Research Designs for Empirical Research

ROBERT C. CALFEE AND MARYL J. CHAMBLISS

In the second edition of the Handbook of Research on Teaching the English Language Arts, our chapter on research methods emphasized four points: (a) framing a research question, (b) principles of research design, (c) constructing a design using factorial techniques, and (d) interpreting findings. This chapter builds on the same foundational elements, updating several points and providing more recent citations. The area of experimental design has remained stable for decades, and so this plan seems sensible.

The previous chapter also included an extended example of the application of factorial design to a complex research problem in the literacy area—teaching argumentative texts to middle grade students. The current revision is considerably abbreviated, and we have shortened the illustrative material. We recommend that readers refer to the earlier version for additional detail.

A significant development since the earlier chapter, the entry of mixed-methods strategies, is a major addition to this chapter. While we mentioned this topic previously, it has become a significant design consideration in the past 10 years. The mixed-methods label first appeared in the mid-1990s (Jaeger, 1997), and the last decade has seen a virtual explosion of activity (Calfee & Sperling, 2010). The original idea was to combine quantitative and qualitative methods for investigating research questions, a departure from earlier practice in which individual problems were explored with one method or the other. Experimental design techniques were traditionally associated with quantitative approaches, while qualitative techniques were employed for in-depth exploratory investigations, often taking shape as case studies. The two strategies differed most obviously in the choice of outcomes, but they also employed different design arrangements and relied on different methodologies and epistemologies. Mixed-methods techniques will be woven throughout this chapter, reflecting their benefits in the literacy and language arenas.

Two key words in the title are design and empirical. We set the stage for the chapter with a brief discussion of the concept of empiricism as an approach to knowledge. We next concentrate on design, where we address the four foundational elements of experimental design mentioned above. Throughout we call upon an abbreviated narrative from the example on teaching and learning argumentative text in the earlier chapter. We assume that a reader is familiar with basic concepts of social science and educational research. For additional detail, consult Berliner and Calfee (1996, especially chapters by Behrens & Smith, Jaeger & Bond, and Hambleton), Gall, Borg, and Gall (2006), Krathwohl (1997), Myers and Well (2003), Osborne (2008), and Alexander and Winne (2006, especially chapters by Nesbit & Hadwin, and Cooper).

Empiricism is a systematic approach to answering certain types of questions. Through the collection of evidence under carefully defined and replicable conditions, social science researchers seek to discover the influence of factors that affect human thought and action, and to understand when and why these influences occur. Nonempirical research spans a wide range of other approaches, including mathematical, logical, historical, and legal, many of which support empirical techniques.

The empirical tradition plays a significant role in creating and validating social and psychological theories about how people think and act. In language arts, for instance, data-based research has supported models that link reading and writing as social acts (e.g., Nystrand, 1989, 1997). No longer are readers and writers perceived as lost in their impenetrable thoughts, but instead as communicating with one another through both oral and written media.

Empirical research also searches for answers to practical questions. A high school English teacher seeks to improve her students’ understanding of formal arguments. A middle school teacher aims to encourage his students toward more strategic comprehension. A remedial reading teacher wants to improve vocabulary instruction to help students
Research Strategy: What Is the Question?

Novices tend to begin a research project by thinking, “I’d like to prove that…” Educators tend to advocate their favored positions and actions. “Spelling tests are bad (or good).” “English teacher should (or should not) know about linguistics.” “Promoting student motivation is (or is not) a critical ingredient in a writing assignment.” We can become passionate about these matters. Such hypotheses are entirely appropriate as starting points for inquiry, but developing a research problem requires a fundamental shift toward “I wonder what will happen…” A small switch, but with major implications. For instance, the earlier proposals now take shape as questions. Under what conditions are spelling tests bad or good? What are the effects of linguistic background on the thinking and behavior of English teachers? In what ways do higher or lower levels of motivation affect students’ responses to different types of writing assignments?

The revised questions can open Pandora’s box because they challenge the researcher to explore a universe of possibilities. No longer is the task to compare one condition with another. Rather, the investigator is led to think about a broad array of situations, outcomes, and individuals. Spelling tests come in many flavors, and may help with some tasks (new spelling tests) and interfere with others (writing assignments), for some students (compulsives) but not others (impulsives). How to grapple with the infinite possibilities? The simple answer emerges from the application of design principles as an essential tool for translating a question into a research study.

Framing an Answerable Question. Empirical research begins with the formulation of a workable scientific question, one that can be answered by objective evidence. For instance, imagine yourself as the high school teacher mentioned at the beginning of the chapter. You want to help your ninth graders write well reasoned and coherent arguments. You are familiar with Toulmin’s (1958) concept of argument, which has become important in your thinking. Toulmin proposed that all arguments have three basic parts: a claim or assertion, the English teacher’s thesis statement; evidence offered in support of the claim; and warrants, which link evidence to claim. Complex arguments may also include qualifications, counterarguments, and rebuttals.

It took you a while to grasp Toulmin’s notion of a “warrant.” Claims, evidence, and even counterarguments and rebuttals seemed fairly straightforward. Toulmin described warrants as general statements, such as “If this evidence, then this claim.” The relationships among claims, evidence, and warrants are clearest in simple arguments. Suppose the claim is that wolves often represent evil in folk tales. The evidence is that in well-known folktales, wolves terrorize and almost kill three little pigs, a little girl wearing a red coat and her sick grandmother, and a little Russian boy and his pets. The warrant fills any gaps in reasoning: Any character that terrorizes and almost kills innocent people and animals represents evil. Warrants become crucial when connections between evidence and claim are not obvious, when they need to be explained or defended against counterarguments.

After reviewing several written arguments, you have become convinced that focusing attention on warrants might help students reason more effectively in reading and writing. In addition to the Toulmin model, you want to explore the impact of social aspects of instruction on reading and writing (Spivey, 1997). Your experience suggests that students learn more effectively when they have opportunities to apply what they are learning; in learning about argument, this means that they have opportunities to critique a writer’s ideas (Mathison, 1998) and to anticipate responses to their writing (Rubin, 1998), which you think could be supported through small-group activities.

You look over a list of questions that you have scribbled during your reading and thinking:
• What is the essence of a good argument?
• What do my students know about the concept of argument?
• How can I effectively teach students to comprehend, critique, and compose arguments?

Let’s consider the researchability of each question. The first cannot be answered empirically because the answer requires value judgments—“good” is the fly in the ointment. The second and third questions, in contrast, both provide starting points for empirical study. Students’ responses to “What makes this argument strong?” can provide evidence about thinking processes. Assessing student writing following different instructional approaches can inform the third question.

