

BY DESIGN

PLANNING RESEARCH ON HIGHER EDUCATION

RICHARD J. LIGHT

JUDITH D. SINGER

JOHN B. WILLETT

We are living in a time of rapid change. The world is becoming more global, more interconnected, and more complex. The challenges we face are unprecedented. We need to think differently about higher education. We need to plan for the future. We need to design a better world.

Harvard University Press
Cambridge, Massachusetts, and London, England
1990

BY DESIGN

WHY DO RESEARCH ON HIGHER EDUCATION?

1

Is your college doing a good job of teaching undergraduates to think critically? Do your students write clear and gracious prose? Which professors are the most effective teachers? What do they do that makes them so effective? Could others become more effective by emulating them? Are students integrating modern technology into the way they work and the way they learn? Do students who use computers learn more than those who do not?

Such questions are not new. But many of our colleagues, including faculty, administrators, students, legislators, and parents, are asking them with new urgency. Perhaps this is because of increasing competition among colleges. Perhaps it is because of a renewed sense among faculty and administrators that, as tuition rises dramatically, they should work harder than ever to deliver the best education

possible. Perhaps it is because consumers are demanding more value for their money. Whatever the reason, more campuses are initiating research and using the results to strengthen educational quality.

We hope to help by making methods for planning good research more easily accessible. We have written this book as a resource for those who want to conduct such research. If ever there was an ideal organization to encourage systematic research, it is the university. Faculty members are generally aware of what it means to do research, even if each professor is not an expert on every intricacy of the empirical method. Faculty members work hard to enhance students' learning, because, despite the cynical views of some of education's critics, most professors take pride in their teaching and work hard to do it well. Organizing systematic ways to use information to improve teaching and learning is a widely shared goal.

Many Questions, Many Options

Research on higher education can address diverse questions. Our goal is to help you design first-rate studies to answer them. We use three general paradigms, which we will call *descriptive*, *relational*, and *experimental* inquiry. Each leads to results with concrete implications for policy and practice.

Descriptive studies are used for doing exactly what their name implies—describing the way things are. They answer questions such as: How well do students write? What are the most popular courses on campus? How many graduates are accepted to medical school? How much money do our graduating seniors owe? Descriptive studies characterize the status quo; they do not tell you *why* things are the way they are.

Beth Schneider (1987) used a descriptive study to esti-

mate the prevalence of sexual harassment at a major Eastern public university. She contacted a random sample of female graduate students, and of the 356 students who returned a mail questionnaire, 60 percent reported having been harassed in some way by a male professor at least once during their graduate career; 10 percent had been sexually propositioned. Schneider's startling results documented the need for university guidelines on sexual harassment and for educational programs designed to ameliorate the problem.

Glenda Rooney (1985) organized a descriptive study at the University of Wisconsin at Madison to profile minority students' participation in campus student organizations devoted specifically to minority concerns. She selected a stratified random sample of minority students, and of the 322 interviewed, 98 percent belonged to at least one campus organization. But fewer than 20 percent belonged to an organization specifically devoted to minority concerns. Rooney's results refuted a *Newsweek* poll suggesting that minority students restrict their campus involvement to minority student groups.

Relational studies are used to examine relationships between two or more factors. You can use them to answer such questions as: Are men more likely than women to persist in studying science? Do dropout rates differ by student socioeconomic status? Do varsity athletes learn as much as their classmates who don't play on a team? In a relational study, you examine natural variation in predictors and outcomes to figure out whether they are associated. Relational studies help you move beyond simple descriptions to understanding *why* things are the way they are.

Christos Theophilides, Patrick Terenzini, and Wendell Lorang (1984) organized a relational study at the State University of New York at Albany to examine the stability of students' choice of major and what characteristics are

associated with the likelihood of change. More than 300 students completed a questionnaire during orientation week and follow-up questionnaires near the end of both freshman and sophomore years. By the end of sophomore year, 77 percent of the students had changed majors: 32 percent changed once and 45 percent twice. Students who changed had lower GPAs and less clear academic objectives than students who did not. Theophilides and his colleagues used these results to highlight the need for early student advising and to suggest methods for identifying students most in need of such services.

Ernest Pascarella, John Smart, and Corinna Ethington (1986) found a relational study helpful for examining the institutional and personal characteristics associated with the likelihood that students attending two-year colleges would eventually pursue and complete a bachelor's degree. Among a national probability sample of 825 students who entered college in 1971 and were followed up each year until 1980, 53 percent had obtained a bachelor's degree and 16 percent were still pursuing a degree. Students were more likely to persist if they were integrated into the academic and social systems of their college. Pascarella and his colleagues concluded that policies and practices that would enhance students' academic and social integration into campus life might increase the likelihood of long-term persistence.

Although relational studies allow you to identify an association between predictor and outcome, the type of relationship you can talk about is limited. With a relational study, you can only talk about *correlation*, not *causation*. Because you are examining *natural variation*, you can never be sure whether a predictor *causes* the outcome to behave the way it does, or whether the effect is caused by some other predictor that you failed to study. Does student integration into academic life cause long-term persistence, or does some other factor, such as prior academic prepara-

tion, cause both academic integration and long-term persistence? Both Theophilides and Pascarella and their colleagues used statistical analyses to rule out many of these rival explanations. But as both teams of researchers concede, relational studies cannot establish causation.

You also can use relational studies to compare the effects of naturally occurring treatments or programs, but the problem of causal attribution persists. Kathleen Berg (1988) conducted a relational study of the association between residence arrangements and eating disorders among 584 female undergraduates at the University of Western Ontario. Fifteen percent of the women met standard criteria for bulimia. Those living on coed floors of coed dorms displayed more bulimic symptomatology than their peers living in women-only residence halls or on women-only floors in coed dorms.

Does coed living *cause* bulimia? It is hard to say, because the students Berg studied chose their own living arrangements. How will we ever know whether it was the living arrangement, and not some other characteristics of their backgrounds associated with choice of living arrangement, such as sex-role development, that caused the increased prevalence of bulimic symptomatology? But even without pinning down a causal link, Berg's compelling results on the prevalence of bulimia across *all* dormitory settings at Western Ontario led to a training program for residence-hall staff on the detection and treatment of eating disorders.

To establish a causal link, you must conduct an *experiment*. In an experiment, you implement a specific treatment, or set of treatments, for the explicit purpose of learning about its efficacy. You intervene in the system, control the experiences of everyone you study, and watch what happens. The statistical principle of random assignment helps you to rule out rival predictors that might "explain away" your findings, eliminating the shadow that always

looms large over relational studies. Of the three research paradigms we discuss, only experimental inquiries allow you to determine whether a treatment *causes* an outcome to change.

