PART

I

THE APPLIED IMPORTANCE OF RESEARCH

1 WHY CONDUCT EXPERIMENTS?

LEARNING OBJECTIVES

After reading this chapter, you will be able to

1.1 Articulate the elements of a systematic experimental inquiry.

1.2 Describe the hallmark characteristics of experimental progress.

1.3 Summarize the assumptions and beliefs that guide researchers in conducting their experiments.

Most gains made in educational practice have resulted from researchers conducting experiments. Whether those gains have been made by group-comparison methods, ethnography, epidemiology, econometrics, or single-case designs, the mechanism by which progress has occurred is *experimentation*. The use of experimental methods to better understand educational practices was championed by a diverse group of scholars in the early 20th century, including John Dewey (1958), George Herbert Mead (see Baldwin, 1987), B. F. Skinner (1954), and John B. Watson (1924). Each of these individuals noted that for systematic progress to occur, research was needed to allow education to rise above the politics, personal biases, and fads that dominate educational policymaking, then and now.

Although there is far to go in creating a universally effective and efficient educational system, the efforts of researchers have transformed educational practice in the last 50 years. What are now considered common practices in schools have emerged from the educational research of previous decades. Examples include peer-mediated instruction, data-based decision making, curriculum-based assessment, systematic instruction, positive behavior supports, token economies, inclusive education, phonics-based reading strategies, and accessing the general education curriculum, among many, many others.

These gains have been made by conducting experiments to answer questions to which there are no effective answers. By carefully crafting a question and then using a set of techniques to methodically study the phenomenon, a clearer idea of how different events interrelate can be revealed. In a sense, researchers ask specific questions regarding "How do things work?" If I want to know how a new teaching technique might improve student learning, my best bet is to conduct an experiment to answer that question. I could simply say "my way is obviously better" or "because we have always done things this way, it is the best way," *but I am more likely to improve student learning if I conduct experiments.*

WHAT IS AN EXPERIMENT?

An experiment is an approach to answering questions. By systematically studying a set of questions, researchers uncover answers that help guide them toward increasingly effective practices. However, there are important differences between how a layperson goes about answering a question and how a researcher answers a question. Humans have used at least three different ways of finding answers to questions.

The approach most familiar to people is the use of *common sense*. I mean to use *common sense* not in a derogatory way, but rather as a label for a set of strategies that are familiar to us. People arrive at assumptions about "how the world works" through everyday experience. For example, a child new to a school may notice that their classmates all raise their hands before answering the teacher's questions or asking for help. If students do not raise their hands, the teacher does not provide them with attention. The new child may quickly learn that they need to raise their hand to get the teacher's attention. If this occurs, they have used their everyday experiences to learn about how to behave in the new classroom and, perhaps, in other settings in the future. Not surprisingly, by and large, this approach works well for us in our daily lives.

However, there is a downside to answering questions through everyday experience. Although accumulated wisdom can often be effective in guiding our daily lives (your grandparents are, indeed, a source of important knowledge about how the world works), common sense also has its limitations. The primary limitation of common sense is that it is derived from correlated events and descriptions of situations. Simply the fact that two events tend to co-occur and there is a pattern to their co-occurrence does not mean that they are related. This places important constraints on how useful common sense is as a tool for understanding the world. For example, a person might observe that each day the sun rises in the east, crosses the sky, and sets in the west. For the majority of recorded history, people used this observation to infer that the sun circles the earth. Such an observation is confirmed by daily experience and makes sense if those experiences are the only basis for drawing conclusions. Although we like to think that as a modern, educated society we no longer believe the sun circles the earth, a recent survey of Americans found that 25% believed that *the sun circles the earth* (Allum et al., 2008; National Science Board, 2014).

Unfortunately, common sense stops at the level of correlation and does not pursue a more rigorous set of tests to verify or discount the nature of covariants. As the dictum taught in every introductory science class says, *"correlation does not imply causation."* Lest the reader think this is not a problem in the field of education, it may be instructive to note that the primary way educational policies are decided in a school district is the "school board." A school board's task is to make decisions about what is taught, who is taught where, how students are taught, and other related issues such as school discipline. However, despite the importance of these decisions, it has been noted that less than 10% of school board members have any type of degree relating to education and virtually none have training in research (Alsbury, 2008; Land, 2002).