The key to evaluating a potential research question is to ask yourself, “Assuming that I collect evidence of some sort, and obtain a particular set of results, to what degree can I use the results to make a convincing argument in relation to the original question?” This exercise requires that you step outside your own convictions and become a self-critic. It also helps to find a friendly enemy along the way, someone interested in your problem, and willing to work hard at destroying your line of argument. Defending your interpretations against alternative explanations is an essential part of the research process, and is a significant theme in this chapter.

Finding Evidence. Once a question has appeared on the scene, the researcher must decide what evidence might be relevant to the question, how to gather it, and how to analyze and interpret the data. It helps to know the territory: What do you already know about research on comprehension of argument texts, about comprehension and composition in general, about effective instructional practices, and so on? Another task is a review of the literature, which can seem a daunting task. By selecting a few best evidence papers as starting points and working backwards (Krathwohl, 1997, chap. 6; Slavin, 1986), you can shape the job into manageable proportions. You should also bring your professional knowledge and experience into the mix.

You must then reach decisions about what data to collect, along with how and where to carry out this task, and from whom. We will cover the what of data collection later as a design task, and a few fundamentals will set the stage for now. First, should you focus on numbers or “stuff” (observations, interviews, videos, documents, and the like) —quantitative or qualitative? In fact, you don’t really have a choice. Empirical data are inherently qualitative, and it takes a uniquely human act—measurement—to assign numbers to observations. A student essay begins as stuff, but you can count the number of words, calculate the average sentence length, or ask a panel of judges to use rubrics to evaluate the work. Whether you decide to measure and how you decide to do it is a conceptual matter.

A second dimension to what is how much? A useful principle is triangulation, which means collecting diverse data for each construct in the study. If you are interested in student writing, then the starting point is to identify a variety of writing assignments appropriate to the research question. We will go into more detail about this task in the design section. You can look at different facets (length, coherence, mechanics) of each composition. You might also explore the value of artifacts (student notes and outlines) and indicators (e.g., ask students to talk about the compositions and how they planned and produced them). If this amalgam of information produces a consistent picture, then you have the basis for a strong argument. The how of data collection encompasses two overlapping strategies; the researcher can either observe or intervene with the intent of describing or experimenting. Imagine a young boy examining an ant hill. One moment he is the naturalist, observing the hectic activity in the insect community. Suddenly driven to intervene, he pokes a twig into the hole and watches the ants’ responses.

To observe or to intervene? Most texts on research methodology separate these two approaches, one section on naturalistic approaches and a second on experiments. Experimental, quantitative, and statistical are often bound together in one package, and contrasted with naturalistic, qualitative, and descriptive. Fortunately, the joining of quantitative and qualitative methods is becoming more commonplace (Calfee & Sperling, 2010). Both approaches are clearly empirical, in the sense that they both rely on evidence. Moreover, the various strategies are independent; you can design a naturalistic investigation that uses quantitative methods, or an experimental study that employs qualitative assessments. Quantifying observations allows the researcher to employ statistical techniques for summarizing information and conducting inferential analyses (how closely related or disparate are two sets of evidence). The richness of qualitative information, on the other hand, may allow the researcher to delve into underlying processes and explore complex hypotheses. For instance, measuring the length of two sets of compositions may reveal substantial and trustworthy differences; students taught about warranting write substantially more than students without such instruction. Student interviews may resist quantification, but suggest to the researcher how instruction led students to write longer essays. For instance, suppose several students tell you something like this: “I knew that if I just wrote my main point and a few details you wouldn’t like my paper, so I just rambled around—that’s kinda what you mean by that warrant thing.” The interview results may not yield what you hoped for, but they connect the quantitative information with the instructional treatment.

The where of data collection is frequently tied to the who. Traditionally, real classroom situations are contrasted with laboratory environments, the latter presumably unreal. More recently, collaboration between teachers and researchers has been contrasted with researcher-imposed designs (Freedman, Simons, & Kalnin, 1999). Simple contrasts may be misleading. Researcher-imposed designs implemented within either classrooms or laboratories are supposed to
eliminate extraneous fluctuations in conditions, because
the classroom can be a wild and crazy place. The practical
value of researcher-imposed designs and laboratory findings
is often questioned, whereas teacher designed, classroom
based research seems directly applicable.

Neither stereotype stands up under close scrutiny. One
can find examples of untrustworthy laboratory research and
excellent instances of classroom-based investigations. The
practical significance of a study depends on the quality of
the research rather than the characteristics of the setting.
An important bridge between these extremes is the design
experiment, in which systematic variations are tried out
in different classrooms through collaborations between
teachers and researchers, using a range of quantitative
and qualitative indicators to inform the teams (Brown, 1992; Collins, 1994). The design experiment technique,
although still undergoing refinement, illustrates the merg-
ing of methodological distinctions that previously seemed
contradictory.

Evidence is trustworthy to the extent that it holds up
against attack from informed critics; research has much
in common with law. Earlier we introduced validity as
centrally important. Informed critics will value signs of
research validity and dismiss research attempts that are
invalid. Closely related is the concept of control. In social
science research, control refers to the researcher’s efforts to
ensure the validity of the interpretations, the trustworthiness
of the argument, the generalizability of the findings. An
essential ingredient for adequate control is design, which
refers to the steps in identifying the contextual factors that
influence performance, planning the conditions of data
collection so that these factors are adequately represented,
and ensuring that the plan supports generalizability of the
findings, so that you can argue that the findings are trust-
worthy, replicable, and broadly applicable. The concepts
of factors and factorial design described later provide one
strategy for establishing adequate control.

Setting the Stage. In this section we expand the earlier
discussion of “teaching argument” into a research scenario
to which factorial design principles can be applied. The
scenario lays out starting points for where and who. Assume
that you are developing a proposal for your masters thesis.
Two teachers in your school employ distinctively different
approaches to argument instruction—one quite traditional,
the other quite innovative. The traditional teacher relies on
lecture and discussion to cover thesis/support forms of argu-
ment. Her basic model is the five-paragraph essay, the claim
or thesis in the introductory paragraph, three paragraphs
of evidence or support, and a conclusion that summarizes
the argument. The second teacher introduces students to
several forms of argument, including counterarguments and
rebuttals. She emphasizes the role of warrants in link-
ing evidence to claims, which serves as a scaffold to help
students identify and question their thinking, which she
refers to as “metacognition.”

