9

MIXED METHODS DESIGN: AN ALTERNATIVE APPROACH

◆ Joseph A. Maxwell
Diane M. Loomis

The explicit use of both quantitative and qualitative methods in a single study, a combination commonly known as mixed methods research, has become widespread in many of the social sciences and applied disciplines during the past 25 years. Tashakkori and Teddlie (1998, p. 14) dated the explicit emergence of mixed methods research to the 1960s, with this approach becoming common by the 1980s with the waning of the "paradigm wars." They also identified a subsequent integration of additional aspects of the qualitative and quantitative approaches—not just methods—beginning during the 1990s, which they called "mixed model" studies (p. 16). Such aspects include epistemological assumptions, types of investigation and research design, and analysis and inference strategies.

However, the practice of mixed methods (and mixed models) research has a much longer history than the explicit discussion of the topic. In natural sciences such as ethology and animal behavior, evolutionary biology, paleontology, and geology, the integration of goals and methods that typically would be considered qualitative (naturalistic settings, inductive approaches, detailed description, attention to context, and the intensive investigation of single cases) with those that are generally seen as quantitative (experimental manipulation; control of extraneous variables; formal hypothesis testing;

theory verification; and quantitative sampling, measurement, and analysis) has been common for more than a century. In addition, many classic works in the social sciences employed both qualitative and quantitative techniques and approaches without deliberately drawing attention to this. (Several of these classic studies are analyzed in detail later in the chapter.) From this broader perspective, mixed methods research is a long-standing (although sometimes controversial) practice rather than a recent development.

Indeed, a case could be made that mixed methods research was *more* common in earlier times, when methods were less specialized and compartmentalized and the paradigm wars were less heated. Staw (1992) observed, "When the field of organizational behavior was beginning in the 1950s, there was less of an orthodoxy in method. People observed, participated, counted, and cross-tabulated. There was ready admission that each methodology was flawed" (p. 136). And Rabinowitz and Weseen (2001) argued,

> There was a time in psychology when qualitative and quantitative methods were more easily combined that they are today. Famous experimental social psychologists such as Solomon Asch, Stanley Milgram, and Leon Festinger combined both approaches in some of their most famous works, although the qualitative aspects of the pieces tend to get lost in textbook accounts of their work, as well as in the minds of many instructors and researchers. (pp. 15-16)

This widespread but relatively implicit use of methods, approaches, and concepts from both the qualitative and quantitative paradigms makes it important, in understanding mixed methods design, to investigate the actual conduct of the study (insofar as this can be determined from the publications resulting from the research) rather than depending only on the authors' own characterization of what they did. Kaplan (1964) coined the terms "logic-in-use" and "reconstructed logic" to describe this difference (p. 8). This issue is magnified when mixed model research is considered because aspects of the study other than methods are often less explicitly identified. It is thus important to pay particular attention to the logic-in-use of mixed methods studies in attempting to understand how qualitative and quantitative methods and approaches can be integrated.

In this chapter, we address this issue by presenting an alternative to the usual ways of thinking about mixed methods design. There are two points on which our position differs from most other approaches to mixed methods studies. First, our concept of "design" is different from that employed in most approaches to designing mixed methods studies. The authors of the latter works have typically taken a typological view of research design, presenting a taxonomy of ways to combine qualitative and quantitative methods. In this handbook, for example, the chapters on design by Morse (Chapter 7) and Creswell, Plano Clark, Gutmann, and Hanson (Chapter 8) both focus on the different *types* of mixed methods research, delineating the dimensions on which such studies can vary and identifying the possible and actual combinations of qualitative and quantitative methods.

We approach the issue of design from a fundamentally different perspective. We see the design of a study as consisting of the different components of a research study (including purposes, conceptual framework, research questions, and validity strategies, in addition to "methods" in a strict sense) and the ways in which these components are integrated with, and

mutually influence, one another. We present what Maxwell (1996) called an "interactive" model for research design and apply this model to mixed methods research, showing how the different components of actual mixed methods studies are integrated. The model is termed *interactive* (*systemic* would also be appropriate) because the components are connected in a network or web rather than a linear or cyclic sequence.

The second way in which our approach is distinctive is that we base our approach to mixed methods research on a conceptual analysis of the fundamental differences between qualitative and quantitative research (Maxwell, 1998; Maxwell & Mohr, 1999; Mohr, 1982, 1995, 1996). This analysis employs a distinction between two approaches to explanation, which we call *variance theory* and *process theory*. The use of this distinction leads to somewhat different definitions of these two types of research from those found in most other works, and thus it leads to a somewhat different idea of what mixed methods research consists of.

Our purpose in this chapter is to provide some tools for analyzing such studies and for developing mixed methods designs. We begin by presenting the contrast between prevalent typological views of design and an interactive approach to research design. We develop the latter approach in detail, explaining the components of the interactive model and the systemic relationships among these components. We then turn to the nature of the qualitative-quantitative distinction, presenting an analysis of this distinction that is grounded in the contrast between two fundamentally different ways of thinking about explanation. This leads to a discussion of paradigms and of whether qualitative research and quantitative research constitute distinct or incompatible paradigms. These two analyses are then combined in a presentation of the ways in which qualitative and quantitative approaches to each of the design components differ and of some of the sources of complementarity that these differences generate. Finally, we apply this approach to a variety of actual studies that combine qualitative and quantitative strategies and methods, providing an in-depth analysis of how the designs of these studies actually functioned and the strengths and limitations of the designs.

In proposing this alternative approach, we are not taking a polemical or adversarial stance toward other approaches to mixed methods design. We see our approach as complementary to others and as providing some tools and insights that other approaches might not as clearly provide. The complementarity that we see between different approaches to design is similar to the complementarity that we advocate in mixed methods research, which Greene and Caracelli (1997) called "dialectical" (p. 8), and we believe that combining typological and systemic strategies for analyzing and creating research designs will be more productive than either used alone.

◆ *Existing Approaches to Mixed Methods Design*

We stated previously that existing approaches to mixed methods design have been primarily typological. This is not to claim that issues other than typology have been ignored. We believe that these issues have generally been framed within an overall typological approach and that the *analysis* of mixed methods studies has focused on the classification of these studies in terms of a typology of mixed methods designs. For example, Caracelli and Greene (1997) identified two basic types

of mixed methods designs, which they called "component" and "integrated" designs. Component designs are ones in which "the methods are implemented as discrete aspects of the overall inquiry and remain distinct throughout the inquiry," while integrated designs involve "a greater integration of the different method types" (pp. 22-23). Within these broad categories, they described seven subtypes, based largely on the purposes for combining methods. Patton (1990) presented a different typology, based on qualitative or quantitative approaches to three key stages of a study (design, data, and analysis); he used this to generate four possible mixed designs, involving a choice of qualitative or quantitative methods at each stage (not all sequences were viewed as possible by Patton). Tashakkori and Teddlie (1998) built on Patton's approach to create a much more elaborate typology. They distinguished mixed methods designs (combining methods alone) from mixed model designs (combining qualitative and quantitative approaches to all phases of the research process) and created an elaborate set of subtypes within these.

Not all work on mixed methods design has been typological. For example, Bryman (1988) focused on identifying the purposes for combining qualitative and quantitative methods, and Brewer and Hunter (1989) took a similar approach, organizing their discussion in terms of the different stages of the research. Creswell (1994) presented three models for mixed methods research but then related these models to each of his "design phases," which correspond roughly to the different sections of a research proposal.

Typologies are unquestionably valuable. They help a researcher to make sense of the diversity of mixed methods studies and to make some broad decisions about how to proceed in designing such a study. In particular, distinctions based on the sequence or order in which approaches are combined, the relative dominance or emphasis of the different approaches, whether the approaches are relatively self-contained or integrated, and the different purposes for combining methods are particularly important in understanding mixed methods design.

However, typological approaches also have their limitations. First, the actual diversity in mixed methods studies is far greater than any typology can adequately encompass; this point was emphasized by Caracelli and Greene (1997) as well as Tashakkori and Teddlie (1998, pp. 34-36, 42). In particular, the recognition of multiple paradigms (e.g., positivist, realist, constructivist, critical, postmodern) rather than only two, the diversity in the aspects of quantitative and qualitative approaches that can be employed, the wide range of purposes for using mixed methods, and differences in the setting where the study is done and the consequences of this for the design all make the actual analysis of a mixed methods design far more complicated than simply fitting it into a taxonomic framework.

Second, most typologies leave out what we feel are important components of design, including the purposes of the research, the conceptual framework used, and the strategies for addressing validity issues. All of these components are incorporated into the interactive design model presented next. Typologies also tend to be linear in their conception of design, seeing the components as "phases" of the design rather than as interacting parts of a complex whole.

Third, typologies by themselves generally do little to clarify the actual functioning and interrelationship of the qualitative and quantitative parts of a design; the

typology presented by Caracelli and Greene (1997) is an exception to this criticism because that typology is based partly on the purposes for which a mixed approach is used. Similarly, Pawson and Tilley (1997, p. 154) argued that a pragmatic pluralism in combining methods leads to no new thinking and does not clarify *how* to integrate approaches or when to stop.

◆ *An Interactive Model of Design*

We believe that an interactive approach to research design can help to address these problems. Rather than seeing "design" as a choice from a fixed set of possible arrangements or sequences in the research process, such approaches (e.g., Grady & Wallston, 1988; Martin, 1982; Maxwell, 1996) treat the design of a study as consisting of the actual components of a study and the ways in which these components connect with and influence one another. This approach to design is consistent with the conception of design employed in architecture, engineering, art, and virtually every other field besides research methods in which the term is used: "an underlying scheme that governs functioning, developing, or unfolding" and "the arrangement of elements or details in a product or work of art" (Merriam-Webster, 1984). A good design, one in which the components are compatible and work effectively together, promotes efficient and successful functioning; a flawed design leads to poor operation or failure.

The interactive model presented here has two essential properties: the components themselves and the ways in which these are related. There are five components to the model, each of which can be characterized by the issues that it is intended to address (Maxwell, 1996, pp. 4-5):

1. Purposes

What are the goals of this study? What issues is it intended to illuminate, and what practices or outcomes is it intended to influence? Why is the study worth doing? These purposes can be personal, practical, or intellectual; all three kinds of purposes can influence the rest of the research design.

