CHAPTER 1

Applied Research Design

A Practical Approach

Leonard Bickman

Debra J. Rog

Planning Applied Social Research

The chapters in this *Handbook* describe several approaches to conducting applied social research, including experimental studies (Boruch, Weisburd, Turner, Karpyn, & Littell, Chapter 5), qualitative research (Maxwell, Chapter 7; Fetterman, Chapter 17), and mixed methods studies (Tashakkori & Teddlie, Chapter 9). Regardless of the approach, all forms of applied research have two major phases—planning and execution—and four stages embedded within them (see Figure 1.1). In the planning phase, the researcher defines the scope of the research and develops a comprehensive research plan. During the second phase the researcher implements and monitors the plan (design, data collection and analysis, and management procedures), followed by reporting and follow-up activities.

In this chapter, we focus on the first phase of applied research, the planning phase. Figure 1.2 summarizes the research planning approach advocated here, highlighting the iterative nature of the design process. Although our chapter applies to many different types of applied social research (e.g., epidemiological, survey research, and ethnographies), our examples are largely program evaluation examples, the area in which we have the most research experience. Focusing on program evaluation also permits us to cover many different planning issues, especially the interactions with the sponsor of the research and other stakeholders.

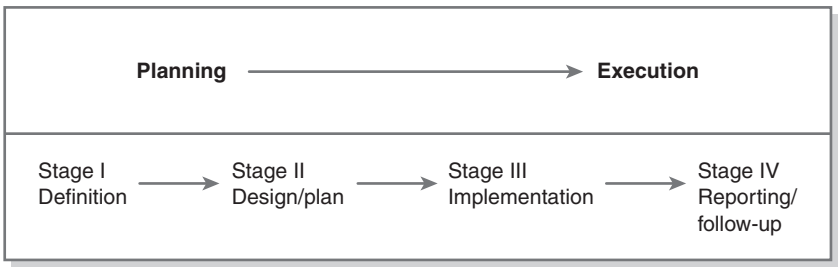

Figure 1.1 The Conduct of Applied Research

Other types of applied research need to consider the interests and needs of the research sponsor, but no other area has the variety of participants (e.g., program staff, beneficiaries, and community stakeholders) involved in the planning stage like program evaluation.

Stage I of the research process starts with the researcher's development of an understanding of the relevant problem or societal issue. This process involves working with stakeholders to refine and revise study questions to make sure that the questions can be addressed given the research conditions (e.g., time frame, resources, and context) and can provide useful information. After developing potentially researchable questions, the investigator then moves to Stage II—developing the research design and plan. This phase involves several decisions and assessments, including selecting a design and proposed data collection strategies.

As noted, the researcher needs to determine the resources necessary to conduct the study, both in the consideration of which questions are researchable as well as in making design and data collection decisions. This is an area where social science academic education and experience is most often deficient and is one reason why academically oriented researchers may at times fail to deliver research products on time and on budget.

Assessing the feasibility of conducting the study within the requisite time frame and with available resources involves analyzing a series of trade-offs in the type of design that can be employed, the data collection methods that can be implemented, the size and nature of the sample that can be considered, and other planning decisions. The researcher should discuss the full plan and analysis of any necessary trade-offs with the research client or sponsor, and agreement should be reached on its appropriateness.

As Figure 1.2 illustrates, the planning activities in Stage II often occur simultaneously, until a final research plan is developed. At any point in the Stage II process, the researcher may find it necessary to revisit and revise earlier decisions, perhaps even finding it necessary to return to Stage I and renegotiate the study questions or timeline with the research client or funder. In fact, the researcher may find that the design that has been developed does not, or cannot, answer the original questions. The researcher needs to review and correct this discrepancy before moving on to Stage III, either revising the questions to bring them in line with what can be done

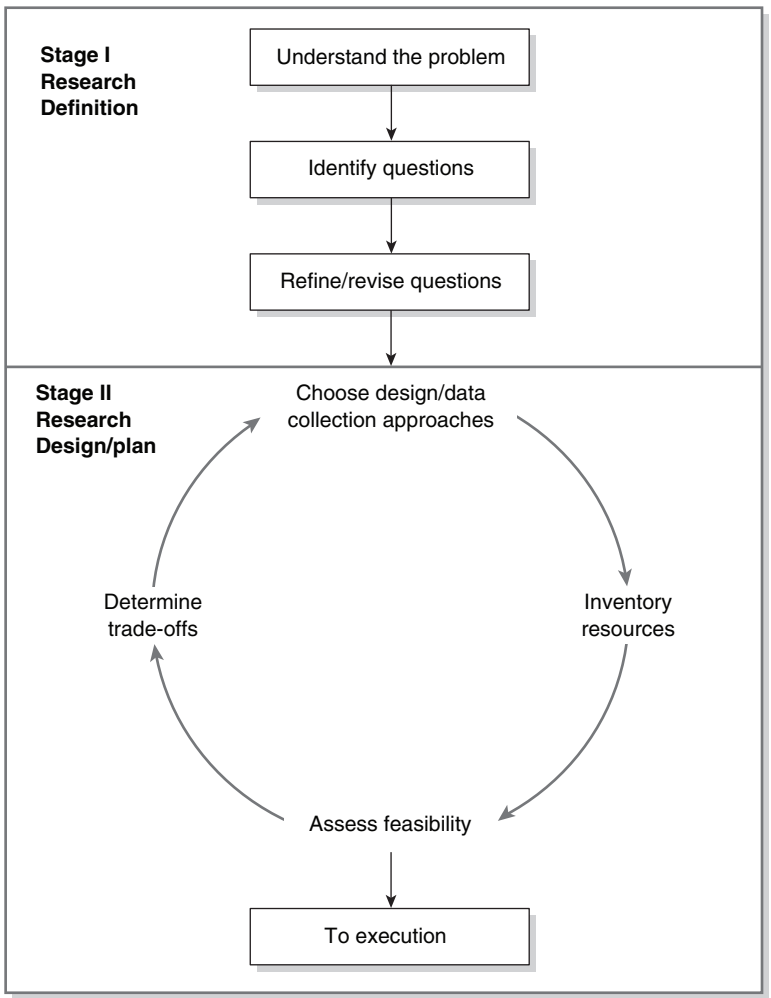

Figure 1.2 Applied Research Planning

with the design that has been developed or reconsidering the design trade-offs that were made and whether they can be revised to be in line with the questions of interest. At times, this may mean increasing the resources available, changing the sample being considered, and other decisions that can increase the plausibility of the design to address the questions of interest.

Depending on the type of applied research effort, these decisions can either be made in tandem with a client or by the research investigator alone. Clearly, involving stakeholders in the process can lengthen the planning process and at some point, may not yield the optimal design from a research perspective. There typically needs to be a balance in determining who needs to be consulted, for what decisions, and when in the process. As described later in the chapter, the researcher needs to have a clear plan and rationale for involving stakeholders in

various decisions. Strategies such as concept mapping (Kane & Trochim, Chapter 14) provide a structured mechanism for obtaining input that can help in designing a study. For some research efforts, such as program evaluation, collaboration, and consultation with key stakeholders can help improve the feasibility of a study and may be important to improving the usefulness of the information (Rog, 1985). For other research situations, however, there may be need for minimal involvement of others to conduct an appropriate study. For example, if access or "buy in" is highly dependent on some of the stakeholders, then including them in all major decisions may be wise. However, technical issues, such as which statistical techniques to use, generally do not benefit from, or need stakeholder involvement. In addition, there may be situations in which the science collides with the preferences of a stakeholder. For example, a stakeholder may want to do the research quicker or with fewer participants. In cases such as these, it is critical for the researcher to provide persuasive information about the possible trade-offs of following the stakeholder advice, such as reducing the ability to find an effect if one is actually present—that is, lowering statistical power. Applied researchers often find themselves educating stakeholders about the possible trade-offs that could be made. The researcher will sometimes need to persuade stakeholders to think about the problem in a new way or demonstrate the difficulties in implementing the original design.

The culmination of Stage II is a comprehensively planned applied research project, ready for full-scale implementation. With sufficient planning completed at this point, the odds of a successful study are significantly improved, but far from guaranteed. As discussed later in this chapter, conducting pilot and feasibility studies continues to increase the odds that a study can be successfully mounted.

In the sections to follow, we outline the key activities that need to be conducted in Stage I of the planning process, followed by highlighting the key features that need to be considered in choosing a design (Stage II), and the variety of designs available for different applied research situations. We then go into greater depth on various aspects of the design process, including selecting the data collection methods and approach, determining the resources needed, and assessing the research focus.

Developing a Consensus on the Nature of the Research Problem

Before an applied research study can even begin to be designed, there has to be a clear and comprehensive understanding of the nature of the problem being addressed. For example, if the study is focused on evaluating a program for homeless families being conducted in Georgia, the researcher should know what research and other available information has been developed about the needs and characteristics of homeless families in general and specifically in Georgia; what evidence base exists, if any for the type of program being tested in this study; and so forth. In addition, if the study is being requested by an outside sponsor, it is important to have an understanding of the impetus of the study and what information is desired to inform decision making.

Strategies that can be used in gathering the needed information include the following:

- review relevant literature (research articles and reports, transcripts of legislative hearings, program descriptions, administrative reports, agency statistics, media articles, and policy/position papers by all major interested parties);
- gather current information from experts on the issue (all sides and perspectives) and major interested parties;
- conduct information-gathering visits and observations to obtain a real-world sense of the context and to talk with persons actively involved in the issue;
- initiate discussions with the research clients or sponsors (legislative members; foundation, business, organization, or agency personnel; and so on) to obtain the clearest possible picture of their concerns; and
- if it is a program evaluation, informally visit the program and talk with the staff, clients, and others who may be able to provide information on the program and/or overall research context.

Developing the Conceptual Framework

Every study, whether explicitly or implicitly, is based on a conceptual framework or model that specifies the variables of interest and the expected relationships between them. In some studies, social and behavioral science theory may serve as the basis for the conceptual framework. For example, social psychological theories such as cognitive dissonance may guide investigations of behavior change. Other studies, such as program and policy evaluations, may be based not on formal academic theory but on statements of expectations of how policies or programs are purported to work. Bickman (1987, 1990) and others (e.g., Chen, 1990) have written extensively about the need for and usefulness of program theory to guide evaluations. The framework may be relatively straightforward or it may be complex, as in the case of evaluations of comprehensive community reforms, for example, that are concerned with multiple effects and have a variety of competing explanations for the effects (e.g., Rog & Knickman, 2004).

In evaluation research, logic models have increased in popularity as a mechanism for outlining and refining the focus of a study (Frechtling, 2007; McLaughlin & Jordan, 2004; Rog, 1994; Rog & Huebner, 1992; Yin, Chapter 8, this volume). A logic model, as the name implies, displays the underlying logic of the program (i.e., how the program goals, resources, activities, and outcomes link together). In several instances, a program is designed without explicit attention to the evidence base available on the topic and/or without explicit attention to what immediate and intermediate outcomes each program component and activity needs to accomplish to ultimately reach the desired longer-term outcomes. The model helps display these gaps in logic and provides a guide for either refining the program and/or outlining more of the expectations for the program. For example, community coalitions funded to prevent community violence need to have an explicit logic that details the activities they are intended to conduct that should lead to a set of outcomes that chain logically to the prevention of violence.

