Review: Research Design, Falsification, and the Qualitative-Quantitative Divide
Reviewed Work(s): Designing Social Inquiry: Scientific Inference in Qualitative Research by Gary King, Robert O. Keohane and Sidney Verba
Review by: James A. Caporaso
Published by: American Political Science Association
Stable URL: http://www.jstor.org/stable/2082441
Accessed: 03-11-2016 15:23 UTC
RESEARCH DESIGN, FALSIFICATION, AND THE QUALITATIVE–QUANTITATIVE DIVIDE

JAMES A. CAPORASO University of Washington

While disagreement may be more interesting than agreement, I preface my remarks by saying that I am broadly sympathetic to the arguments of Designing Social Inquiry by King, Keohane, and Verba. The authors have tried, with considerable success, to provide unifying principles and research strategies for qualitative and quantitative research. The central argument is that the rules for descriptive and causal inference have been unnecessarily restricted to quantitative designs. Good qualitative designs also profit from variance in the explanatory variables, proper measurement strategies, and control of extraneous variation. While there are legitimate differences between qualitative and quantitative research, KKV debunk the polarized images of the systematic quantitative researcher reducing politics to rows of equations versus the qualitative scholar giving solo performances with nonreproducible insight and Fingerspitzengefühl. In short, the authors' “reconciliation project” provides a methodological bridge connecting qualitative and quantitative research. Some may see reconciliation as conquest, since the unity achieved does take place on a particular turf, with particular standards. Yet the results are impressive.

KKV place strategies of inquiry, or research design, at the center of the book. In teaching research methods, what I find most useful are books that encourage us to construct research strategies that bring some probing value to our questions. Some books see all problems as resolvable by sophisticated statistical manipulation. The spirit of this book is quite different. With design at the center, the central issues are choosing appropriate units, ensuring variation in explanatory variables, and controlling for confounding influences. Weak (or indeterminate) designs cannot be salvaged by clever data analysis (p. 120). Research that is structurally defective in the sense that there is no variation in explanatory variables (or more explanations than observations) is doomed to fail, no matter how insightful the analyst. Without an appropriate organizing structure, additional data and even sophisticated analysis can tell us little.

My differences with the work are framed by the fact that my methods education was heavily influenced by two people never mentioned in this book: Hubert Blalock and Donald Campbell. Blalock's sociological contributions are premised on the notion that the ecology of social science is characterized by many independent variables, all intercorrelated and imperfectly measured, with feedback effects from the dependent variable. Many methodological problems are diagnosed in this notion: overdetermination, multicollinearity, error in variables (fallible measures), and endogeneity. Each of these problems is dealt with by KKV. Thus, while Blalock's enormous contributions are not noted (perhaps because they have been assimilated into modern statistical theory), the spirit of his work is well represented in this book. On the other hand, Campbell's quasi-experimental orientation is not only omitted but rejected early in the book (p. 7). Since Campbell and Stanley's Experimental and Quasi-experimental Designs for Research (1963) was—and to some extent still is—an important reference for empirical researchers spanning sociology, education, psychology, political science, and policy analysis, I take up KKV's categorical rejection of this type of research.

My reactions focus on three points: (1) the nature of qualitative research, (2) the meaning of falsificationism, and (3) the usefulness of quasi-experimental designs. The first two points are largely agreeable elaborations of positions taken in the book. The third represents a disagreement.

The Nature of Qualitative Research

KKV strenuously argue that the same rules of inference apply to qualitative and quantitative research. While I am inclined to agree, it is because I share the authors' definition of qualitative research as research based on in-kind rather than in-degrees differences. With this distinction, variance can be of two types: across categories (e.g., types of government, gender) and across quantities of the same variable (income, degree of labor repression). In measurement theory, qualities are represented as nominal variables, and quantities, as ordinal, interval, and ratio measures. Qualitative variation is not variation in magnitude, quantitative variation is. This characterization shows that it is not really numbers that are at issue (nominal measures are assigned numbers, too) but the issue of magnitude versus quality.

With this definition of qualitative research in place, the authors easily show that a sound qualitative research strategy requires attention to the same rules of inference as a quantitative strategy ("if x, then y" is not logically different from "as x increases, y increases"). But qualitative work can be conceived differently and in ways that are more resistant to KKV's reconciliation project.

For some, qualitative research signifies something different from explanations of in-kind variation. Indeed, the whole idea of systematic research harnessed to the goal of explanation is put into question. Thick description and interpretation may serve as ends, not merely as spadework preparatory to explanation. Scholars may be interested in empathetic understanding, the interpretation of meanings, and detailed investigation of single (nonvarying) cases. Some of the book's arguments (e.g., the rules of
A related point is that KKV’s arguments about differences and similarities between qualitative and quantitative research take place in a variable-centered world. This is not the only starting point. A variable-centered approach is already one in which variable properties have been abstracted from things, concrete names, and places. In the classroom, I find that the most difficult argument to make is not the unity of qualitative and quantitative research once a variable-centered model has been accepted but how one makes the transition from instances and concretely experienced sense data to variables. On this crucial issue, I know of no methodological guides. Between “Jumbo the elephant sliding down a grassy hill at Gasworks Park” and “a certain mass moving down an inclined plane with a given coefficient of friction” there is a gap. Neither logic nor observation obliges us to accept the second statement once we accept the first. Yet the leap has to be taken to reach the abstract world of variables. Hitler’s Reich as totalitarian regime; Austria, Norway, and Sweden as small, open, corporatist social democracies; and Brazil, Argentina, and South Korea as late-developing bureaucratic-authoritarian polities all represent examples of concept formation not forced by deductive or inductive logic (assuming one believes in the latter).