2. Conceptual Framework

What theories and beliefs about the phenomena studied will guide or inform the research? These theories and beliefs may be drawn from the literature, personal experience, preliminary studies, or a variety of other sources. This component of the design contains the *theory* that the researcher has developed, or is developing, about the setting or issues being studied.

3. Research Questions

What specifically does the researcher want to understand by doing this study? What questions will the research attempt to answer?

4. Methods

How will the study actually be conducted? What approaches and techniques will be used to collect and analyze the data, and how do these constitute an integrated strategy? There are four distinct parts of this component of the model: (a) the relationship that the researcher establishes with the participants in the study; (b) the selection of settings, participants, times and places of data collection, and other data sources such as documents (what is

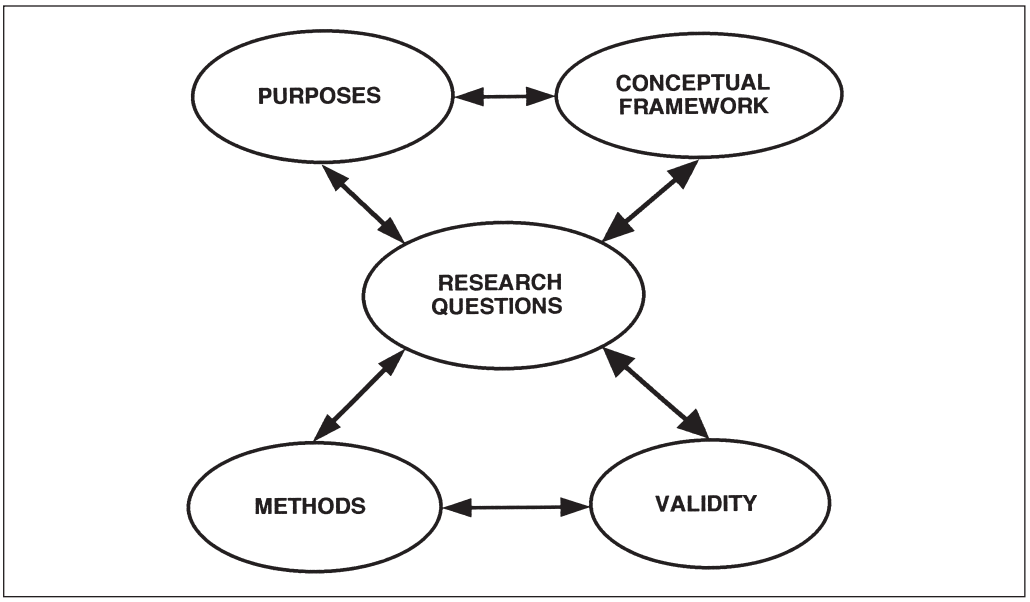

Figure 9.1. An Interactive Model of Research Design

often called "sampling"); (c) data collection methods; and (d) data analysis strategies and techniques.

5. Validity

How might the conclusions of the study be wrong? What plausible alternative explanations and validity threats are there to the potential conclusions of the study, and how will these be addressed?

These components are not radically different from the ones presented in many other discussions of research design (e.g., LeCompte & Preissle, 1993; Miles & Huberman, 1994; Robson, 1993). What is innovative is the way in which the relationships among the components are conceptualized. In this model, the components form an integrated and interacting whole, with each component closely tied to several others rather than being linked in a linear or cyclic sequence. Each of the five components can influence and be influenced by any of the other components. The key relationships among the components are displayed in Figure 9.1. In this diagram, the most important of these relationships are represented as two-way arrows. There is considerable similarity to a systems model of how the parts of a system are organized in a functioning whole.

While all of the five components can influence other components of the design, the research questions play a central role. In contrast to many quantitative models of design, the research questions are not seen as the starting point or guiding component of the design; instead, they function as the hub or heart of the design because they form the component that is most directly linked to the other four. The research questions need to inform, and be responsive to, all of the other components of the design.

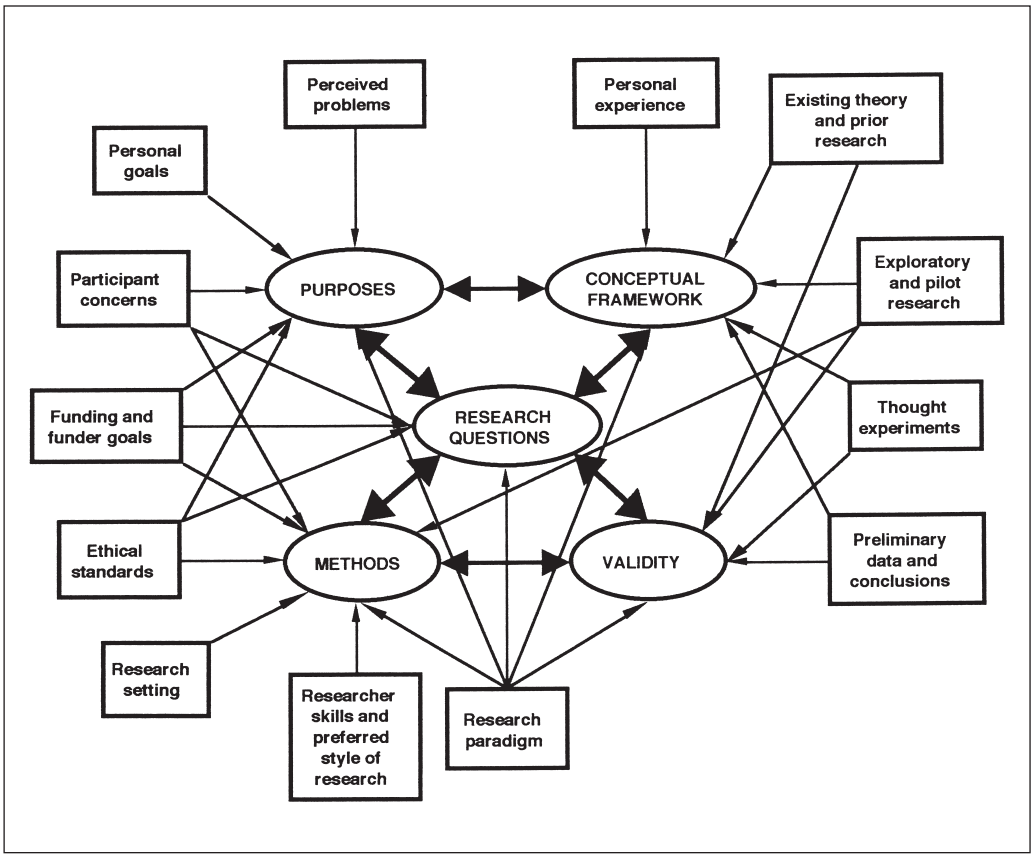

Figure 9.2. Contextual Factors Influencing a Research Design

There are many other factors besides these five components that can influence the design of a study. These include the resources available to the researcher, the researcher's abilities and preferences in methods, perceived intellectual or practical problems, ethical standards, the research setting, the concerns and responses of participants, and the data that are collected. These additional influences are best seen not as part of the design itself but rather either as part of the *environment* within which the research and its design exist or as *products* of the research (Maxwell, 1996, pp. 6-7). Figure 9.2 presents some of the factors in the environment that can influence the design and conduct of a study. The five components of this design model, by contrast, represent issues that are not external to the design of the study but rather are integral parts of it; they represent decisions and actions that must be addressed, either explicitly or implicitly, by the researcher.

One way in which this design model can be useful is as a tool or template for conceptually mapping the design of a study, either as part of the design process or in analyzing the design of a completed study. This involves filling in the boxes or circles for the five components of the model with the *actual* components of a

particular study's design. We apply this technique specifically to a variety of mixed methods studies later in the chapter.

◆ *The Qualitative-Quantitative Distinction*

Because there are so many points of difference between the qualitative and quantitative approaches, there has been considerable variation in how the distinction between the two has been framed. Early work often based this distinction simply on the kind of data employed (textual or numerical). Creswell (1994), by contrast, saw the distinction between inductive and deductive approaches as most important, while Tashakkori and Teddlie (1998, p. 55) distinguished three different stages or dimensions for which the distinction can be made: type of investigation (exploratory or confirmatory), data collection (qualitative or quantitative), and analysis and inference (qualitative or statistical). Guba and Lincoln (1989) made the distinction at a more philosophical level, as a distinction between constructivism and positivism.

In our view, the qualitative-quantitative distinction is grounded in the distinction between two contrasting approaches to explanation, which Mohr (1982) termed variance theory and process theory. Variance theory deals with *variables* and the correlations among them; it is based on an analysis of the contribution of *differences* in values of particular variables to differences in other variables. Variance theory, which ideally involves precise measurement of differences on and correlations between variables, tends to be associated with research that employs extensive prestructuring of the research, probability sampling, quantitative measurement, statistical testing of hypotheses, and experimental or correlational designs. As Mohr noted, "The variance-theory model of explanation in social science has a close affinity to statistics. The archetypal rendering of this idea of causality is the linear or nonlinear regression model" (p. 42).

Process theory, by contrast, deals with *events* and the processes that connect them; it is based on an analysis of the causal *processes* by which some events influence others. Because process explanation deals with specific events and processes, it is much less amenable to quantitative approaches. It lends itself to the in-depth study of one or a few cases or a small sample of individuals and to textual forms of data that retain the contextual connections between events. Weiss (1994) provided a concrete example of this strategy:

> In qualitative interview studies the demonstration of causation rests heavily on the description of a visualizable sequence of events, each event flowing into the next. . . . Quantitative studies support an assertion of causation by showing a correlation between an earlier event and a subsequent event. An analysis of data collected in a large-scale sample survey might, for example, show that there is a correlation between the level of the wife's education and the presence of a companionable marriage. In qualitative studies we would look for a process through which the wife's education or factors associated with her education express themselves in marital interaction. (p. 179)

Mohr (1996) has more recently extended his original distinction between process theory and variance theory to identify two conceptions of causation that he has called "factual causation" and "physical causation." Factual causation is

the traditional mode of reasoning about causes in quantitative research, where the argument for causality is based on the comparison of situations in which the presumed causal factor is present or absent or has different values. Physical causation, by contrast, does not rely on such comparative logic; it is based on a notion of a mechanical connection between a cause and its effect (p. 16). Similar distinctions have been developed by realist philosophers such as Harre (1972; see also Harre & Madden, 1975) and Salmon (1984, 1989, 1998). While factual causation is an appropriate concept for comparative studies with large *N*'s, physical causation is appropriate for case studies or qualitative interview studies that do not involve formal comparisons.