The use of logic modeling in program evaluation is an outgrowth of the evaluability assessment work of Wholey and others (e.g., Wholey, 2004), which advocates describing and displaying the underlying theory of a program as it is designed and implemented prior to conducting a study of its outcomes. Evaluators have since discovered the usefulness of logic models in assisting program developers in the program design phase, guiding the evaluation of a program's effectiveness, and communicating the nature of a program as well as changes in its structure over time to a variety of audiences. A program logic model is dynamic and changes not only as the program matures but also may change as the researcher learns more about the program. In addition, a researcher may develop different levels of models for different purposes; for example, a global model may be useful for communicating to outside audiences about the nature and flow of a program, but a detailed model may be needed to help guide the measurement phase of a study.

In the design phase of a study (Stage II), the logic model will become important in guiding both the measurement and analysis of a study. For these tasks, the logic model needs to not only display the main features of a program and its outcomes but also the variables that are believed to mediate the outcomes as well as those that could moderate an intervention's impact (Baron & Kenny, 1986). Mediating variables, often referred to as intervening or process variables, are those variables through which an independent variable (or program variable) influences an outcome. For example, the underlying theory of a therapeutic program designed to improve the overall well-being of families may indicate that the effect of the program is mediated by the therapeutic alliance developed between the families and the program staff. In other words, without the development of a therapeutic alliance, it is not expected that the program can have an effect. Often, mediators are short-term outcomes that are believed to be logically necessary for a program to first accomplish in order to achieve the longer-term outcomes.

Moderators are those variables that explain differences in outcomes due to preexisting conditions. For example, demographic variables, such as gender, age, income, and others are often tested as moderators of a program's effects. Contextual variables also can act as moderators of the effects of a program; for example, a housing program for homeless families is expected to have greater effect on housing stability in communities that have higher housing vacancy rates than those with lower rates (i.e., less available housing).

Identifying the Research Questions

As noted in the introduction to this *Handbook*, one of the major differences between basic research and applied research is that the basic researcher is more autonomous than the applied researcher. Basic research, when externally funded, is typically conducted through a relatively unrestricted grant mechanism; applied research is more frequently funded through contracts and cooperative agreements. Even when applied research is funded through grant mechanisms, such as with foundations, there is usually a "client" or sponsor who specifies (or at least guides) the research agenda and requests the research results. Most often, studies have multiple stakeholders: sponsors, interested beneficiaries, and potential users (Bickman

& Rog, 1986). The questions to be addressed by an applied study tend to be posed by individuals other than the researcher, often by nontechnical persons in nontechnical language.

Therefore, one of the first activities in applied research is working with the study clients to develop a common understanding of the research agenda—the research questions. Phrasing study objectives as questions is desirable in that it leads to more clearly focused discussion of the type of information needed. It also makes it more likely that key terms (e.g., welfare dependency, drug use) will be operationalized and clearly defined. Using the logic models also helps focus the questions on what is expected from the program and to move to measurable variables to both study the process of an intervention or program as well as its expected outcomes. Later, after additional information has been gathered and reviewed, the parties will need to reconsider whether these questions are the "right" questions and whether it is possible, with a reasonable degree of confidence, to obtain answers for these questions within the available resource and time constraints.

Clarifying the Research Questions

In discussing the research agenda with clients, the researcher will usually identify several types of questions. For example, in a program evaluation, researchers are frequently asked to produce comprehensive information on both the implementation ("what actually is taking or took place") and the effects ("what caused what") of an intervention. When the research agendas are broad such as those in the example, they pose significant challenges for planning in terms of allocating data collection resources among the various study objectives. It is helpful to continue to work with the sponsors to further refine the questions to both more realistically plan the scope of the research and to also ensure that they are specific enough to be answered in a meaningful way and one that is agreed on by the clients.

The researcher should guard against biasing the scope of the research. The questions left unaddressed by a study can be as or more important than the questions answered. If the research addresses only questions likely to support only one position in a controversy and fails to develop information relevant to the concerns voiced by other interested parties, it will be seen as biased, even if the results produced are judged to be sound and conclusive. For example, an evaluation that is limited to measuring just the stated goals of a program may be biased if any possible unintended negative side effects of the program are not considered. Thus, the research agenda should be as comprehensive as is necessary to address the concerns of all parties. Resource constraints will limit the number and scope of questions that may be addressed, but at minimum the researcher should state explicitly what would be necessary for a comprehensive study and how the research meets or does not meet those requirements. Resources will also determine the degree of certainty one can have in an answer. Thus, a representative survey is much more expensive to conduct than sampling by convenience, but the generalizability of the results will be much stronger in the representative sample.

Ideally, the development of the conceptual framework/logic model will occur simultaneously with the identification of the research questions. Once the

conceptual framework has been agreed on, the researcher can further refine the study questions—grouping questions and identifying which are primary and secondary questions. Areas that need clarification include the time frame of the data collection (i.e., "Will it be a cross-sectional study or one that will track individuals or cohorts over time; how long will the follow-up period be?"); how much the client wants to generalize (e.g., "Is the study interested in providing outcome information on all homeless families that could be served in the program or only those families with disabilities?"); how certain the client wants the answers to be (i.e., "How precise and definitive should the data collected be to inform the decisions?"); and what subgroups the client wants to know about (e.g., "Is the study to provide findings on homeless families in general only or is there interest in outcomes for subgroups of families, such as those who are homeless for the first time, those who are homeless more than once but for short durations, and those who are 'chronically homeless'?"). The levels of specificity should be very high at this point, enabling a clear agreement on what information will be produced. As the next section suggests, these discussions between researcher and research clients oftentimes take on the flavor of a negotiation.

Negotiating the Scope of a Study

Communication between the researcher and stakeholders (the sponsor and all other interested parties) is important in all stages of the research process. To foster maximum and accurate utilization of results, it is recommended that the researcher regularly interact with the research clients—from the initial discussions of the "problem" to recommendations and follow-up. In the planning phase, we suggest several specific communication strategies. As soon as the study is sponsored, the researcher should connect with the client to develop a common understanding of the research questions, the client's time frame for study results, and anticipated uses for the information. The parties can also discuss preliminary ideas regarding a conceptual model for the study. Even in this initial stage, it is important for the researcher to begin the discussion of the contents and appearance of the final report. This is an opportunity for the researcher to explore whether the client expects only to be provided information on study results or whether the client anticipates that the researcher will offer recommendations for action. It is also an opportunity for the researcher to determine whether he or she will be expected to provide interim findings to the client as the study progresses.

At this juncture, the researcher also needs to have an understanding of the amount of funds or resources that will be available to support the research. Cost considerations will determine the scope and nature of the project, and the investigator needs to consider the resources while identifying and reviewing the research questions. In some studies, the budget is set prior to any direct personal contact with the research client. In others, researchers may help to shape the scope and the resources needed simultaneously or there may be a pilot effort that helps design the larger study.

Based on a comprehensive review of the literature and other inputs (e.g., from experts) and an initial assessment of resources, the researcher should decide if the

research questions need to be refined. The researcher and client then typically discuss the research approaches under consideration to answer these questions as well as the study limitations. This gives the researcher an opportunity to introduce constraints into the discussion regarding available resources, time frames, and any trade-offs contemplated regarding the likely precision and conclusiveness of answers to the questions.

In most cases, clients want sound, well-executed research and are sympathetic to researchers' need to preserve the integrity of the research. Some clients, however, have clear political, organizational, or personal agendas, and will push researchers to provide results in unrealistically short time frames or to produce results supporting particular positions. Other times, the subject of the study itself may generate controversy, a situation that requires the researcher to take extreme care to preserve the neutrality and credibility of the study. Several of the strategies discussed later attempt to balance client and researcher needs in a responsible fashion; others concentrate on opening research discussions up to other parties (e.g., advisory groups). In the earliest stages of research planning, it is possible to initiate many of these kinds of activities, thereby bolstering the study's credibility, and often its feasibility.

Stage II: The Research Design

Having developed a preliminary study scope during Stage I, the researcher moves to Stage II, developing a research design and plan. During this stage, the applied researcher needs to perform five activities almost simultaneously: selecting a design, choosing data collection approaches, inventorying resources, assessing the feasibility of executing the proposed approach, and determining trade-offs. These activities and decisions greatly influence one another. For example, a researcher may revisit preliminary design selections after conducting a practical assessment of the resources available to do the study, and may change data collection plans after discovering weaknesses in the data sources during planning.

The design serves as the architectural blueprint of a research project, linking design, data collection, and analysis activities to the research questions and ensuring that the complete research agenda will be addressed. A research study's credibility, usefulness, and feasibility rest with the design that is implemented. *Credibility* refers to the validity of a study and whether the design is sufficiently rigorous to provide support for definitive conclusions and desired recommendations. Credibility is also, in part, determined by who is making the judgment. To some sponsors, a credible project need only use a pre-post design. Others may require a randomized experimental design to consider the findings credible. Credibility is also determined by the research question. A representative sample will make a descriptive study more credible than a sample of convenience or one with known biases. In contrast, representativeness is not as important in a study designed to determine the causal link between a program and outcomes. The planner needs to be sure that the design matches the types of information needed. For example,

under most circumstances, the simple pre-post design should not be used if the purpose of the study is to draw causal conclusions.

Usefulness refers to whether the design is appropriately targeted to answer the specific questions of interest. A sound study is of little use if it provides definitive answers to the wrong questions. *Feasibility* refers to whether the research design can be executed, given the requisite time and other resource constraints. All three factors—credibility, usefulness, and feasibility—must be considered to conduct high-quality applied research.

Design Dimensions

Maximizing Validity

In most instances, a credible research design is one that maximizes validity—it provides a clear explanation of the phenomenon under study and controls all plausible biases or confounds that could cloud or distort the research findings. Four types of validity are typically considered in the design of applied research (Bickman, 1989; Shadish, Cook, & Campbell, 2002).

- *Internal validity:* the extent to which causal conclusions can be drawn or the degree of certainty that "A" caused "B," where A is the independent variable (or program) and B is the dependent variable (or outcome).
- *External validity:* the extent to which it is possible to generalize from the data and context of the research study to other populations, times, and settings (especially those specified in the statement of the original problem/issue).
- *Construct validity:* the extent to which the constructs in the conceptual framework are successfully operationalized (e.g., measured or implemented) in the research study. For example, does the program as actually implemented accurately represent the program concept and do the outcome measures accurately represent the outcome? Programs change over time, especially if fidelity to the program model or theory is not monitored.
- *Statistical conclusion validity:* the extent to which the study has used appropriate sample size, measures, and statistical methods to enable it to detect the effects if they are present. This is also related to the statistical power.

All types of validity are important in applied research, but the relative emphases may vary, depending on the type of question under study. With questions dealing with the effectiveness of an intervention or impact, for example, more emphasis should be placed on internal and statistical conclusion validity than on external validity. The researcher of such a study is primarily concerned with finding any evidence that a causal relationship exists and is typically less concerned (at least initially) about the transferability of that effect to other locations or populations. For descriptive questions, external and construct validity may receive greater emphasis. Here, the researcher may consider the first priority to be developing a comprehensive and rich picture of a phenomenon. The need to make cause-effect attributions is not relevant. Construct validity, however, is almost always relevant.