John Belland and his colleagues (1985) designed an experiment to determine whether a moderate amount of external pacing improved a microcomputer-based instructional program for teaching undergraduate biology. They randomly assigned 100 freshmen at Ohio State University among three instructional programs and a control group. Comparison of student knowledge after completion of the instruction revealed that, while all three experimental groups differed from the control group, students working with a moderate level of external pacing learned the most. Because they conducted an experiment, the researchers were able to conclude that external pacing *caused* the improved performance.

Larry Weber, Janice McBee, and Jean Krebs (1983) conducted an experiment at Virginia Polytechnic Institute to investigate the effects of test administration ("in-class closed book" versus "in-class open book" versus "take-home") on student achievement. Sixty-four students were randomly assigned to three groups, and each group took three tests, one of each type. When students took the take-home tests, their scores on knowledge items were higher and their levels of anxiety were lower than when they took the in-class tests. No evidence of rampant cheating was found with any test format. Because the researchers conducted an experiment, they were able to conclude that differences in test format *caused* the differential results.

The beauty of experimental studies stems from the strong, clean inferences you can draw from their results. An experiment reduces ambiguity. Causal attribution is clear. Of course, not all research questions can be addressed experimentally—you can't randomly assign students to different sexes, for example—but when experiments are feasible, they are preferable to relational studies.

Our Philosophy of Research Design

Through our own research experience and work with colleagues, we have developed several principles for designing research. We offer four maxims to reveal our biases and to foreshadow the type of advice we give throughout the book.

Our basic tenet is that *your study's design is the single most important factor that determines whether your findings will be scientifically first-class*. When colleagues seek us out for assistance, they often have their data in hand and they ask us to suggest appropriate statistical analyses. We nearly always find that their entire project would have yielded more useful and more convincing results if they had thought through their design more carefully before collecting data. Elaborate statistical analyses rarely, if ever, can retrospectively correct weak project design. Taking extra care at the design stage is well worth the extra effort.

A corollary is that you should *explore many design options before adopting any plan, especially a weak one*. When practical constraints prevent you from implementing the ideal design, don't say: "If I can't do it right, it doesn't really matter how I do it. I'll just do something quick and dirty." Research designs form a hierarchy and, if your first-choice design isn't feasible, there is often a second choice distinctly preferable to an entirely uncontrolled investigation. Explore all the possibilities, and then decide.

Our third maxim—*pay enormous attention to detail*—may seem trivial, little more than common sense, but, as we constantly remind ourselves and our colleagues, the little details determine the ultimate credibility of a project. If you want to evaluate a new curriculum, for example, you can choose among many different designs, several of which use randomization in the selection and assignment of students. But there are many different ways to randomize. Depending upon the method you choose, your ultimate analyses may have very high power or very low power, all

at the same cost and with similar numbers of professors and students involved. Paying attention to detail always has a high payoff.

Our final maxim urges an attitude for approaching research in higher education. Most of this research requires student participation, and administrative and faculty cooperation too. You should continually remind yourself, as well as your administrative and faculty colleagues, that *research must respect collaboration and cooperation*. This attitude has an important implication: treat people's participation with respect, don't settle for a weak research design. Otherwise, in Frederick Mosteller's words, you will be doing little more than "fooling around with people."

How This Book Is Organized

This book is written in the format of a discussion with colleagues. In its pages we ask the very questions we ask colleagues who come to us for advice on research design. They are questions you should ask yourself when planning your study. The answers should guide you toward an effective plan. One good way to use this book is to read it with a concrete research problem in mind. If you don't have a specific problem, you may still find it useful for assessing the work of others. Although you can read any chapter on its own, we believe you will get the most from the book if you read it sequentially.

In Chapter 2, we ask: *What are your questions?* Before you can design your project, you must decide exactly what you want to know. Well-crafted research questions guide the systematic planning of research; without research questions, you will not be able to manipulate those facets of design that increase your ability to learn what you want to know. In this chapter, we suggest specific strategies for writing research questions and show how to use previous research to refine those questions.

In Chapter 3, we ask: *What groups do you want to study?* To select a sample of individuals, you must specify precisely whom you want to investigate. Are you interested in students in general, or just freshmen and sophomores? What about faculty members? How you specify your target population determines the generalizability of your study—the extent to which your results are applicable to other persons, places, and times. In this chapter, we discuss how to decide whom to study, and how to select a sample of people who meet your criteria.

In Chapter 4, we encourage you to identify the factors or programs you want to study by asking: *What predictors do you want to study?* Good research designs incorporate and manipulate the predictors of primary interest, the ones you will build into your work, thereby facilitating the detection of their effects. Organizing a first-class design requires you to identify the important predictors *before* collecting data. In this chapter, we present strategies for identifying predictors and for incorporating them into a design.

We devote Chapter 5 to a special kind of predictor—one distinguishing a program or treatment group from another group receiving no special treatment—when we ask: *Compared to what?* People in the comparison group are a baseline or standard against whom you compare people in the treatment group. The particular comparison group you choose determines how well your inferences will hold up. In this chapter, we present the pros and cons of eight alternative comparison groups, and show how you can choose the best one for the research question you want to answer.

In Chapter 6 we ask: *What are your outcomes?* College changes people in many different ways. Students learn new facts, new ways of thinking, and new ways of viewing the world. Which of these outcomes do you want to focus on? In addition, do you want to emphasize students' status at a particular point in time, or changes in their status over time? How do you know you are measuring what you *think* you are measuring? In this chapter, we describe how you

WHAT ARE YOUR QUESTIONS?

2

Anecdotal reports of sexual harassment fill the student newspaper. Formal charges have been brought against two professors, and an accompanying editorial implies these two are just the tip of the iceberg. The president asks the assistant dean for student affairs to “study the problem” of sexual harassment on campus. How should the dean begin?

Three years ago, faced with a declining number of applications, the dean of admissions recommended relaxing admission standards. Without this change, she argued, the size of the entering class would decline. The faculty reluctantly agreed. The registrar now reports that more students dropped out last year than ever before. Is this the result of lowered admissions criteria, or of some other cause? How can the dean find out?

Each of these scenarios identifies a broadly stated re-

search theme—a dilemma to investigate. The research theme in the first example is the extent of sexual harassment on campus; the research theme in the second example is the link between admissions criteria and student persistence. Both themes are good candidates for detailed investigation.

How can you take a broad theme and actually plan a study in detail? What is the first step you should take? Identify available data? Ask experts for advice? Ask colleagues on other campuses what they know? Send a research assistant to the library to review past literature?

Information-gathering is essential, but it should not be your first step. Your first step should be to *articulate a set of specific research questions*. Good design flows from clear goals. Do you want to know how many undergraduates have been sexually harassed? Conduct a confidential student survey. Do you want to know whether the problem is worse at the graduate level? Include graduate students as well. Do you want to know whether a workshop for faculty and students would decrease the incidence of harassment and increase the chances that people who were harassed would come forward? Offer a workshop and evaluate what happens. *Well-crafted questions guide the systematic planning of research. Formulating your questions precisely enables you to design a study with a good chance of answering them.*

It is challenging to move from broad research themes to specific research questions. Many prospective researchers say: “I’m just interested in the general topic; I don’t have specific questions. I need to collect some data; the questions will arise from those data. If I knew the questions, I wouldn’t need to do the research.” Although you should always be open to new ideas generated by data, these views are a woefully inadequate basis for *planning* research. If your research is not grounded in specific questions, you court the serious risk of not finding anything. Your design won’t be targeted to a precise purpose. If questions are not

posed, you have no basis for manipulating features of your project's design to help you find answers.

