet been presented). Frameworks are built on research and theory *already* published. This means the redesigned framework is built from exactly the same empirical findings and theoretical arguments available before the study was conducted. The redesigned framework also is constrained by needing to justify exactly those methods used in the study. The redesigned framework cannot justify different methods or even slightly altered methods. The task for researchers is to show how the new theoretical framework *necessarily* generates, using the same methods, the predictions they present in the research paper. Just as before, it is unlikely this framework can account for all the findings. Just as before, after presenting the results the researchers should explain why they believe particular hypotheses were confirmed and why others should be revised, even in small ways, based on the findings reported. Researchers can now use these findings to revise the hypotheses presented in the paper. The point we are making is that we believe it is acceptable to reconstruct frameworks before writing research reports if doing so would be more educative for the reader.

Finally, the third path becomes appropriate when researchers, in reexamining their theoretical framework, trace the problem to a misalignment between the methods they used and the theoretical framework or the research questions. Perhaps the researchers recognize that the tasks they used did not yield data that could test the predictions and address the research questions. Or perhaps the researchers realize that the sample they selected would likely have been heavily influenced by a factor they failed to take into account. In other words, the researchers decide that the unexpected findings were due to a problem with the methods they used, not with the framework or the accompanying predictions. In this case, we recommend that the researchers correct the methodological problems and conduct the study again.

Part V. A Few Suggestions for Structuring Your Discussion Section

Writing the Discussion section of your research paper can be overwhelming given all our suggestions about what to include in this section. Here are a few tips that might help you create a simple template for this section.

We recommend the Discussion begin with a brief summary of the main results, especially those you will interpret in this section. This summary should *not* contain new data or results not previously presented in the paper.

The Discussion could then move to presenting the contributions in the ways we have described. To do this you could point out the ways in which the results differed from the predictions and suggest revisions to your rationales that would have better predicted the results. Doing this will show how the contributions of your study extend what is known beyond the research you drew on to build your original rationale. You can then propose how to extend your contributions to research by proposing future research studies that would test your new predictions. If you believe the revisions you make to your rationales produce new insights or understandings that could be helpful for educational practitioners, you can identify these contributions to practice as well. This comprises the bulk of the Discussion section.

If you have embedded the limitations in earlier sections of the paper, you will have presented your results and interpreted your findings constrained by these limitations. If you choose (or are asked) to describe limitations in the Discussion, you could identify the limitations and then point to the ways they affected your interpretations of the findings. Finally, the Discussion could conclude with the implications of the study for methodological choices that could improve research in the domain in which your study is located or how future studies could overcome the limitations you identified.

Because we are providing guidance on writing your research paper for publication, we will reiterate here that you should investigate the expectations and conventions of the journal to which you will submit your paper. Usually, it will be acceptable to use the terms "significance," "contributions," and "implications" as we have defined them. However, if the editors expect you to use the terms differently, follow the editors' expectations. Our definitions in this chapter are meant to help you think clearly about the different ways you can make a case for the importance of your research. What matters is that you have carefully built and described a coherent chain of scientific inquiry that allows your study to translate the significance of your research problem into contributions to the field.

We began the chapter with the "So what?" question. The question looks simple and straightforward but is challenging and complicated. Its simple appearance can lead researchers to believe it should have a simple answer. But it almost never does. In this chapter, we tried to address the many complications that arise when answering the question. We hope you now have some new insights and new tools for answering the question in your next study.

Epilogue

Mathematicians appreciate solutions or proofs that are elegant—neat, tidy, tasteful, stylish, refined. But also, parsimonious—frugal, thrifty, sparing. These solutions, say mathematicians, are aesthetically pleasing, even beautiful. These solutions also are often the most useful. Although it might strike readers as strange to characterize research in this way, beauty is what we see when research is conceived as formulating, testing, and revising hypotheses—the central concept of this book. It is a tidy, refined concept that is frugal and sparing—it includes no more than necessary. Its parsimony comes from the fact that this one concept anchors the entire research process. All decisions that must be made when conceptualizing and conducting a study can be made by connecting them back to formulating, testing, and revising hypotheses. This one concept is all you need to make all research decisions.