Operationalizing the Key Variables and Concepts

The process of refining and revising the research questions undertaken in Stage I should have yielded a clear understanding of the key research variables and concepts. For example, if the researcher is charged with determining the extent of high school drug use (a descriptive task), key outcome variables might include drug type, frequency and duration of drug use, and drug sales behavior. Attention should be given at this point to reassessing whether the researcher is studying the right variables—that is, whether these are "useful" variables.

Outlining Comparisons

An integral part of design is identifying whether and what comparisons can be made—that is, which variables must be measured and compared with other variables or with themselves over time. In simple descriptive studies, there are decisions to be made regarding the time frame of an observation and how many observations are needed. Typically, there is no explicit comparison in simple descriptive studies. Normative studies are an extension of descriptive studies in that the interest is in comparing the descriptive information to some appropriate "standard." The decision for the researcher is to determine where that standard will be drawn from or how it will be developed. In correlative studies, the design is again an extension of simple descriptive work, with the difference that two or more descriptive measures are arrayed against each other to determine whether they covary. Impact or outcome studies, by far, demand the most judgment and background work. To make causal attributions (*X* causes *Y*), we must be able to compare the condition of *Y* when *X* occurred with what the condition of *Y* would have been without *X*. For example, to know if a drug treatment program reduced drug use, we need to compare drug use among those who were in the program with those who did not participate in the program.

Level of Analysis

Knowing what level of analysis is necessary is also critical to answering the "right" question. For example, if we are conducting a study of drug use among high school students in Toledo, "Are we interested in drug use by individual students, aggregate survey totals at the school level, aggregate totals at the school district, or for the city as a whole?"

Correct identification of the proper level or unit of analysis has important implications for both data collection and analysis. The Stage I client discussions should clarify the desired level of analysis. It is likely that the researcher will have to help the client think through the implications of these decisions, providing information about research options and the types of findings that would result. In addition, this is an area that is likely to be revisited if initial plans to obtain data at one level (e.g., the individual student level) prove to be prohibitively expensive or unavailable. A design fallback position may be to change to an aggregate analysis level (e.g., the school), particularly if administrative data at this level are more readily available and less costly to access.

In an experiment, the level of analysis is typically determined by the level that the intervention is introduced. For example, if the intervention was targeted at individual students, then that should usually be the level of analysis. Similarly, a classroom intervention should use classroom as the level and a schoolwide intervention should use the school. It is tempting to use the lowest level with the largest sample size because that provides the most statistical power—that is, ability to find an effect if one is there. For example, if an intervention is at the school level and there is only a treatment and control school then the sample size is two, not the total number of students. Statistical programs that take into account multilevel designs are easily accessible (Graham, Singer, & Willett, 2008). However, the real challenge with multilevel designs is finding enough units (e.g., schools) to cooperate as well as enough resources to pay for the study.

Population, Geographic, and Time Boundaries

Population, geographic, and time boundaries are related to external validity issues. Each can affect the generalizability of the research results—for instance, whether the results will be representative of all high school students, all high school students graduating within the past 3 years, all students in urban areas, and so on. Population generalizability and geographic generalizability are probably the most commonly discussed types of generalizability, and researchers frequently have heated debates concerning whether the persons or organizations that they have studied and the locations where they conducted their studies will allow them to use their findings in different locations and with different populations. In basic research, generalizability or external validity is usually not considered but in applied research some may rate it more important than internal validity (Cronbach et al., 1980).

Time boundaries also can be crucial to the generalizability of results, especially if the study involves extant data that may be more than a few years old. With the fast pace of change, questions can easily arise about whether survey data on teenagers from even just 2 years prior are reflective of current teens' attitudes and behaviors.

The researcher cannot study all people, all locations, or all time periods relevant to the problem/program under scrutiny. One of the great "inventions" for applied social research is sampling. Sampling allows the researcher to study only a subset of the units of interest and then generalize to all these units with a specifiable degree of error. It offers benefits in terms of reducing the resources necessary to do a study; it also sometimes permits more intensive scrutiny by allowing a researcher to concentrate on fewer cases. More details on sampling can be found in Henry (1990; see also Sieber, Chapter 4, this volume).

Level of Precision

Knowing how precise an answer must be is also crucial to design decisions. The level of desired precision may affect the rigor of the design. When sampling is used, the level of desired precision also has important ramifications for how the sample is drawn and the size of the sample used. In initial discussions, the researcher and the

client should reach an understanding regarding the precision desired or necessary overall and with respect to conclusions that can be drawn about the findings for specific subgroups. The cost of a study is very heavily influenced by the degree of precision or certainty required. In sampling, more certainty usually requires a bigger sample size, with diminishing returns when samples approach 1,000. However, if the study is focused on subgroups, such as gender or ethnicity, then the sample at those levels of analysis must also be larger.

Another example of precision is the breadth and depth of a construct that need to be measured in a study. More breadth usually requires more questions, and greater depth often requires the use of in-depth interviewing, both likely increasing the costs of data collection especially if administered in person or with a telephone interview. The level of precision is discussed later in the section dealing with trade-offs as level of precision is often a trade-off decision that must be made within the budget of a study.

Choosing a Design

There are three main categories of applied research designs: descriptive, experimental, and quasi-experimental. In our experience, developing an applied research design rarely allows for implementing a design straight from a textbook; rather, the process more typically involves the development of a hybrid, reflecting combinations of designs and other features that can respond to multiple study questions, resource limitations, dynamics in the research context, and other constraints of the research situation (e.g., time deadlines). Thus, our intent here is to provide the reader with the tools to shape the research approach to the unique aspects of each situation. Those interested in more detailed discussion should consult Mark and Reichardt's work on quasi-experimentation (Chapter 6) and Boruch and colleagues' chapter on randomized experiments (Chapter 5). In addition, our emphasis here is on quantitative designs; for more on qualitative designs, readers should consult Maxwell (Chapter 7), Yin (Chapter 8), and Fetterman (Chapter 17).

Descriptive Research Designs

Description and Purpose. The overall purpose of descriptive research is to provide a "picture" of a phenomenon as it naturally occurs, as opposed to studying the effects of the phenomenon or intervention. Descriptive research can be designed to answer questions of a univariate, normative, or correlative nature—that is, describing only one variable, comparing the variable to a particular standard, or summarizing the relationship between two or more variables.

Key Features. Because the category of descriptive research is broad and encompasses several different types of designs, one of the easiest ways to distinguish this class of research from others is to identify what it is not: It is not designed to provide information on cause-effect relationships.

Variations. There are only a few features of descriptive research that vary. These are the representativeness of the study data sources (e.g., the subjects/entities)—that is, the manner in which the sources are selected (e.g., universe, random sample, stratified sample, nonprobability sample); the time frame of measurement—that is, whether the study is a one-shot, cross-sectional study, or a longitudinal study; whether the study involves some basis for comparison (e.g., with a standard, another group or population, data from a previous time period); and whether the design is focused on a simple descriptive question, on a normative question, or on a correlative question.

When to Use. A descriptive approach is appropriate when the researcher is attempting to answer "what is," or "what was," or "how much" questions.

Strengths. Exploratory descriptive studies can be low cost, relatively easy to implement, and able to yield results in a fairly short period of time. Some efforts, however, such as those involving major surveys, may sometimes require extensive resources and intensive measurement efforts. The costs depend on factors such as the size of the sample, the nature of the data sources, and the complexity of the data collection methods employed. Several chapters in this volume outline approaches to surveys, including mail surveys (Mangione & Van Ness, Chapter 15), internet surveys (Best & Harrison, Chapter 13), and telephone surveys (Lavrakas, Chapter 16).

Limitations. Descriptive research is not intended to answer questions of a causal nature. Major problems can arise when the results from descriptive studies are inappropriately used to make causal inferences—a temptation for consumers of correlational data.

Experimental Research Designs

Description and Purpose. The primary purpose in conducting an experimental study is to test the existence of a causal relationship between two or more variables. In an experimental study, one variable, the independent variable, is systematically varied or manipulated so that its effects on another variable, the dependent variable, can be measured. In applied research, such as in program evaluation, the "independent variable" is typically a program or intervention (e.g., a drug education program) and the "dependent variables" are the desired outcomes or effects of the program on its participants (e.g., drug use, attitudes toward drug use).

Key Features. The distinguishing characteristic of an experimental study is the random assignment of individuals or entities to the levels or conditions of the study. Random assignment is used to control most biases at the time of assignment and to help ensure that only one variable—the independent (experimental) variable—differs between conditions. With well-implemented random assignment, all individuals have an equal likelihood of being assigned either to the treatment group or to the control group. If the total number of individuals or entities assigned to the treatment and control groups is sufficiently large, then any differences between the groups should be small and due to chance.

Variations. The most basic experimental study is called a *post-only design,* in which individuals are randomly assigned either to a treatment group or to a control group, and the measurement of the effects of the treatment is conducted at a given period following the administration of the treatment. There are several variations to this simple experimental design that can respond to specific information needs as well as provide control over possible confounds or influences that may exist. Among the features that can be varied are the number and scheduling of posttest measurement or observation periods, whether a preobservation is conducted, and the number of treatment and control groups used. The post-only design is rarely used because faulty random assignment may result in the control and treatment groups not being equivalent at the start of the study. Few researchers are that (over) confident in the implementation of a field randomized design to take the chance that the results could be interpreted as being caused by faulty implementation of the design.

When to Use. An experimental study is the most appropriate approach to study cause-effect relationships. There are certain situations that are especially conductive to randomized experiments (Boruch et al., Chapter 5, this volume; Shadish et al., 2002) when random assignment is expected (i.e., certain scarce resources may already be provided on a "lottery" or random basis), when demand outstrips supply for an intervention, and when there are multiple entry groups over a period of time.

Strengths. The overwhelming strength of a randomized experiment is its control over threats to internal validity—that is, its ability to rule out potential alternative explanations for apparent treatment or program effects. This strength applies to both the variables that are measured and, more important, the variables that are not measured and, thus, are unknown to the researcher but continue to be controlled by the design.

Limitations. Randomized experiments can be difficult to implement with integrity, particularly in settings where the individuals responsible for random assignment procedures lack research training or understanding of the importance of maintaining compliance with the research protocol (Bickman, 1985; Cook, 2002; Riccio & Bloom, 2002). In addition, random assignment does not control for all biases such as participant preference for one condition over the other (Macias, Hargreaves, Bickman, Fisher, & Aronson, 2005) or local history where some external event occurs for one group but not for the other.

Quasi-Experimental Designs

Description and Purpose. Quasi-experimental designs have the same primary purpose as experimental studies—to test the existence of a causal relationship between two or more variables. They are used when random assignment is not feasible or desired.

Key Features. Quasi-experiments attempt to approximate randomized experiments by substituting other design features for the randomization process. There are generally two ways to create a quasi-experimental comparison base—through the

addition of nonequivalent comparison groups or through the addition of pre- and posttreatment observations on the treated group; preferably, both methods should be used.

If comparison groups are used, they are generally referred to as *nonequivalent* comparison groups based on the fact that they cannot be equivalent with the treatment group as in a randomized experiment. The researcher, however, strives to develop procedures to make these groups as equivalent as possible to provide necessary information and control so that competing explanations for their results can be ruled out.