Fortunately, people have developed other ways of answering questions that apply more rigorous tests about events before arriving at conclusions. An alternative to everyday experience as a means of understanding nature emerged in Ancient Greek and Persian cultures (Kantor, 1963). This approach, which we will call *logical analysis*, uses formal mathematical systems to test and arrive at conclusions. When conducting a logical analysis, a person needs to clearly define the question, use an established set of procedures to test possible answers to the question, and then arrive at a conclusion (Marr, 1986). For example, one could develop the following proposition: "All behavior occurs for a reason; reading is a form of behavior; therefore, the reasons why people read can be identified." A rule set could then be used to test the logical adequacy of the proposition and its conclusion: "The reasons why people read can be identified" (Cohen & Nagel, 1962). A number of individuals have attempted to show that such mathematical analyses are the basis of philosophical knowledge, sometimes referred to as "refined knowledge" (e.g., Whitehead & Russell, 1925).

However, there is a very important limitation to the logical approach. Although the system is clear and rigorous in what it does, the system is purely linguistic. That is, it never actually makes contact with natural phenomena and demonstrates the existence of the logical outcome. Although it presents hypotheses to test, the tests themselves are only verbal arguments. This has led to concerns that logical analysis is an inwardly defined system that is not tested in the real world. Harking back to our solar system question, one could formally propose that the sun circles the earth and develop a mathematical model of how it works (see Kuhn, 1957). Such a system, referred to as Ptolemaic astronomy, was the sine qua non of understanding our solar system for centuries. However, such a logical argument does not demonstrate the existence of the phenomenon; it demonstrates only that the phenomenon could exist as a logical outcome of a verbal argument. Such limitations to logical analysis set the occasion for the development of experimentation as an alternative way of learning about how the world works.

What makes experimentation different from common sense or logic is that it requires an individual to systematically test their assumptions (Bristow, 2010). Indeed, there are a set of characteristics that distinguish experimentation from other human endeavors that might not be immediately obvious to those not trained as researchers. First, a clear *experimental question* needs to be asked (see Chapter 5). Such questions are often referred to as "hypotheses" (see Box 1.1). However, there are different types of hypotheses, with some being more specific than others. For example, I might ask any of the following questions regarding a particular teaching technique:

1. "Does the use of time delay as a prompting technique lead to children acquiring addition and subtraction skills?"
2. "Does the use of time delay as a prompting technique result in faster acquisition of addition or subtraction skills than trial-and-error feedback?"
3. "Are parameters of the matching law, such as magnitude of reinforcement or latency of reinforcement, the reason that time delay is more effective than trial-and-error feedback in teaching basic mathematics skills?"

BOX 1.1: VARIOUS TYPES OF HYPOTHESES

An experimental question can be considered a type of hypothesis. However, in many areas of the social and health sciences, hypotheses have a more formal meaning and role in experimentation. This role can be traced to the early 20th century and the emergence of inferential statistics as a tool for agrarian research (Street, 1990). Tests using inferential statistics require the statement of a formal hypothesis (often referred to as a null hypothesis). In this sense, a hypothesis is not so much an experimental question as a statement about the anticipated results of the study. Once a formal hypothesis has been stated, the study can be conducted, the results statistically analyzed, and the statistical results used to either confirm the hypothesis or fail to reject the null hypothesis. However, with single-case designs such an arrangement is sometimes deemed unnecessary to competently conduct a study. Instead of creating formal hypotheses to confirm or reject, researchers focus on developing appropriate experimental questions and techniques to analyze them. The focus is not on the adequacy of an experimenter's prediction about the outcomes of their research (i.e., formal hypothesis testing), but on allowing the phenomenon to be revealed through careful experimentation. Researchers using single-case designs may use inferential statistics (see Chapter 17), but the emphasis is on experimental control and internal validity.