Maxwell and Mohr (1999) used this distinction to identify two aspects of a study that can be productively denoted by the terms *qualitative* and *quantitative:* data and design/analysis.

> We define quantitative data as *categorical* data, with either enumeration or measurement within categories. A conceptual dimension that is itself a category subdivided by measurement, or that is divided into subcategories for enumerative or frequency data, is generally called a "variable," which is a hallmark of the quantitative approach. Qualitative data, in contrast, are typically *textual* in nature, consisting of written or spoken words, but may include video recordings and photographs as well as narrative text. (p. 2)

Categorical data lend themselves to aggregation and comparison, and they are easily quantified. Textual data, on the other hand, lend themselves to investigation of the processes by which two events or characteristics are connected.

> In addition, we propose that quantitative design/analysis is research design and consequent analysis that rely in a variety of ways on the *comparison* of frequencies or measurements across subjects or across categories. Such designs focus on identifying differences between groups or correlations between variables. In contrast, qualitative design/analysis is design and analysis that rely in various ways on the treatment of focal entities as singular wholes in context, with an emphasis on the identification of meaning and process.
>
> With these definitions of secondary terms in mind, the two fundamentally distinct ways of understanding the world can be specified as two distinct combinations of types of data on the one hand with types of design/analysis on the other. Thus, a quantitative way of understanding the world is a way that views the world in terms of categorical data, featuring the comparison of frequencies and measurements across subjects and categories. A qualitative way of understanding is a way that views the world in terms of textual data, featuring the treatment of focal entities as singular wholes in context. (p. 2)

◆ *Paradigmatic Unity and Compatibility*

This analysis of the qualitative-quantitative distinction reframes the nature of the qualitative and quantitative paradigms but does not address the issue of paradigmatic unity or of the compatibility of different paradigms. This unity is often assumed to be a critical issue in combining methods. For example, Patton (1980, p. 110) emphasized the "integrity" of

each approach, and Morse (Chapter 7, this volume) argues,

> When using mixed methods, it is important that methodological congruence be maintained, that is, that all of the assumptions of the major method be adhered to and that components of the method (such as the data collection and analytical strategies) be consistent.

However, the need for such paradigmatic integrity cannot be assumed. McCawley (1982) examined the debate between two positions in linguistics, generative semantics and interpretive semantics, that had generally been seen as unitary paradigms. He showed that both of these approaches in fact consisted of two packages of positions on a large number of issues, with each package corresponding to the views of some prominent members of two communities of linguists. However,

> neither of these communities was completely homogeneous, no member of the community retained exactly the same set of views for very long, . . . and the relationships among the views that were packaged together as "generative semantics" or as "interpretive semantics" were generally far more tenuous than representative members of either community led people (including themselves) to believe. (p. 1)

Pitman and Maxwell (1990) similarly argued that the supposed paradigmatic unity of one area of qualitative research, qualitative evaluation, is largely illusory and that major figures in this field hold widely divergent and conflicting views on many of the fundamental issues regarding the use of qualitative approaches for program evaluation. On the quantitative side, the recent debate over null hypothesis significance testing has revealed how the development of this approach incorporated fundamentally incompatible assumptions from different schools of statistics.

Such a position does not entail that there is *no* relationship among the different aspects of each paradigm, as Reichardt and Cook (1979, p. 18) appeared to argue. We agree with Sayer (1992) that there are "resonances" among the different components of each paradigm that "encourage the clustering of certain philosophical positions, social theories, and techniques" (p. 199). The relationship is simply not a necessary or invariant one. Each paradigm constitutes a "loosely bundled innovation" (Koontz, 1976, cited in Rogers, 1995, p. 178), and researchers often resemble the innovation adopters described by Rogers (1995), "struggling to give their own unique meaning to the innovation as it is applied in their local context" (p. 179).

Thus, we do not believe that there exist uniform, generic qualitative and quantitative research paradigms. Despite the philosophical and methodological resonances among the components of each paradigm, both of these positions include a large number of distinct and separable components, and there is disagreement even within each approach over the nature, use, and implications of some of the different components. The classic qualitative approach includes the study of natural real-life settings, a focus on participants' meanings and context, inductive generation of theory, open-ended data collection, analytical strategies that retain the textual nature of the data, and the frequent use of narrative forms of analysis and presentation. The quantitative approach includes the formulation of prior hypotheses, the use of experimental interventions, a comparison of treatment and control groups,

random sampling or assignment, standardization of instruments and data collection, quantitative data, statistical hypothesis testing, and a focus on causal explanation. Each of these (and other variations too numerous to list) is a separable module with its own requirements and implications rather than an integral and inseparable part of a larger methodological and epistemological whole (Maxwell, Sandlow, & Bashook, 1986; Patton, 1990; Pitman & Maxwell, 1992). While the connections among these components are crucial to the overall coherence of a particular research design (Maxwell, 1996), the possible legitimate ways of putting together these components are multiple rather than singular and, to a substantial extent, need to be discovered empirically rather than logically deduced (Maxwell, 1990).

However, we also agree with Kidder and Fine's (1987) statement, "We share the call for 'synthesis,' but at the same time, we want to preserve the significant differences between the two cultures. Instead of homogenizing research methods and cultures, we would like to see researchers become bicultural" (p. 57). Our view of mixed methods design includes the position that Greene and Caracelli (1997) termed "dialectical" in which differences between the paradigms are viewed as important and cannot be ignored or reconciled. Bernstein (1983), in discussing the differences between Habermas and Derrida, provided a clear statement of what we advocate:

> I do not think there is a theoretical position from which we can reconcile their differences, their otherness to each other—nor do I think we should smooth out their "aversions and attractions." The nasty questions that they raise about each other's "project" need to be relentlessly pursued. One of the primary lessons of "modernity/postmodernity" is a radical skepticism about the possibility of a reconciliation—an *aufhebung,* without gaps, fissures, and ruptures. However, *together,* Habermas/Derrida provide us with a force-field that constitutes the "dynamic, transmutational structure of a complex phenomenon"—the phenomenon I have labeled "modernity/postmodernity." (p. 225)

From this perspective, the "compatibility" of particular qualitative and quantitative methods and approaches becomes a much more complex issue than either paradigmatic or pragmatist approaches usually suggest. Maxwell (1990) claimed that "the theoretical debate about combining methods has prevented us from seeing the different ways in which researchers are *actually* combining methods, and from understanding what works and what doesn't" (p. 507). What we want to do here is use the interactive design model to understand how qualitative and quantitative approaches can productively be combined.

◆ *Qualitative and Quantitative Approaches to the Design Components*

In this section, we identify the distinctive properties of the quantitative and qualitative approaches to each of the components of design described previously: purposes, conceptual framework, research questions, methods, and validity. The ways in which the two paradigms typically frame each of the components are described briefly and are summarized in Table 9.1. A more detailed discussion of each of the components, focusing mainly on qualitative research but also contrasting this with

TABLE 9.1 Possible Quantitative and Qualitative Elements of the Design Components

	Quantitative	*Qualitative*
Purposes	Precise measurement and comparison of variables Establishing relationships between variables Inference from sample to population	Meaning Context Process Discovering unanticipated events, influences, and conditions Understanding single cases Inductive development of theory
Conceptual framework	Variance theories	Process theories
Research questions	Variance questions Truth of proposition Presence or absence Degree or amount Correlation Hypothesis testing Causality (factual)	Process questions How and why Meaning Context (holistic) Hypotheses as part of conceptual framework Causality (physical)
Research methods		
Relationship	Objectivity/reduction of influence (researcher as extraneous variable)	Use of influence as tool for understanding (researcher as part of process)
Sampling	Probability sampling Establishing valid comparisons	Purposeful sampling
Data collection	Prior development of instruments Standardization Measurement/testing—quantitative/categorical	Inductive development of strategies Adapting to particular situation Collection of textual or visual material
Data analysis	Numerical descriptive analysis (statistics, correlation) Estimation of population variables Statistical hypothesis testing Conversion of textual data into numbers or categories	Textual analysis (memos, coding, connecting) Grounded theory Narrative approaches
Validity		
Internal validity	Statistical conclusion validity Construct validity Causal validity (control of extraneous variables)	Descriptive validity Interpretive validity Construct validity Causal validity (identification and assessment of alternative explanations)
Generalizability	External validity (comparability)	Transferability Generalizing to theory

quantitative research, is provided in Maxwell (1996).

PURPOSES

The possible purposes of a study are too numerous and disparate to list, and specific personal and practical purposes are usually not tightly linked to one or the other approach. Intellectual purposes, in contrast, do tend to segregate into qualitative and quantitative categories. Quantitative purposes include precise measurement and comparison of variables, establishing relationships between variables, identifying patterns and regularities that might not be apparent to the people in the settings studied, and making inferences from the sample to some population. Qualitative purposes include understanding the context, process, and meaning for participants in the phenomena studied; discovering unanticipated events, influences, and conditions; inductively developing theory; and understanding a single case.

CONCEPTUAL FRAMEWORK

The conceptual framework for a study consists of the theory (or theories) relevant to the phenomena being studied that inform and influence the research. The key issue for mixed methods studies, then, is the nature of these theories. Are they variance theories, process theories, some combination of these, or theories that do not fit neatly into this dichotomy? A mismatch between the conceptual framework and the research questions or methods used can create serious problems for the research; a variance theory cannot adequately guide and inform a process-oriented investigation and vice versa. Mismatches between the conceptual framework and the purposes or validity strategies are less common but can also be problematic. A mixed methods study is often informed by both variance and process theories, and the main design issue is sorting out specifically how different parts of the conceptual framework are integrated with one another and how they are linked to the other design components.

RESEARCH QUESTIONS

As with conceptual frameworks, research questions can usually be categorized as variance questions or process questions. The research questions in a quantitative study typically are questions about the measurement or analysis of variation—the amount or frequency of some category, the value of some variable, or the relationship between two or more variables. Such questions are usually framed in terms of the values of key variables, and specific hypotheses are often stated. The questions and hypotheses are nearly always specifically formulated (or presented as if they were formulated) in advance of any data collection, and they are frequently framed in "operational" terms, connecting directly to the measurement or data collection strategies. In a qualitative study, by contrast, the research questions typically deal with the verbal description of some event, phenomenon, or process (What is happening here? What are the characteristics of this phenomenon?); its meaning to participants in the setting studied; or the process by which some events or characteristics of the situation influence other events or characteristics. The questions might not be explicitly stated, and when they are, they might include only the broad initial questions with which the study began and not the more focused questions that developed during the research.