Here you have the makings of a natural experiment.
The plan seems simple enough on first glance; your task as
researcher will be to visit classrooms and document what
you see. On reflection, however, you to realize that reality
is more complex. For instance, your questions and your
presence may influence both teachers and students. Such
effects are not necessarily bad, but they show that research
almost always entails some intervention.

You next consider a planned experiment, with one class
assigned to an innovative treatment, the other to business
as usual. This plan also seems simple enough at first. You
need to construct materials for the innovative treatment, find
out what happens in business-as-usual, identify measures to
assess performance at the beginning and end of the study,
and work out details of the statistical analysis, your advi-
or raises several questions. First, she warns that the two
treatments appear to be confounded. She is not denouncing
your ideas! Confounding is a technical term describing a
condition where two or more dimensions or factors vary
simultaneously. In your plan, the two treatments differ in
several ways, including the goals (five-paragraph com-
positions vs. analytic essays), the reading materials (none
vs. some), the teaching approach (lecture vs. small-group
discussion), and student activities (individual vs. group
assignments), to name a few. These contrasts co-vary; they
change from one class to another, along with differences in
the teachers and the students. If the results favor the innova-

tive approach, what are the critical elements?

Confounding is serious enough, but your advisor raises
other concerns. How can you be sure that the treatments
will be implemented as you intend? What if the measures
do not tap into critical instructional elements? You begin
to understand that, even in a planned experiment, you still
have to play the naturalist’s role, documenting in detail what
happens during instruction in the two classes. Your advisor
has caused you to pause for reflection—her mantra is pay
now or pay later.

Principles of Factorial Design

This section of the chapter develops the foundational
concepts of research design. Any field of study evolves in
stages or paradigms, beginning with careful examination
of intuitive experiences and ideas and then increasingly
careful collection of evidence. Data patterns emerge, of-
ten to transmute or vanish. Eventually the patterns lead
toward formulation of theoretical ideas, which explain and
enlighten the evidence. Along the way, investigators rely
on informed guesses. Educational research is in the middle
stage of educated guesses today. Educators do not yet have
powerful theories and so must still rely on experience and
hunches to guide much of their work. Disciplined planful-
ness is crucial, however, and hence the focus in this chapter
on research design.

We first lay out three fundamental barriers that design
techniques help surmount: inadequate construct validity,
confounding, and extraneous variability. Then we discuss
four fundamental principles of factorial design: the concept
of design, the elements of factorial design, connections among the elements, and integration around a theme. We rely on a technical vocabulary that has evolved over the past several decades; the critical terms are shown in Table 56.1. This table should be helpful as you proceed through the chapter. As we lay out the factorial system, we will also present ways in which mixed-methods bridges can connect quantitative and qualitative approaches.

Three Barriers. The construct validity of a research study, as for an achievement test, refers to the trustworthiness of various interpretations of the evidence; does the finding mean what you think it means, where “it” is the construct? Validity can be compromised in several ways, but most shortcomings arise from a failure to think through the path that leads from initial question to final interpretation. The concept of test validity offers a helpful metaphor.

As for an achievement test, the construct validity of a set of research findings depends not only on the data but on the interpretation. Is the plan of the study adequate? To what extent does the context allow generalization to other situations? To what extent do findings agree with other studies? What are the cost-benefit implications of various decisions springing from the study; when the decision is correct or when it is wrong? The more you know about the answers to such questions, the more secure will be the construct validity. One purpose of research design is to increase the chances of trustworthy outcomes. As Cronbach (1988) puts it, “Validators should do what the detached scientist does; [the key ingredient is] a vigorous, questing intellect….” (p. 14).

The second barrier, confounding, occurs when the effect of the primary factor cannot be separated from the confounded factor(s), in which instance the findings are significantly compromised. Confounding can arise in your study if you select two teachers, assigning each (and their students) to one of the instructional approaches. Suppose you find a striking difference in writing outcomes. The cause might be the teacher, the students, the program, or any combination of the three. Given these possibilities, the evidence is difficult to interpret with any confidence, and this difficulty is virtually impossible to repair after the fact.

Confounding is the major shortcoming of designs that contrast an innovative approach and a traditional method, the classical experimental-control technique. A quarter century ago, Cronbach (1963) pointed out the limitations of this design, but it is still commonplace in the empirical literature, and is viewed as the “gold standard” by the Department of Education (see National Research Council, 2002, for a discussion of federal policies in educational research). Any comparison between two groups is necessarily confounded, and hence subject to multiple interpretations. Our advice, if you are contemplating such a study, is to consider options. A more complex design can separate the confounded variables. Qualitative descriptions of classroom life under the innovation and the control group can often help untangle confoundings.

Uncontrolled variability, the third concern, occurs when unintended fluctuations obscure patterns in the data. Reducing or identifying extraneous variability is essential because of the critical importance of variance in educational research. On the one hand, systematic or explainable variance is the payoff from a study. You predict that performance under the novel treatment will differ from the traditional approach, presumably because of the treatment. You hope the difference will be large. On the other hand, unexplained variance is the gauge against which systematic differences are measured; large differences in student performance within the two conditions may completely obscure a treatment effect.

This chapter is not about statistics, but when designing a study it helps to understand the statistics game. Your job is to construct a design in which systematic variability is maximized and unexplained variability is minimized. Suppose, for instance, that writing scores (rated on a 1–10 rubric scale) for one approach range from 8 to 10, while they range from 3 to 5 in the other approach. This difference clearly passes the eyeball test. On the other hand, if scores range from 6 to 10 in the treatment group and 5 to 9 in the control group, the differences may be due to chance. In this second situation, however, suppose that many girls score 9 or 10 in the first group, while boys in both groups range around 5 to 8. Now you can tell a different story; the treatment makes a difference, but only for girls.

The concept of control encompasses the full range of methods employed to strengthen validity. Chief among these methods is design, although other techniques are also important. For instance, in order to generalize your findings
to other situations, the data should be based on a random sample from some population of interest. This standard is difficult if not impossible to meet. Social science research typically relies on handy random samples. You call upon your professional network to gain access to a particular school. It is not a chance selection, but perhaps you can argue that it is typical of schools in the area. Some teachers cooperate with you, others do not. You may locate a purpose sample, selected because it meets conditions important for your hypotheses. Such constraints and decisions limit the generalizability of your findings. The important point is to be aware of these constraints, and to document them for yourself and your audience. The reader can then assess the degree to which nonrandomness compromises your argument.