Hypothesis 1 is a general question about whether a particular teaching technique is effective. This is a very general type of hypothesis, but one that proposes a potentially important experimental question. Hypothesis 2 is more specific and asks whether one type of teaching technique is better than another. Hypothesis 3 asks a very specific question regarding what *mechanism* is responsible for the effectiveness of a particular teaching strategy. All of these are valid experimental questions that vary in specificity. Indeed, hypotheses can range from open-ended questions to very precise predictions about how things work. However, regardless of the level of specificity, *all hypotheses specify an experimental question in an objective manner that can be tested* (see Curry et al. [2020] and Mager [1962] for more on objective statements).

A second characteristic of research is that a clear plan is developed for objectively *measuring the events of interest*. That is, a researcher needs to identify what needs to be measured to adequately study the experimental question. For example, if I am studying the effects of a teaching technique on mathematics skill acquisition, one thing I need to measure is mathematics performance. In addition, based on the nature of my experimental question, I may also want to gather information on the types of errors that are made, the time that elapses between correct responses, performance on novel types of mathematics problems, and/or the occurrence of competing behaviors. Typically, *the events of interest are formalized into an observational code and a set of procedures outlined for when and how the events of interest will be measured* (see Part III of this book for more on *measurement*).

Another aspect of research that differs from other ways of exploring nature focuses on conducting an *experimental analysis*. The world is full of events that are constantly changing. This flux of activity is part of how our everyday lives function. However, it makes the systematic study of *causes* very difficult. If things are continually changing, it is hard to ascertain what are simply correlated events and what are causal relations among events. Therefore, researchers use experimental designs to distinguish between what is correlated from what is causal. To accomplish this, all events—also referred to as *variables*—are held constant except for one. This one variable—referred to as an *independent variable* (see Chapter 3)—is then allowed to operate, then withdrawn, then allowed to operate again, and so on. For example, if an investigator wants to study how teacher attention influences the challenging behavior of a child, adult attention could be selectively presented when challenging behavior occurs, then withdrawn for a period of time, and then re-presented. In addition, the researcher would need to hold all other potentially influential events constant (e.g., task type, task difficulty, and the presence of peers) while varying teacher attention so that there are no correlated events co-occurring with changes in teacher attention. If challenging behaviors (in this case the *dependent variable*) increase when attention is provided and decrease when attention is withdrawn, this pattern suggests that the behavior is related to teacher attention. By systematically presenting and withdrawing an independent variable while holding other variables constant and measuring changes in the dependent variable, experimental control can be demonstrated. By demonstrating experimental control through the use of research designs, an estimation of the degree to which a particular variable influences behavior can be established (see Part IV of this book for a discussion of types of experimental designs).

Once the experiment is conducted, a researcher needs to *analyze* the results. Typically, a series of exploratory analyses of the data are conducted to find out what types of patterns exist. These exploratory analyses seek to reveal how the independent variable influenced the dependent variable(s). Once the nature of the data patterns have been established, a researcher then seeks to summarize them for presentation to an audience. This is done so that the patterns that were found to occur as a result of the experiment can be clearly and concisely presented to other researchers. In essence, exploratory data analysis is a method for finding out "what the data have to say." These patterns are then summarized in tables and graphs so that information can be clearly communicated to a larger audience (see Part V of this book for a discussion of data analysis).

Finally, once a researcher has identified an experimental question, developed a measurement system, conducted an experimental analysis, and analyzed the results, one critical, final step needs to be completed. That is, *the results of the experiment need to be publicly reported and subjected to peer review.* A hallmark of experimentation is that the process is very transparent, meaning that any other person interested in what the researcher has done can obtain information about what occurred, where it was done, how it was done, when it was done, what resulted from those efforts, and how the researchers interpret their findings. A general rule of thumb is that the experiment is described in sufficient detail so that another person can read the paper reporting the study, replicate the procedures, and see if they obtain the same results. A manuscript reporting an experiment is then submitted to a *peer-reviewed journal* where experts evaluate the believability of the experiment and decide whether it was competently conducted and should be published. Finally, any experimental outcome is suspect until another group of researchers *replicate* the procedures and findings (see Chapter 4). This focus on public reporting, evaluation by experts, and replication by independent research groups makes the research process unlike most other human activities. There is no room in the process for vagueness, deceit, or false claims because everything is available for public scrutiny and rigorous evaluation (see the *Publication Manual* of the American Psychological Association [2020] for more information on the preparation and peer review of research papers). However, replication in its various instances is a challenging endeavor, and much attention is focused on the degree to which findings in the social sciences are replicable (see Box 1.2 and Chapter 4).