METHODS

As described previously, "methods" as a design component include (a) the relationship that the researcher establishes with individuals and groups being studied; (b) the selection of sites, participants, settings, and times of data collection; (c) the methods used for data collection; and (d) the strategies used for data analysis.

Research Relationship. The relationship the researcher has with participants in the study, or with others who control access to these individuals or groups or that may influence the conduct of the study, is a key component of the research design and can have a major impact on the conduct and results of a study. This aspect of design tends to be treated very differently in quantitative and qualitative studies. Quantitative researchers tend to see the research relationship as an extraneous variable—something to be controlled. This can be done either to prevent the relationship from influencing the results or affecting the variables studied or to prevent *variance* in the relationship from introducing confounding variance in the dependent variables (e.g., standardizing survey interview procedures so that differences in procedures, either within or between interviewers, do not create an additional source of variation in the results). Qualitative studies, on the other hand, typically treat the research relationship not as a variable but rather as a process, one that can have important positive as well as negative consequences for the research. The goal is not to create a standardized relationship but rather to create a relationship that maximizes the understanding gained from each participant interviewed or each situation observed. Such a relationship is often much more personal and informal than is the case in quantitative studies.

Sampling. The two main strengths of quantitative sampling (and for experimental research, this can be extended to include assignment of participants to conditions) are to establish valid comparisons and to allow generalization from the sample to the population of interest. Some form of probability sampling (or random assignment) is usually the preferred method; in the absence of this, post hoc strategies (matching or analytical techniques such as analysis of covariance) can be used to increase comparability and generalizability. Qualitative research normally places less emphasis on formal comparisons, and the usual sampling strategy is some form of purposeful sampling. In this approach, participants are selected because they are most likely to provide relevant and valuable information or to allow the researcher to develop or test particular theoretical ideas (in grounded theory research, the latter strategy is called theoretical sampling).

Data Collection. Quantitative data collection is typically preplanned, structured, and designed to ensure comparability of data across participants and sites. The data are normally collected in numerical or categorical form, using instruments or procedures that have been designed and tested to ensure reliability and validity. Qualitative data collection is typically more open-ended, flexible, and inductive, and the data are usually textual descriptions, either written notes or recorded verbal data that are converted to textual form by transcribing (increasingly, visual means such as videotaping are being used).

Data Analysis. Quantitative analysis can be descriptive (assigning numbers or category labels to data or aggregating data on particular variables) or relational (investigating the relationship between two or more variables in the sample). Quantitative analysis can also make inferences to

the population from which the sample was drawn, either estimating the values of population variables or testing hypotheses about the relationship of variables in the population. In addition, textual data can be converted into categorical or numerical form for analysis. Qualitative analysis is more diverse but typically addresses the goals listed under purposes (meaning, context, process, inductive theory development, and in-depth understanding of single cases). The analysis can involve the categorization (coding) of the textual data, but the purpose is quite different from that of quantitative categorization. Rather than being a preliminary step to counting instances of something or aggregating measurements on some variable, the function of qualitative categorization is to collect all of the instances of some type of phenomenon for further qualitative comparison and investigation. The goals of this strategy are to develop an in-depth description of this phenomenon, to identify key themes or properties, and to generate theoretical understanding. The categories are often inductively developed during the analysis rather than systematically formulated prior to the analysis. Both quantitative and qualitative analysis can be either exploratory (on exploratory quantitative data analysis, see Tukey, 1977) or confirmatory, although qualitative researchers usually do not simply test a prior theory without further developing that theory.

VALIDITY

Under validity, we include both causal (internal) validity and generalizability (external validity). Quantitative researchers, most notably Campbell and Stanley (1963) and Cook and Campbell (1979), have developed a detailed typology of validity issues, validity threats, and strategies for addressing these threats. In addition to causal validity and generalizability, Cook and Campbell identified statistical conclusion validity (the validity of inferences from the sample to the population sampled) and construct validity (the validity of the theoretical constructs employed) as distinct issues. There is less agreement on classifying validity issues in qualitative research. Maxwell (1992) distinguished four main categories of validity in qualitative research: descriptive validity (the validity of the descriptions of settings and events), interpretive validity (the validity of statements about the meanings or perspectives held by participants), explanatory (or theoretical) validity (the validity of claims about causal processes and relationships, including construct validity as well as causal validity proper), and generalizability.

Inferences about causality are controversial in qualitative research. Some researchers (e.g., Guba & Lincoln, 1989) deny that causality is an appropriate concept in qualitative research, and this view has been widely accepted. In contrast, Sayer (1992, 2000) and Maxwell (1998), taking a critical realist perspective, argue that causal explanation not only is legitimate in qualitative research but is a particular strength of this approach, although it uses a different strategy from quantitative research, based on a process rather than a variance concept of causality. Construct validity is similar for both approaches, although quantitative research may use quantitative means of assessing the construct validity of instruments. Generalizability is also similar (statistical generalization to the population sampled is included under statistical conclusion validity) and is always a matter of transferring the conclusions of a study to other situations, an inherently judgmental process; Guba and Lincoln (1989) referred to this as "transferability." However, in quantitative research, generalizability is usually seen as a matter of the results of the study

being valid in other settings (replicability). Qualitative researchers, by contrast, tend to "generalize to theory" (Yin, 1984, pp. 39-40)—developing a theory and then applying that theory to other settings that may be dissimilar but that can be illuminated by the theory in question, appropriately modified (Becker, 1990).

In Table 9.1, we have tried to summarize the typical features of both quantitative and qualitative research as these involve the five design components of the interactive model.

ANALYSIS OF SELECTED EXAMPLES OF MIXED METHODS DESIGNS

Uncovering the actual integration of qualitative and quantitative approaches in any particular study is a considerably more complex undertaking than simply classifying the study into a particular category on the basis of a few broad dimensions or characteristics. It requires an understanding of each of the five components of the study's design and of the ways in which each component incorporates quantitative elements, qualitative elements, or both. In addition, as stated previously, it is important to examine the actual conduct of the study rather than simply depending on the author's assertions about the design. This issue is illustrated by Blumstein and Schwartz's (1983) study of American couples, which used both survey questionnaires and open-ended interviews. The authors described the results of their study as based entirely on the statistical analysis of the survey data, while the qualitative data were relegated to providing illustrative instances:

> we use the phrase "we find . . ." in presenting a conclusion based on statistical analysis of data from the questionnaires. . . . The interview data help us interpret our questionnaire findings, but unless we are using one of the parts of the interview that is readily quantifiable, we do not afford them the same degree of trust we grant to information derived from the questionnaires.
>
> The interviews serve another purpose. We use the interview materials to illustrate both majority patterns and important exceptions. (p. 23)

And the authors characterize the chapters in their book that deal with relationship histories, which are based mainly on the interviews, by stating, "In these chapters, which have nothing to do with our analysis of the data but are included only for their illustrative value . . ." (p. 22).

However, this does not explain why Blumstein and Schwartz (1983) conducted in-depth interviews, lasting 2.5 to 4.0 hours, with both partners, separately and together, for 300 couples; transcribed and coded these interviews; and followed up with questionnaires to fill in any gaps. It also seems inconsistent with the fact that, in addition to their extensive use of quotes in the thematically organized sections of the book, they devoted 213 pages, nearly half of the results section of the book, to detailed case studies of 20 couples' relationships. A closer analysis of their account reveals that triangulation of methods was an important feature of the study so as to "see couples from several vantage points" (p. 15) and that the case studies "helped to illuminate some of the ways in which money, sex, and work shape the nature of [the partners'] relationships" (p. 332). It appears that the "reconstructed logic" of the design was heavily influenced by a quantitative ideology of what counts as "results," distorting the study's logic-in-use and the actual contribution of the qualitative component.

The main purpose of this section is to present in-depth analyses of well-documented, complex examples of mixed model research, illustrating the numerous ways in which qualitative and quantitative approaches to each of the design components can be combined. We discuss these studies in terms of Caracelli and Greene's (1997) distinction between "component" and "integrated" mixed methods designs, moving from studies that resemble component designs to those that resemble integrated designs. Component designs are those in which the different methods remain discrete throughout the study and only the results of the methods are combined (p. 22). Integrated designs, by contrast, are those in which there is "a greater integration of the different method types" (p. 23); such designs involve the use not of relatively self-contained qualitative and quantitative methods modules but rather of qualitative and quantitative elements or strategies integrated within a single phase or strand of the research; the elements occur concurrently and in constant interaction with one another rather than as conceptually separate enterprises that are later linked together.

Their distinction is most useful when applied to methods; it is less meaningful when applied to the other components of a research design, and in fact the use of both qualitative and quantitative elements of components other than methods seems to have been treated by Caracelli and Greene (1997) as an "integrated" design almost by definition. In addition, Caracelli and Greene's two types are not categorically distinct; actual studies exhibit a continuum of the amount of integration of methods and also a variety of different strategies for integration. We have nonetheless organized the studies in this order for two reasons. First, doing so provides a clearer organization to this section. Second, it allows us to address the design features of particular *types* of mixed methods studies as well as the specific studies we describe.

A common approach to using both quantitative and qualitative methods is to use them sequentially. Sutton and Rafaeli (1992) provided an unusually detailed and candid account of such a design, a study of the relationship between expressed emotion and sales in convenience stores (see also Sutton & Rafaeli, 1988). They began their research with a well-developed theory of the expression of emotion by employees, based not only on published literature but also on informal querying of waitresses, clerks, and telephone operators. They had numerous ideas for possible empirical studies, but no actual research in progress, when they unexpectedly gained access to a quantitative data set derived from covert observations of employees and from company sales records, with detailed data on numerous control variables. Although one of the authors had considerable experience with qualitative research, this study was originally designed as a purely quantitative multiple regression analysis of this data set.

Sutton and Rafaeli's statistical analysis of this data was intended to achieve two main purposes. First, it would support their theory and further develop their scholarly agenda on expressed emotion. Second, it would advance their careers without involving all the work of collecting their own data. Unfortunately, the analysis flatly contradicted their hypotheses; expressed positive emotions had a consistently *negative* correlation with sales. They tried tinkering with the analysis, but to no avail; they could find no errors, and dozens of runs using different combinations of variables gave the same result. Their validity checks were unable to resolve the contradiction between their

theory and their results. It was clear that they needed to revise their conceptual framework.