A second nondesign control issue is the establishment of uniformity during data collection. A well constructed design provides control over certain variables, but others will be free floating. For instance, suppose your study spans a 5-week period. Consult the calendar—what coming events may influence instruction or assessment? If the critical posttest is scheduled on the day before a big football game, students may not give full attention to the task. What is happening in the lives of students and teachers throughout the study? If several students know that their family is moving, they may not try their best. If one teacher is in the midst of a divorce and another is fighting with the Internal Revenue Service, this may not be the best time for them to take on a new program—nor, for that matter, to handle a traditional approach.

These scenarios illustrate the difficulty of establishing uniformity. You should nonetheless do your best to keep conditions constant, attend to discrepancies, and take care to document everything. Like randomness, complete uniformity is seldom attainable. Problems arise when you do not detect unusual happenings and fail to report them. Performance outcomes can fluctuate because of these events, and these fluctuations can cloud the picture if you ignore them.

The Concept of Design. A well-planned design can be critical for separating treatment effects from background noise. It is the best protection against threats to validity, confoundings, and extraneous variability. Textbooks on research design often stress the procedures and mechanics of the design task, including statistical methods like analysis of variance. We will start instead with the underlying principles of design, which apply equally to descriptive and experimental investigations, to quantitative and qualitative approaches.

Many human endeavors rely on the concept of design, sometimes through recognition and appreciation of naturally occurring patterns, more often through creation and construction. As Simon (1981) notes, design is the feature that distinguishes between the natural and the artificial, between happenstance and the artifacts of humankind. All designs have three essential ingredients (Chambliss & Calfee, 1998). First is a set of distinctive elements, what Simon calls “nearly decomposable components.” Second are the linkages that bind individual elements together. Third is the theme that gives overall shape and meaning to the enterprise.

We turn next to the application of design principles in a research study. The elements include the factors that influence performance: treatment or environmental variations, differences between individuals, and various plans for assessing performance. The elements are linked by one of two relationships, crossing or nesting. The theme encompasses the overarching research objectives, as these are represented by specific questions and hypotheses. A design that covers these three domains should generate a data structure that informs your research questions in a well-controlled—i.e., trustworthy and generalizable—fashion.

Factorial Elements. A factor is a variable that the researcher defines and controls in order to evaluate its influence on performance. Some factors can be directly controlled; others spring from natural variations. Your study, for instance, provides several candidates as factors for inclusion in the design: argument type, instructional method, prior student experience with arguments, age and sex of students, teacher experience with the genre, and choice of a written or oral test. As suggested earlier, your best strategy at the outset is to cast a wide net—brainstorm, think divergently. The idea is not to create a shopping list of every conceivable variable, but to identify a range of factors that may substantially influence performance or inform your understanding of the phenomenon.

Novice researchers tend to identify one or two factors of central interest, and then rely on randomness to handle other effects. Such a strategy leaves too much to chance. Keep in mind the following principle: If you ignore factors that influence performance, variability from these sources does not disappear; instead, it remains to cloud the picture. In a well-controlled study, the researcher pins down important sources of variability, ensuring that systematic effects stand out clearly against background noise.

For practical purposes, we distinguish three primary types of factors: (a) treatment factors; (b) person or individual-difference factors; and (c) outcome factors. A fourth category, nuisance or control factors, is also useful in preparing a design.

A treatment factor is an environmental variation directly controllable by the researcher. Argument type and task might serve as treatment factors in your study. You decide to introduce students to two types of arguments: a simple version where all the evidence supports a single claim and a complex form where different facets of a claim are supported by different pieces of evidence. You arrange two types of classroom interaction: one in which students work together in small groups to analyze the two types of arguments, and another in which the teacher models the analysis through lectures. You have defined two treatment factors, each with two variations.
The primary goal in identifying factors is to assess the importance of each variation in its own right, the main effect of the variation. To what extent do students perform differently on simple and complex arguments? What is the effect when students work in small groups? Main effects are the primary goal from most studies, and it usually makes sense to include two or more primary factors in the design. In a complete factorial design, the research plan includes all combinations of the factors, which can increase the cost of the study, but not by much, and sometimes not at all. For instance, you could conduct two separate studies, one for each factor, yielding a total of four different conditions. If you combine the two factors, then two times two equals four, which is the same number of conditions as for two separate studies.

The real payoff from a factorial design is that you can also assess the interaction among factors. An interaction occurs when the effect of one variable depends on conditions associated with a second factor. For instance, simple arguments may not benefit from social interaction, but complex arguments may be more readily comprehended under this condition. Interactions are not bad, nor are they necessarily complicated. They often play an essential role in providing contextual information, in helping the researcher understand the conditions under which a treatment is more or less effective.

A person factor is an intrinsic characteristic of an individual or group. Age, sex, ability, and prior experience are examples. These factors guide the selection of teachers, students, and classes, either because of theoretical interest in the effects or to control extraneous variability. For instance, if you know that some students have learned about arguments while others are unfamiliar with the concept, then you should include student experience as a design factor. If some teachers understand the argument genre better than others, then you should include teacher understanding as a factor. Interactions can occur among person factors. Students with no prior instruction about arguments might benefit from a teacher who understands the argument genre, but having a teacher who understands the genre well might matter less for students with prior knowledge. Interactions can also emerge from combinations of treatment and person factors. For instance, more experienced students may not benefit from small-group interactions, while novices do better in a small group than when left on their own.

Outcome factors direct the choice of measures in an investigation. Like treatment factors, they can be directly manipulated by the researcher, although this opportunity is frequently overlooked. The tendency is to select an off-the-shelf instrument without thinking about its relation to the research questions. Your school probably administers a standardized comprehension test. Why not use this test to assess the relative effectiveness of the two programs? In making this decision, you face trade-offs. On the one hand, most standardized tests use rather vague expository passages, seldom in the argument genre, and they tap the students’ ability to recognize rather than to produce.

Given such limitations, you should think about constructing measures that directly assess students’ ability to handle argument structures, that demonstrate their ability to craft a persuasive text, and that reveal attitudes and confidence about these tasks. On the other hand, standardized tests are proven instruments with established reliability estimates, while your measure has neither established reliability nor validity. You might use a standardized test as an index of general student ability, and rely on your measure for a more focused look at students’ composition of arguments.

Once you have chosen the factors for your design, you then need to decide on the levels for each factor. Sometimes the decision is straightforward; if sex is a factor, then male and female are obvious choices. For a factor like undergraduate major, the range of options is greater, and the selection requires more thought. If revision time is a treatment variable, the number of options is virtually infinite. Think first about the relation between this factor and performance. For instance, does performance increase steadily with time? Might it increase for a while and then decline? Beyond a certain point, further time might actually lead to a poorer outcome. Each factor in the design requires careful consideration along these lines.