BOX 1.2: "THE REPLICATION CRISIS" AND SINGLE-CASE DESIGNS

Over the last decade there has been much discussion, some experimental analysis, and many emerging recommendations regarding the "replicability" of contemporary science. Much of that discussion has focused on the social sciences and particularly psychological research (Pashler & Wagenmakers, 2012; Simmons et al., 2011). A primary focus of these concerns has been the degree to which statistically significant findings from an initial experiment can be replicated by other researchers using similar procedures. From a single-case design perspective, it is important to frame this discussion about replicability in the context of experimental and quasi-experimental group design logic. The large majority of psychological research randomly assigns participants to an individual condition and then compares the central tendency and dispersion of the data from each of the groups. If the groups are shown to be statistically distinct, then an experimental effect is claimed. This is the core logic of most psychological experiments. However, note that there *is no requirement for replication to occur within an individual experiment*. Replication of findings in group-comparison studies requires additional experiments, and a concerning percentage of those follow-up studies fail to replicate the original findings. *The concern about replicability has not emerged as a strong concern in single-case design research.* Thus, not all psychological research seems to have a "replication crisis." The reasons for this are likely based on the strong emphasis on internal validity through replication used in single-case research. In any particular experiment, the effect of the independent variable is replicated at least once and is often replicated multiple times within an individual experiment (see Chapter 4). Single-case experiments that fail to replicate their findings are studies that fail to show experimental control, and the results are interpreted accordingly. Thus, the emphasis in single-case designs on internal validity likely reduces the risk of reporting false positive findings.

EXPERIMENTAL PROGRESS

Just as the process of conducting experiments differs substantially from everyday activities, so does the course of experimental progress. The stereotype most people have of research would suggest that it is an efficient, rigorous process that makes linear progress toward a particular goal. For instance, a researcher decides to solve a particular problem, devises an appropriate experimental question, and then conducts an experiment that solves the puzzle. *Such a streamlined process is rare.* This is because research is a very messy, inefficient, and nonlinear process.

One of the most interesting aspects of experimental progress is the *unpredictability* of the endeavor. Rarely do experimental questions result in precisely what was hypothesized to occur. That is why strict hypothesis testing (Box 1.1) is not always employed in experiments. For example, a researcher may set out to study curricular issues relating to student misbehavior. The experimental question might focus on whether task difficulty is related to increased levels of student noncompliance—the harder the work, the less compliant the student. However, during the course of establishing a baseline, the researcher might notice variability from one day to the next in student behavior, even though task difficulty is being held constant from day to day. Upon further investigation the researcher might discover an unanticipated event covarying with noncompliance. For instance, the student may not have the opportunity to eat breakfast, and on days when they arrive late for school, the student misses the opportunity to have breakfast in the school cafeteria. This missing of a meal may be associated with noncompliance apart from, or in conjunction with, task difficulty. To adequately address this question a competent researcher would need to analyze not only task difficulty in relation to noncompliance, but also the role of missing breakfast in relation to these variables. Although this might seem unrelated to the original hypothesis, to adequately answer the question a somewhat different experimental question would need to be studied.