Variations. Quasi-experiments vary along several of the same dimensions that are relevant for experiments. Overall, there are two main types of quasi-experiments: those involving data collection from two or more nonequivalent groups and those involving multiple observations over time. More specifically, quasi-experimental designs can vary along the following dimensions: the number and scheduling of pre- or postobservation periods; the nature of the observations—whether the preobservation uses the same measurement procedure as the postobservation, or whether both are using measures that are proxies for the real concept; the manner in which the treatment and comparison groups are determined; and whether the treatment group serves as its own comparison group or a separate comparison group or groups are used.

Some of the strongest time-series designs supplement a time series for the treatment group with comparison time series for another group (or time period). Another powerful variation occurs when the researcher is able to study the effects of an intervention over time under circumstances where that intervention is both initiated and later withdrawn. A third strong design is the regression discontinuity design in which participants are assigned to a treatment or comparison group based on a clearly designated pretest score. Although this design has been used in clinical screening (e.g., CATS Consortium, 2007), it is rarely used as most studies do not involve the use of a pretest score as a cutoff.

When to Use. A quasi-experimental design is not the method of choice but rather a fallback strategy for situations in which random assignment is not feasible. Situations such as these include when the nature of the independent variable precludes the use of random assignment (e.g., exposure or involvement in a natural disaster); retrospective studies (e.g., the program is already well under way or over); studies focused on economic or social conditions, such as unemployment; when randomization is too expensive, not feasible to initiate, or impossible to monitor closely; when there are obstacles to withholding the treatment or when it seems unethical to withhold it; and when the timeline is tight and a quick decision is mandated.

Strengths. The major strength of the quasi-experimental design is that it provides an approximation to the experimental design and supports causal inferences. Although often open to several types of threats to internal validity (see Mark & Reichardt, Chapter 6), the quasi-experiment does provide a mechanism for chipping away at the uncertainty surrounding the existence of a specific causal

relationship. Additional nonequivalent comparison groups also can bolster an experimental design, particularly if it is narrowly focused.

Limitations. The greatest vulnerability of quasi-experimental designs is the possibility that the comparison group created is biased and that it does not give an accurate estimate of what the situation would have been in the absence of the treatment or program. This is especially a concern when participants self-select into treatment or control groups. Although not a perfect remedy, propensity score matching is increasingly used as a technique for helping to correct for selection bias between treatment and comparison groups (Foster, 2003; Rosenbaum & Rubin, 1983, 1984; Rubin, 1997). A propensity score is a composite of variables that controls on known differences between two groups by creating matches or subgroups of cases that are similar on this score.

Selecting Data Collection Approaches

Concurrent with deciding on a design, the researcher should investigate possible data collection approaches. Most applied research studies, particularly those investigating multiple research questions, often encompass several data collection efforts. We begin this section with a discussion of the data collection issues that the researcher must consider during the planning stage, including the sources of data available, the form in which the data are available, the amount of data needed, the accuracy and reliability of the data, and whether the data fit the parameters of the design. We then review the major methods of data collection that are used in applied research and discuss the need for an analysis plan.

Sources of Data

The researcher should identify the likely sources of data to address the research questions. Data sources typically fall into one of two broad categories: primary and secondary. Among the potential primary data sources that exist for the applied researcher are people (e.g., community leaders, program participants, service providers, the general public), independent descriptive observations of events and activities, physical documents, and test results. These data are most often collected by the investigator as part of the study through one or more methods (e.g., questionnaires, interviews, observations). Secondary sources can include administrative records, management information systems, economic and social indicators, and various types of documents (e.g., prior research studies, fugitive unpublished research literature) (Gorard, 2002; Hofferth, 2005; Stewart & Kamins, 1993). Typically the investigator does not collect these data but uses already existing sources such as census data, program administrative records, and others. In recent years, there has been an increasing emphasis on performance-monitoring systems and the implementation of management information systems, especially in agencies and organizations that receive government funding. These systems can be often considered potential sources to tap in applied research projects depending on the quality and completeness of the data collected (as discussed below).

Form of the Data

The form in which the data are found is a very important factor for any applied research project and may even determine the overall feasibility of the study. Some projects are easy to conduct—the data sources are obvious and the data are already gathered, archived, and computerized. The researcher need only request access to the files and have the ability to transfer them. However, even these data may not be easy to use if the data have problems such as missing or duplicated cases or are composed of different files that require matching clients across files. Other projects are extremely difficult—identifying appropriate sources for the needed information may be confusing, and it may turn out that the procedures necessary for obtaining the information are expensive and time-consuming. Gathering data may sometimes be so difficult that the study is not feasible—at least not within the available level of resources and other constraints. For example, a study of several school systems required that the researchers have access to the student achievement data. Obtaining these data sets actually took several years because the researchers' needs were not a high priority in the school systems relative to other priorities. Moreover, one of the school districts was changing computer software, further delaying the process. The lesson here is what seems like a simple request is usually not that simple.

Possible forms of data include self-reports (e.g., attitudes, behaviors and behavioral intentions, opinions, memories, characteristics, and circumstances of individuals), computerized or manual (i.e., hard copy) research databases or administrative records, observations (e.g., events, actions, or circumstances that need to be described or recorded), biobehavioral measures (e.g., urinalysis to measure drug use), and various kinds of documentary evidence (e.g., letters, invoices, receipts, meeting minutes, memoranda, plans, reports).

Self-Report Data

When dealing with self-reported data, the researcher may ask individual research participants to provide, to the best of their ability, information on the areas of interest. These inquiries may be made through individual interviews, through telephone or mail surveys, Web-based surveys, or through written corroboration or affirmation. Self-report data may be biased if the questions deal with socially desirable behavior, thoughts, or attitudes. In general, people like to present themselves in a positive way. Making the data collection anonymous may improve the accuracy of these data, especially about sensitive topics. However, anonymous data can be difficult, but not impossible, to use in the conduct of longitudinal studies.

Extant Databases

When dealing with extant data from archival sources, the researcher is generally using the information for a purpose other than that for which they were originally collected. There are several secondary data sources that are commonly used, such as those developed by university consortia, federal sources such as the Bureau of the

Census, state and local sources such as Medicaid databases, and commercial sources such as Inform, a database of 550 business journals.

Given the enormous amount of information routinely collected on individuals in U.S. society, administrative databases are a potential bonanza for applied researchers. More and more organizations, for example, are computerizing their administrative data and archiving their full databases at least monthly. Management information systems, in particular, are becoming more common in service settings for programmatic and evaluation purposes as well as for financial disbursement purposes.

Administrative data sets, however, have one drawback in common with databases of past research—they were originally constructed for operational purposes, not to meet the specific objectives of the researcher's task. When the data are to be drawn from administrative databases, the researcher should ask the following questions: Are the records complete? Why were the data originally collected? Did the database serve some hidden political purpose that could induce systematic distortions? What procedures have been used to deal with missing data? Do the computerized records bear a close resemblance to the original records? Are some data items periodically updated or purged from the computer file? How were the data collected and entered, and by whom?

Biobehavioral Data

Biobehavioral measures are becoming increasingly important, especially in health and health-related research. Body mass index, for example, is often used in research on obesity as a measure of fitness (Flegal, Carroll, Ogden, & Johnson, 2002). Increasingly, in studies of illegal behavior, such as drug use, biobehavioral measures using urinalysis are viewed as more valid than self-reports due to the stigma associated with the behavior (e.g., Kim & Hill, 2003). Many of the measures, however, require the use of advanced technology and can increase the expense of data collection.

Observational Data

Observational procedures become necessary when events, actions, or circumstances are the major form of the data. If the events, actions, or circumstances are repetitive or numerous, this form of data can be easier to collect than data composed of rare events that are difficult to observe. Because the subject of the data collection is often complex, the researcher may need to create detailed guidelines to structure the data collection, coding, and analysis (see Maxwell, Chapter 7, for more detail on qualitative data categorization and analysis).

Documents

Documentary evidence may also serve as the basis for an applied researcher's data collection. Particular kinds of documents may allow the researcher to track

what happened, when it happened, and who was involved. Examples of documentary data include meeting minutes, journals, and program reports. Investigative research may rely on documentary evidence, often in combination with data from interviews.

Amount of Data

The research planner must anticipate the amount of data that will be needed to conduct the study. Planning for the appropriate amount involves decisions regarding the number and variety of data sources, the time periods of interest, and the number of units (e.g., study participants), as well as the precision desired. As noted earlier, statistical conclusion validity concerns primarily those factors that might make it appear that there were no statistically significant effects when, in fact, there were effects.

Effect size is defined as the proportion of variance accounted for by the treatment, or as the difference between a treatment and control group measured in standard deviation units. The purpose of using standard deviation units is to produce a measure that is independent of the metric used in the original variable. Thus, we can discuss universal effect sizes regardless of whether we are measuring school grades, days absent, or self-esteem scores. This makes possible the comparison of different studies and different measures in the same study. Conversion to standard deviation units can be obtained by subtracting the mean of the control group from the mean of the treatment group and then dividing this difference by the pooled or combined standard deviations of the two groups.

There are several factors that could account for not finding an effect when there actually is one. As Lipsey and Hurley (Chapter 2) indicate, there are four factors that govern statistical power: the statistical test, the alpha level, the sample size, and the effect size. Many researchers, when aware of power concerns, mistakenly believe that increasing sample size is the only way to increase statistical power. Increasing the amount of data collected (the sample size) is clearly one route to increasing power; however, given the costs of additional data collection, the researcher should consider an increase in sample size only after he or she has thoroughly explored the alternatives of increasing the sensitivity of the measures, improving the delivery of treatment to obtain a bigger effect, selecting other statistical tests, and raising the alpha level. If planning indicates that power still may not be sufficient, then the researcher faces the choice of not conducting the study, changing the study to address more qualitative questions, or proceeding with the study but informing the clients of the risk of "missing" effects below a certain size. (More information on how to improve the statistical power of a design can be found in Lipsey & Hurley, Chapter 2.)

With qualitative studies, the same set of trade-offs are made in planning how much data to collect—that is, consideration of the number and variety of data sources available, the time periods of interest, and the number of units, as well as the precision desired (see Harrison, Chapter 10). Precision in qualitative studies, however, does not refer to statistical power as much as the need for triangulation to establish the validity of conclusions. Triangulation refers to the use of multiple data sources and/or methods to measure a construct or a phenomenon in order to see if

they converge and support the same conclusions. The more diverse the sources and methods, the greater confidence there is in the convergence of the findings. Maxwell (Chapter 7) describes a number of strategies, including triangulation, for ensuring and assessing the validity of conclusions from qualitative data.

Accuracy and Reliability of Data

Data are not useful if they are not accurate, valid, and reliable. The concept of construct validity (i.e., Are we measuring what we intend to measure?) is relevant whether one is using extant data or collecting primary data. The researcher is concerned that the variables used in the study are good operationalizations of key variables in the study's conceptual framework.

The researcher must also be concerned with the possibility of large measurement errors. Whenever there is measurement of a phenomenon, there is some level of error. The error may be random or systematic. It is important for the researcher to remember that just about all measures contain some degree of error; the challenge is to minimize the error or understand it sufficiently to adjust the study. If the error is systematic (i.e., not random), the researcher may be able to correct statistically for the bias that is introduced. However, it is often difficult for the researcher to discover that any systematic error exists, let alone its magnitude. Random error can best be controlled through the use of uniform procedures in data collection.