Building a Factorial Design. The simplest way to construct a design from factors is to treat them like Lego blocks—simply snap the pieces together. This strategy works well as a start, but needs additional refinements. Two factors can be joined in either of two ways: crossed (every level of the first factor is combined with every level of the second factor) or nested (the levels of the second factor differ for each level of the first factor). The two plans are illustrated in Figure 56.1, where you can see the parallel between a matrix and a hierarchy. In a matrix, every level on the first dimension is combined with every level of the second dimension. In a hierarchy, the lower levels may have a common thread, but they do not connect to other points at the same level. When a set of factors is crossed, you can evaluate both the main effects of each factor and the interactions among them. When factors are nested, only the main effects of the primary factor can be evaluated, because the design does not include combinations of the two factors.

As illustrated in Figure 56.1, nesting may be required by the nature of two factors. When the researcher decides to vary text type, then different works are needed within each type. You could try to rewrite Three Little Pigs as a compare-contrast or argument, but the result would lose something in translation.

Factorial methods have two advantages. First, like Lego toys, they can combine large numbers of factors in simple but flexible ways. Second, factorial methods offer a strong defense against inadvertent confounding, because the effects of any two factors are independent of one another. This assurance has two caveats. First, each combination must include an equal (or proportionate) number of observations. For instance, suppose you divide a writing class into high and low achievers (the achievement factor) which
you cross with boys and girls (the sex factor). You are likely to find more high girls and low boys than the other two combinations; the design is partly confounded because the two person factors are correlated; the boys level is also low achieving, and contrariwise. This problem can be serious, and needs to be fixed. One approach is to divide boys and girls into two groups of equal size, using the achievement midpoint (the median), which reduces the confounding.

Second, the factorial strategy offers no guarantee that a particular factor will not be confused with other factors not in the design. Low-achieving may also mean from poor families, for instance, a problem that can be dealt with in a similar manner.

An important consideration in planning a design is deciding how to assign individuals or groups to various treatment combinations. The issue is generally discussed in research texts as the choice of a between-subject or within-subject plan, but it is better thought about as a decision about crossing versus nesting persons with other design factors. The decision to nest or cross persons with other factors reflects both practical and theoretical considerations. The researcher sometimes has little choice for practical reasons. For instance, individual-difference factors like sex or personality dictate that individuals be nested within the levels of a factor. A person is either male or female, impulsive or reflective. Treatment factors can generally be crossed with person factors, and sometimes it makes sense to do so. If a treatment combination takes only a minute or two to administer, and the student is available for an hour, the researcher should probably administer as many conditions as possible, which means crossing student with several factors.

CROSSED AND NESTED PERSON DESIGNS

CROSSED DESIGN

<table>
<thead>
<tr>
<th>Factor B</th>
<th>Factor A</th>
</tr>
</thead>
<tbody>
<tr>
<td>B&lt;sub&gt;1&lt;/sub&gt;</td>
<td>A&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>B&lt;sub&gt;2&lt;/sub&gt;</td>
<td>A&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

NESTED DESIGN

<table>
<thead>
<tr>
<th>Factor B</th>
<th>Factor A</th>
</tr>
</thead>
<tbody>
<tr>
<td>B&lt;sub&gt;1&lt;/sub&gt;</td>
<td>A&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>B&lt;sub&gt;2&lt;/sub&gt;</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>B&lt;sub&gt;3&lt;/sub&gt;</td>
<td>A&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Figure 56.1 Two types of factorial linkage: Crossed and nested.

Research Designs for Empirical Research

Theme. The final ingredient in a design is the conceptual framework that guides selection of factors and decisions about how to combine them. While we have placed this topic at the end of our list, it is of paramount importance. The thematic foundations of a research study require knowledge of the territory, experience in dealing with the issues, and a large dollop of intuition and art. On the other hand, the task can also be guided by a systematic strategy, for which Simon (1981) again offers counsel. Although some systems appear complex on the surface, Simon argues that all are fashioned around a relatively small set of separable components, each with a distinctive internal structure, each linked in simple ways to one another. We applied this notion earlier to the composing of a written argument and the planning of the sample study. It also applies more generally to the conceptualizing of virtually any research problem.

The key is to look for the joints that divide a complex system into a few simpler entities. Carving a turkey provides a metaphor. A turkey can pose quite a challenge to the novice carving the Thanksgiving bird. The trick is to find the joints, so the carver can divide a big job into a few relatively small ones. Think about messes, lumps, chunks. If you carve a problem into a lot of little pieces, you will be overwhelmed by the details. If you try to handle the problem as a whole, you will be confused by the apparent complexity. Human beings can most effectively handle a few items at a time; the key is to keep it simple—more to the point, make it simple.

How do you know when you have hit a joint in a conceptual domain? We suggest that when the technical language and relations in one chunk differ from those in another chunk, you have found a joint. The previous discussion about selecting treatment, person, and outcome factors illustrates the point; we talked differently about the choices within each of these domains. Locating the chunks, then, is the key to analysis of a complex question; it also lays the foundation for synthesis, for bringing the chunks back into relation to one another.

Let us apply this reasoning to the previous vignette. Your initial thinking about argument was fuzzy and complex. You saw the issues as one dimensional, and the best method seemed obvious. But then you were soon burdened by technical details of control. Looking for a few joints allows you to divide the big problem into manageable chunks that help with management of the details. You have already moved in this direction by focusing on two thematic areas: forms of argument and styles of integrated instruction. Both areas have a distinctive technical base; each can be considered as an entity in its own right.

You can then apply the divide-and-conquer principle to each of the two domains. For instance, how might you subdivide the complexities of instruction—pedagogical method, materials, and management? The answer is implicit in the context.
question. Divide each big chunk into a handful of distinctive subchunks, and decide which are critical to your research question. The chunks will then need to be re-related to one another, but the capacity to assess interactions is inherent in the technology of factorial design.

Creating the Design

This section addresses ways in which the previous concepts and procedures apply to construction of a specific research plan. Here we move from divergent to convergent thinking, from strategy to tactics. Assume that you have identified two thematic issues: how well can students comprehend arguments that include warrants, and in what ways do opportunities for small-group interaction help support both comprehension and composition. You plan to use multiple-choice tests to assess comprehension and writing assignments to tap into productive competence. Your research questions are reflected in the design: What is the impact of small-group activities on students’ ability to handle more or less complex arguments when they are assessed by recognition and performance tasks? You have created a learning model that describes your hunches about the underlying mechanisms at work under the various conditions, for students with more or less experience in dealing with arguments, from which you have ventured several expectations about the pattern of the results.