Fortunately, a colleague suggested an alternative theory, which came to be called the "Manhattan effect": that in busy stores, employees did not have time and/or were too harassed to express positive emotions. This theory was consistent with their data, and the authors' initial inclination was to simply revise their hypotheses and submit the paper for publication, having learned from experienced colleagues that this was common practice in both the natural and social sciences. There were two reasons why they did not do this. First, it would contradict their previously published theoretical work, potentially impairing their career advancement. Second, they wanted to write a paper that conveyed their actual process, believing that, although it would be harder to publish, it would be a better paper. To do this, however, they needed a clearer theoretical understanding of their findings. This led to the qualitative phase of the study, which consisted of interviews with managers and executives, four case studies, informal observations in stores, and one of the authors working for a day as a store clerk. Sutton and Rafaeli (1992) stated,

> These qualitative data proved to be essential for helping us to refine our revised conceptual perspective. For example, while we had thought about how a crowded store suppresses the display of positive emotion, we had not thought about the ways in which a slow store supports the display of good cheer. During the day that Bob spent working as a clerk, he learned that customers are an important source of entertainment, and that clerks are more friendly during slow times because they are genuinely pleased to see customers and want to encourage customers to engage in conversation. (p. 123)

Their revised and elaborated theory was used to develop a different hypothesis, which was supported by a further quantitative analysis of the original data set.

This research thus involved two cycles of induction and deduction. The first cycle was typical of quantitative research; it began with informal data collection and literature-based theorizing about how the display of positive emotion influences sales, and it ended with the statistical test of a hypothesis derived from this theory. The failure of the study to support the hypothesis forced the authors into a second cycle, beginning with a colleague's suggestion and continuing with a diverse array of qualitative data collection and analysis, which eventually led to the inductive development of a new conceptual framework that emphasized the reverse process: how store pace has a negative effect on the display of positive emotion. This conceptual framework was used to generate a new quantitative hypothesis, which was then tested statistically.

In this study, the quantitative and qualitative phases were relatively distinct. The qualitative phase was largely self-contained, and its purpose was nearly exclusively to revise and develop the conceptual framework, incorporating a process model of how the pace of work affects displayed emotion. This framework was then used to generate a variance theory hypothesis that was tested with quantitative data. Figure 9.3 provides a design map of the study.

In other component studies, rather than shifting from one approach to another in sequence, the two approaches are used concurrently, although separately, and integrated only in drawing conclusions.

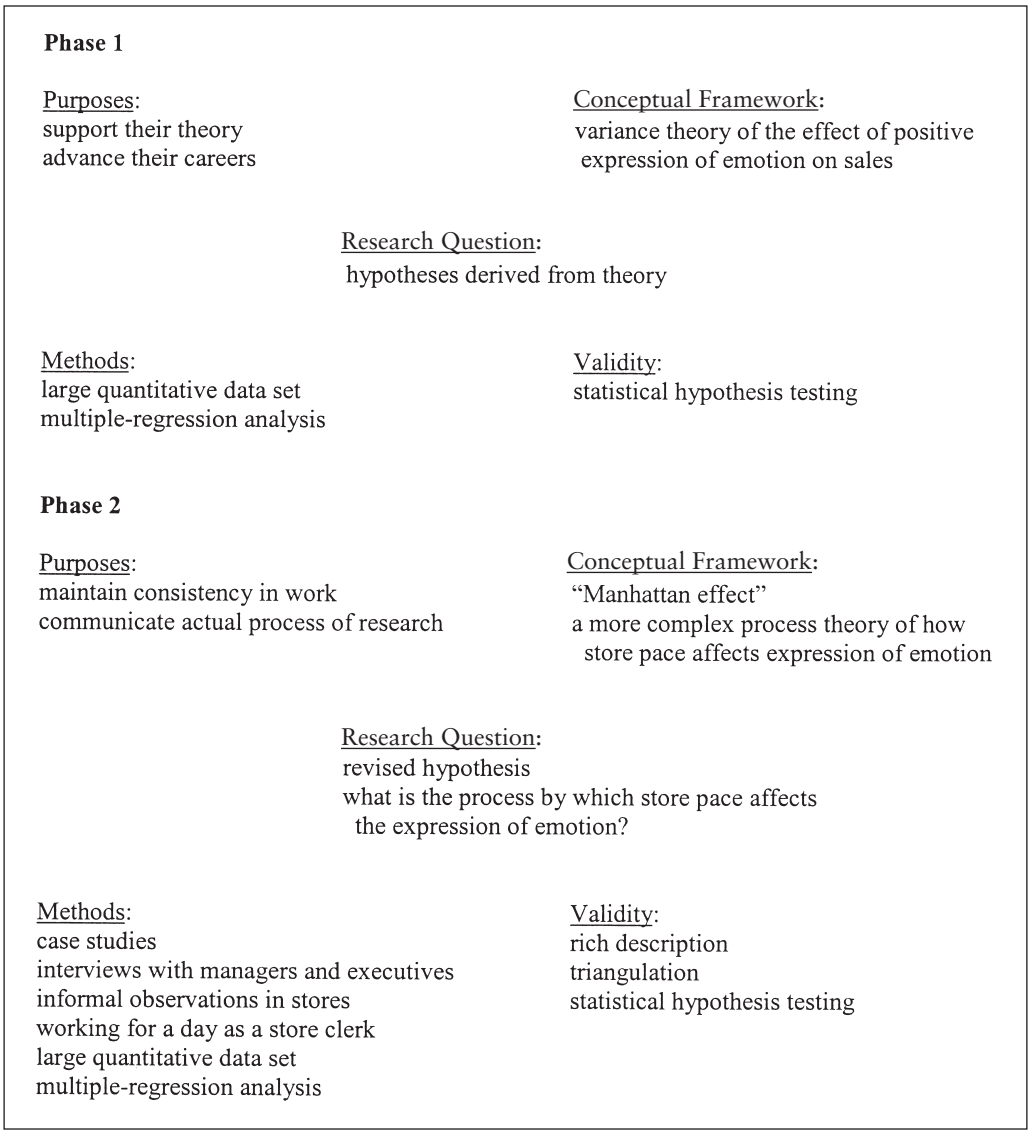

Figure 9.3. Design Map of Sutton and Rafaeli (1988, 1992) Study

Trend (1978/1979) gave an account of such a study, an evaluation of an experimental federal housing subsidy program involving both quantitative and qualitative data collection and analysis. Trend described the study as a "naturalistic experiment" (p. 69), but it would more accurately be called a "pre-experiment" in Campbell and Stanley's (1963) typology because it did not involve a control group. Extensive quantitative data were collected on agency activities, expenses, demographic characteristics of clients, and housing quality, mainly through surveys.

In addition, each site had an observer (usually an anthropologist) who prepared a qualitative case study of that site, using field observations, interviews, and documents. The intent was that program outcomes would be determined through analysis of the quantitative data, while the case studies would provide a holistic picture of program process (Trend, 1978/1979, p. 70).

However, this plan began to unravel when the conclusions of an observer in one site directly contradicted the results of the quantitative analysis of program outcomes in that site. While neither side doubted "the facts" produced by the other, the two interpretations of these facts differed radically. The agency conducting the evaluation sided with the quantitative results, and the observer was repeatedly told to rewrite his analysis to fit the quantitative conclusions. Finally, Trend and the observer made a sustained effort to get at what had really been going on, using both the quantitative and qualitative data. They eventually came up with a coherent process explanation for nearly all of the data that went well beyond either the quantitative or the initial qualitative conclusions and that revealed serious shortcomings in both accounts.

Although this study clearly fits into the "component" type of design in that the quantitative and qualitative data were collected and analyzed separately and were combined only in developing conclusions, it also resembles the most developed subtype of integrated design described by Caracelli and Greene (1997), the transformative design. In such designs, the value commitments of different traditions are integrated, giving voice to different ideologies and interests in the setting studied. (In the interactive design model, these value commitments can form part of both the purposes and the conceptual framework.) The quantitative analysts tended to represent the views of the program managers and funders, while the observer was an advocate for the agency staff and clients. These differing value stances, as well as the separation of the quantitative and qualitative strands of the study, led to polarization and conflict; "each side held so tightly to its own views that it was impossible to brush aside the lack of congruence" (Trend, 1978/1979, p. 84).

Trend (1978/1979) concluded that multiple methods might not lead to an easy integration of findings and that "unanimity may be the hallmark of work in which other avenues to explanation have been closed off prematurely" (p. 68). If the discrepancy between the qualitative and quantitative accounts had been discovered earlier, or if the two approaches had been more closely integrated, then it is possible that the observer would have been subtly or overtly coerced into making his conclusions fit the "hard" data (p. 84). Trend thus argued that "the proliferation of divergent explanations should be encouraged" (p. 68) but also that an effort should be made to develop an account that does justice to all of the conflicting perspectives.