Moreover, when pressed, most researchers actually do have specific questions in mind. Something—an *observation* about the world, a *theory* or *hypothesis* about how the world works, or the need to know about the effectiveness of a *new policy*—led them to pursue a project. These observations, theories, hypotheses, or policies lead to specific questions. The goal of this chapter is to help you move from the generalities of research themes to the specifics of research questions. In subsequent chapters, your research questions will provide the foundation for making decisions about your design. By the end of this chapter, you should have some ideas about how to:

- *Articulate clearly specified research questions* linked to hypotheses, theories, observations about the world, or problems in practice. Research questions form the basis for making subsequent design decisions.
- *Understand the link between research questions and methodology.* Although some research questions can be addressed using a variety of research designs, others require the use of particular types of designs.
- *Learn from the work of others and refine your research questions accordingly.* Review other people's research on the questions you want to address. Learn from their successes and from their failures. Knowing what has gone before helps you to avoid pitfalls and to identify new directions.

Why Are Research Questions So Important?

To design a project you must make some decisions. Time and time again, you will have a choice and you will have to determine the best course of action. Whom do you want to study—freshmen, all undergraduates, or doctoral stu-

dents? How should you collect data—using registrar's records, tests, independent evaluations, or interviews? What time frame is appropriate and feasible—the past, the present, or the future? The quality of your decisions shapes the quality of your study. Make good decisions, and your study will be first-rate; make poor decisions, and your study will be second-rate at best. The dilemma you face is: On what bases should I make these decisions?

We believe that clearly specified *research questions* are the only basis for making sensible planning decisions. Think about what you want to know. Ask yourself: "If I make Choice A, will I be able to answer my research questions? What if I make Choice B?" Every design decision has consequences—some trivial, some monumental. Considering your research questions, and understanding the ramifications of your decisions, can help you make intelligent choices.

To illustrate how research questions help inform design decisions, suppose your broad research theme is concerned with the effectiveness of your college's faculty advising system for undergraduates. You want to know more about the present system and to consider the possibility of a new, intensive mentoring system. You have yet to specify precise research questions. How will these questions shape subsequent design decisions?

Research questions identify the target population from which you will draw a sample. Should you study both students and faculty? Freshmen and sophomores only, or juniors and seniors, too? You should decide whom to study only after considering exactly whom you want to make policy decisions about. If you do not identify the people you are most interested in before collecting data, you risk omitting important respondents from your study.

Research questions determine the appropriate level of aggregation. Should you measure efficacy at the level of student, advisor, department, or institution? Are you inter-

ested in the characteristics of advisors that make *them* particularly effective, or in the characteristics of students that make *them* particularly easy to advise? Research questions can be framed at different levels of aggregation. If you do not think about the issue of aggregation before you collect data, you risk not having enough data at a crucial level of aggregation to answer your research questions.

Research questions identify the outcome variables. What do you mean by the effectiveness of advising? Are you interested in student perceptions or in objective measures, such as the number of student-advisor contact hours? In short-term or long-term success? You can determine appropriate outcomes only after considering exactly what you want to know. If you do not define the outcome variables before data collection, you may fail to collect data on the most important outcomes.

Research questions identify the key predictors. Does efficacy differ by student gender? By advisor gender? By the match between advisor and student gender? By the advisor's academic rank? You can determine the important predictors only after thinking about all the things that might be associated with your outcomes. If you do not think about predictors before data collection, you may fail to measure essential variables.

Research questions determine how much researcher control is needed and whether a descriptive, relational, or experimental study is most appropriate. When the influence of mentoring is compared to traditional advising, will you study the new system the first year it is implemented or after it has been in place for three years? Can you randomly assign students to advising systems, or must you study them after they have selected the system they prefer? If you have a great deal of control over the research setting, you can draw strong inferences. By not deciding how much control you need, you risk not having enough.

Research questions identify background characteristics

that might be related to the outcome. Should you account for differences in faculty burden due to different numbers of advisees per advisor? Should you incorporate different students' goal orientation? Random assignment of students to advisors will eliminate most potential biases, but if you cannot use random assignment, differences in background characteristics such as faculty advising loads may distort your findings. Disentangling the effects of different predictors is often very difficult during analysis. It is much easier to control background influences by design.

Research questions raise challenges for measurement and data collection. Are there published instruments that assess advisor effectiveness? Do they require individual administration? Can they be mailed, or must they be filled out in person? Are the measures appropriate for the students and faculty members at your school? Is one measure of effectiveness sufficient, or should you use several? Data collection is expensive, so spend your resources wisely. If you do not think about measurement and data collection at the outset, you may never gather the key information you require.

Research questions influence the number of people you must study. Not only is research design guided by your questions, so is statistical analysis. Different questions require different analyses which, in turn, require specific sample sizes to ensure adequate statistical power to detect effects. Once the data have been collected, it is too late to add respondents—you have to make this decision in advance.

Our message is simple: *Research questions determine every facet of research design.* If your questions are not precisely stated, you have little basis for making crucial decisions. Today's naive choices may have dire consequences tomorrow. If you plan your study with your research questions in mind, you can ensure that your project will be able to answer those questions at the end.

EXAMPLE: Linking research questions to design decisions: Geographic mobility for academic men and women.

The conventional wisdom of academic life suggests that (1) career advancement often requires geographic mobility, and (2) women advance more slowly and occupy less prestigious positions than men. Rachel Rosenfeld and Jo Ann Jones (1987) asked whether these phenomena are related. Can sex differences in mobility explain sex differences in career progress?

Rosenfeld and Jones developed three specific research questions: (1) Does geographic mobility differ by sex? (2) Does the relationship between mobility and career attainment differ by sex? (3) Have these patterns changed over time? Based on a random sample of 311 women and 311 men, they found that (1) early in their careers, women are less mobile than men, but later on, this differential diminishes; (2) although geographic mobility is related to career advancement, this relationship does not differ by sex; and (3) these patterns have not changed over time.

