

❖

METHODOLOGY MATTERS: DOING RESEARCH IN THE BEHAVIORAL and SOCIAL SCIENCES

JOSEPH E. MCGRATH

PSYCHOLOGY, UNIVERSITY OF ILLINOIS, URBANA

23 OCTOBER 1994

❖

“Doing research” simply means the systematic use of some set of theoretical and empirical tools to try to increase our understanding of some set of phenomena or events. In the social and behavioral sciences, the phenomena of interest involve states and actions of human systems — of individuals, groups, organizations, and larger social entities — and the by-products of those actions.

The meaning of research evidence, in any area of science, is inherently tied to the means or methods by which that evidence was obtained. Hence, to understand empirical evidence, its meaning, and its limitations, requires that you understand the concepts and techniques on which that evidence is based.

This chapter is about some of the tools with which researchers in the social and behavioral sciences go about “doing” research. It raises some issues about strategy, tactics and operations. Especially, it points out some of the inherent limits, as well as the potential strengths, of various features of the research process by which behavioral and social scientists do research.

SOME BASIC FEATURES OF THE RESEARCH PROCESS

Doing research, in the behavioral and social sciences, always involves bringing together three sets of things:

- (a) some *content* that is of interest,
- (b) some *ideas* that give meaning to that content, and
- (c) some *techniques or procedures* by means of which those ideas and contents can be studied.

For example, the contents of a study might involve the behavior of a jury, conversations in a family about buying a new car, the voting behavior of members of a community, littering in a park, courtship patterns in a small town, and so forth. The ideas might include the concept of attitudes, the notion that education affects political preferences, the concept of conformity, the hypothesis that groups whose members like one another perform tasks better than groups whose members do not like each other, and so forth. The techniques might include a questionnaire to assess individual attitudes, toward a car or a candidate or group mates; a set of procedures for observing family discussions about cars and money; a means to gather election returns; a plan to evaluate the quality of group task products; and so forth.

I will refer to these three sets of things more formally, as three distinct, though inter-related, domains:

- (a) The *Substantive* domain, from which we draw contents that seem worthy of our study and attention;
- (b) The *Conceptual* domain, from which we draw ideas that seem likely to give meaning to our results; and
- (c) The *Methodological* domain, from which we draw techniques that seem useful in conducting that research.

Furthermore, research always deals with several levels of phenomena: With *relations* between units or *elements* within a context or *embedding system*. The elements, relations, and embedding systems have different forms in each of the three domains [See Figure 1].

SUBSTANTIVE DOMAIN

In the substantive domain, I will call the units or elements Phenomena, and the relations among them Patterns of phenomena. These Phenomena, and Patterns of them, are the object of our study. For the behavioral and social sciences, the phenomena of interest involve the states and actions of some human systems — individuals, groups, organizations, communities, and the like — and the conditions and processes that give rise to and follow from those states and actions.

Another way to say this is to say that the behavioral and social sciences study “actors behaving toward objects in context”. An example would be “an individual casting a vote in a county election”. Another example would be “the number of units produced in the week of April 12th by group 32 of the production division of the Danville plant”.

It must be understood that both “actors” and “context” here refer to human systems at any of a number of system levels — individual, group, community, organization, and so on. Different behavioral and social sciences specialize in the study of different human systems — that is, in the study of phenomena and patterns at different levels and of different kinds. The rest of this book presents material that illustrates many of the substantive phenomena and patterns that have been studied within the field of human-computer interaction.

CONCEPTUAL DOMAIN

For the social and behavioral sciences, the elements of interest in the conceptual domain are properties of the states and actions of those human systems that are the focus of study — properties of “actors behaving toward objects in context”. These might include such familiar ideas as “attitude,” “cohesiveness,” “power,” “social pressure,” “status,” as well as many others that are used in social and behavioral science research. *Relations* in the conceptual domain refer to any of a variety of possible ways in which two or more elements can be connected. Some of those ways are viewed as “causal” connections. Some are logical relations. Some simply are chronological relations. For example, two elements can be equal or unequal, they can be related linearly or non-linearly, one can be a necessary or sufficient cause of the other, one can include the other, the relation between them can be one way or reciprocal, and many more. Materials from the conceptual domain — properties, and relations among those properties — are the “ideas” that can give meaning to the phenomena and patterns that we study in the substantive domain.