A third study that initially appears "component-like," in that the quantitative and qualitative elements are conceptually distinct phases of the research, is the research described in Festinger, Riecken, and Schachter's (1956) book, *When Prophecy Fails*. This was a psychological study of an end-of-the-world cult and the consequences for cult members of the failure of its predictions. The study began with a variable-oriented theory and a hypothesis about the conditions under which disconfirmation of belief will paradoxically be followed by increased commitment. The data were collected entirely through participant observation; a number of researchers pretended to be con-

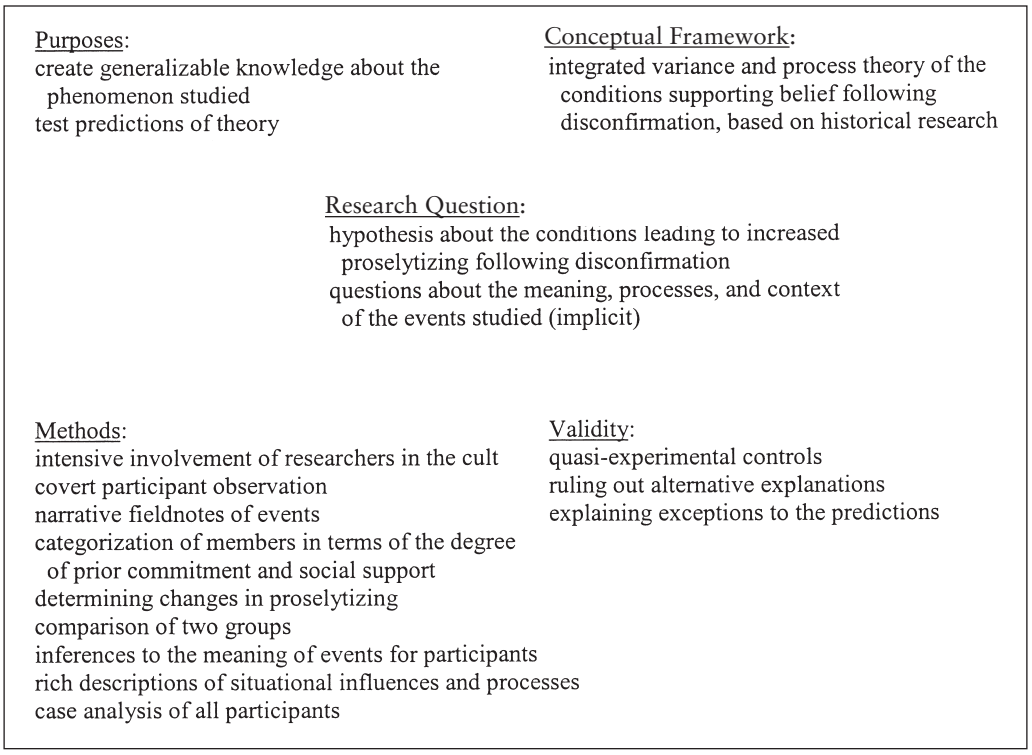

Figure 9.4. Design Map of Festinger et al. (1956) Study

verts to the cult and covertly amassed detailed descriptive notes on what happened as the day of judgment approached and then passed. However, to test the hypothesis, these observational data were analyzed primarily by categorizing members in terms of the degree of prior commitment and social support (the two key independent variables) and measuring changes in proselytizing (the indicator of subsequent commitment) following disconfirmation. Figure 9.4 depicts the design of the study.

This study differs from a component study such as Sutton and Rafaeli's in that the "components" are different aspects of a single research design rather than separate quantitative and qualitative strands or phases of a larger study. At first glance, it seems to fit one of Patton's (1990) types of "methodological mixes"—experimental design, qualitative data, and statistical analysis (p. 191)—and would thus be considered a mixed model design. The main differences from Patton's type are that the study was a "natural" experiment (more accurately, a quasi-experiment) rather than a manipulated intervention and that the analysis was hypothesis testing, variable focused, quantitative, and based on prior analytical categories but not specifically statistical due to the small number of participants. However, the design is more complex than this categorization suggests, and we want to analyze the study to reveal some of these complexities.

The purposes and explicit research questions for Festinger et al.'s (1956) study were predominantly quantitative—a goal of testing the predictions of their

theory of how people with a strongly held belief respond to disconfirmation of that belief; a hypothesis, deductively generated from this theory, about the effect of social support following disconfirmation on the key measure of commitment (proselytizing); and the testing of this hypothesis, with the goal of creating generalizable knowledge. However, their conceptual framework addressed both the process by which the predicted outcome (disconfirmation leads to increased commitment) could occur and the variables that could influence this outcome, and some implicit process questions became apparent in the conclusions section.

In terms of methods, the study could be seen as a quasi-experiment, with a naturally occurring intervention, pre- and post-intervention data collection, and a comparison of two parts of the group that differed in the degree of social support. However, with the detailed qualitative data collection, the logic also resembled a qualitative case study. The research relationships and data collection involved covert participant observation, intensive involvement of the researchers in the cult, and narrative fieldnotes of events. It is unclear what formal qualitative analysis techniques, if any, were used. In the narrative of the study, the researchers made frequent inferences to the meaning of events for participants, and there were rich descriptions of situational influences and processes.

In the concluding chapter of their book, Festinger et al. (1956) first gave a case-by-case analysis of all participants in terms of the hypothesized preconditions (independent variables) and outcomes. Participants were then categorized in terms of these variables, and the authors tallied the confirmations and exceptions to the predictions and compared the two situations that differed in the key independent variable (social support). This argument is essentially quantitative. However, it is extensively supplemented by a process analysis of the sequence of events for each individual; this is used to explain apparent exceptions and to modify the hypotheses to some extent. The authors also made use of unanticipated outcomes (e.g., the persistence of predictions of disaster, the identification of visitors as spacemen) that were relevant to their conclusions.

This was a coherent and workable mixed methods design because the different components were compatible and complementary in this particular situation, not because they derived from a single paradigm or were consistent with a single set of assumptions. Testing Festinger et al.'s (1956) specific hypothesis (the primary aim of the study) would ideally have involved an experimental design. However, the nature of the phenomenon addressed by the theory made an experimental test of the hypothesis impossible. The only real alternative was a kind of "natural experiment," and one was dropped into the researchers' laps. The authors noted, somewhat apologetically, that the situation that was available to them precluded the sort of formal standardized methods that "the orthodoxy of social science" would normally require (pp. 248-249); consequently, the sampling and data collection were almost purely qualitative. These consisted mainly of the use of participant observers, who gathered whatever data they could that related to the theory and research questions—including data on the meaning, context, and process of the group's activities—and produced a detailed narrative account of the events leading up to and following the disconfirmation of the group's predictions.

The crucial links in making this a coherent design were the analysis and validity procedures employed to connect the authors' qualitative data to their research questions, hypotheses, theories, and pur-

poses. This was accomplished in two ways. One of these involved quantifying the qualitative data to adapt these to the logical requirements of hypothesis testing. The two groups of believers, which differed in the value of the major independent variable (social support), were compared in terms of the main outcome variable (extent of proselytizing) as well as on other indicators of the strength of commitment (a key mediating variable) both before and after the disconfirmation.

If this were the entire analysis, however, the research results would have been far less convincing than they were given that the number of participants (17) on whom sufficient data existed was quite small. The study's conclusion that social support was essential to strengthened belief and proselytizing was buttressed by a second qualitative analysis that examined the data on each group member for evidence relevant to the hypothesis and constructed a "mini-case study" of each member. These cases relied heavily on inductive identification of relevant data, attention to meaning and context, and a process account that elucidated the mechanisms by which belief was strengthened or weakened—all features that are characteristic of qualitative research. In addition, most of the report was such a "case study" of the entire phenomenon, revealing in rich detail how the group developed and how it responded to the disconfirmation of its predictions. These analyses reveal (or create) a set of implicit qualitative research questions about the meaning, processes, and context of the events studied that parallel the quantitative hypothesis and connect to qualitative aspects of the authors' conceptual framework. This dual analysis was facilitated by the conceptual framework for the study, which included both variance and process aspects of the phenomenon.

The validity of Festinger et al.'s (1956) conclusions is vulnerable to the fact that traditional experimental controls were impossible, that data were not collected in a structured way that would ensure reliability and facilitate comparison, and that the sample was quite small and self-selected. The researchers' main strategy for dealing with these validity issues was to explicitly identify plausible alternative explanations and to use their data to argue that these are not credible explanations for their results. This strategy draws on both qualitative and quantitative approaches.

We believe that few, if any, sequentially "mixed" designs of the type described by Patton (1990) maintain a complete sequential separation of the qualitative and quantitative elements of the research. As in this example, the different components tend to grow "tendrils" backward and forward, integrating both qualitative and quantitative elements into all components of the research. This is understandable given the "resonance" among the components of each approach; qualitative data collection tends to generate qualitative analysis, research questions, conceptualizations, and validity strategies, and the same is true of quantitative components, while a qualitative component of the conceptual framework tends to generate qualitative research questions and methods.

Another approach that blurs the distinction between component and integrated designs is to conduct the quantitative and qualitative data collection strands in parallel, as in the studies by Trend (1978/1979) and Festinger et al. (1956), but to embed these within an overall experimental or quasi-experimental design, one that involves a deliberate intervention as well as establishing experimental and control conditions. This sort of design has been employed by Lundsgaarde, Fischer, and Steele (1981) and by Maxwell et al. (1986), among others. Such studies are classed as integrated designs by Caracelli

and Greene (1997, p. 26) and would be considered mixed model designs by Tashakkori and Teddlie (1998) because they go beyond the mixing of methods in a strict sense. However, the actual methods components can range from largely separate (as in the study by Maxwell et al., 1986) to closely integrated (as in the study by Milgram, 1974 [discussed later]). The study by Lundsgaarde et al. (1981), conducted during 1976-1977, illustrates some of the possible complexities of such designs.

These researchers, all anthropologists, carried out what they described as an "ethnographic" study of the effect of a computerized medical information system (known as PROMIS) on the functioning of a hospital ward. They did this by studying two hospital wards prior to the implementation of PROMIS and then continuing this research while PROMIS was introduced on one of the wards, using the other ward as a control group. They described this ethnographic study as one "component" of the PROMIS evaluation; it was designed to complement the other components of the evaluation, which employed a quantitative analysis of medical records to determine the impact of PROMIS on the health care delivery process. The context in which PROMIS was implemented was politically charged, and the developers of the overall evaluation strategy were concerned that variation in human and situational variables might make it difficult to interpret the overall quantitative results. The goals of the ethnographic study were to document the events surrounding the implementation of the PROMIS system, and the experiences of the health care providers using this system, using a more descriptive and inductive approach so as to characterize the context in which the system was developed and demonstrated (Lundsgaarde et al., 1981, pp. 10-11).

However, the "ethnographic" component itself involved a mix of qualitative and quantitative elements. The purposes (described previously) were mainly qualitative but included the explicit comparison of the experimental and control wards so as to determine the effects of the PROMIS implementation on the experimental ward. The conceptual framework for the study was largely drawn from innovation theory (expressed in 20 "propositions" that were a mix of variance and process statements) and the sociology of medical practice. No research questions were explicitly stated; although some process questions can be clearly inferred from the study's purposes, the overall evaluation was guided by a specific variance hypothesis about the effect of PROMIS on patient care behavior, a hypothesis that was tested by the quantitative components of the evaluation. The ethnographic component relied heavily on participant observation, informal interviewing, and document analysis, and Lundsgaarde et al. (1981) presented an explicit defense of such qualitative methods in terms of the goals of context and meaning (p. 16). However, the study also included a questionnaire, a more structured interview, and a comparative observational assessment (following the introduction of PROMIS) of the amount of time spent generating and using medical records on the two wards, using matched pairs of residents or interns observed on a randomized schedule. (The latter task was required by the funding institution [p. 11] and forced a reallocation of much of the later qualitative data collection time from participant observation to in-depth interviews.) In addition, midway through the study, the observers on the two wards switched settings so as to gain a comparative perspective and to control for biases.