We have relied on this one concept to guide our entire presentation. We have claimed that scientific inquiry, in any field, *is* formulating, testing, and revising hypotheses. We have proposed that this core concept or process (it can be used as a noun or a verb) can be appropriated to guide the design and conduct of research in education.

Educational research is a complicated business. It is full of twists and turns. You will have lots of questions along the way, questions that cannot all be anticipated in a book on conducting research. The beauty of this concept—formulating, testing, and revising hypotheses—is that you can use it to answer every question, even those not considered in this book.

Of course, making research choices and answering questions about how to conduct research requires much more than repeating the mantra "formulating, testing, and revising hypotheses." Lots of detailed knowledge is needed to tell whether your choices help you formulate, test, and revise your hypotheses. Indeed, that is what much of this book has been about. But the point we are making here is that this mantra captures the single, final arbiter against which you can judge the appropriateness of your choices.

The beauty of this approach extends beyond its arbiter function. A fortunate consequence of using this concept to guide research decisions is the inevitability of a

J. Hiebert et al., *Doing Research: A New Researcher's Guide*,
Research in Mathematics Education,
https://doi.org/10.1007/978-3-031-19078-0

more coherent study. Using the concept as a guiding principle encourages, even forces, a coherence to your work that often eludes beginning researchers and even experienced researchers. If you are formulating, testing, and revising a hypothesis, the hypothesis ties everything together. From the goal of the study to the rationale for conducting it, to the methods used, to the interpretation of the data, and to the claims about its contributions, everything connects to the hypotheses you are testing. This is true for conducting your study and also for writing your evolving research paper.

Reflecting Back on the Three Phases of Scientific Inquiry

We suspect there are aspects of each phase of scientific inquiry—formulating hypotheses, testing them, and then revising them—that have been especially hard to appreciate or digest. These are the aspects with which our four protagonists (Martha, Sam, Adrian, and Corin) were struggling. Now that you have in mind the whole story, it is worth thinking about these aspects again.

Formulating Hypotheses—Making Predictions and Explaining Why

Is it really possible, or necessary, to always make a prediction of what I will find? Did I not ask a research question because I didn't know the answer? Can't I do research without making a prediction?

We will respond first to the last question. Sure, you can do research without making predictions and developing rationales for them. Even though we wrote this book, we have done that. Chances are, however, that the study will not be as informative as it could have been. You are likely to realize partway through the study that you forgot to take something into account or did not include a task you now wish you had. These things might happen anyway, but they are less likely if you think through as much as you can before you begin. And, a really good trigger to begin this process is to try to make a prediction, and then iteratively ask yourself why you made that prediction, then ask yourself why you think the reason you gave is a good one, and so on.

Our second response is that we contend it is always possible to make predictions. Predictions might not be correct; in fact, they probably will not be correct. But that is not the important thing. What is important is that predictions force you to make explicit everything you know at the time. Said another way, they allow you to take advantage of everything you know about the topic. It is by developing and revising your rationales and refining your predictions that you think through exactly what you are trying to find and what conditions might affect the results. There is a big advantage to doing the hard intellectual work before gathering data by formulating hypotheses that can guide *all* of this work.

If you determine you cannot make a prediction, we suggest taking one of three actions: (1) conduct pilot studies until you can, (2) revise your research question so you can predict an answer, or (3) choose another research question.

