

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

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 researchers 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 is 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.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