Lundesgaarde et al. (1981) justified this mix of quantitative and qualitative methods as a means of triangulating data and resolving contradictions between data sources (p. 16). The concerns of the evalu-

ation planners were well-founded; the quantitative analysis of medical records found no statistically significant advantages of PROMIS over its manual counterpart, while the ethnographic study showed that "many of the clinicians who were required to use the system were unwilling participants in the experiment and even unsympathetic to many of the goals of those who developed it" and that "many of the human and organizational problems . . . could have been avoided, or at least neutralized, if the developers had paid more attention to contextual social variables affecting system users" (p. 2). The authors stated,

> It is the unpredictability of the temporal characteristics of all innovations that presents researchers with the most thorny problems of analysis. The objective measurement of the rate of acceptance, and the estimation of the potential rate of diffusion, has proved the most difficult analytical problem in our study of the PROMIS innovation. (p. 4)

For this reason, they emphasized "the importance of a multifaceted and flexible research design for the study of the many social and operational problems created by the installation of [PROMIS]" (p. 9). The presentation of the results of the study demonstrated a close integration of the quantitative and qualitative elements in drawing conclusions and addressing validity threats. For example, their discussion of the effect of PROMIS on house staff physicians (pp. 61-91) closely integrated the data from participant observations, qualitative interviews, and the systematic time-sampling observations of residents and interns. This presentation embedded the statistical analysis of the quantitative behavioral data in a descriptive account of these activities, one that clarifies the contextual variations in, and influences on, these behaviors. They noted that the quantitative data did not support the widespread perception on the experimental ward that residents and interns spent more time entering medical data into patients' records, and the authors devoted considerable space to discussing possible reasons for this misperception, drawing on their interviews and observations (pp. 86-90).

While this evaluation superficially resembles a component design, with separate qualitative and quantitative components of the evaluation, a detailed examination reveals a much more integrated design. Some of this integration may initially have been externally imposed but was fully incorporated into the analysis, validity procedures, and conclusions. The triangulation of different methods was the result of not only the different purposes of the evaluation but also the validity concerns that would have threatened a purely quantitative study. The presence of quantitative elements in the ethnographic part of the study was partly the result of an implicit purpose (the researchers' need to satisfy the external funder) that had little intrinsic connection to the study's conceptual framework. However, these elements were closely integrated into the study's analysis, using a validity approach based on both quantitative and qualitative concepts (experimental controls and statistical tests and a process approach to ruling out alternative explanations). Figure 9.5 provides a design map of this study.

The quantitative and qualitative elements can be even more closely integrated than in this example. Milgram's (1974) *Obedience to Authority* is a report of an experimental study (carried out between 1960 and 1963) of how people respond when they are ordered by authorities to inflict pain and possible serious harm on others. Milgram and his associates designed a series of laboratory situations in

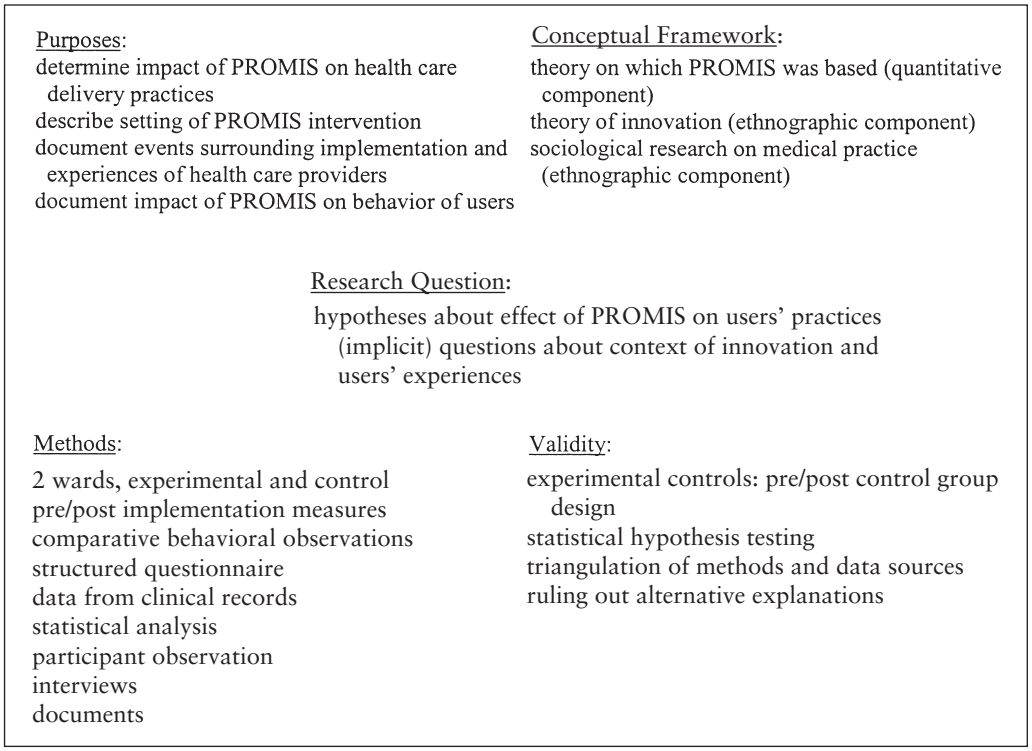

Figure 9.5. Design Map of Lundsgaarde et al. (1981) Study

which participants were deceived into believing that they were part of a study of the effects of punishment on learning and were then told to give increasingly severe electrical shocks to a supposed "subject" who was actually an accomplice of the researchers and who feigned pain and eventual refusal to cooperate. Unlike Festinger et al. (1956), Milgram (1974) explicitly grounded this study in the experimental tradition in social psychology (p. xv). The researchers employed numerous different experimental conditions designed to determine the effect of different variables on the degree of obedience (the dependent variable), and they collected quantitative data about the level of shock that participants administered (the main measure of obedience) in each of the different conditions.

However, the researchers were also concerned with the *process* by which people responded to the researchers' directions: how the participants made sense of and reacted to these directions and why they complied with or resisted the orders. In introducing the individual case studies, Milgram (1974) stated,

> From each person in the experiment we derive one essential fact: whether he has obeyed or disobeyed. But it is foolish to see the subject only in this way. For he brings to the laboratory a full range of emotions, attitudes, and individual styles. . . . We need to focus on the individuals who took part in the study not only because this provides a personal dimension to the experiment but also because the quality of each

> person's experience gives us clues to the nature of the process of obedience. (p. 44)

The researchers covertly recorded the participants' behavior during the experiment, interviewed some participants at length after the experiment was over to determine their reasons for compliance or refusal, and sent a follow-up questionnaire to all participants that allowed expression of their thoughts and feelings. The analysis of these data is primarily qualitative but is closely integrated with the quantitative data. The results chapters of the book present a fine-grained blending of quantitative tables and graphs with observational notes, excerpts from recorded observations and interviews, and case studies of particular participants' responses to the experimental situation.

In addition, the theoretical model developed from the study is not a pure "variance" model, restricted to the different variables that affect obedience; as in the study by Festinger et al. (1956), it incorporates extensive discussion of the social processes and subjective interpretations through which obedience and resistance to authority develop. And in discussing potential validity threats to the study's conclusions, Milgram (1974) used both the quantitative results from the experimental manipulations and qualitative data from the observations to rule out these threats. In this study, experimental intervention, laboratory controls, and quantitative measurement and analysis were integrally combined with qualitative data collection and analysis to answer both qualitative and quantitative research questions. Although Milgram himself said virtually nothing explicitly about the integration of quantitative and qualitative elements in this study, Etzioni (1968) claimed that this research "shows that the often stated opposition between meaningful, interesting humanistic study and accurate, empirical quantitative research is a false one: The two perspectives can be combined to the benefit of both" (cited in Milgram, 1974, p. 201). Figure 9.6 provides a design map of the study.

◆ *Conclusions and Implications*

In this chapter, we have tried to show the value of a broader and more interactive concept of research design for understanding mixed methods research. We have also argued for a broader and more fundamental concept of the qualitative-quantitative distinction, one that draws on the idea of two different approaches to explanation as well as two different types of data. Through detailed examination of particular studies, we have tried to demonstrate how these tools can be used to attain a better understanding of mixed methods research. We draw several implications from these arguments and examples.

First, the logic-in-use of a study can be more complex, and can more closely integrate qualitative and quantitative elements of the study, than an initial reading of the report would suggest. The studies by Blumstein and Schwartz (1983), Lundsgaarde et al. (1981), Festinger et al. (1956), and Milgram (1974) all involved a greater integration of qualitative and quantitative approaches than one would guess from their explicit descriptions of their methods, and the two other studies presented (Sutton & Rafaeli, 1988, 1992; Trend, 1978/1979) may be exceptions only because the authors had published candid in-depth accounts of their studies' designs and methods, including aspects rarely addressed in research reports.