Testing Hypotheses

Once you have formulated a hypothesis, all the methodological work of the study should be designed to test it. This is the only thing you need to worry about with respect to methods. The most appropriate methods are those that will provide the clearest, cleanest test of your hypotheses. Remember, testing hypotheses means determining in what ways the predictions were correct and in what ways they were incorrect, *and* assessing the adequacy of the rationales. Comparing predicted outcomes to results is usually more straightforward than assessing adequacy of rationales. If your predictions are not entirely correct (and they usually aren't), you need to re-examine the reasons for your predictions. Once you find the glitch in your reasoning or identify something you overlooked from the literature, you are ready to revise your rationale and use it to create new predictions.

Revising Hypotheses

Based on the outcomes of testing your predictions and re-examining your rationales, you will need to refine your rationales and predictions because they will be incorrect, at least in some ways. Revising your hypotheses completes the scientific inquiry because it requires using the results to determine what your study contributes to the knowledge of the profession. The difference between the original hypotheses and the revised ones is the new information and new understandings you can present to your audience. Your revised hypotheses are your primary contribution to the educational community.

The Challenging Life of an Educational Researcher

Even with the elegant anchor concept of formulating, testing, and revising hypotheses, conducting educational research is incredibly challenging. Education is a complicated system; trying to carve out a piece of it to study is, itself, a challenging task. Once you decide what to study, you will need to make decision after decision about how to phrase a hypothesis, how to test it, and then how to revise it based on the evidence.

After you make good decisions and conduct your study, and simultaneously create multiple drafts of your research report, you will submit your paper to a journal and wait for the reviews. Rarely will you hear your paper has been accepted for publication. You might hear that your paper will be reviewed again if you revise the paper in ways outlined by the editor. Often, the revisions require substantive additional work. Sometimes you will hear your paper was rejected. The lives of most researchers include many rejects. The challenge is to persevere. Recognize this as the common experience of beginning researchers. Talk with colleagues and share the challenges of doing research and getting published. Use the opportunity to "fail productively." Continue to pursue your research agenda. The field depends on your persistence to grow its knowledge and improve its practices.

From our point of view, despite the many setbacks along the way, the challenges make the successes that much more satisfying and rewarding. Contributing to the growth of knowledge that ultimately leads to richer learning opportunities for all students is a reward that everyone reading this book can recognize. But we believe the *process* of doing research can also be enjoyable and rewarding. Solving challenging intellectual problems brings its own kind of enjoyment. We sincerely hope the information in this book will help you experience both kinds of rewards—improving the quality of the knowledge you contribute to the profession and increasing the fun of doing the research that generates it.

References

Agnes, M., & Guralnik, D. B. (Eds.). (2008). *Hypothesis. In Webster's new world college dictionary* (4th ed.). Wiley.

Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). (2006). *Educational design research*. Routledge.

Atkinson, P., Coffey, A., Delamont, S., Lofland, J., & Lofland, L. (2007). *Handbook of ethnography*. SAGE.

Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics, 102*, 1–8.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*, 586–598. https://doi.org/10.1037/0022-3514.41.3.586

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Biddle, B. J., & Anderson, D. S. (1986). Theory, methods, knowledge, and research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching: A project of the American Educational Research Association*. Macmillan.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education, 29*, 41–62.

Britannica. (n.d.). Scientific method. In *Encyclopaedia Britannica*. Retrieved July 15, 2022 from https://www.britannica.com/science/scientific-method

Brownell, W. A., & Moser, H. E. (1949). *Meaningful vs. mechanical learning: A study in grade III subtraction*. Duke University Press..

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard University Press.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., & Hiebert, J. (2019a). Choosing and justifying robust methods for educational research. *Journal for Research in Mathematics Education, 50*(4), 342–348. https://doi.org/10.5951/jresematheduc.50.2.0114

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., & Hiebert, J. (2019b). Posing significant research questions. *Journal for Research in Mathematics Education, 50*(2), 114–120. https://doi.org/10.5951/jresematheduc.50.2.0114

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., & Hiebert, J. (2019c). Theoretical framing as justifying. *Journal for Research in Mathematics Education, 50*(3), 218–224. https://doi.org/10.5951/jresematheduc.50.3.0218