This aspect of the research process has led to the observation that "any experiment worth its salt will raise more questions than it answers" (Sidman, 1960, p. 8). Sometimes those new questions can be predicted before the experiment is conducted, and at other times, as in the breakfast example, those new questions will only be revealed once the experiment is in progress. This truism suggests that a wise researcher should stay vigilant during an experiment to a range of events that might influence behavior. B. F. Skinner (1983) once remarked that his most important scientific discoveries were due to serendipity and might have been missed had he not been willing to follow his data even though they did not fit his original hypothesis. For example, his discovery of the behavioral process we now call "superstitious behavior" (i.e., responding maintained by an adventitious reinforcement contingency) occurred because the apparatus he was using to condition behavior malfunctioned. The apparatus failure inadvertently produced a contingency that created response-independent reinforcement, thus demonstrating how superstitious behavior can be shaped and maintained (Skinner, 1948; Vyse, 2022).

Similarly, the course of a *program of research* rarely goes precisely in the direction a researcher anticipates. Consider the example of sleep deprivation and problem behavior. Several research groups in the 1990s identified correlations between sleep deprivation and increases in challenging behavior (Fisher et al., 2002; Horner et al., 1997; Kennedy & Itkonen, 1993; O'Reilly, 1995; Symons et al., 2000). This finding came about when researchers were attempting to account for day-to-day variability in challenging behavior that could not be understood from events being manipulated during functional behavioral assessments. If environmental events were held constant, challenging behavior still fluctuated from day to day. Upon further investigation, sleep deprivation emerged as an influential variable in its own right, although the researchers had not set out to study sleep (Kennedy, 2021).

In our own research (Kennedy et al., 2000; Kirby & Kennedy, 2003), this line of inquiry took an unexpected turn. Findings across researchers (Horner et al., 1997; Kennedy & Itkonen, 1993; O'Reilly, 1995) suggested that negatively reinforced behavior was being affected by sleep deprivation, but it was unclear whether positively reinforced behaviors were similarly affected. Two issues needed to be analyzed to answer this question. First, a range of variables might have been co-occurring with sleep deprivation that were influencing behavior. Second, the specific types of reinforcers maintaining behavior needed to be explicitly controlled. These concerns required complete control of the environment to isolate single reinforcer functions, while holding other variables constant, and the direct manipulation of sleep. Such requirements dictated that a model system be used to clarify questions regarding sleep deprivation. This led us to conduct a series of laboratory experiments with nonhumans that revealed that sleep deprivation increased negatively reinforced behaviors (e.g., Kennedy et al., 2000), but decreased or did not change positively reinforced behaviors (Kirby & Kennedy, 2003). Thus, given the questions that emerged from our initial research findings, the direction of subsequent research was adjusted accordingly, along with its implications for applied intervention.

This example illustrates that research is a highly inductive endeavor. Only by conducting experiments can we get clear answers to our questions, but at the same time the answers are often a surprise. Experimentation in many respects is like exploration. There are no signposts to guide a researcher; instead, they push forward into the unknown and create a road map for those who follow. This observation highlights the cumulative nature of experimental findings. Most "discoveries" are the result of dozens of experiments, often conducted by several different research groups (Simon, 1987, 2013). The reason for this can be described in a metaphor. Think of each experiment that people conduct as an individual piece of a large jigsaw puzzle. Each piece needs to be fit into place, but no single piece defines what the final product is. In the long term, the critical outcome is not the fitting of a single piece into the puzzle, but the completion of the entire puzzle.

Similarly, individual research studies replicate and build on each other. One research group may conduct an experiment demonstrating that students with learning disabilities can learn new skills, such as phonological decoding, in general education settings. They may then conduct a second study to extend this finding by comparing the rate of learning in special versus general education settings and find that they are similar. Another research group may conduct a related study asking a similar comparative question regarding the quantity and quality of social interactions in different settings and find that general education participation produces superior outcomes. Yet another research group may read all of these experimental findings and ask about the impact of academic and social development on students in general education settings who do not have disabilities. Other research groups may replicate these studies for students with moderate disabilities, another research group may focus on students with gifts and talents, and so on. The net result of these studies is a clearer picture of the strengths and limitations of educating students with and without disabilities in general education settings. No single study could answer all the relevant questions, but the conduct of a range of studies, each asking a slightly different question, both replicates the results of previous studies and extends those studies into new directions. The cumulative result of this process is improved knowledge about educational practices, but the development of such a knowledge base can take years and sometimes decades (see Chapter 4).