Researchers should be cautious in the development of their own measures. As noted in other chapters in the *Handbook* (Fowler & Cosenza, Chapter 12), developing a good questionnaire requires more than writing some questions. In one of our projects, we needed to use instruments that were short, valid, reliable, and free. Unfortunately, such measures are rare in the child and adolescent mental therapeutic alliance and session impact. Developing these measures was a yearlong activity that consumed a great deal of time and money. Creating the questions was the easy part. We needed to conduct cognitive testing to determine if the respondents were interpreting the instructions and questions as expected, piloting for length, and then intensive psychometric testing that included collecting data from more than 1,000 respondents and analyzing the data using both classical and item response theory approaches. The test battery is available free at http://peabody.vanderbilt.edu/ptpb.

Design Fit

Even when accurate and reliable data exist or can be collected, the researcher must ask whether the data fit the necessary parameters of the design. Are they available on all necessary subgroups? Are they available for the appropriate time periods? Is it possible to obtain data at the right level of analysis (e.g., individual student vs. school)? Do different databases feeding into the study contain comparable variables? Are they coded the same way?

If extant databases are used, the researcher may need to ask if the database is sufficiently complete to support the research. Are all variables of interest present? If an interrupted time-series design is contemplated, the researcher may need to make

sure that it is possible to obtain enough observations prior to the intervention in question and that there has been consistency in data reporting throughout the analytic time frame.

Types of Data Collection Instruments

Observational Recording Forms

Observational recording forms are guides to be used in the requesting and documenting of information. The subjects may be events, actions, or circumstances, whether live or re-created through discussions or review of written documentation. Observational recording forms are needed when there is substantial information to be collected through observational means or when there are multiple data collectors. When a study employs multiple data collectors, creating a recording guide can help the researcher make sure that all areas have been covered and can eliminate the need for recontacting research participants. Also, when there are multiple data collectors, the use of a recording form provides necessary structure to the data collection process, thereby ensuring that all collectors are following similar procedures and employing similar criteria in choosing to include or exclude information. There are several programs available that increase the ease of data collection through the use of laptops or personal digital assistants (Eid & Diener, 2006; Felce & Emerson, 2000).

Tests

In applied studies, researchers are more likely to make use of existing instruments to measure knowledge or performance than to develop new ones. Whether choosing to use a test "off the shelf" or to capitalize on an existing database that includes such data, it is very important that the researcher be thoroughly familiar with the content of the instrument, its scoring, the literature on its creation and norming, and any ongoing controversies about its accuracy. There are several compendiums of tests available that describe their characteristics (e.g., Robinson, Shaver, & Wrightsman, 1999).

Data Extraction Forms/Formats

Frequent reliance on administrative records and documents is a major factor underlying the use of this type of data collection. Whether obtaining information from manual case records or computerized data tapes, the researcher needs to screen the data source for the key variables and record them into the research database.

A data extraction form may be a manual coding sheet for recording information from a paper file folder (e.g., medical chart) or the data collector may use a portable computer to enter information directly into a preformatted research database.

Even when the original source is computerized, the researcher will still likely need to create a data extraction format. The format should identify the relevant variables on the computerized file and include a program to extract the appropriate

information into the research file. In circumstances where there are multiple sources of data (e.g., monthly welfare caseload data tapes), it may be necessary to apply these procedures to multiple data sources, using another program to merge the information into the appropriate format for analysis.

Structured Interview Guides

Whenever a research project requires that the same information items be obtained from multiple individuals, it is desirable for the researcher to create a structured interview guide. The need for structured data collection processes becomes even greater when multiple data collectors are being used (see Fowler & Cosenza, Chapter 12, on standardized survey interviewing). Computer-assisted personal interviewing (CAPI) has become increasingly popular for more structured personal interviewing. With CAPI, interviewers use portable computers rather than paper questionnaires to collect and enter the data. CAPI is particularly useful for large-scale surveys and especially those with complex question patterns.

A structured interview guide may begin with an explanation of the purpose of the interview and then proceed to a set of sequenced inquiries designed to collect information about attitudes, opinions, memories of events, characteristics, and circumstances. The questions may be about the respondents themselves or about activities occurring in their environment (e.g., individual dietary habits, housing history, program activities, world events). The guide itself is typically structured to interact with the individual's responses branching from one area to the next based on the individual's previous answer. There are also instances in which semistructured or even unstructured interviews (or parts of the interview) may be appropriate. These approaches are generally appropriate for the conduct of descriptive, exploratory research in a new area of inquiry or when the construct is difficult to measure in a close-ended, structured format. For example, in collecting data on homeless families' history of residential arrangements, a semistructured residential follow-back tool (New Hampshire-Dartmouth Psychiatric Research Center, 1995; Tsemberis, McHugo, Williams, Hanrahan, & Stefancic, 2006) is commonly used to walk a person through a calendar, keying on dates that will spark the person's memory of where the person may have been living at different points in time. Some people respond better to walking backward in remembering their residential arrangements, others are more comfortable beginning at a selected starting point and progressing to the present time. Flexibility in administration is important, therefore, to obtain complete data from a variety of individuals.

Mail and Telephone Surveys

Mail and telephone surveys are used when the researcher needs to obtain the same information from large numbers of respondents. There are many parallels between these methods and structured in-person interview data collection, with

the key difference being the mode of data collection. In Chapter 16, Lavrakas describes telephone survey methods, including issues of sampling and selection of respondents and supervision of interviewers. Computer-assisted telephone interviewing (CATI), the oldest form of computer-assisted interviewing, allows interviewers to ask questions over the telephone and key the data directly into the computer system. As with CAPI, CATI has a strong advantage in situations where the interview has a complex structure (e.g., complicated skip patterns) and also provides the ability to reconcile data inconsistencies at the point of data collection (e.g., Fowler, 2002). In Chapter 15, Mangione and Van Ness provide more detail on the use of mail surveys.

Web-Based or Online Surveys

Web-based surveys are becoming more popular with the advent of inexpensive software and Web storage space. This approach is excellent when surveying a specific group such as employees of a company or college students. It is typical that these groups will have access to computers and feel comfortable in their use. There are several advantages to this approach. First, the data can be collected very rapidly, clearly more so through mail and phone surveys. Second, there are no data entry costs since the respondent enters his or her own data. Third, the data are almost immediately available to the researcher. With the development of sophisticated software, the survey can be programmed with skipping and branching where questions are given to the respondent based on their previous responses. This ability is also available in CATI and computerized surveys but not in written questionnaires. Finally, the researcher can track the completion rate and respond while the survey is still in the field to increase that rate.

Audio Computer-Assisted Self-Interview

Another approach to automating the data collection process is the use of audio computer-assisted self-interview (ACASI) software. With this approach, people with lower literacy are able to participate in such interviews, since the entire interview and instructions are heard instead of just read. The research participant listens to digitally recorded question items over a headset and, if desired, can simultaneously read the questions on the computer screen. The participant responds by pressing a number key or using a touch sensitive screen.

As Dillman (2006) notes, there are often situations in which we have the need to change data collection modes or mix modes (e.g., enhancing response rates of telephone surveys by contacting individuals by Internet or in person). He cautions that the accuracy of data collection from mixed mode efforts cannot be assumed, due to, for example, unintentional differences in the question stimulus presented to respondents and differences in social desirability. Attention to potential differences in the nature of responses due to data collection mode should be considered in the design stage and checked in analysis.

Resource Planning

Before making final decisions about the specific design to use and the type of data collection procedures to employ, the investigator must take into account the resources available and the limitations of those resources. Resource planning is an integral part of the iterative Stage II planning activities (see Figure 1.2). Resources important to consider are the following:

- *Data:* What are the sources of information needed and how will they be obtained?
- *Time:* How much time is required to conduct the entire research project, including final analyses and reporting?
- *Personnel:* How many researchers are needed and what are their skills?
- *Money:* How much money is needed to implement the research and in what categories?

Data as a Resource

The most important resource for any research project consists of the data needed to answer the research question. As noted, data can be obtained primarily in two ways: from original data collected by the investigator and from existing data. We discuss below the issues associated with primary data collection and the issues involved in the use of secondary data.

Primary Data Collection

There are five major issues that the researcher needs to consider in planning for primary data collection: site selection, authorization, the data collection process, accessibility, and other support needed.

Site Selection. Applied research and basic research differ on several dimensions, as discussed earlier, but probably the most salient difference is in the location of the research. The setting has a clear impact on the research, not only in defining the population studied, but also in the researcher's formulation of the research question, the research design, the measures, and the inferences that can be drawn from the study. The setting can also determine whether there are enough research participants available.

Deciding on the appropriate number and selection of sites is an integral part of the design/data collection decision, and often there is no single correct answer. Is it best to choose "typical" sites, a "range" of sites, "representative" sites, the "best" site, or the "worst" site? There are always more salient variables for site selection than resources for study execution, and no matter what criteria are used, some critics will claim that other more important site characteristics were omitted. For this reason, we recommend that the researcher make decisions regarding site selection in close

coordination with the research client and/or advisory group. In general, it is also better to concentrate on as few sites as are required, rather than stretching the time and management efforts of the research team across too many locations.

There is another major implication connected with site selection. As noted earlier, multilevel designs have implications for the number and type of sites selected. In hierarchical designs, if the research intervention is at the site level (as in the earlier school example), then the investigator needs to have a sufficient number of sites in each experimental condition to maintain enough statistical power to detect a meaningful effect. For example, if a drug prevention program is instituted at the school level, then the number of schools, not classes or students, is what is important. One of the problems of using units lower in the hierarchy, such as classes, is that there may be concern about contamination from one condition to another. In the case where teachers are delivering the intervention and they teach in more than one classroom, then it should be obvious that classroom is not a suitable unit of analysis. Even if there is little or no chance of contamination, the observations still may be correlated and not independent of each other. This correlation, sometimes called the design effect, reduces the statistical power by reducing in effect the number of participants or units. Proper design and analysis requires multiple units, with the implication that enough units have to exist in the environment to do the study. In the case of schools, there may be a sufficient number in a given city. The same may not be true for hospital emergency rooms, public housing units, or mental health centers. Studies with these organizations will typically require the participation of multiple cities. More about designing and analyzing these site-based hierarchical designs can be found in Raudenbush and Bryk (2002) and Graham et al. (2008).

The distinction between "frontstage" and "backstage" made by Goffman (1959) also helps assess the openness of the setting to research. Frontstage activities are available to anyone, whereas backstage entrance is limited. Thus in a trial, the actions that take place in the courtroom constitute frontstage activity, open to anyone who can obtain a seat. Entrance to the judge's chambers is more limited, presence during lawyer-client conferences is even more restricted, and the observation of jury deliberations is not permitted. The researcher needs to assess the openness of the setting before taking the next step—seeking authorization for the research.

Authorization. Even totally open and visible settings usually require some degree of authorization for data collection. Public space may not be as totally available to the researcher as it may seem. For example, it is a good idea to notify authorities if a research team is going to be present in some public setting for an extended period of time. Although the team members' presence may not be illegal and no permission is required for them to conduct observations or interviews, residents of the area may become suspicious and call the police.

If the setting is a closed one, the researcher will be required to obtain the permission of the individuals who control or believe they control access. If there are several sites that are eligible for participation and they are within one organization, then it behooves the researcher to explore the independence of these sites from the parent organization. For example, in doing research in school systems, it might also be advisable to approach a principal to obtain preliminary approval that then can

be presented to central administration for formal approval. Most school systems have written procedures that investigators must follow if they are going to gain access to the schools.