Now you need to formalize a plan of action. You have several options, which can be guided by a few principles. First, two thematic chunks—how students process arguments and how instructional approaches produce learning—need to be expanded into operational factors. Second, you should think about starting with two or three bite-sized investigations rather than putting all of your eggs into a single basket. Third, you need to keep the ultimate goal in mind, and take care not to drown in details. The factors selected for the design support the thematic foundations of the study, while ensuring that the design controls significant sources of extraneous variability. The following sections offer practical advice about refining the design and preparing the details of the research plan. We have selected a few items for emphasis and illustration.

**Big Picture and First Steps.** The first word of advice is to keep your primary goals in mind, and keep moving in that direction—unless you have a good reason to change direction. You have shaped the elements of a plan; an image of the research problem is taking shape in your mind’s eye. How should you proceed next? As noted above, one option is to proceed with a full-scale experiment. Another is to conduct a series of mini-studies. A third, returning to your first idea, is to pursue a naturalistic investigation based on observation, interview, and assessment. Our suggestion is that you keep in mind all levels of this continuum, but especially the middle. You can collect preliminary data while also refining your thoughts about the big picture. Immersing yourself in actual data will have both practical and theoretical benefits.

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Comprehension Processes</th>
<th>Method of Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>Argument differences</td>
<td>Small Groups vs. Teacher Lectures Comprehension &amp; composition</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Different levels of expertise</td>
<td>Different levels of expertise</td>
</tr>
<tr>
<td></td>
<td>Performance measures</td>
<td>Performance measures Transfer measures Satisfaction measures Field notes</td>
</tr>
</tbody>
</table>

**Figure 56.2** Design for study of argument comprehension and social context of instruction.

You began the project with a literature review, which helped you shape a theoretical perspective. Refining a conceptual framework requires abstract thinking, but it also benefits from practical experiences. The challenge is to translate your ideas into concrete form. A graphic organizer may help you document your evolving plan. Figure 56.2 shows a midstream road map for moving from Post-It notes to the next step in your design. The matrix presents the two thematic elements as column headings; the rows show the factorial categories central to most research plans. The sketch provides a structure for capturing ideas; the entries in the figure are only illustrative. Creating the plan is a dynamic enterprise; use Post-Its, or record your thoughts on a word processor. Ask colleagues for comment and criticism. Be flexible; the one constancy in research design is change.

Our second recommendation is that, as your plans take shape, spend time in your research context (e.g., classrooms, teachers, and students), looking and listening, trying out your ideas and procedures and materials in realistic settings. What you see and hear will be guided—consciously or unconsciously—by the conceptual framework that fills the walls of your study area. Before the design is cast in stone, recheck the context. What are potential sources of evidence? What variations are especially critical? Where are you least certain and most confused? What questions should you pose to informants? What answers do you expect, and how can you follow up for clarification? These initial forays into the field, though tentative, should be guided by your design. You are still framing the research question. You are still developing instruments. The decisions springing from ongoing descriptive work can continue to shape the study.

**Data Analysis** The design proposal was finally approved, you decided on your strategic course, came to know the territory, and the data are in hand. This chapter is about the role of design in the research enterprise, and we have neither the mandate nor space to say much about the tasks of data analyses. Some remarks about the connection between design and analysis do warrant attention, however. Whatever
the nature of the evidence, one basic principle should be followed, whether the evidence is quantitative or qualitative, whether in the form of numbers, field notes, interviews, pictures, videotapes, or whatever. The principle is to fully immerse yourself in the data. This activity is commonplace when the data are qualitative and the design is a case study, but less so with a factorial design and quantitative data. Here the usual approach is to load the data into a statistical program and see what emerges—hopefully including some statistically significant findings.

This approach is a mistake for a variety of reasons, which we will not explore in detail. But here are some guidelines for an alternative strategy. One job is to summarize, where you move through the basic observations looking for eye-catching trends. For numbers, the search for trends falls under labels like mean, variance, and correlation. The usual tactic is to ask the computer for descriptive statistics. For the complex observations typical in qualitative analyses, the approach is to immerse yourself in raw observations, transcribing recordings, filling thick binders with Post-Its and multicolored highlightings. The researcher, following the tenets of grounded theory (Glaser & Strauss, 1967), is also looking for trends, springing from consistencies in the evidence, based on informed judgment and a search for consistencies. In an ongoing debate in Educational Researcher on epistemological differences between quantitative and qualitative analyses, Ercikan and Roth (2006) argued that the differences in methodology might not really be all that great; that all research is based on perception, all perception has both qualitative and quantitative aspects, and all types of methodological justification must lead to logical/reasonable conclusions (Chambless, Alexander, & Price, in press).

Both tactics share important strategic components. The researcher dealing with numbers is now advised to explore the data, to study frequency distributions, prepare scatterplots, and be on the lookout for unusual events. The field of exploratory data analysis (Behrens & Smith, 1996) includes a range of systematic techniques for tasks similar in many respects to the work of qualitative researchers. Likewise, the qualitative researcher can call upon computer programs like NVivo (www.qsrinternational.com) and QDA Miner (www.provalisresearch.com) to replace Post-It notes and highlighters for data management but, more importantly, to handle the tasks of classification and summarization. The new programs even provide the qualitative researcher with the capacity to count and to measure. The researcher exploring qualitative evidence is well advised to look for trends analogous to those found in statistical methods. Central tendencies—what are the typical elements in the data set? Variability—what deviations from typicality do you find? Correlations—in what ways do parallel trends seem to emerge? You may not be able to attach precise indices to these trends, but you can convince your readers of their existence. To be sure, arguments based solely on anecdotes rest on perilous ground.