J. Hiebert et al., *Doing Research: A New Researcher's Guide*,
Research in Mathematics Education,
https://doi.org/10.1007/978-3-031-19078-0

Cambridge University Press. (n.d.). Hypothesis. In *Cambridge dictionary*. Retrieved July 15, 2022 from https://dictionary.cambridge.org/us/dictionary/english/hypothesis

Campbell, D. T. (1961). The mutual methodological relevance of anthropology and psychology. In F. L. K. Hsu (Ed.), *Psychological anthropology: Approaches to culture and personality* (pp. 333–352). Dorsey Press.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). *Experimental and quasi-experimental designs for research*. Houghton.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*(4), 385–531.

Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. Educational Researcher, 32(1), 9–13. http://doi.org/10.3102/0013189X03200100

Cobb, P., Jackson, K., & Sharpe, C. D. (2017). Conducting design studies to investigate and support mathematics students' and teachers' learning. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 208–233). National Council of Teachers of Mathematics.

Cobb, P., & Steffe, L. P. (1983). The constructivist researcher as teacher and model builder. *Journal for Research in Mathematics Education, 14*(2), 83–94.

Cohen, E. G., & Lotan, R. A. (2014). *Designing groupwork: Strategies for the heterogenous classroom* (3rd ed.). Teachers College Press.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology*. Springer.

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the academy of management journal. *The Academy of Management Journal, 50*(6), 1281–1303.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Cronbach, J. L. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*, 116–127.

Cronbach, L. J. (1986). Social inquiry by and for earthlings. In D. W. Fiske & R. A. Shweder (Eds.), *Metatheory in social science: Pluralisms and subjectivities* (pp. 83–107). University of Chicago Press.

de Freitas, E., Lerman, S., & Parks, A. N. (2017). Qualitative methods. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 159–182). NCTM.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2017). *The SAGE handbook of qualitative research*. SAGE.

Design-Based Research Collaborative. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040

Eisenhart, M. A. (1991). Conceptual frameworks for research circa 1991: Ideas from a cultural anthropologist; implications for mathematics education researchers. In R. G. Underhill (Ed.), *Proceedings of the thirteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. I, pp. 202–219). Blacksburg, VA: Division of Curriculum & Instruction.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403–434.

Firestein, S. (2015). *Failure: Why science is so successful*. Oxford University Press.

Fisher, R. A. (1935). *The design of experiments*. Oliver & Boyd

Fishman, B., & Penuel, W. (2018). Design-based implementation research. In F. Fischer, C. Hmelo-Silver, P. Reimann, & S. Goldman (Eds.). *International handbook of the learning sciences* (pp. 393–400). Routledge.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Pearson.

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Sociology Press.

Glaser, B. G., & Holton, J. (2004). Remodeling grounded theory. *Forum: Qualitative Social Research,* 5(2). https://www.qualitative-research.net/index.php/fqs/article/view/607/1316

Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education, 44*(1), 218–243. https://doi.org/10.3102/0091732X20903302

Gorsuch, R. L. (2014). *Factor analysis* (Classic 2nd Edition). Routledge.

Gournelos, T., Hammonds, J. R., & Wilson, M. A. (2019). *Doing academic research: A practical guide to research methods and analysis*. Routledge.

Grant, C., & Osanloo, A. (2014). Understanding, selecting, and integrating a theoretical framework in dissertation research: Creating the blueprint for your "house." *Administrative Issues Journal: Connecting Education, Practice, and Research, 4*(2), 12–26.

Gravemeijer, K., & van Eerde, D. (2009). Design research as a means for building a knowledge base for teachers and teaching in mathematics education. *Elementary School Journal, 109*(5), 510–524.

Hart, L. C., Smith, S. Z., Swars, S. L., & Smith, M. E. (2008). An examination of research methods in mathematics education (1995–2005). *Journal of Mixed Methods Research, 3*(1), 26–41.