METHODOLOGICAL DOMAIN

In the methodological domain, elements are methods. I will call the methods *Modes of Treatment* (of properties of phenomena). Modes of Treatment are different ways by which a researcher can deal with a particular feature of the human systems that are to be studied.

One set of such Modes of Treatment include various *techniques for measuring* some feature (that is, for assessing the state or magnitude of some property of some actors-behaving- in-context), so that the researcher can determine what value or level that feature has for each “case” to be studied. Measurement methods include such things as: a questionnaire, a rating scale, a personality test, instruments for observing and recording communications, techniques for assessing the quality of some products resulting from individual or group task performance, and the like. (More is said about kinds of measures near the end of this chapter.)

Modes of Treatment also include various *techniques for manipulating* some feature of a research situation (that is, some property of an actor-behavior-context). To carry out an experimental manipulation of a feature of the situation (sometimes referred to as “manipulating a variable”) means making that feature have one particular predetermined value or level for certain “cases” to be studied and another specific preordained value or level for certain other “cases,” so that the effect of differences in that property can be assessed by comparing those two sets of “cases.” For example: You might want to study the effectiveness of a particular human-computer system by studying two sets of work groups, one set of groups working with that computer system and the other set doing the same tasks “manually”. Social psychologists have tried to manipulate features of the systems they study by a number of techniques, such as:

- (a) **giving instruction** to participants (e.g., trying to motivate them to try hard by telling them that there will be a valuable prize for the best product);
- (b) **imposing constraints** on features of the environment (e.g., providing some participants with a particular software program that may help task performance, and providing other participants with a different or no program to carry out that function);
- (c) **selecting materials** for use (e.g., trying to produce differences in task difficulty by giving some participants very difficult word problems to complete, and giving other participants easier problems of the same type);
- (d) **giving feedback** about prior performances (e.g., trying to induce feelings of success or failure by telling some participants they did well, and telling others they did poorly, on a previous task);
- (e) **using experimental confederates** (e.g., trying to establish different degrees of liking for fellow group members by having an experimental assistant who is pretending to be a normal participant work very hard in some groups and act indifferent in others).

(More is said about techniques for manipulating variables near the end of this chapter).

Modes of Treatment of variables also include a set of *techniques for controlling the impact* of various “extraneous” features of the situation — features that are important but that you are not going to measure or manipulate in a particular study. These include: techniques for *experimental control*, by which you make certain features take the same predetermined value for all cases in the study (e.g., study only 6-year-olds to control on

age); techniques for *statistical control* by which you try to nullify the effects of variations in a given property within a study by “removing” those variations by statistical means; and techniques for *distributing the impact* of a number of features of the system and its context—without directly manipulating or controlling any one of them—so that such impact can be taken into account in interpretation of results. The most prominent means for distributing impact of a number of features is called *randomization*, and refers to procedures for the allocation of “cases” among various conditions within the study. These Modes for dealing with various features of the human systems to be studied — *measuring, manipulating, controlling and distributing impact* — are the basic sets of elements or “tools” by which social and behavioral scientists systematically gather empirical information.

Relations in the methodological domain have to do with the application of various *Comparison Techniques*. These are methods or techniques by means of which the researcher can assess relations among the values of two or more features of the human system under study. Such comparisons involve three sets of features of the systems under study: (a) the features that have been measured, and that are regarded as measures of the phenomena of interest (these are sometimes called “dependent variables”); (b) the features that have been measured or manipulated, and that are regarded as potential covariates of, or antecedents to, the phenomena of interest (these are sometimes called “independent variables”); and (c) all of the other features of the system that are relevant to the relations of interest (between dependent and independent variables), and that you have (or have failed to) control, or whose impact you have (or have failed to) distribute or otherwise take into account. Comparisons assess the covariation or association between the values of the first two sets (the dependent and independent variables), against the backdrop of the third set (i.e., other relevant features that were not studied directly but that nevertheless are a part of the meaning of results).