Exploration and summarization can both be linked to design issues in factorial plans. We mentioned earlier that factorial designs allow the researcher to evaluate both main effects and interactions. To remind you, main effects reflect differences that emerge as you move from one level to another of a factor, such as the differences between males and females, or between writing by individuals versus small group assignments. Interactions describe patterns associated with factorial combinations; girls do better than boys when writing individually, while boys do better than girls in small group settings. Statistical procedures like analysis of variance generate indices for identifying reliable differences associated with main effects and interactions. An exact parallel does not exist for qualitative methodologies, but the researcher can still examine the evidence for such patterns and develop arguments to support various conclusions. In doing so, the researcher may find himself or herself falling back on numbers, as in the following example:

In 70% of the small group protocols, boys expressed a competitive stance on the writing task, whereas girls voiced a more cooperative slant. These trends were supported in the interview data. When I talked with students after individual writing assignments, competitive, cooperative motivations were mentioned by only 15% of the students. The notion of including both quantitative and qualitative evidence in the overall design of research studies has a long history, but has advanced in the past two decades in new and invigorating forms. Classic research methods textbooks began to emphasize a mixed strategy in the 1980s (e.g., Jaeger, 1988), usually by recommending that a primary method be complemented by secondary methods. For example, a large-scale survey of writing practice (quantitative) might provide direction for selection of a set of individual case studies (qualitative). Creswell and his associates (Creswell, 1994, 2003, Creswell & Plano Clark, 2007) expanded this idea, developing a taxonomy of mixed-methods combinations applicable to a variety of settings. These treatments tended to keep the two basic methodologies—quantitative and qualitative—at arm’s length. The reasons for maintaining the separation are understandable. The two methods approach research problems from different epistemological stances—different ways of thinking about the nature of knowledge and the tasks of science. They tend to rely on different design strategies—case studies versus factorial designs and large-scale surveys. They obviously differ in the type of basic evidence upon which they rely—numbers versus words—although Ercikan and Roth (2006) point out many underlying epistemological similarities.

Even if an underlying logic could be used to integrate qualitative and quantitative approaches, quantitative and qualitative researchers differ in the language that they use to talk about their work. Calfee and Sperling (2010) liken the situation to the Tower of Babel, and propose an integrated model for mixed-methods research strategies. They acknowledge that epistemological differences pose a fundamental challenge to a genuine merger of the fields. Nonetheless, by proposing a bridging vocabulary, they show...
how activities such as exploration and summarization can be connected within a study, not as parallel play, but as genuinely integrated endeavors. This approach begins during the framing of the research problem, where methods emerge as appropriate complements to the problem. They present illustrations from previous studies in the language-literacy field demonstrating the possibilities springing from an integrated approach, including connections with factorial design strategies. Greene (2001) has also argued for the power of integrating the two approaches, explaining that “Points of tension between the different methods...are juxtaposed and analytically blended into a dialectic synthesis. It is this synthesis in integrated designs that fulfills the fundamental mixed method purpose of understanding more fully” (p. 57). If a similar chapter is prepared a decade hence, the author may no longer preface the work by describing such designs as primarily within the purview of quantitative outcomes, but as a general strategy for research planning.

Interpretation

The study is complete. You have implemented the design, have completed the analyses, and the data are in the bag. Unfortunately, data do not answer questions; people do. For evidence to have meaning, you must deal with several more questions. How far can you trust the evidence; how far can you generalize the findings; how convincingly can you persuade others of your interpretation? You hopefully reflected earlier on your response to various outcomes—before you collected the data. The findings either support your expectations or they have surprised you.

Suppose the results turn out as you predicted. You show that students in the novel treatment for analyzing and composing arguments are more likely to participate vigorously in classroom discussions about one another’s arguments and to prepare coherent arguments for their classroom assignments than those in the traditional approach. What does this result mean? Your argument appears straightforward; the innovative approach is superior, supporting your convictions about what students need to learn and how they can best learn it.

Suppose the results do not come out as expected? You may have difficulty imagining this outcome. Given all your planning, thinking, and work, how could this happen? But it does. The most frequent disappointment occurs when an innovative treatment produces little or no effect, when the null hypothesis (no difference) cannot be rejected. This result can come about for either or both of two reasons. First, the treatment may actually not be effective—hard for you to accept, but possible. Second, student performance may vary so widely that random fluctuations swamp the treatment effect. It is like a slot machine, which costs you on each play; you do not immediately notice the loss because sometimes you win and sometimes you lose. A well-conceived research design allows you to identify extraneous sources of variability in performance, so that you can tell whether you have won or lost.

At the core of interpretation is establishing the validity of the findings, ensuring that your argument holds up under close scrutiny. You are about the concept of validity as it applies to testing: Validity is the extent to which a test measures what it is supposed to measure. In fact, recent thinking (Messick, 1995; Kane, 2006; Lissitz, 2009) about validity has taken a different turn: “Validity is the strength of the argument that a particular test outcome means what the tester says that it means” (Lissitz, 2009, p. 742). Test theorists approach validity from two perspectives: internal validity and external validity (Campbell & Stanley, 1966; Cook & Campbell, 1979; Porter, 1997). Internal validity responds to the question, “To what degree can I trust the evidence that is within my grasp” External validity asks, “To what degree can I extend the findings to other situations?”

The matrix in Figure 56.3 links internal and external validity with the design perspectives from this chapter. The matrix looks in turn at factors controlled in the design and at uncontrolled or free floating factors. The cell to the upper left corresponds to the original notion of internal validity. The other three cells are variations on external validity.

The first test of validity, conceptual clarity, centers on the adequacy of the factorial design. Now that the data are in, how clearly can you tell what happened? To what degree do the factors appear as compelling representations of the constructs (the underlying concepts) that are incorporated in the conceptual framework and the research questions? To what degree can you make sense of patterns in the data? Complex interactions may have appeal when you first think about a problem, but they can also render interpretation difficult. To what extent did the treatments work as intended? Most significantly, to what degree might the study be replicated with some assurance of obtaining similar results? Secrest, West, Phillips, Redner, and Yeaton (1979) refine these points: “The essence of construct validity is that one has a good understanding of the conceptual meaning of the treatment...It refers to our interpretation of the treatments, not the treatments themselves” (p. 17). For instance, you may discover that when you form small group writing teams, student exchanges do not play out as you intended. You imagined a cooperative enterprise, but qualitative observations reveal cooperation, along with competition and a lot of parallel play. This finding is not a failure—you have learned something from the results.

The second validity test, situational stability, is the degree to which the evidence allows you to project the basic findings with confidence to other contexts, based on the original design. What about the influence of factors that you decided to ignore; how do they appear to influence the outcome, directly or through interactions? If the sample of participants is too small or too homogeneous, then you may not feel confident about extending the findings. If the instruments are too specialized, you may be hesitant to recommend your results to others. Replication is again a key consideration; how confident do you feel that you would obtain similar results if someone else repeated the
study, keeping the design intact, but with different samples of participants and in different contexts?

The next two categories go beyond the details of your original design to extensions of the underlying constructs. Researchers seldom limit their interpretive scope to a particular study. You are interested not just in the program that you have developed, whatever shape it may take in the final design, but in the concepts that undergird the program. Researchers aspire to broadly generalizable statements, and here the issue of external validity comes to the fore.