Despite its nonlinear nature, however, experimentation does result in progress. It may be difficult to predict from one experiment to the next the particular course a line of research will take. However, the way the research process is oriented seems to ensure that progress is made. By requiring researchers to make public their procedures, findings, and interpretation of experimental results, the process is open to others for critique, debate, and replication. The accumulated result of this process is an improved understanding of an educational problem. The result at any single point in time may be more effective educational procedures, a better understanding of the complexity of the problem, or the realization that a particular line of research is not productive. Whatever the outcome, the process results in knowledge advancing beyond what could be known from common sense or logical analysis.

ASSUMPTIONS OF RESEARCHERS

Researchers approach experimentation with a different set of assumptions than most people use in their personal or professional lives. Often these assumptions are not explicitly recognized and, instead, are learned through the research apprenticeship process referred to as graduate training and postdoctoral study. Although most researchers do not spend a great deal of time contemplating *epistemological* assumptions relating to scientific inquiry (instead, they are likely engaged in the act of conducting research), there is a consistent set of beliefs that researchers hold. These assumptions tend to be very robust and occur across a broad range of disciplines and approaches to research (Lehr & Schauble, 2015; Linn, 1990; Underwood, 1957).

One assumption held by researchers was discussed at length in the previous section. *Everyday experience and even stringent logical analysis are not enough to understand the world.* Instead, *systematic inquiry* is needed to parse out correlation from causation. By engaging in carefully described and arranged procedures that others can replicate, researchers learn more about how the world works than by other approaches to acquiring knowledge.

A second assumption relates to the lawfulness of the world. Typically referred to as *determinism*, this belief postulates that events have identifiable causes. Apples fall from trees toward the ground, gasoline ignites at a certain temperature, and behaviors occur in certain patterns as a function of their consequences. If Behavior X occurs in a particular pattern, there must be a set of events related to Behavior X that cause it to occur in such a pattern. For example, if a child cries every time a parent drops them off at preschool, there must be something about the antecedent and consequent events that surround that episode that causes the child to cry. For nonresearchers, determinism is easier to accept and, perhaps, understand for the physical sciences than for educational or psychological phenomena (Goldenweiser, 1938; Pearl, 2009). Nevertheless, all events have a cause, and identifying those causes is the foundation of experimentation, whether a particular researcher articulates this assumption or not.

Closely related to the notion of determinism is the assumption of *material causes*. Hundreds of years ago, when asked why water turns from a liquid to a gas, most educated people would have invoked a metaphysical explanation—for example, that the essential spirits in the water had become excited and left for heaven. Or, a person acted the way they did because a "homunculus" in their head directed them to act that way. This type of explanation—still alive and well in our contemporary society—invokes causes that do not physically exist (Burgos, 2021; MacCorquodale & Meehl, 1948; Morey, 1991). To say the least, metaphysical causes are difficult to experimentally investigate.

By focusing on material causes, researchers are forced to deal with physical events as causal entities. Things that are being studied need to be operationally specified and accurately measured. In addition, to find the source of the occurrence of those events, some other event must be identified and tracked in relation to that which is the event of interest. If the occurrence of scolding by a parent tends to follow the yelling of a child and the nonoccurrence of scolding is related to the nonoccurrence of yelling, then scolding might be causally related to yelling. Additional manipulations of scolding as a consequence for yelling may suggest that the two events are so closely related, and in a particular pattern of occurrence, that we would say that yelling is caused by scolding. No appeal to forces that exist in some other place, time, or physical dimension is needed to explain the behavior (Skinner, 1950).