Most of the rest of this chapter will deal with features of the research process that emphasize the methodological domain, without much systematic consideration of either conceptual or substantive matters. The reader should keep in mind, though, that the research process, like a three-legged stool, always depends on materials from *all three domains* — content, ideas, and techniques.

DOMAINS

LEVELS	SUBSTANTIVE	CONCEPTUAL	METHOD- OLOGICAL
ELEMENTS	Phenomena	Properties	Modes of Treatment
RELATIONS	Patterns	Relations	Comparison Techniques
EMBEDDING SYSTEMS	Ongoing systems [e.g. human-computer systems	Conceptual Systems (e.g., field theory)	Research Strategies (e.g. laboratory experiment)

Figure 1:
Domains and levels of concepts in behavioral and social science research.

RESEARCH METHODS AS OPPORTUNITIES AND LIMITATIONS

Methods are the tools —the instruments, techniques and procedures — by which a science gathers and analyzes information. Like tools in other domains, different methods can do different things. Each method should be regarded as offering potential opportunities not available by other means, but also as having inherent limitations. You cannot pound a nail if you don’t have a hammer (or some functional equivalent). But if you do have a hammer, that hammer will not help you much if you need to cut a board in half. For that you need a saw (or the functional equivalent). And, of course, the saw would not have helped to drive the nail. So it is with the tools or methods of the social and behavioral sciences.

All research methods should be regarded as *bounded opportunities* to gain knowledge about some set of phenomena, some substantive domain. Knowledge in science is based on use of some combination of substance, concepts and methods. The meaning of that knowledge, and the confidence we can have in it, both are contingent on the methods by which it was obtained. All methods used to gather and to analyze evidence offer both opportunities not available with other methods, and limitations inherent in the use of those particular methods.

One good example of this dual nature of methods —both opportunities for gaining knowledge and limitations to that knowledge — is the widespread use of questionnaires and other forms of self-report in many areas of the social and behavioral sciences. On the one hand, self-report measures (questionnaires, interviews, rating scales, and the like) are a direct way, and sometimes the only apparent way, to get evidence about certain kinds of variables that are worthy of study: attitudes, feelings, memories, perceptions, anticipations, goals, values, and the like. On the other hand, such self-report measures have some serious flaws. For example: Respondents may try to appear competent, to be consistent, to answer in socially desirable ways, to please (or frustrate) the researcher. Sometimes respondents are reactive on such self-report measures without even being aware of it. These flaws limit, and potentially distort, the information that can be gained from such self-report measures. Other approaches to data collection, such as observation of visible behavior, may be difficult or impossible to use when studying particular kinds of variables. For example: How do you go about observing anxiety, or sadness, or some other emotion? In any case, while such methods may avoid some of the particular weaknesses of self-reports, those methods will have other different weaknesses.

Such is the dilemma of empirical science: All methods have inherent flaws, though each has certain potential advantages. You cannot avoid these flaws; but you can bring more than one approach, more than one method, to bear on each aspect of a problem. If you only use one method, there is no way to separate out the part that is the “true” measure of the concept in question from the part that reflects mainly the method itself. If you use multiple methods, carefully picked to have different strengths and weaknesses, the methods can add strength to one another by offsetting each other’s weak-

nesses. Furthermore, if the outcomes of use of different methods are consistent, this way of proceeding can add credibility to the resulting evidence. If the outcomes differ across different methods, then you can avoid misinterpretation of the resulting evidence by properly qualifying your conclusions.

This same general problem (that methods are inherently flawed, though each is flawed differently), and this same general prescription for dealing with it (by use of multiple methods), hold, as well, for research strategies, for comparison techniques and for research designs, all of which will be discussed subsequently in this chapter. For example, the research strategy called the laboratory experiment has some important strengths. It can permit precise measurement of effects resulting from deliberate manipulation of presumed causes, and therefore the drawing of strong inferences about cause-effect relations. But laboratory experiments also have some serious flaws. Researchers using laboratory experiments often greatly narrow the scope of the problem; they study it in artificial settings; and they are likely to use procedures and measures that make the situation seem even more artificial to the participants.