The planner needs to know not only at which level of the organization to negotiate but also which individuals to approach. Again, this will take some intelligence gathering. Personal contacts help a great deal, because authorities are usually more likely to meet and be cooperative with the researcher if he or she is recommended by someone they know and trust. Thus, the investigator should search for some connection to the organization. If the researcher is at a university, then it is possible that someone on the board of trustees is an officer of the organization. If so, contact with the university's development office is advisable. In sum, it is best for the researcher to obtain advance recommendations from credible sources and, hence, to avoid approaching an organization cold.

Permission from a central authority, however, does not necessarily imply cooperation from the sites needed for data collection. Nowhere is this more evident than in state/county working relationships. Often, central approval will be required just for the researcher to approach local sites. However, the investigator should not assume that central approval guarantees cooperation from those lower down on the organization's hierarchy; this belief can lead the investigator to behave in an insensitive manner. Those at the upper levels of an organization tend to believe that they have more power than they actually wield. A wise investigator will put a great deal of effort into obtaining cooperation at the local level, where he or she will find the individuals who feel they control that environment and with whom he or she will be interacting during the data collection phase. A good example is the school superintendent saying that he or she strongly supports the research but in reality, each principal will have to decide to participate.

Some closed organizations have procedures that must be followed before they can issue permission to conduct research in their settings (e.g., prisons and schools). Confidentiality and informed consent are usually significant issues for any organization. Will participants be identified or identifiable? How will the data be protected from unauthorized access? Will competitors learn something about the organization from this research that will put it at a disadvantage? Will individuals in the organization be put in any jeopardy by the project? The researcher needs to resolve such issues before approaching an organization for permission.

Organizations that have experience with research usually have standard procedures for working with researchers. For example, school systems typically have standard forms for researchers to complete and deadlines by which these forms must be submitted. These organizations understand the importance of research and are accustomed to dealing with investigators. In contrast, other organizations may not be familiar with applied research. Most for-profit corporations fall into this category, as do many small nonprofit organizations. In dealing with such groups, the investigator will first have to convince the authorities that research, in general, is a good idea and that their organization will gain something from their participation. In some cases, the researcher may also have to obtain the support of staff within the participating organizations, if they are needed to collect data or to obtain access to research participants. In conducting research on programs for

homeless families, for example, researchers often have to convince program staff that the research will be worthwhile, will not place the families in the position of "guinea pigs," and will treat the families with respect and dignity. Most important, an organization's decision makers must be convinced that the organization will not be taking a significant risk or taking up valuable time in participating in the study. The planner must be prepared to present a strong case for why a nonresearch-oriented organization should want to involve itself in a research project.

Finally, any agreement between the researcher and the organization should be in writing. This may take the form of a letter addressed to the organization's project liaison officer (there may be one) for the research. The letter should describe the procedures that will take place and indicate the dates that the investigator will be on-site. The agreement should be detailed and should include how the organization will cooperate with the research.

The importance of site cooperation cannot be stressed too much. Lack of cooperation or dropping out of the study are some of the major factors that cause studies to fail. It is better to recruit more sites than you think you will need because invariably some will drop out before the study starts, and others will not have the client flow that they assured you that they had. This is discussed more in the next section.

Data Collection Process. The primary purpose of obtaining access to a site is to be able to collect data from or about people. The researcher should not assume that having access ensures that the target study participants will agree to participate in the study. Moreover, the researcher should be skeptical regarding assurances from management concerning others' availability and willingness to participate in a study. In a review of 30 randomized studies in drug abuse, Dennis (1990) found that 54% seriously underestimated the client flow by an average of 37%. Realistic and accurate participant estimates are necessary for the researcher to allocate resources and to ensure sufficient statistical power. Many funding agencies require power analyses as part of submitted grant proposals. These power analyses should be supported by evidence that the number of cases in these analyses are valid estimates. Dillman's (1978, 2000) total design method has been used successfully to improve recruitment rate (Records & Rice, 2006).

A planner can try to avoid shortfalls in the number of cases or subjects needed by conducting a small pilot study. In a pilot study, the researcher can verify client flow, enrollment and attendance data, program or service capacity, and willingness to participate. In cases where potential subjects enter into some program or institution, it will be important to verify the actual subject flow (e.g., number per week). This type of study is often called a pipeline study. In some circumstances, the flow into the program is affected by seasonal issues, contextual factors, organizational changes, and other factors. In addition, program capacity also can change and affect the size of the potential study participant pool. For example, in an evaluation of a newly developed service program for homeless families, initial sample size estimates were derived by program estimates that each of 6 case managers would be working with an average of 15 families at a time for an average of 9 months. Therefore, over an 18-month period, it was expected that there would be approximately 180 families in the participant pool. However, this estimate did not account

for delays in hiring the full set of case managers as well as other times when one or more positions were unfilled, delays in enrolling families, and difficulties in both having full caseloads and moving families out of service in the 9-month time period due to the problems that families faced. Therefore, with the slippage of each part of the equation, the number of potential families for the study (before even considering eligibility criteria and refusal rates) was considerably smaller than initial expectations.

Care must be especially used in defining exactly who is eligible to participate in the study. For example, a pipeline study found that there were more than enough potential participants. However, the participant sample was limited to one child per family. It was not known until the study was underway that 30% of the potential participants had a sibling receiving treatment from the same organization.

Related to the number of participants is the assurance that the research design can be successfully implemented. Randomized designs are especially vulnerable to implementation problems. It is easy to promise that there will be no new taxes, that the check is in the mail, and that a randomized experiment will be conducted—but it is often difficult to deliver on these promises. In an applied setting, the investigator should obtain agreement from authorities in writing that they will cooperate in the conduct of the study. This agreement must be detailed and procedurally oriented and should clearly specify the responsibilities of the researcher and those who control the setting. While a written document may be helpful, it is not a legal contract that can be enforced. The organization leadership can change and with it the permission to conduct the study.

The ability to implement the research depends on the ability of the investigator to carry out the planned data collection procedures. A written plan for data collection is critical to success, but it does not assure effective implementation. A pilot study or walk-through of the procedure is necessary to determine if it is feasible. In this procedure, the investigator needs to consider both accessibility and other support. Written plans agreed to before the start of the study are helpful but not the final word. The researcher needs to monitor the implementation of the research. Studies can be sabotaged by resentful employees. For example, children eligible for services were recruited from a mental health center by the staff person who determined the severity of each case on a 10-point scale. The staff person was instructed that the mild cases, rated 4 or less, or the emergency cases, rated 10, were not eligible for the study. That left us cases rated in the range of 5 to 9, which would supply the needed number of participants. In the first month, much fewer children entered the study than expected. It was discovered that the person answering the phone was rating much fewer cases in the range than needed because she didn't think the study should be done. Once the director of the center talked to her, the situation was resolved.

Accessibility. There are a large number of seemingly unimportant details that can damage a research project, if they are ignored. Will the research participants have the means to travel to the site? Is there sufficient public transportation? If not, will the investigator arrange for transportation? Will families need child care to participate? If the study is going to use an organization's space for data collection, will the investigator need a key? Is there anyone else who may use the space? Who controls

scheduling and room assignments? Has this person been notified? For example, a researcher about to collect posttest data in a classroom should ensure that he or she will not be asked to vacate the space before data collection is completed.

Other Support. Are the lighting and sound sufficient for the study? If the study requires the use of electrical equipment, will there be sufficient electrical outlets? Will the equipment's cords reach the outlets or should the researcher bring extension cords? Do the participants need food or drink? Space is a precious commodity in many institutions; the researcher should never assume that the research project will have sufficient space.

Secondary Data Analysis

The use of existing data, compared with collecting primary data, has the advantage of lower costs and time savings, but it may also entail managing a large amount of flawed and/or inappropriate data. In some cases, these data exist in formats designed for research purposes; for example, there are a number of secondary data sources developed by university consortia or by federal agencies such as the Bureau of the Census. Other kinds of data exist as administrative records (e.g., mental health agency records) that were not designed to answer research questions.

In the planning process, the investigator must establish with some confidence that the records to be used contain the information required for the study. Sampling the records will not only provide the researcher with an indication of their content, it will give an idea of their quality. It is frequently the case that clinical or administrative records are not suitable for research purposes. The planner must also have some confidence in the quality of the records. Are the records complete? Why were the data originally collected? The database may serve some hidden political purpose that could induce systematic distortions. What procedures are used to deal with missing data? Are the same procedures used for all variables or only selected variables? Do the computerized records bear a close resemblance to the original records (if available)? Are some data items periodically updated or purged from the computer file? How were the data collected and entered, and by whom? What quality control and verification checks are used? To assess the quality of the database, the planner should try to interview the data collectors and others experienced with the data, observe the data entry process, and compare written records to the computerized version. Conducting an analysis of administrative records seems easy only if it is not done carefully.

The investigator should not assume that the level of effort needed to process extant data will be small or even moderate. Data sets may be exceedingly complex, with changes occurring in data fields and documentation over time. In many cases, there may be very poor documentation, making interpretation of the data difficult. Moreover, if the researcher is interested in matching cases across existing data sets (as in tracking service used across multiple county databases), he or she will need to ensure that identification fields are available in each data set to match individuals' records. Often, matching alone can take a considerable amount of time and resources.

Finally, once the researcher has judged the administrative records or other database to be of sufficient quality for the study, he or she must then go through the necessary procedures to obtain the data. In addition to determining the procedures for extracting and physically transferring the data, the investigator also must demonstrate how the confidentiality of the records will be protected. For example, school systems may want a formal contractual agreement between the university and the school system before they would release identifiable student achievement data. Knowledge of relevant laws and regulations are important. In this example, the researchers had legitimate right to the identifiable data under federal regulations, namely, the Family Educational Rights and Privacy Act (FERPA) and the Protection of Pupil Rights Amendment (PPRA). While it may seem to be a simple request, it took over a year to obtain the data.

Time as a Resource

Time takes on two important dimensions in the planning of applied research: calendar time and clock time. Calendar time is the total amount of time available for a project, and it varies across projects.

Time and the Research Question

The calendar time allotted for a study should be related to the research questions. Is the phenomenon under study something that lasts for a long period or does it exist only briefly? Does the phenomenon under study occur in cycles? Is the time allocated to data collection sufficient?

Time and Data Collection

The second way in which the researcher needs to consider time is in terms of the actual or real clock time needed to accomplish particular tasks. For example, the event that is being studied might exist infrequently and only for a short period of time; thus, a long period of calendar time might need to be devoted to the project, but only a short period of clock time for data collection. Having established the time estimates, the investigator needs to estimate how long actual data collection will take. In computing this estimate, the researcher should consider how long it will take to recruit study participants and to gain both cooperation and access. The researcher should also attempt to estimate attrition or dropout from the study. If high attrition is predicted, then more recruitment time may be needed for data collection for the study to have sufficient statistical power. Thus, in computing the time needed, the investigator should have an accurate and comprehensive picture of the environment in which the study will be conducted.

Time Budget

In planning to use any resource, the researcher should create a budget that describes how the resource will be allocated. Both calendar and clock time need to be budgeted.