These findings are persuasive because Rosenfeld and Jones linked their research design to their questions in six specific ways. First, because of interest in "academics," their target population was men and women who received a doctorate and who worked at a college or university immediately after receiving their degree. Second, because they needed published data on career histories (to save money) and a sufficient number of women (to examine their questions), they studied psychologists. (Thirty percent of the members of the American Psychological Association are female and the membership directory includes extensive employment data.) Third, because of their focus on sex differences, they oversampled women and undersampled men, selecting equal numbers for their project. Fourth, because of their interest in changes over time, they did not restrict attention to psychologists who graduated in any one year, but included people who received a doctorate between 1965 and 1974. Fifth, because their questions emphasized career *progress*, not career *status at a given time*, they followed psychologists from doctoral degree until 1981, creating a study period as long as 16 years and as short as 7 years. Sixth, because of their focus on career success as measured by standard academic barometers, they used the Social Science and Science Citation Indices to determine the number of articles published each year, the National Union Catalog to determine the number of books published each year, and academic rank as a measure of career progress.

Getting Specific

Research questions rarely come in a single burst of inspiration. Do not expect to sit down for an hour and produce an elaborate list of specific questions. Although you must take the time to do just that—sit down and write—your initial list will not be your final list. Expect to iterate. A good set of research questions will evolve, over time, after you have considered and reconsidered your broad research theme.

Begin by asking yourself the simple question: "What do I want to find out?" Be as general as you like—the broader, the better. Do not worry about data collection; that comes later. Do not worry about measuring fuzzy concepts; that comes later, too. Imagine you have access to any resource you need. Is that really all you want to know? Have you left something out? Will you be content if that is all you can say? Write down all your ideas, however grandiose, however small.

Now comes the hard part: *Get specific*. Examine every outcome on your list. Define it. Refine that definition. Think about how you would measure it. Do the same for every predictor. Identify the target population of greatest interest. Should it include anyone else? Be as precise as you can—the clearer, the better.

Make sure your questions are not inward, incomprehensible to colleagues. After all, your findings must be persuasive to external readers, too. Ask colleagues to review your list. Do they understand every question? Do they see something you overlooked?

Be wary of the desire to push forward before going through this process. It is all too easy to write instruments, select respondents, and collect data before articulating specific questions. Many researchers believe they are productive only when they write interview schedules, gather data, or do statistical analysis. Don't fall into this trap. You may end up with long, rambling questionnaires and aimless

interviews that alienate respondents. You may omit respondents who will be crucial later. Ultimately, you risk becoming stranded with key questions unanswered and key issues unresolved.

EXAMPLE: *Stating research questions: The effects of a career development course for undecided freshmen.*

One way to foster career development is to offer, for academic credit, a one-semester course in career choices. David Carver and David Smart (1985) examined the effects of one such course: "Career and Self-Exploration (CSE)" at the University of Northern Colorado.

The researchers' broad theme was an evaluation of "the effectiveness of the fall 1981 sections of CSE in promoting four major areas of student development: (a) academic and career decisionmaking; (b) career maturity; (c) positive self-concept; and (d) interaction between the student and the campus environment" (p. 38). But Carver and Smart did not stop there; they pushed themselves to become even more precise. Building upon a review of the literature, their specific research questions were:

whether freshmen students completing CSE would score significantly higher than a comparison group in the following seven areas: (a) career decidedness (as measured by the CDS, Item 1), (b) academic major certainty (as measured by the CDS, Item 2), (c) reduction in academic and career indecision (CDS, Items 3-18), (d) maturity of career attitudes (CMI, Attitude Scale), (e) overall level of self-esteem (TSCS, total P score), (f) use of academic advising, personal counseling, career planning, placement, and tutorial services (as measured by the Student Involvement Survey), and (g) involvement in student organizations, university programming and student government (Student Involvement Survey). (p. 39)

CDS (Career Decision Scale), CMI (Career Maturity Inventory), and TSCS (Tennessee Self-Concept Scale) are standardized, widely available instruments.

Carver and Smart used their research questions to make many design decisions. Their questions guided their measurement choices; the instruments, and even specific items, are cited in the questions. The questions also guided their choice of a comparison group. Carver and Smart wanted

a comparison group of students who did not take CSE, but were otherwise similar to the CSE group. They were especially concerned about differential motivation to explore career options. Random assignment was not possible because of the university's policy of admitting students to courses on a first-come, first-served basis. Their very good compromise for a control group was students who expressed interest in taking CSE but who could not enroll, because of schedule conflicts or because all sections were filled.

Building on the Work of Others

Having specified your research questions in as much detail as possible, you should spend time examining previous research on these and related questions. Reviewing and synthesizing the work of others can help both to clarify and to broaden your questions. A good literature review can identify unexamined target populations, important predictors and outcomes, and tried and tested measurement techniques. A thorough review has another advantage as well: by learning from the work of others, you can avoid repeating their mistakes.

The Goals of a Research Review

The primary purpose of a research review is to learn what is already known so that you can build on it. The most helpful reviews identify corroborating *and* conflicting evidence. Corroborating evidence suggests "facts" that you must take into account when designing your new project. Conflicting evidence generates new questions for future research.

Occasionally, every study you examine will suggest the same clear and unambiguous answers to your research questions. If this happens, think about exploring new un-

charted territory, or modifying your questions. Why conduct yet another study documenting for the eleventh time what has already been replicated ten times?

But unanimous agreement is rare. All the studies in a group of studies probably used different research methods, and may have reached somewhat different conclusions. They may have sampled different target populations, used different measuring instruments, applied different analytic techniques, or simply found different answers. But whatever the differences are, you can learn from them. These differences, and any systematic patterns in them, can tell you a lot about how to design any new study. In fact, although conflicting findings prevent unambiguous answers from shining through, they actually tell you more about future design than you would learn from a world where there was unanimous agreement. You can use conflicting evidence to generate new, more fine-grained questions and to construct more powerful research designs.

EXAMPLE: What can you learn from previous research? The effects of financial aid on student persistence.

Does financial aid help students to stay in, and eventually graduate from, college? Tullise Murdock (1987) examined more than 500 studies of the effect of financial aid on persistence. She carefully synthesized a subset of 62 of these, and found that students receiving financial aid were slightly more likely to persist than similar students without aid. Specifically, "financial aid could be expected to move the typical person from the 50th to the 55.2 percentile of persistence for the nonfinancial aid population" (p. 84).

From a research design standpoint, Murdock's findings on the variation across studies of the impact of financial aid were even more informative. Her review showed that the impact of financial aid differed depending upon the following factors: (a) *The type of institution*: effects were larger at two-year schools than at four-year schools, and larger at private colleges than at public colleges. (b) *The definition of persistence*: effects were larger if

persistence was defined in terms of graduation, as opposed to reenrollment from semester to semester or year to year. (c) *Student enrollment status*: effects were larger for full-time students than for part-time students. Murdock's careful review turned up another critical finding: effects differed depending upon whether differences in academic ability between recipients and nonrecipients were controlled. In the seven studies that matched groups of students by ability, financial aid had no effect on persistence.

Does Murdock's review establish definitively that financial aid and persistence are related, and therefore that no additional research is needed? She found a "small" effect. Because money woes make it difficult for needy students to persist, and because financial aid converts this considerable disadvantage into a slight advantage, you might argue that even a small effect "in the right direction" establishes the efficacy of financial aid. But this conclusion may be premature. Effects differed across institutions, students, and measures, and the seven studies that controlled for academic ability found no positive effect on persistence. Additional research would certainly help to clarify these issues.