Second, the interactive design model that we have presented can be a valuable tool in understanding the integration of

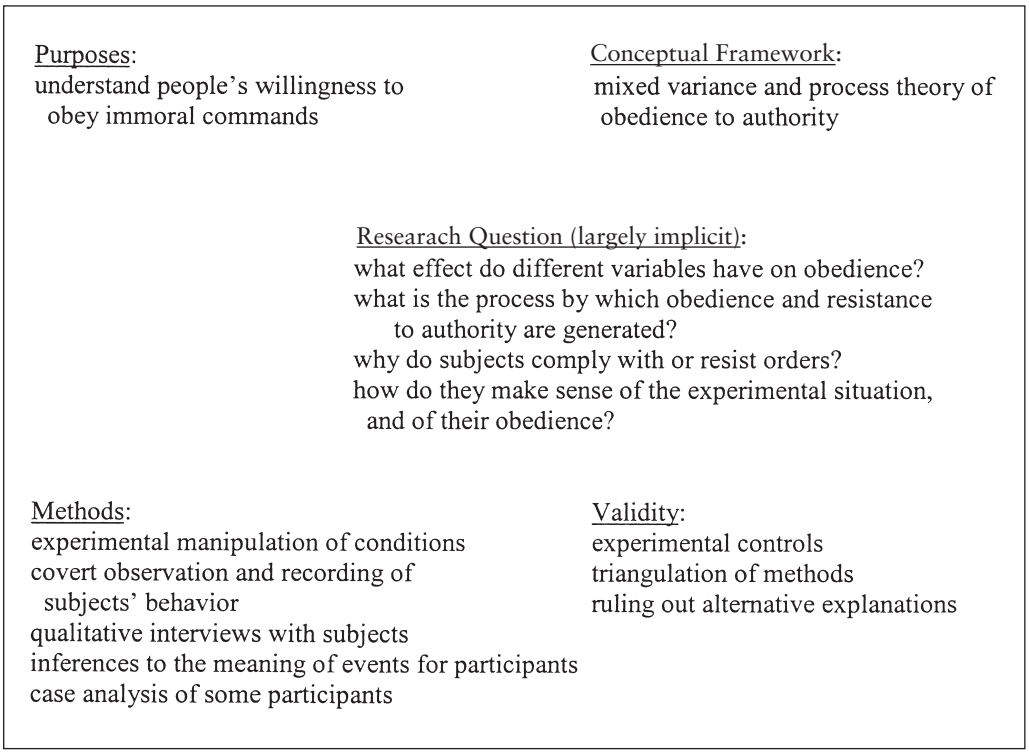

Figure 9.6. Design Map of Milgram (1974) Study

qualitative and quantitative approaches and elements in a particular study. For example, the conceptual framework of a study may be largely variance theory (Sutton & Rafaeli, 1988, 1992, Phase 1), largely process theory (Sutton & Rafaeli, 1988, 1992, Phase 2), a combination of both types of theories (Trend, 1978/1979; Lundsgaarde et al., 1981), or an integration of the two in a single theory (Festinger et al., 1956; Milgram, 1974).

Third, there is considerable value in a detailed understanding of how qualitative and quantitative methods are actually integrated in particular studies. For example, the degree of integration of qualitative and quantitative elements in the conceptual framework, analysis, or validity components of a study might not correspond to the integration of data collection methods. The study by Lundsgaarde et al. (1981) has more integration in the methods and validity components than in the conceptual framework, while the study by Festinger et al. (1956) has more integration in the conceptual framework and validity than in methods. In addition, the actual integration among different components of the design is often essential to understanding how a particular combination of quantitative and qualitative elements is or is not coherent. For example, the integrated process/variance conceptual framework of Milgram's (1974) study played a key role in the integration of methods and analysis.

Fourth, we do not believe that typological models by themselves provide adequate guidance for designing mixed methods research. The examples and analyses

of specific studies provided by Greene and Caracelli (1997) and by Tashakkori and Teddlie (1998) are essential complements to their typologies; these provide both a concrete realization of how the types play out in practice and an illustration of aspects of mixed methods design that are not captured in the typology.

Fifth, we also believe, however, that there is no easy generalizability or transferability of the analysis of particular studies; the actual integration of the components of a study is influenced by a wide range of conditions and factors and is not dictated by the category in which it fits. The design model that we have presented is a tool for designing or analyzing an actual study rather than a template for designing a particular *type* of study. In a sense, we are presenting a more qualitative approach to mixed methods design, emphasizing particularity, context, holistic understanding, and the process by which a particular combination of qualitative and quantitative elements plays out in practice, in contrast to a more quantitative approach based on categorization and comparison. As with quantitative and qualitative approaches in general, we advocate an integration of the two approaches.

■ *References*

Becker, H. S. (1990). Generalizing from case studies. In E. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 233-242). New York: Teachers College Press.

Bernstein, R. J. (1992). *The new constellation: The ethical-political horizons of modernity-postmodernity.* Cambridge, MA: MIT Press.

Blumstein, P., & Schwartz, P. (1983). *American couples.* New York: Simon & Schuster.

Brewer, J., & Hunter, A. (1989). *Multimethod research: A synthesis of styles.* Newbury Park, CA: Sage.

Bryman, A. (1988). *Quantity and quality in social research.* London: Unwin Hyman.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally. (Reprinted in 1966 as *Experimental and quasi-experimental designs for research*)

Caracelli, V. J., & Greene, J. C. (1997). Crafting mixed-method evaluation designs. In J. C. Greene & V. J. Caracelli (Eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* (New Directions for Evaluation, No. 74, pp. 19-32). San Francisco: Jossey-Bass.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Creswell, J. W. (1994). *Research design: Quantitative and qualitative approaches.* Thousand Oaks, CA: Sage.

Etzioni, A. (1968). A model of significant research. *International Journal of Psychiatry, 6,* 279-280.

Festinger, L., Riecken, H. W., & Schachter, S. (1956). *When prophecy fails.* Minneapolis: University of Minnesota Press.

Grady, K. A., & Wallston, B. S. (1988). *Research in health care settings.* Newbury Park, CA: Sage.

Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. In J. C. Greene & V. J. Caracelli (Eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms* (New Directions for Evaluation, No. 74, pp. 5-17). San Francisco: Jossey-Bass.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Harre, R. (1972). *The philosophies of science.* Oxford, UK: Oxford University Press.

Harre, R., & Madden, E. H. (1975). *Causal powers: A theory of natural necessity.* Oxford, UK: Basil Blackwell.

Kaplan, A. (1964). *The conduct of inquiry.* San Francisco: Chandler.

Kidder, L. H., & Fine, M. (1987). Qualitative and quantitative methods: When stories converge. In M. M. Mark & R. L. Shotland (Eds.), *Multiple methods in program evaluation: New directions in program evaluation* (pp. 57-75). San Francisco: Jossey-Bass.

Koontz, V. (1976). Innovation: Components of bundles. *Medical Care, 7*(4).

LeCompte, M. D., & Preissle, J. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). San Diego: Academic Press.

Lundsgaarde, H. P., Fischer, P. J., & Steele, D. J. (1981). *Human problems in computerized medicine* (Publications in Anthropology, No. 13). Lawrence: University of Kansas.

Martin, J. (1982). A garbage can model of the research process. In J. E. McGrath, J. Martin, & R. Kulka (Eds.), *Judgment calls in research* (pp. 17-39). Thousand Oaks, CA: Sage.

Maxwell, J. A. (1990). Response to "Campbell's Retrospective and a Constructivist's Perspective." *Harvard Educational Review, 60,* 504-508.

Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62,* 279-300.

Maxwell, J. A. (1996). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.

Maxwell, J. A. (1998). *Using qualitative research to develop causal explanations*. Working paper, Harvard Project on Schooling and Children, Harvard University.

Maxwell, J. A., & Mohr, L. B. (1999). *Quantitative and qualitative: A conceptual analysis*. Paper presented at the annual meeting of the American Evaluation Association, Orlando, FL.

Maxwell, J. A., Sandlow, L. J., & Bashook, P. G. (1986). Combining ethnographic and experimental methods in evaluation research: A case study. In D. M. Fetterman & M. A. Pitman (Eds.), *Educational evaluation: Ethnography in theory, practice, and politics* (pp. 121-143). Newbury Park, CA: Sage.

McCawley, J. (1982). *Thirty million theories of grammar.* Chicago: University of Chicago Press.

Merriam-Webster. (1984). *Webster's ninth new collegiate dictionary.* Springfield, MA: Author.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative research design* (2nd ed.). Thousand Oaks, CA: Sage.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper & Row.

Mohr, L. (1982). *Explaining organizational behavior.* San Francisco: Jossey-Bass.

Mohr, L. (1995). *Impact analysis for program evaluation* (2nd ed.). Thousand Oaks, CA: Sage.

Mohr, L. (1996). *The causes of human behavior: Implications for theory and method in the social sciences.* Ann Arbor: University of Michigan Press.

Patton, M. Q. (1980). *Qualitative evaluation and research methods.* Beverly Hills, CA: Sage.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation.* London: Sage.

Pitman, M. A., & Maxwell, J. A. (1992). Qualitative approaches to evaluation. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 729-770). San Diego: Academic Press.

Rabinowitz, V. C., & Weseen, S. (2001). Power, politics, and the qualitative/quantitative debates in psychology. In D. L. Tolman & M. Brydon-Miller (Eds.), *From subjects to subjectivities: A handbook of interpretive and participatory methods* (pp. 12-28). New York: New York University Press.

Reichardt, C. S., & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in*

program evaluation (pp. 7-32). Beverly Hills, CA: Sage.

Robson, C. (1993). *Real world research: A resource for social scientists and practitioner-researchers.* Oxford, UK: Blackwell.

Rogers, E. C. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Salmon, W. C. (1989). Four decades of scientific explanation. In P. Kitcher & W. C. Salmon (Eds.), *Scientific explanation* (pp. 3-219). Minneapolis: University of Minnesota Press.

Salmon, W. C. (1998). *Causality and explanation*. New York: Oxford University Press.

Sayer, A. (1992). *Method in social science: A realist approach*. London: Routledge.

Sayer, A. (2000). *Realism and social science*. Thousand Oaks, CA: Sage.

Staw, B. M. (1992). Do smiles lead to sales? Comments on the Sutton/Rafaeli study. In P. Frost & R. Stablein (Eds.), *Doing exemplary research* (pp. 136-142). Newbury Park, CA: Sage.

Sutton, R. I., & Rafaeli, A. (1988). Untangling the relationship between displayed emotions and organizational sales: The case of convenience stores. *Academy of Management Journal, 31,* 461-487.

Sutton, R. I., & Rafaeli, A. (1992). How we untangled the relationship between displayed emotions and organizational sales: A tale of bickering and optimism. In P. Frost & R. Stablein (Eds.), *Doing exemplary research* (pp. 115-128). Newbury Park, CA: Sage.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Applied Social Research Methods, No. 46). Thousand Oaks, CA: Sage.

Trend, M. G. (1979). On the reconciliation of qualitative and quantitative analyses: A case study. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 68-86). Newbury Park, CA: Sage. (Original work published 1978)

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Weiss, R. S. (1994). *Learning from strangers: The art and method of qualitative interviewing*. New York: Free Press.

Yin, R. K. (1984). *Case study research: Design and methods*. Beverly Hills, CA: Sage.