To budget calendar time, the researcher must know the duration of the entire project. In applied research, the duration typically is set at the start of the project, and the investigator then tailors the research to fit the length of time available. There may be little flexibility in total calendar time on some projects. Funded research projects usually operate on a calendar basis; that is, projects are funded for specific periods of time. Investigators must plan what can be accomplished within the time available.

The second time budget a researcher must create concerns clock time. How much actual time will it take to develop a questionnaire or to interview all the participants? It is important for the investigator to decide what units of time (e.g., hours, days, months) will be used in the budget. That is, what is the smallest unit of analysis of the research process that will be useful in calculating how much time it will take to complete the research project? To answer this question, we now turn to the concepts of tasks.

Tasks and Time

To "task out" a research project, the planner must list all the significant activities (tasks) that must be performed to complete the project. The tasks in a project budget serve a purpose similar to that of the expense categories—rent, utilities, food, and so on—used in planning a personal financial budget. When listing all these expense items, one makes implicit decisions concerning the level of refinement that will be used. Major categories (such as utilities) are usually divided into finer subcategories. The degree of refinement in a research project task budget depends on how carefully the investigator needs to manage resources.

To construct a time budget, the investigator first needs to consider the time required to manage the overall process; keep various stakeholders informed as needed either through meetings, monthly reports, update telephone calls, and/or other mechanisms; maintain connections with other members of the team in team meetings, conference calls (especially if the team is in more than one location); and other activities that maintain the integrity of the project over the entire study time frame. Second, the researcher should list all the tasks that must be accomplished during the research project. Typically, these tasks can be grouped into a number of major categories. The first category usually encompasses conceptual development. This includes literature reviews and thinking and talking about the problem to be investigated. Time needs to be allocated also for consulting with experts in areas where investigators need additional advice. The literature reviews could be categorized into a number of steps, ranging from conducting computerized searches to writing a summary of the findings.

The second phase found in most projects is instrument development and refinement. Regardless of whether the investigator plans to do intensive face-to-face interviewing, self-administered questionnaires, or observations, he or she needs to allocate time to search for, adapt, or develop relevant instruments used to collect data. The researcher also needs to allocate time for pilot testing of the instruments. Pilot testing should never be left out of any project. Typically, a pilot test will reveal "new" flaws that were not noted by members of the research team in previous applications of the instrument. If multiple data collection sites are involved, it is often

important to pilot the procedures in all the sites or at least a sample that represents the range of sites involved. If the data collection approach involves extracting information from administrative records, the researcher should pilot test the training planned for data extractors as well as the data coding process. Checks should be included for accuracy and consistency across coders.

When external validity or generalizability is a major concern, the researcher will need to take special care in planning the construction of the sample. The sampling procedure describes the potential subjects and how they will be selected to participate in the study. This procedure may be very complex, depending on the type of sampling plan adopted.

The next phase of research is usually the data collection. The investigator needs to determine how long it will take to gain access to the records as well as how long it will take to extract the data from the records. It is important that the researcher not only ascertains how long it will take to collect the data from the records but also discovers whether information assumed to be found in those records is there. If the researcher is planning to conduct a survey, the procedure for estimating the length of time needed for this process could be extensive. Fowler and Cosenza (Chapter 12) describe the steps involved in conducting a survey. These include developing the instrument, recruiting and training interviewers, sampling, and the actual collection of the data. Telephone interviews require some special techniques that are described in detail by Lavrakas (Chapter 16). Time must also be allotted to obtain institutional review board's approval of the project if it involves human subjects. If a project is involved in federal data collection, review may also be required by the Office of Management and Budget (OMB), which, depending on the size of the project, can involve a considerable effort to develop the OMB review package and up to 4 months for the review to occur.

The next phase usually associated with any research project is data analysis. Whether the investigator is using qualitative or quantitative methods, time must be allocated for the analysis of data. Analysis includes not only statistical testing using a computer but also the preparation of the data for computer analysis. Steps in this process include "cleaning" the data (i.e., making certain that the responses are readable and unambiguous for data entry personnel), physically entering the data, and checking for the internal consistency of the data (Smith, Breda, Simmons, Vides de Andrade, & Bickman, 2008). Once the data are clean, the first step in quantitative analysis is the production of descriptive statistics such as frequencies, means, standard deviations, and measures of skewness. More complex studies may require researchers to conduct inferential statistical tests. As part of the design, a clear and comprehensive analysis plan should be developed that includes the steps for cleaning the data as well as the sequence of analyses that will take place, including analyses that may be needed to test for possible artifacts (e.g., attrition).

Finally, time needs to be allocated for communicating the results. An applied research project almost always requires a final report, usually a lengthy, detailed analysis as well as one or more verbal briefings. Within the report itself, the researcher should take the time needed to communicate the data to the audience at the right level. In particular, visual displays can often communicate even the most complex findings in a more straightforward manner than prose.

Because most people will not read the entire report, it is critical that the researcher include a two- or three-page executive summary that succinctly and clearly summarizes the main findings. The executive summary should focus on the findings, presenting them as the highlights of the study. No matter how much effort and innovation went into data collection, these procedures are of interest primarily to other researchers, not to typical sponsors of applied research or other stakeholders. The best the researcher can hope to accomplish with these latter audiences is to educate them about the limitations of the findings based on the specific methods used.

The investigator should allocate time not only for producing a report but also for verbally communicating study findings to sponsors and perhaps to other key audiences. Moreover, if the investigator desires to have the results of the study used, it is likely that he or she needs to allocate time to work with the sponsor and other organizations in interpreting and applying the findings of the study. This last utilization-oriented perspective is often not included by researchers planning their time budgets.

Time Estimates

Once the researcher has described all the tasks and subtasks, the next part of the planning process is to estimate how long it will take to complete each task. One way to approach this problem is to reduce each task to its smallest unit. For example, in the data collection phase, an estimate of the total amount of interviewing time is needed. The simplest way to estimate this total is to calculate how long each interview should take. Pilot data are critical for helping the researcher to develop accurate estimates.

The clock-time budget indicates only how long it will take to complete each task. What this budget does not tell the researcher is the sequencing and the real calendar time needed for conducting the research. Calendar time can be calculated from clock-time estimates, but the investigator needs to make certain other assumptions as well. For example, calendar conflicts need to be considered in the budgeting. Schools, for example, have a restricted window of time for data collection, usually avoiding the month around school entry and any testing. As another example, some service programs have almost no time for researchers around the busy holiday times, making December a difficult time to schedule any onsite data collection.

Another set of assumptions is based on the time needed for data collection. For example, if the study uses interviewers to collect data and 200 hours of interviewing time are required, the length of calendar time needed will depend on several factors. Most clearly, the number of interviewers will be a critical factor. One interviewer will take a minimum of 200 hours to complete this task, whereas 200 interviewers could theoretically do it in 1 hour. However, the larger number of interviewers may create a need for other mechanisms to be put into place (e.g., interviewer supervision and monitoring) as well as create concerns regarding the quality of the data. Thus the researcher needs to specify the staffing levels and research team skills required for the project. This is the next kind of budget that needs to be developed.

Each research project has unique characteristics that make it difficult to generalize from one project to another. Estimating time and expenses is an inexact art. In

most cases the research underestimates the time and cost of a project. Unexpected events that disrupt the research should be expected. Since research budgets typically do not permit funds to be reserved for unforeseen events the planner is advised to build in some aspect of the project that could be sacrificed without affecting the central features of the research. The time and funds allocated to that task can usually be used to provide the additional support needed to complete the research.

Personnel as a Resource

Skills Budget

Once the investigator has described the tasks that need to be accomplished, the next step is to decide what kinds of people are needed to carry out those tasks. What characteristics are needed for a trained observer or an interviewer? What are the requirements for a supervisor? To answer these questions, the investigator should complete a skills matrix that describes the requisite skills needed for the tasks and attaches names or positions of the research team to each cluster of skills. Typically, a single individual does not possess all the requisite skills, so a team will need to be developed for the research project. As noted earlier, in addition to specific research tasks, the investigator needs to consider management of the project. This function should be allocated to every research project. Someone will have to manage the various parts of the project to make sure that they are working together and that the schedule is being met.

Person Loading

Once the tasks are specified and the amount of time required to complete each task is estimated, the investigator must assign these tasks to individuals. The assignment plan is described by a person-loading table that shows how much time each person is supposed to work on each task.

At some point in the planning process, the researcher needs to return to real, or calendar, time, because the project will be conducted under real-time constraints. Thus the tasking chart, or Gantt chart, needs to be superimposed on a calendar. This chart simply shows the tasks on the left-hand side and the months of the study period at the top. Bars show the length of calendar time allocated for the completion of specific subtasks. The Gantt chart shows not only how long each task takes, but also the approximate relationship in calendar time between tasks. Although inexact, this chart can show the precedence of research tasks and the extent to which some tasks will overlap and require greater staff time. One of the key relationships and assumptions made in producing a plan is that no individual will work more than 40 hours a week. Thus the person-loading chart needs to be checked against the Gantt chart to make sure that tasks can be completed by those individuals assigned to them within the periods specified in the Gantt chart. Very reasonably priced computer programs are available to help the planner do these calculations and draw the appropriate charts.

Financial Resources

Usually, the biggest part of any research project's financial budget is consumed by personnel—research staff. Social science research, especially applied social science, is very labor-intensive. Moreover, the labor of some individuals can be very costly. To produce a budget based on predicted costs, the investigator needs to follow a few simple steps.

Based on the person-loading chart, the investigator can compute total personnel costs for the project by multiplying the hours allocated to various individuals by their hourly costs.

The investigator should compute personnel costs for each task. In addition, if the project will take place over a period of years, the planner will need to provide for salary increases in the estimates. Hourly cost typically includes salary and fringe benefits and may also include facilities and administration (F&A) or overhead costs. (In some instances, personnel costs need to be calculated by some other time dimensions, such as daily or yearly rates; similarly, project costs may need to be categorized by month or some time frame other than year.)

After the budget has been calculated, the investigator may be faced with a total cost that is not reasonable for the project, either because the sponsor does not have those funds available or because the bidding for the project is very competitive. If this occurs, the investigator has several alternatives. Possible alternatives are to eliminate some tasks, reduce the scope of others, and/or shift the time from more expensive to less expensive staff for certain tasks where it is reasonable. The investigator needs to use ingenuity to try to devise not only a valid, reliable, and sensitive project, but one that is efficient as well. For example, in some cases this may mean recommending streamlining data collection or streamlining the reporting requirements.

The financial budget, as well as the time budget, should force the investigator to realize the trade-offs that are involved in applied research. Should the investigator use a longer instrument, at a higher cost, or collect fewer data from more subjects? Should the subscales on an instrument be longer, and thus more reliable, or should more domains be covered, with each domain composed of fewer items and thus less reliable? Should emphasis be placed on representative sampling as opposed to a purposive sampling procedure? Should the researcher use multiple data collection techniques, such as observation and interviewing, or should the research plan include only one technique, with more data collected by that procedure? These and other such questions are ones that all research planners face. However, when a researcher is under strict time and cost limitations, the salience of these alternatives is very high.

Making Trade-Offs and Testing Feasibility

Before making a firm go/no-go decision, it is worthwhile for the researcher to take the time to assess the strengths and weaknesses of the proposed approach and decide whether it is logistically feasible. This section returns to a discussion of the iterative process that researchers typically use as they assess and refine the initial design approach. Two major activities take place: (a) identifying and deciding on

design trade-offs and (b) testing the feasibility of the proposed design. These activities almost always occur simultaneously. The results may require the researcher to reconsider the potential design approach or even to return to the client to renegotiate the study questions.