What direction might new research take, and how can it build upon the thorough review? Murdock addresses this very question:

For more accurate measures of financial aid effect on persistence, researchers should try to include part-time students, transfers, and stopouts in their study population. These three groups compose a large percentage of the total student population, and how a study treats their persistence mediates the effect size. To adopt this recommendation, studies will have to measure persistence over a longer period of time than just one semester, one year, or even two years. Only thus will the true effect of financial aid on the total student population be determined. (p. 96)

A good "next" study could build upon Murdock's work by including a broader cross-section of students and by collecting data for a longer period of time. Her careful review has refined the design of future research.

How to Conduct a Review

Carol Weiss has said that, until the 1970s, the best practical advice on how to review an extensive research literature was: read everything you can find, think carefully, and be smart. In a survey of 39 books, 87 review articles,

and 2050 article abstracts discussing methods for reviewing the literature, Gregg B. Jackson (1980) concluded that systematic procedures were rarely presented or recommended.

What Jackson describes is the classic, narrative literature review. Narrative reviewers collect as many studies of a particular research question as they can find and try to synthesize the disparate results into a coherent story. Sometimes reviewers present lists displaying each study's results, often restricting these lists to studies that meet specific criteria, such as those with true experimental designs, those supporting a specific point of view, or those carried out by particularly highly respected investigators. Narrative reviewers use common sense to distill essential findings from the literature. Conflicting findings are pitted against each other. Often a reviewer simply declares a few very well-done studies "the winners" and encourages her readers to believe only them.

But is this the best way to carry out a narrative review? Just because the review is qualitative, must it be nonsystematic? Obviously not. A good narrative literature review is like any other piece of research: if it is to be successful, it must be methodical and systematic. Without systematic procedures for identifying existing studies and comparing their findings, the biases of any particular reviewer will affect the overall conclusions. Another reviewer examining available reports on the same research question could reach an entirely different conclusion.

You should be as systematic in your narrative research as you are in your regular research. Your review should adhere to the same high standards and adopt the same systematic strategies. The only difference is that your unit of analysis is "study," not "respondent." Try to be as objective as possible; do whatever you can to avoid letting your own biases influence your findings. Your review should include at least *five* steps. You should: (1) establish one or two questions that will drive your review; (2) examine a

representative sample of studies from the full "population"; (3) record the study findings and characteristics; (4) "analyze" the recorded "data"; (5) interpret your "analyses."

Questions. Focus your review on two types of questions: questions about *substantive detail* (e.g., What student characteristics are related to persistence?) and questions about *research design and methodology* (e.g., What characteristics of research design are related to each study's findings about persistence?). You are reviewing the literature to learn about both *substance* and *method*.

Sampling. Be sure to include in your review a representative sample of all the studies that have been carried out (the population of available studies). Do not restrict your attention to published research. Because of journal acceptance and rejection practices, published research has an overabundance of "statistically significant" findings, and by examining only published results you can overestimate the frequency of truly positive findings. Take the time to track down unpublished studies, dissertations, technical reports, and internal memoranda. There is much research done in higher education at the institutional level that is never even submitted for publication. You should examine this too. *Fugitive documents* balance the picture of what has been done and what has been learned.

Data collection. In tables, charts, and lists systematically code and record the characteristics of every study and its findings. Pay special attention to methodology—who was sampled, how they were assigned to groups, what measures were used, what background influences were controlled, and so on. The relationships between different research designs and different findings provide valuable information for you when you design your study.

Data analysis. Examine the recorded "data" and try to generalize across studies. Synthesizing the literature is never easy. Findings will vary considerably. Use conflicts between findings to examine the effects of methodology. Examine the strengths and weaknesses of the studies, de-

termine whether similarly labeled programs or treatments differ in important ways, and assess the impact of differences among the studies in settings and respondents. Take special care before concluding that an answer is "known."

Interpretation. Use your research review in several ways. First, consider the *focus* of previous research. Identify areas that have been thoroughly investigated versus those that need additional work. Direct your research to uncharted areas or to areas where conflicts still exist. Try to predict how your study will fit into an accumulating knowledge base when the next person reviews the literature, next year. Second, consider the *strengths and weaknesses* of previous research. Learn from both mistakes and successes. Were findings more impressive when certain types of measures were used? Be sure to ask why. Were findings larger for certain types of respondents? Be sure you understand why. Don't reinvent the wheel; improve it. Third, use *discrepancies* among studies to generate new hypotheses for your work. Have studies at public colleges found larger effects than similar studies at private colleges? Why do you think this differential exists? Is it attributable to different organizational structures? To admissions criteria? To curricula? Do you think that private colleges could learn something from what public colleges are doing? The other way around? Perhaps you should include these new hypotheses in your project.

EXAMPLE: How to review the literature systematically: Developing a model of nontraditional undergraduate student attrition.

Why do nontraditional students—older, part-time, and commuter students—drop out of school? Is it because they cannot find the on-campus supports

they need, or because their off-campus commitments present major obstacles to persistence?

John Bean and Barbara Metzner's (1985) work on attrition of nontraditional students is an excellent illustration of how to review the existing literature for a specific research question. Noting that the literature on nontraditional students was sparse, Bean and Metzner carefully searched books, articles, dissertations, and ERIC clearinghouse documents. They identified 56 useful studies. They constructed tables summarizing the substantive findings and methodological details of each study. Their summary gives information on each study's location, target population, sample size, definition of "dropout," and statistical method.

Then, in a series of narrative sections, Bean and Metzner describe the relationship between the decision to drop out and five sets of "predictors"—intent to leave, and academic, psychological, environmental, and background factors. Each set of predictors includes several specific variables. For example, the environmental cluster includes information on finances, hours of employment, outside encouragement, opportunity to transfer, and family responsibilities. The authors discuss the effects of each predictor in turn, describing studies that focused on it, noting what was found, and identifying discrepancies among studies and relating them to the type of institution or analytic method.

Borrowing from theoretical models of attrition among traditional students, Bean and Metzner synthesize their findings into a comprehensive model of nontraditional student attrition. Within this model, they comment specifically on the limitations of previous research—they identify, for instance, that "more than half of the 40 studies from two-year colleges were exit or autopsy studies . . . [in which] control groups of persisters and tests of statistical significance were lacking" (p. 528).

Bean and Metzner's careful attention to the needs of future research makes their review particularly helpful to anyone planning a new project. For example, they argue that research at commuter schools should not emphasize variables measuring social integration, but rather students' external environment. They discuss the need to distinguish between part-time and full-time students and between older and younger students. They also suggest that the variables associated with persistence will differ depending upon student demographics. They stress how future researchers must be sensitive to these demographics. Their comprehensive review, and their suggestions for future research, are an asset for researchers planning to investigate attrition for nontraditional students, and an example to those intending a narrative literature review.