Earlier in this chapter I alluded to an aspect of experimentation that was different from most other ways of knowing—*replication*. In Chapter 4 we will discuss different types of replication in detail, but here I would like to present the idea that *independent replication* is one of the foundations of research. Any published report of research should explain why a study was conducted, exactly what was done, what the results were, and how those data might be interpreted. This allows other researchers to attempt to independently replicate the experiment to see if similar results can be obtained. In general, there is a consensus among researchers that any finding is suspect until it has been replicated (Haack, 2011; Maxwell et al., 2015). The infamous case of *"cold fusion"* is an exemplar of this issue (see Taubes, 1993). Briefly, a pair of physicists claimed to be able to initiate nuclear fission under low temperatures—something that was fundamentally inconsistent with what was known about this phenomenon through thousands and thousands of studies. Research groups from around the world attempted to replicate this finding, but even after many years and many experiments, no other research group could replicate the findings (indeed, the original researchers could not replicate their own findings!). Because no one could replicate the original results, researchers have come to regard the finding as an error in experimentation (i.e., poor experimental methods and/or inaccurate interpretation of results). The ability of others to replicate new findings is a critical component of the research process.

A belief in the *cumulative nature of research findings*, then, is an important assumption among researchers. As was previously discussed, each study is like an individual piece of a jigsaw puzzle, which is not complete until a range of experiments have been completed. This process (a) serves to check on the veracity of individual research findings, (b) is self-correcting in that errors will be found and alternative findings/interpretations publicly presented, and (c) results in an increased understanding of the phenomenon being investigated. In applied areas, such as educational research, there is also an implicit assumption that this whole process results in better educational practices (Chapter 6). The result is that students, teachers, and community members benefit from the work that we refer to as experimentation.

Along with a more complete understanding of a phenomenon, there is also an expectation that at some point a more *parsimonious* understanding of it will result. By parsimonious, what is meant is that a set of findings can be summarized in as simple a manner as does justice to the phenomenon. For example, when the initial finding emerged showing that *time delay* as a technique for transferring stimulus control (Touchette, 1971) could be extended to educational contexts (Halle et al., 1979), nobody knew exactly what would result. However, after many decades of research by multiple research groups, a great deal is known about when, where, how, and to what degree time delay is an effective teaching strategy. Not only are the general parameters of time delay well understood, but the techniques can be summarized as a handful of procedures for practitioners to use (Horn et al., 2023). In this case, parsimony resulted from a more complete knowledge of how the behavioral processes worked and how they could be organized. What resulted was not only a greater understanding of what comprised a certain area of research, but also an efficient way of organizing those findings.

CONCLUSION

Experimentation as an approach for answering educational questions emerged in the early 20th century (Tomlinson, 1997). Since then, experimentation has provided tremendous insight into processes that improve educational practices and outcomes for a wide variety of students. However, research itself is a difficult concept for most people to grasp, in part because the typical citizen has little knowledge of, and no direct experience with, the process. As a matter of course we propose questions and find answers to them in our everyday lives and, to the extent that things unfold as we anticipate, are psychologically satisfied with the results. There is no obvious need to pursue issues further, as long as things work as we expect. Despite the general success of common sense in our day-to-day experience, it often falls short when confronted with complex causes.

It is because of this limitation to common sense that people have developed a set of techniques for answering complex or nonintuitive problems that are referred to as experimentation. The research process is rigorous, not easily understood, and effortful. However, when done properly, it is an invaluable tool for knowledge generation. Experimentation is not a "thing" to be reified and kept at a distance, but a tool set for asking questions about the world. At its most utilitarian it is something to be used to solve people's problems.

In the remainder of this book we will explore one approach to experimentation, referred to as *single-case designs.* These designs embody the quintessential properties of experimental methods and are ideally suited for a range of questions relevant to educational contexts. They are an exciting set of tools that allow people to ask questions that can be answered using individual students, classrooms, or schools, with no need for "control" or "contrast" groups as comparisons. These designs have a rich history in educational, psychological, and health sciences research, as well as an exciting future for exploring currently unsolved questions relating to education.

REFLECTION QUESTIONS

1. What makes experimentation different from everyday experience?
2. Why is objective measurement a central focus of experimentation?
3. In what ways is a hypothesis an experimental question?
4. Why might there be so few failures to replicate when investigators use single-case designs for research?
5. Can you think of a systematic line of research in your area of interest and describe how it evolved over time?