Making Design Trade-Offs

Examples of areas where design trade-offs often occur include external generalizability of study results, conclusiveness of findings, precision of estimates, and comprehensiveness of measurement. Trade-offs are often forced by external limitations in dollar and staff resources, staff skills, time, and the quality of available data.

Generalizability

Generalizability refers to the extent to which research findings can be credibly applied to a wider setting than the research setting. For example, if one wants to describe the methods used in vocational computer training programs, one might decide to study a local high school, an entire community (including both high schools and vocational education agencies and institutions), or schools across the nation. These choices vary widely with respect to the resources required and the effort that must be devoted to constructing sampling frames. The trade-offs here are ones of both resources and time. Local information can be obtained much more inexpensively and quickly than can information about a larger area; however, one will not know whether the results obtained are representative of the methods used in other high schools or used nationally.

Generalizability can also involve time dimensions, as well as geographic and population dimensions. Moreover, generalizability decisions need to have a clear understanding of the generalizability boundaries at the initiation of the study.

Conclusiveness of Findings

One of the key questions the researcher must address is how conclusive the study must be. Research can be categorized as to whether it is exploratory or confirmatory in nature. An exploratory study might seek only to identify the dimensions of a problem—for example, the types of drug abuse commonly found in a high school population. More is demanded from a confirmatory study. In this case, the researcher and client have a hypothesis to test—for example, among high school students use of marijuana is twice as likely as abuse of cocaine or heroin. In this example, it would be necessary to measure with confidence the rates of drug abuse for a variety of drugs and to test the observed differences in rate of use.

Precision of Estimates

In choosing design approaches, it is essential that the researcher have an idea of how small a difference or effect it is important to be able to detect for an outcome

evaluation or how precise a sample to draw for a survey. This decision drives the choice of sample sizes and sensitivity of instrumentation, and thus affects the resources that must be allocated to the study.

Sampling error in survey research poses a similar issue. The more precise the estimate required, the greater the amount of resources needed to conduct a survey. If a political candidate feels that he or she will win by a landslide, then fewer resources are required to conduct a political poll than if the race is going to be close and the candidate requires more precision or certainty concerning the outcome as predicted by a survey.

Comprehensiveness of Measurement

The last area of choice involves the comprehensiveness of measurement used in the study. It is usually desirable to use multiple methods or multiple measures in a study (especially in qualitative studies, as noted earlier) for this allows the researcher to look for consistency in results, thereby increasing confidence in findings. However, multiple measures and methods can sometimes be very expensive and potentially prohibitive. Thus researchers frequently make trade-offs between resources and comprehensiveness in designing measurement and data collection approaches.

Choosing the most appropriate strategy involves making trade-offs between the level of detail that can be obtained and the resources available. Calendar time to execute the study also may be relevant. Within the measurement area, the researcher often will have to make a decision about breadth of measurement versus depth of measurement. Here the choice is whether to cover a larger number of constructs, each with a brief instrument, or to study fewer constructs with longer and usually more sensitive instrumentation. Some trade-off between comprehensiveness (breadth) and depth is almost always made in research. Thus, within fixed resources, a decision to increase external validity by broadening the sample frame may require a reduction in resources in other aspects of the design. The researcher needs to consider which aspects of the research process require the most resources, often in consultation with the research sponsor or other possible users of the study findings.

Feasibility Testing of the Research Design/Plan

Once researchers have tentatively selected a research design, they must determine whether the design is feasible. Areas to be tested for feasibility include the assessment of any secondary data, pilot tests of data collection procedures and instruments, and pilot tests of the design itself (e.g., construction of sampling frames, data collection procedures, and other study procedures). Additionally, efforts may be needed to explore the likelihood of potential confounding factors—that is, whether external events are likely to distort study results or whether the study procedures themselves may create unintended effects. The process of feasibility testing may take as little as a few hours or may involve a trial run of all study procedures in a real-world setting and could last several weeks or months.

The premise of feasibility testing is that, although sometimes time-consuming, it can greatly improve the likelihood of success or, alternatively, can prevent

resources from being wasted on research that has no chance of answering the posed questions. A no-go decision does not represent a failure on the part of the researcher but rather an opportunity to improve on the design or research procedures, and it ultimately results in better research and hopefully better research utilization. A go decision reinforces the confidence of the researcher and others in the utility of expending resources to conduct the study.

Once the researcher has appropriately balanced any design trade-offs and determined the feasibility of the research plan, he or she should hold final discussions with the research client to confirm the proposed approach. If the client's agreement is obtained, the research planning phase is complete. If agreement is not forthcoming, the process may start again, with a change in research scope (questions) or methods.

Conclusion

The key to conducting a sound applied research study is planning. In this chapter, we have described several steps that can be taken in the planning stage to bolster a study and increase its potential for successful implementation. We hope that these steps will help you to conduct applied research that is credible, feasible, and useful.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bickman, L. (1985). Randomized experiments in education: Implementations lessons. In R. F. Boruch (Ed.), *Randomized field experiments* (pp. 39–53). San Francisco: Jossey-Bass.

Bickman, L. (1987). The functions of program theory. In L. Bickman (Ed.), *Using program theory in evaluation* (pp. 5–18). San Francisco: Jossey-Bass.

Bickman, L. (1989). Barriers to the use of program theory: The theory-driven perspective. *Evaluation and Program Planning, 12,* 387–390.

Bickman, L. (Ed.). (1990). *Advances in program theory.* San Francisco: Jossey-Bass.

Bickman, L., & Rog, D. J. (1986). Stakeholder assessment in early intervention projects. In L. Bickman & D. Weatherford (Eds.), *Evaluating early childhood intervention programs.* Austin, TX: PRO-ED.

CATS Consortium. (2007). Implementing CBT for traumatized children and adolescents after September 11th: Lessons learned from the Child and Adolescent Trauma Treatments and Services (CATS) project. *Journal of Clinical Child & Adolescent Psychology, 36,* 581–592.

Chen, H. (1990). *Theory-driven evaluations.* Newbury Park, CA: Sage.

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis, 24,* 175–199.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., et al. (1980). *Toward reform of program evaluation.* San Francisco: Jossey-Bass.

Dennis, M. L. (1990). Assessing the validity of randomized field experiments: An example from drug treatment research. *Evaluation Review, 14,* 347–373.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method.* New York: Wiley-Interscience.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method.* New York: Wiley.

Dillman, D. A. (2006). Why choice of survey mode makes a difference. *Public Health Reports, 191,* 11–13.

Eid, M., & Diener, E. (Eds.). (2006). *Handbook of multimethod measurement in psychology.* Washington, DC: American Psychological Association.

Felce, D., & Emerson, E. (2000). Observational methods in assessment of quality of life. In T. Thompson, D. Felce, & F. J. Symons (Eds.), *Behavioral observation: Technology and applications in developmental disabilities* (pp. 159–174). Baltimore: Paul Brookes.

Flegal, K. M., Carroll, M. D., Ogden, C. L., & Johnson, C. L. (2002). Prevalence and trends in obesity among US adults, 1999–2000. *Journal of the American Medical Association, 288,* 1723–1727.

Foster, E. M. (2003). Propensity score matching: An illustrative analysis of dose response. *Medical Care, 41,* 1183–1192.

Fowler, F. J., Jr. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Frechtling, J. A. (2007). *Logic modeling in program evaluation.* San Francisco: Jossey-Bass.

Goffman, E. (1959). *The presentation of self in everyday life.* Garden City, NY: Doubleday.

Gorard, S. (2002). The role of secondary data in combining methodological approaches. *Educational Review, 54,* 231–237.

Graham, S. E., Singer, J. D., & Willett, J. B. (2008). An introduction to the multilevel modeling of change. In P. Alasuutari, L. Bickman, & J. Brannen (Eds.), *The SAGE handbook of social research methods* (pp. 869–899). London: Sage.

Henry, G. T. (1990). *Practical sampling.* Newbury Park, CA: Sage.

Hofferth, S. L. (2005). Secondary data analysis in family research. *Journal of Marriage and Family, 67,* 891–907.

Kim, M. T., & Hill, M. N. (2003). Validity of self-report of illicit drug use in young hypertensive urban African American males. *Addictive Behaviors, 28,* 795–802.

Macias, C., Hargreaves, W., Bickman, L., Fisher, W., & Aronson, E. (2005). Impact of referral source and study applicants' preference in random assignment on research enrollment, service engagement, and evaluative outcomes. *American Journal of Psychiatry, 162,* 781–87.

McLaughlin, J. A., & Jordan, G. B. (2004). Using logic models. In H. P. Hatry, J. S. Wholey, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (2nd ed., pp. 7–32). San Francisco: Jossey-Bass.

New Hampshire-Dartmouth Psychiatric Research Center. (1995). *Residential follow-back calendar.* Lebanon, NH: Dartmouth Medical School.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Records, K., & Rice, M. (2006). Enhancing participant recruitment in studies of sensitive topics. *Journal of the American Psychiatric Nurses Association, 12,* 28–36.

Riccio, J. A., & Bloom, H. (2002). Extending the reach of randomized social experiments: New directions in evaluations of American welfare-to-work and employment initiatives. *Journal of the Royal Statistical Society: Series A* (Statistics in Society), *165,* 13–30.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (Eds.). (1999). *Measures of political attitudes.* San Diego, CA: Academic Press.

Rog, D. J. (1985). *A methodological analysis of evaluability assessment.* PhD dissertation, Vanderbilt University, Nashville, TN.

Rog, D. J. (1994). Expanding the boundaries of evaluation: Strategies for refining and evaluating ill-defined interventions. In S. L. Friedman & H. C. Haywood (Eds.),

Developmental follow-up: Concepts, genres, domains, and methods (pp. 139–154). New York: Academic Press.

Rog, D. J., & Huebner, R. (1992). Using research and theory in developing innovative programs for homeless individuals. In H. Chen & P. H. Rossi (Eds.), *Using theory to improve program and policy evaluations* (pp. 129–144). Westport, CT: Greenwood Press.

Rog, D. J., & Knickman, J. (2004). Strategies for comprehensive initiatives. In M. Braverman, N. Constantine, & J. Slater (Eds.), *Foundations and evaluations: Contexts and practices for effective philanthropy* (pp. 223–235). San Francisco: Jossey-Bass.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of propensity score in observational studies of causal effects. *Biometrica, 70,* 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79,* 516–524.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127,* 757–763.

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

Smith, C. M., Breda, C. B., Simmons, T. M., Vides de Andrade, A. R., & Bickman, L. (2008). Data preparation and data standards: The devil is in the details. In A. R. Stiffman (Ed.), *The nitty gritty of managing field research.* New York: Oxford University Press.

Stewart, D. W., & Kamins, M. A. (1993). *Secondary research: Information sources and methods* (2nd ed). Newbury Park, CA: Sage.

Tsemberis, S., McHugo, G., Williams, V., Hanrahan, P., & Stefancic, A. (2006). Measuring homelessness and residential stability: The residential time-line follow-back inventory. *Journal of Community Psychology, 35,* 29–42.

Wholey, J. S. (2004). Evaluability assessment. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of Practical Program Evaluation* (2nd ed., pp. 33–61). San Francisco: Jossey-Bass.