Hay, C. M. (Ed.). (2016). *Methods that matter: Integrating mixed methods for more effective social science research*. University of Chicago Press.

Hohensee, C. (2014). Backward transfer: An investigation of the influence of quadratic functions instruction on students' prior ways of reasoning about linear functions. *Mathematical Thinking and Learning, 16*(2), 135–174.

Husserl, E. (1973). Cartesian meditations: An introduction to phenomenology (D. Cairns, Trans.). Martinus Nijhoff. (Original work published 1929).

Jansen, A. (2012). Developing productive dispositions during small-group work in two sixth-grade mathematics classrooms: Teachers' facilitation efforts and students' self-reported benefits. *Middle Grades Research Journal, 7*(1), 37–56.

Jonson, J. L., & Geisinger, K. F. (Eds.). (2022). *Fairness in educational and psychological testing: Examining theoretical, research, practice, and policy implications of the 2014 standards*. American Educational Research Association.

Kelly, A. E., & Lesh, R. A. (Eds.). (2000). *Handbook of research design in mathematics and science education*. Erlbaum.

Kerlinger, F. N. (1964). *Foundations of behavioral research: Educational and psychological inquiry*. Holt, Rinehardt, & Winston.

Kersting, N. B., Givvin, K., Sotelo, F., & Stigler, J. W. (2010). Teacher's analysis of classroom video predicts student learning of mathematics: Further explorations of a novel measure of teacher knowledge. Journal of Teacher Education*,* 61(1–2), 172–181. https://doi.org/10.1177/0022487109347875, 172.

Kersting, N. B., Givvin, K. B., Thompson, B., Santagata, R., & Stigler, J. (2012). Developing measures of usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568–590. https://doi.org/10.3102/0002831212437853

Lester, F. K., Jr. (2005). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. *ZDM Mathematics Education, 37*(6), 457–467. https://doi.org/10.1007/BF02655854

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board Task Force report. *American Psychologist, 73*(1), 26–46.

Makel, M. C., Hodges, J., Cook, G. B., & Plucker, J. A. (2021). Both questionable and open research practices are prevalent in education research. *Educational Researcher, 50*(8), 493–504.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*(6), 304–316.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*(2), 3–11.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Sage.

Medawar, P. (1982). *Pluto's republic* [no typo]. Oxford University Press.

Merriam-Webster. (n.d.). Explain. In *Merriam-Webster.com dictionary*. Retrieved July 15, 2022, from https://www.merriam-webster.com/dictionary/explain

Merton, R. K. (1968). *Social theory and social structure* (Enlarged edition). Free Press.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (4th ed.). Sage.

Morris, A. K., & Hiebert, J. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher, 40*, 5–14.

National Research Council. (2002). *Scientific research in education*. National Academy Press.

Nemirovsky, R. (2011). Episodic feelings and transfer of learning. *Journal of the Learning Sciences, 20*(2), 308–337. https://doi.org/10.1080/10508406.2011.528316

Nuthall, G. (2004). Relating classroom teaching to student learning: A critical analysis of why research has failed to bridge the theory-practice gap. *Harvard Educational Review, 74*(3), 273–306. https://doi.org/10.17763/haer.74.3.e08k1276713824u5

Nuthall, G. (2005). The cultural myths and realities of classroom teaching and learning: A personal journey. *Teachers College Record, 107*(5), 895–934.

O'Donnell, C. A. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*, 33–84.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Oxford University Press. (2021). Theory. In *Lexico.com dictionary*. Lexico.com. https://www.lexico.com/en/definition/theory

Popper, K. R. (1968). *Conjectures and refutations: The growth of scientific knowledge*. Harper & Row.

Praetorius, A.-K., & Charalambous, C. (2022). Moving the field forward with respect to theorizing teaching: An introduction. In A.-K. Praetorius & C. Charalmabous (Eds.), *Expert perspectives on theorizing teaching: Current status, problematic issues, and aspirations*. Springer.