Meta-Analysis

In recent years, methodologists have developed more rigorous ways of *coding* and *analyzing* the evidence collected in a literature review. The new approach, which uses the power of *quantitative methods* to systematize the narrative review, is called *meta-analysis* (Glass, McGaw, and Smith, 1981; Cooper, 1984; Light and Pillemer, 1984; Rosenthal, 1984; Hedges and Olkin, 1985).

Rather than creating narrative tables of summary findings and then informally synthesizing these "data" by counting the number of studies "for" and "against" a particular finding, meta-analysis summarizes each study *numerically*. Study findings—either the relationship between predictor and outcome, or the differences among treatment groups—are recorded as "effect sizes." Study characteristics are recorded as categorical and continuous variables. Then, over all coded studies, meta-analysis uses descriptive and inferential statistical techniques to explore the dependency of effect size on the study characteristics. Larry Hedges and Ingram Olkin (1985) provide a compendium of useful strategies.

Two types of effect size are most common. If most of the studies you review have estimated correlations between predictor and outcome, then the *correlation coefficients* themselves can be used as effect sizes to summarize the study finding. Bigger correlations indicate stronger relationships. On the other hand, if most studies have compared means for groups of respondents, say students getting a new curriculum versus students getting an old one, then a *standardized mean difference*—the difference between the two group averages, expressed in standard deviation units—is the appropriate effect size. Bigger effect sizes usually indicate more successful innovations or curricula or treatments. Whether you compute the summary effect sizes as correlation coefficients or as standardized

mean differences is largely irrelevant, because each of them can be converted into the other by simple arithmetic manipulation (see Glass, 1976; Hedges, 1983; Hedges and Olkin, 1985).

Once you have coded the effect size and study characteristics for each of the studies, you can begin the "analysis" phase of your review by exploring your coded data from study to study. What type of analyses should you perform? The typical quantitative research synthesis will address four questions.

1. *Are effect sizes consistent across studies?* If not, how do they vary? You can examine the empirical distribution of effect sizes over studies. Do they follow a recognizable pattern? Are they centered around zero or around some non-zero value? Is the distribution of findings dispersed or tightly clustered? Is it symmetric or asymmetric? (Light and Smith, 1971; Rosenthal and Rubin, 1986.)

2. *What is the average effect size over studies, and is it different from zero?* If the empirical distribution of effect sizes does not have an unusual shape, or worrisome outliers, then the average effect size across all studies may be a useful summary statistic. It gives you an overall impression of the relationships or group differences that other researchers have found. If it is non-zero, the new curriculum you are looking at is better, or worse, than the old. If it is approximately zero, the collective evidence is that the innovation you are examining is no different from the "old way." Take special care if the effect sizes have an unusual distribution. In these instances, the average effect size is less useful and must be interpreted extra carefully. For example, perhaps an innovation works especially well for freshmen but not for seniors; the empirical distribution of effect size may be bimodal.

3. *Are the effect sizes related to study characteristics?* When study findings conflict, their effect sizes differ. Are there some very large or very small effect sizes? Is there

anything unusual about these particular studies? Are differences in effect sizes related to how studies were designed? Do effects differ at private colleges versus public colleges? At four-year colleges versus two-year colleges? For freshmen versus upperclassmen? Do well-controlled studies show a stronger or a weaker impact for a new curriculum or advising system? (Light, 1979; 1984).

4. *Does publication bias exist?* Publication bias can seriously affect the findings of your review. If you synthesize results from published studies only, you will be excluding, unintentionally but systematically, all those studies that never found their way into print.

If findings in the "omitted" papers are similar to the findings in those that were published, you will be fine. If they are not, then looking only at published work means you are reviewing a biased subset of findings. Typically, this will cause you to overestimate effect size. This is because *statistically significant* findings are more likely to be submitted to, and accepted by, a refereed journal than *non-significant* findings (Greenwald, 1975; Rosenthal, 1978).

You should take two steps to deal with possible publication bias. First, because publication bias arises when only those studies that appear in journals are reviewed, we urge you to include unpublished work in your review. Second, you should estimate the size of the publication bias. You can separate published from unpublished sources in your review, and explore whether one type of source reports noticeably different findings from the other. If it does, then you should think hard about informally "adjusting" the recommendations of your reviews.

The strength of meta-analysis is its systematic nature. Another reviewer using meta-analysis to synthesize the same literature should arrive at the same results as you (Light, 1983).

EXAMPLE: Using meta-analysis to refine your questions: The effectiveness of special programs for high-risk and disadvantaged college students.

Since 1894, when Wellesley College implemented what may have been the first remedial course for college students, many colleges have instituted programs designed to help students who might have problems. Given all the resources expended, just how successful are such programs?

Chen-lin Kulik, James Kulik, and Barbara Shwalb (1983) conducted a meta-analysis of research on college programs for high-risk and disadvantaged students. They focused on programs defining risk in terms of low test scores, low high school achievement, low college achievement, or membership in a socioeconomically disadvantaged group. Using computer searches of three major bibliographic databases, they identified 504 available documents. Sixty studies that systematically compared similar groups of students who participated and did not participate in an intervention program were selected for meta-analysis.

Effect sizes were computed to summarize group differences (membership in remedial program versus no membership) on two outcomes: achievement (college GPA) and persistence (reenrollment during the study period). The authors also coded 12 study characteristics: 3 describing the program (e.g., intervention mode), 2 describing the setting (e.g., type of college), 5 describing the design (e.g., random assignment), and 2 describing publication history (e.g., published or unpublished document). We describe results for the 57 studies reporting data on GPA.

Effect sizes varied from a low of $-.41$ to a high of 1.00 , and their empirical distribution was fairly symmetric with no unusual values. The mean effect size was $.27$, indicating that in the average study, students getting special help had GPAs $.27$ standard deviations higher than students not getting such help. Kulik and her colleagues translate these effect sizes back into GPAs as follows: "In the typical report, the GPA for students from the special programs was 2.03 in the latest semester studied; the GPA for control students was 1.82 . Although positive and statistically reliable, the overall effect of special programs on GPA was therefore clearly small in size" (p. 401).

Examining the large amount of variation among effect sizes, the authors explored relationships between study findings and the 12 study characteristics. They discovered four systematic patterns. Effects were *smaller* among established programs, programs that looked at long-term GPAs, and programs that used remediation. A potential publication bias also was discovered: unpublished studies had a lower mean effect size ($.07$) than published studies ($.31$) and dissertations ($.30$).