Several strategies that are alternatives to laboratory experiments are discussed later in this chapter. They include: field studies, sample surveys, and several others. Each of these other strategies offers different strengths, some of them offsetting the weaknesses of the laboratory; but each also has different inherent weaknesses, some of these being the very strengths of the laboratory strategy. No one strategy, used alone, is very useful; each of them is far too flawed. But again, the researcher needs to take advantage of multiple approaches. Usually, this cannot be done within a single study — often, the researcher must use a single strategy as a practical matter. But multiple strategies can be used over several studies of the same problem. The approaches need to be chosen so that the weaknesses of each strategy can be offset by the strengths of another. If we obtain consistent outcomes across studies using different strategies, we can be more confident that those outcomes have to do with the phenomena we are studying, and not just with our methods.

To summarize:

- (a) Methods enable but also limit evidence.
- (b) All methods are valuable, but all have weaknesses or limitations.
- (c) You can offset the different weaknesses of various methods by using multiple methods.
- (d) You can choose such multiple methods so that they have patterned diversity; that is, so that strengths of some methods offset weaknesses of others.

Given these principles, it should be why it is not appropriate to ask whether any given study is flawless, and therefore to be believed (as in the query, “But is that study valid?”). Rather, we should ask whether the evidence from any given study is consistent with other evidence on the same problem, done by the same or other researchers using other strategies and other methods. If two sets of evidence based on different methods are consistent, both of those sets of evidence gain in credibility. If they are not consistent, that inconsistency raises doubts about the credibility of both sets. How

much doubt we may have about the two sets of evidence depends on what else is known about the problem and the methods from still other studies. On the other hand, if all of the studies of a given problem have been based on the same methods, then that body of information is very much contingent on, and limited by, the flaws of those methods. Such a body of information must be regarded with some skepticism until you know whether it holds for a broader array of methods.

It should be noted here, though, that no one investigator is apt to be trained in the use of all methods, nor to have access to the resources needed for all of them. For example, some researchers have access to use of extensive and well designed laboratory facilities and are well trained in those methods but do not have ready access to the resources needed for a full scale sample survey, or for an elaborate field study. Other researchers may be in the reverse situation, with poor or no laboratory facilities but with excellent survey facilities and field study opportunities. What is crucial is not that a given researcher be able to use all methods on his or her research problem, but rather that the field as a whole make such use of diverse methods on each of its key problem areas. The fundamental principle, in behavioral and social science is that *credible empirical knowledge requires consistency or convergence of evidence across studies based on different methods*. These issues and their implications for behavioral and social science are discussed further in the parts of this chapter to follow, along with more detailed descriptions of strategies, comparison techniques, designs and methods.

RESEARCH STRATEGIES: CHOOSING A SETTING FOR A STUDY

Research evidence, in the social and behavioral sciences, always involves *somebody doing something, in some situation*. We can always ask about three facets: Who [which actors], what [which behaviors] and when and where [which contexts]. [The terms “actor”, “behavior” and “context” are used here as technical terms with meanings somewhat different from ordinary usage. Actor refers to those human systems, at whatever level of aggregation (e.g., individuals, groups, organizations, communities) whose behavior is to be studied. Behavior refers to all aspects of the states and actions of those human systems that might be of interest for such study. Context refers to all the relevant temporal, locational and situational features of the “surround” within which those human systems are embedded.]

When you gather a batch of research evidence, you are always trying to maximize three desirable features or criteria:

- A. **Generalizability** of the evidence over the populations of Actors.
- B. **Precision** of measurement of the behaviors that are being studied (and precision of control over extraneous factors that are not being studied).
- C. **Realism** of the situation or Context within which the evidence is gathered, in relation to the contexts to which you want your evidence to apply.

Although you always want to maximize all three of these criteria, A, B and C simultaneously, you cannot do so. This is one fundamental dilemma of the research process. The very things you can do to increase one of these three features reduces one or both