Figure 56.3 has two entries in this column. First is conceptual match. In going beyond the original conditions, while staying close to the original conception, how safe are you in projecting your results? A great deal depends on the clarity of the original conceptualization, and the degree to which the basic conditions can be implemented in a similar manner under variations in the specifics of the design. The argument program showed considerable promise on its first test flight. A colleague plans to implement the program in a different setting, but must modify it to fit local conditions. What are the boundaries? The small-group activity seemed to be a critical component. You developed a series of group planning guides with scripts and graphic organizers, which appeared quite effective in scaffolding students’ efforts to lay out the claim-evidence-warrant trilogy in complex arguments. Duplicating the guides was expensive, and your friend is planning on a much reduced version. You provided a week of professional development before the start of the school year. Your friend is thinking about two hour-long after-school meetings. What should she keep and what can be jettisoned? Changes of this sort can be thought of as “dosage” issues, as changes in the intensity of the treatment. But as you can imagine, the changes also influence the conceptual structure of the program in ways that are difficult to assess.

Finally there is situational match, which is related to what Cronbach (Cronbach, Glesser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 2006) labeled generalizability. Suppose a user wants to change the program and then implement it in a different situation; what are the chances that the results apply under these circumstances? Your program has been tried out under one set of conditions, with certain factors under control. The students are from middle-class backgrounds, the classes are relatively small, the teachers are experienced professionals, and resources are available for staff development and collegial interactions. Do the findings hold up in situations where these conditions do not hold? If the treatment is powerful, then the variation in program specifics and local contexts should not matter. An investigation should ideally provide linkages that inform judgments about the transferability of the findings.

The two dimensions in Figure 56.3 are represented by two binary levels; in fact, they are better thought of as end points on a continuum. Any effort to replicate a study necessarily entails changes of various sorts. Some will be obvious, others more subtle. Some will be well planned, others more accidental and incidental. Research depends on informed judgment, which depends in turn on continuously wrestling with the conceptual issues. Interpretation is a form of pattern detection; fortunately, the human mind excels at this task.

### Concluding Thoughts

Research is problem solving, and the problems are real. Empirical data are part of the process, though not always the most significant element. Educational and social science research are particularly demanding because the theoretical foundations are weak and researchers sometimes overlook the theoretical tools that are available (Suppes, 1974). But the times are a-changing, and rapidly. Cognition and social cognition, the practical emphasis on educating rather than

<table>
<thead>
<tr>
<th>Conditions Within Immediate Domain of Study (design remains constant)</th>
<th>Conditions Beyond Domain of Study (design changes; underlying concepts remain constant)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factors Within the Design</strong>&lt;br&gt;CONCEPTUAL CLARITY (Affected by the design factors)</td>
<td><strong>CONCEPTUAL MATCH</strong> (Affected by the replicability of underlying concepts)</td>
</tr>
<tr>
<td>* Clarity of concepts</td>
<td>* Repeatability of plan</td>
</tr>
<tr>
<td>* Size and simplicity of effects</td>
<td>* Interactions</td>
</tr>
<tr>
<td>* Inadvertent confounding</td>
<td>* Faithfulness to plan</td>
</tr>
<tr>
<td><strong>Factors Outside the Design</strong>&lt;br&gt;SITUATIONAL STABILITY (Affected by the effect of factors in the situation)</td>
<td><strong>SITUATIONAL MATCH</strong> (Affected by the degree of similarity between situations)</td>
</tr>
<tr>
<td>* Extraneous factors, direct and interactive</td>
<td>* Presence of new conditions and contexts</td>
</tr>
<tr>
<td>* Reliability of measures</td>
<td>* Reliance on new tests</td>
</tr>
<tr>
<td>* Size of sample; number of observations</td>
<td>* Different people and groups</td>
</tr>
</tbody>
</table>

*Figure 56.3* Matrix displaying impact of design and context factors on internal and external validity of a research study.
training, and the challenge of helping every individual realize his or her full potential—the road ahead is exciting and demanding.

Educational and social science research nonetheless remains in the sleepwalking phase (Koestler, 1968). Even the best of our theories are more heuristic than formal, and we must often rely on experience and intuition. Success frequently depends on doing several things right rather than pursuing the one best answer (Slavin, 1986; Tyack, 1974). Cronbach (1975) paints a gloomy prospect for generalizable research in education, portraying a hall of mirrors with infinitely complex and intricate interactions. In fact, where research designs allow evaluations to be evaluated, the evidence suggests that simple interactions are fairly commonplace, most often in conjunction with substantial main effects. Complex interactions, however, are actually quite rare.

Although the tasks may appear daunting, we are optimistic about the enterprise. Researchers are well advised to bite off only as much as they can chew comfortably when first approaching a problem, while also avoiding taking merely baby bites. Whether as producer or consumer of empirical research, you might consider the divide and conquer strategy. A series of modest but well-designed studies is likely to be more informative than a single humongous effort. Critical experiments are rare in our business; a researcher has made a contribution if his or her investigation provides one or two insights—which may spring from a mistake that suggests what not to do. When you read a research report, it may resemble bowling; the investigator sets the pins, throws the ball, and counts how many pins fall. Reality is different. “Some of the most excellent inquiry is free ranging and speculative in its initial stages, trying what might seem to be bizarre combinations of ideas and procedures, restlessly casting about...” (Cronbach & Suppes, 1969, p. 16). But threading through all the elements is one critical theme—design.

We have not recommended a fixed algorithm for planning empirical research, but the strategy exemplified in the earlier vignette often works quite well. First, learn as much as you can about the territory through a descriptive study. Your goal is often driven by person factors such as motivation and psychological processes. Then, move on to experiment and innovate; try out a series of instructional treatments, perhaps one or two chunks at a time. Innovations are difficult to implement, and success is more likely by proceeding in phases. Plan to assess the implementation and to examine in detail the full range of potential effects (positive and negative). You may not be able to complete an in-depth evaluation of every participant; a better mixed-method strategy is to select a few individuals for “thick” study, for contrast with the thinner data from the entire group.

Our main message throughout has been the essential importance of design—basic building blocks, linkages, and an overarching theme. We have focused on factorial designs because they are simple, flexible, and have been thoroughly studied for almost a century. The various components assume different shapes in different stages of an investigation, but if you build on them consistently, they provide coherence and unity for the effort. You will learn something from the experience, and will gain satisfaction from the enterprise—whether you take on the task yourself, or move through an investigation vicariously.

References