What can a researcher interested in evaluating the effects of special-help programs learn from this meta-analysis? First, remedial programs can help, although their impact may be smaller than an administrator might hope. Second, different kinds of programs have different effect sizes. Programs using guidance sessions had the largest mean effect size (.41); programs using academic training and comprehensive support services followed (mean effect sizes of .29 and .26); and programs emphasizing remedial studies were the least successful (mean effect size of .05). Third, be sure to look at both long-term and short-term success. Short-term evaluations give an overly optimistic picture of efficacy; follow students for several semesters if you want to learn whether the program is really effective.

Correlation versus Causation

Many research questions in higher education ask about the association between predictors and outcomes. Joseph Seneca and Michael Taussig (1987) studied the relationship between an offer of financial aid (the predictor) and the decision to enroll at Rutgers University (the outcome). Sophia Mahler and Dan Benor (1984) examined the relationship between attending a teacher-training workshop (the predictor) and instructional behavior of faculty members (the outcome). Larry Weber, Janice McBee, and Jean Krebs (1983) examined the relationship between the type of examination given—in-class versus take-home—the predictor) and the amount of “rampant cheating” (the outcome).

What does it mean to examine the relationship between a predictor and an outcome? When you speak of relationships, do you mean *correlation* or *causation*? Do you want to know whether the predictor and outcome are simply associated or whether a change in the predictor will actually change the outcome?

If you could observe a correlation only when variables were causally linked, this distinction would be unneces-

sary. Questions about causality would be answered by examining correlations. But for years, nonstatisticians such as George Bernard Shaw (1911) have pointed out that many things in life that are correlated are anything but causally linked:

It is easy to prove that the wearing of tall hats and the carrying of umbrellas enlarges the chest, prolongs life, and confers comparative immunity from disease . . . A university degree, a daily bath, the owning of thirty pairs of trousers, a knowledge of Wagner's music, a pew in church, anything, in short, that implies more means and better nurture . . . can be statistically palmed off as a magic-spell conferring all sorts of privileges . . . The mathematician whose correlations would fill a Newton with admiration, may, in collecting and accepting data and drawing conclusions from them, fall into quite crude errors by just such popular oversights.

The moral: correlation does not imply causality.

Yet, if your work is to influence policymaking, then it must examine causal linkages. Does a generous offer of financial aid *cause* students to enroll in our school? If it does, we might change our school's financial aid policy with the hope of improving our yield rate. Does teacher training *cause* professors to teach better? If it does, we might require all professors to participate in workshops with the hope of improving their performance.

At other times, we want to know only whether a predictor and an outcome are correlated, but we want to be sure the correlation is attributable to a direct relationship, not to other predictors we failed to study. We would want to be sure, for example, that the correlation between attendance at a teacher training workshop and performance in the classroom was not an artifact of differential attendance at the workshops caused by a tendency of professors who were better instructors to begin with to be more likely to participate in the training.

Establishing a Causal Link

Frederick Mosteller and John Tukey (1977) identify four conditions for demonstrating a causal link. First, you must show that a change in the predictor produces a change in the outcome: the outcome must be *responsive* to changes in the predictor. Second, you must show that there is *no plausible alternative explanation*: no rival predictor must be able to explain the correlation you have observed. One way to ensure this is to assign study respondents randomly to the various levels of the predictor. Random assignment makes this group membership uncorrelated with all other predictors, thereby ensuring that any effect you observe is attributable to group membership. Without random assignment, you must systematically examine, and rule out, all plausible rival predictors. Third, you must have some idea what *mechanism* explains how a change in a predictor produces a change in an outcome. If you have a theory describing how the predictor affects the outcome, then when you find a relationship you can tell an appropriate story. Without a theory, you should not be looking for a causal link. Fourth, you must be able to replicate the correlation in different populations with different characteristics. If a link is found time and time again, the consistent pattern is far more compelling than each isolated result.

Can research on higher education meet these four criteria? The answer is yes. The challenge is to plan research that can uncover strong linkages supported by well-crafted theory. We address this challenge throughout this book.

When investigating causation, the key criterion is responsiveness. Responsiveness is the \$64,000 question of applied research, especially policy research. It is not enough to show that a policy choice is *correlated* with an outcome. You should demonstrate that a *change* in policy—be it a change in teaching style, dormitory assignment, or class size—*changes* the outcome. After all, if a policy

change will not change the outcome, why bother changing the policy?

Active intervention is the only way to demonstrate responsiveness. Relational studies cannot do so. Comparing the enrollment rates of students receiving financial aid with those of students who are not cannot tell us, with confidence, whether offering financial aid to more students would improve yield rates. Determining the effects of a policy change requires changing the policy and seeing what happens.

Because of the intimate link between causation and responsiveness, you should always *answer questions about causality, if at all possible, with randomized experiments*. When you conduct a randomized experiment, you are an “agent of change,” intervening in the system and observing what happens. Randomization has another advantage as well; its “balancing” properties help rule out rival explanations of the correlation between predictors and outcomes. These properties combine to make randomized experiments the best, and to many methodologists the only, way to go.

EXAMPLE: Establishing a causal link: Can you train teaching assistants to be better teachers?

Widespread complaints about the poor teaching of teaching assistants (TAs) has led many colleges to offer programs designed to improve TAs' classroom skills. Kathleen Dalgaard (1982) evaluated the effectiveness of one such program by conducting an experiment with 22 graduate-student TAs in the business administration, economics, and geology departments of the University of Illinois, Urbana-Champaign. She divided the TAs by department and randomly assigned half from each department to a training group, which participated in six two-hour seminars and an individual videotape critique session with the seminar instructor, and half to a no-training group. All TAs were videotaped once before the training began and once five

weeks later. Independent raters (unaware of group membership) gave TAs in the training group better ratings of overall teaching quality than TAs in the control group.

Dalgaard's study provides convincing evidence that TAs can be trained to teach better—that training causes improvement. By offering a program and randomly assigning TAs to groups, she acted as an “agent of change,” intervening in the system and observing what happened. This allowed her to demonstrate responsiveness. Random assignment also ensured that the two groups of TAs were similar with respect to background characteristics such as initial teaching skill and motivation to participate in a training program. (Comparing initial teaching skill ratings for the two groups confirmed this comparability.) This helped Dalgaard eliminate plausible rival explanations of group differences. The training program was based upon well-known theories of teaching and teacher training. This helped her posit a mechanism whereby the training caused behavior to change. The only criterion she was unable to comment on was consistency, for her study was limited to 22 TAs. Replication was left to future researchers.

If randomized experiments enable you to make strong inferences about relationships between predictors and outcomes, why conduct relational studies at all? One reason is that some predictors simply cannot be manipulated. You cannot randomly assign participants to levels of sex, age, race, or class year, and so to learn about the relationships between these predictors and an outcome, you must examine them as they occur in nature. A second reason is that logistical, practical, or ethical constraints often preclude randomization. For example, you may not have enough dollars to randomly assign students to different financial aid packages. Without random assignment, you can only examine the statistical association between measures.

EXAMPLE: Conducting a relational study: The effects of freshman orientation on student persistence.