Puntambekar, S. (2018). Design-based research. In Fischer, F., Hmelo-Silver, C. E., Goldman, S. R., & Reimann, P. (Eds.). *International handbook of the learning sciences* (pp. 383–392). Retrieved from http://ebookcentral.proquest.com

Sechrest, L., Babcock, J., & Smith, B. (1993). An invitation to methodological pluralism. *Evaluation Practice, 14*(3), 227–235.

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2001). *Mathematics teacher noticing: Seeing through teachers' eyes*. Routledge.

Sloane, F. C. (2008). Randomized trials in mathematics education: Recalibrating the proposed high watermark. *Educational Researcher, 37*(9), 624–630. https://doi.org/10.3102/0013189X08328879

Sloane, F. C., & Wilkins, J. L. M. (2017). Aligning statistical modeling with theories of learning in mathematics education research. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 183–207). NCTM.

Small, M. L. (2011). How to conduct a mixed methods study: Recent trends in a rapidly growing literature. *Annual Review of Sociology, 37*, 57–86.

Spangler, D. A., & Williams, S. R. (2019). The role of theoretical frameworks in mathematics education research. In K. R. Leatham (Ed.), *Designing, conducting, and publishing quality research in mathematics education* (pp. 3–16). Springer.

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 266–287). Lawrence Erlbaum.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation, 9*, Article 4. https://doi.org/10.7275/96jp-xz07

Stirman, S. W., & Beidas, R. S. (2020). Expanding the reach of psychological science through implementation science: Introduction to the special issue. *American Psychologist, 75*(8), 1033–1037.

Tolman, C. W., Cherry, F., van Hezewijk, R., & Lubek, I. (1996). *Problems of theoretical psychology*. Ontario, CA.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. Falmer Press.

Vygotsky, L. (1987). The development of scientific concepts in childhood: The design of a working hypothesis. In A. Kozulin (Ed.), *Thought and language* (pp. 146–209). The MIT Press.

Wang, G., Zhang, S., & Cai, J. (2021). How are parental expectations related to students' beliefs and their perceived achievement? *Educational Studies in Mathematics., 108*, 429–450. https://doi.org/10.1007/s10649-021-10073-w

Weis, L., Eisenhart, M., Duncan, G. J., Albro, E., Bueschel, A. C., Cobb, P., Eccles, J., Mendenhall, R., Moss, P., Penuel, W., Ream, R. K., Rumbaut, R. G., Sloane, F., Weisner, T. S., & Wilson, J. (2019a). Mixed methods for studies that address broad and enduring issues in education research. *Teachers College Record, 121*, 100307.

Weis, L., Eisenhart, M., Weisner, T. S., Cobb, P., Duncan, G. J., Albro, E., Mendenhall, R., Penuel, W., Moss, P., Ream, R. K., & Rumbaut, R. G. (2019b). Exemplary mixed-methods research studies compiled by the mixed methods working group. *Teachers College Record, 121*, 100308.

Weisner, T. S. (Ed.). (2005). *Discovering successful pathways in children's development: Mixed methods in the study of childhood and family life*. University of Chicago Press.

Weisner, T., Ryan, G. W., Reese, L., Kroesen, K., Bernheimer, L., & Gallimore, R. (2001). Behavior sampling and ethnography: Complementary methods for understanding home-school connections among Latino immigrant families. *Field Methods, 13*(1), 20–46.

Wolf, C., Joye, D., Smith, T. W., & Fu, Y.-C. (2016). *The SAGE handbook of survey methodology*. SAGE.

Zieky, M. J. (2013). Fairness review in assessment. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 293–302). American Psychological Association. https://doi.org/10.1037/14047-017

Index

J. Hiebert et al., *Doing Research: A New Researcher's Guide*,
Research in Mathematics Education,
https://doi.org/10.1007/978-3-031-19078-0