Many colleges and universities invite incoming freshmen to an orientation session held before the beginning of classes. Ernest Pascarella, Patrick Terenzini, and Lee Wolfe (1986) conducted a relational study at a private residential university to examine the impact of freshman orientation on decisions to drop out before sophomore year. The researchers studied 763 students who were part of a random sample of incoming freshmen, and who also filled out two questionnaires—one in early fall and one in late spring of freshman year. Information on persistence came from university records. Seventy percent of the incoming students attended orientation; 88 percent of them persisted until sophomore year. Orientation and persistence were strongly correlated: students attending orientation were more likely to continue than those not attending.

Had the researchers estimated only simple correlations, their results would not have been convincing, for common sense suggests that background characteristics of students, associated with the decisions to attend orientation and to persist in school, may have spuriously created this correlation. The authors also hypothesized that the effect of orientation on persistence might not be direct, but might operate by influencing social and academic integration, which in turn affect persistence.

To investigate these alternative hypotheses, Pascarella and his colleagues estimated a series of statistical models incorporating rival predictors, such as the students' gender, major, socioeconomic status, and academic preparation, and moderating predictors, such as freshman-year social integration, academic integration, and institutional commitment. They found that presence at orientation did not have a direct effect on persistence; the correlation between the two variables diminished after the other predictors were included. But orientation did have an *indirect* effect on persistence, through its strong relationship with social integration and institutional commitment, which, in turn, predicted persistence. They conclude that attendance at orientation has a positive, albeit indirect, effect.

Because attendance at orientation was voluntary, this relational study cannot demonstrate responsiveness, and as a single study, it cannot demonstrate consistency. But the researchers did an excellent job of considering the remaining two criteria. Their research was well-grounded in theories explaining the persistence decisions of traditional students, including the work of Vincent Tinto (1975), William Spady (1970), and J.P. Bean (1985). They clearly articulated and explored the mechanism whereby orientation affects persistence. Using sophisticated statistical analyses, they eliminated

many *plausible alternative predictors*, thereby suggesting that the effects they found were not artifacts.

But the nagging question remains: Does orientation *cause* persistence? Pascarella and his colleagues are direct and to the point:

Finally, and perhaps most importantly, the study has the obvious . . . problems inherent in correlational data. Students in the sample self-selected themselves into the orientation and nonorientation group. This necessitated statistical controls . . . While such analytical models are useful in portraying what the patterns of causal influence might look like, they do not provide the same order of control as that achieved by a randomized experiment . . . [T]he estimation of the effects of orientation experience . . . under more controlled experimental conditions is a fruitful area for future inquiry. (p. 172)

Unfortunately, such speculation always haunts relational studies.

You must decide, early in the research planning phase, whether you are interested in correlation or causation. This decision, more than almost any other, determines much about your final design. If you want to establish causality, conduct an experiment. If you are content to establish an association, conduct a relational study.

The Wheel of Science

Throughout this chapter, we argue that effective empirical research must be guided by specific questions. You will not find what you have not sought. You will not be able to provide answers if you have not directly looked for them.

But isn't this premise a bit too simplistic? Isn't it possible to have tacit knowledge without having asked a question first? We often learn by direct experience unprovoked by interrogation. We observe first and then ask why. Can't we learn from data even if we haven't posed a specific question? Can't research proceed without questions?

Of course it can. At issue here, though, is how successful your project will be if you plan it without attention to specific questions. Research does not require the articulation of specific questions; the systematic *planning* of research does.

Philosophers of science distinguish between two modes of inquiry, one based on inductive logic and the other based on deductive logic. When you use inductive logic, you begin with observations and then explain what you have observed by generalizing. You move from particular instances to general principles, from facts to theories. In contrast, when you use deductive logic, you begin with a conjecture—a theory, hypothesis, or law—and then you collect data to test the accuracy of the conjecture. You move from general principles to particular instances, from theories to facts.

Is one mode of inquiry “better” than the other? Does “true science” require deduction? Is induction second-rate? For centuries, philosophers have pondered these questions, some arguing for induction, some arguing for deduction. Most agree that deductive research, so common in the natural and physical sciences, *is* scientific. But fewer agree about inductive research, so widespread in the social sciences. Some argue that induction cannot be scientific, others argue that as long as induction is accompanied by rigorous methods, it, too, *is* scientific.

We take a middle position. We believe that practical scientific inquiry blends deduction and induction, cycling endlessly between the two extremes. In the inductive phase, you reason from data; in the deductive phase, you reason toward data. Both modes of inquiry are essential. Induction helps us generalize and build new theories, which in turn generate new hypotheses for future deductive research. On and on the circle turns, new knowledge building on old in an endless spiral—theory to data, data to theory. Practical scientific inquiry becomes a wheel—the “wheel of science” (Wallace, 1971).

We do not argue that our position is the best possible one, but rather that it is practical and effective. *Inductive*

inquiry must precede and support your design plans. Exploratory research focuses ideas and helps build theory. But by framing specific questions and testing particular hypotheses derived from theory you gain irrefutable knowledge about how the world actually works.

Even statistics, a field wedded to deduction and hypothesis testing, has spawned disciples of induction. John Tukey, the father of the movement known as Exploratory Data Analysis, has said: "We need both exploratory and confirmatory [research] . . . [new] ideas come from previous exploration" (1980, p. 124). Use data not only to *test* theory, but to *develop* theory. Good statistical analysis combines exploratory model-building and confirmatory testing of hypotheses.

A close reading of almost any empirical research report reveals an implicit acceptance of this position. Consider Carver and Smart's study of the effects of career classes on career development. This was a solid piece of deductive research, yet they based their study on theories of career development, vocational choice, and student-institutional fit that were exploratory results from previous studies. And Carver and Smart discuss new exploratory findings, such as the effect of a student's advisor on her career choice, that future researchers can reframe as testable hypotheses.

Our philosophy has two practical consequences. First, *design your study only after doing a healthy dose of exploratory work.* Use your eyes and ears. Use informal contacts with administrators, faculty members, and students. Use colleagues on other campuses. Use the research of others. *Be inductive.* Second, *design your study with clearly stated research questions in mind.* Questions are crucial to your project's ultimate success. They fuel the engine and turn the wheel of science. They may be broad in a descriptive study, or narrow in an experimental study, but they are still questions. *Be deductive.*

WHAT GROUPS DO YOU WANT TO STUDY?

3

Before you can begin, you must ask: Whom should I study? The answer comes directly from your research questions. Suppose, for example, you want to examine the effectiveness of a new approach to teaching expository writing. To supplement individual writing outside of class, each student will write collaboratively with a classmate for one hour during each two-hour class. You hypothesize that collaboration will help students develop skills necessary for revising and improving not only their joint writing, but their individual writing as well.

At first glance, the goal here is to examine *students*. The purpose of the innovation is to improve students' writing skills, and the most direct way to assess improvement is to study the changes in their skills over time.

But identifying a target group with a label as broad as