It may be that the urge to abstract is irresistible. Campbell was fond of arguing that theory and concept formation are “hard-wired in our retina,” reflecting the absence of theoretical innocence in our sensory equipment. In the end, I am in agreement with the authors’ but they have more careful work to do before the quantitative-qualitative gaps are bridged—and some will never be.

### The Meaning of Falsificationism

Science proceeds not only by hypothesis and conjecture but also by relentless attempts to reject our own theories. This does not mean that we hope our theories are wrong but that we believe them to the extent they survive difficult tests. The falsificationist perspective is important because it emphasizes the pruning-editing-winnowing side of science (Campbell and Stanley 1963, 35) in contrast to the confirmatory perspective that attempts to assess hypotheses by discovering confirming instances.

I accept KKV’s starting point—that confirmation and rejection are logically asymmetric. But the authors tend to see falsificationism in terms of deriving many implications of a theory, to increase the theory’s exposure to evidence. The problem with this criterion is that there is no guarantee (or greater likelihood) that the additional derivations will be any riskier than the initial hypotheses. A developed falsificationist perspective would add three points.

First, we should consider which of our theories’ implications are least likely to be confirmed if the theory is not true. This is another way of asking what the most distinctive explanatory-predictive content of the theory is. To predict that it will rain in Seattle during November is not risky. Similarly, to explain why strong states win out over weaker ones (using standard definitions of capabilities) is not risky. Anomalies are those outcomes which go against the grain. They are not what our prevailing intuitions and theories would have us believe. A recent case study of bargaining outcomes illustrates the point. Lisa Martin and Kathryn Sikkink (1993) compared U.S. pressure on Argentina and Guatemala to improve their human rights records. The puzzle motivating the study was that Argentina (larger, more powerful, more autonomous) caved in to U.S. pressure, while Guatemala successfully resisted (p. 332). The authors’ theory relied on a number of factors, among which was the strength of transnational human rights lobbies and organizations. Their theory is riskier in the sense that it explains an outcome different from what we would otherwise expect.

The intuitions embedded in this example is that many theories are compatible with a particular outcome. Outcomes are overdetermined. In this sense, confirmation is highly equivocal in theoretical terms. Theoretically consistent outcomes are a necessary but hardly sufficient aspect of a good research strategy. Instead of finding data that correspond to theory, why not first ask which of the outcomes implied by the theory are least likely to be true if the theory is not true? This question forces us to find the “reduced set” of outcomes that are most distinctively implied by the theory. The art of good research design is to identify those cases which can tell us the most in terms of distinct theoretical content.

The second point is derivative. The authors argue that testing our theories in alternative settings is a good idea. But what guides do we have for how to conduct these tests? The falsificationist perspective provides a criterion. Elaborate the implicative core of the theory in such a way that the multiple tests reduce the set of rival hypotheses (competing explanations) as much as possible. Carrying out the same test in the same setting provides little additional support for the theory. The same test in a different setting expands the scope of a theory and may add confirmatory weight if additional factors thought to influence the outcome are taken into account. But this is hit or miss. The researcher should isolate the set of implications that has the greatest nonoverlap in competing explanations. If a theory holds across highly diverse settings, this is more impressive than confirmation under similar conditions. The presumption is that rival explanations have a greater opportunity to register their influence under diversity. This point is crucial to the most different system design (Przeworski and Teune 1970). Using Durkheim’s theory of suicide as an example, Stinchcombe convincingly outlines the logic of this procedure (1968, 15–22).

Third, KKV could improve their argument by drawing out the links between falsification, quantitative reasoning, and the theoretical development of
our discipline. In part, this relates to Rogowski’s argument about strong theory (in the present symposium). One advantage of quantitative research is that it generates more precise predictions (often numerical values or ranges within which such values fall), which increase the difficulty of a test. Much of social science is at least implicitly about expectational standards. In statistics, one weak standard is the nondirectional null model. Findings departing from chance expectations in any direction are sufficient to reject the model. Another standard is provided by substantive theory. Do the results differ from what one expects after taking into account x, y, and z (this does not rule out a differently specified null model)?

My overall point is as follows. Improvements in measurement accuracy, theoretical specification, and research should yield a smaller range of allowable outcomes consistent with the predictions made. Cumulative improvements in knowledge should make our predictions riskier, more falsifiable. This seems to me to happen all too rarely in political science, in part because we are anxious to move on to new topics (skimming the cream from little investigated areas) and in part because we are more interested in presenting a “fresh look” or “new paradigm” than in using our collective achievements to define novel yet cumulative departures. We rarely report results in incremental (value-added) fashion, as additions to the existing capital stock. Instead, our results are presented as separate “findings.” We are confronted with perverse incentives. To take seriously the *acquis* of social sciences has the effect of increasing the difficulty of our tests in the sense of raising the “observational hurdles” required to accept a hypothesis (Meehl 1967, 103). Conversely, to ignore past achievements makes our hypotheses easier to accept—but at great costs in terms of lowered standards and cumulation of knowledge.

Quasi-experimental Analysis

The experimental method is often considered too narrowly as a battery of techniques applicable in a laboratory but irrelevant to the “real” world. By elevating experimental *procedures* over its *logic*, we lose the opportunity to learn what experimentation implies for ex post facto research. In broad terms, the biggest achievement of experimental design is the preexperimental equivalence of groups through random assignment (Campbell and Stanley 1963, 2). The power of random assignment is often not fully appreciated in social science research. The important distinction between random assignment and random sampling is elided. Random sampling does not solve the problems of drawing inferences when numerous causal factors are associated with outcomes. By contrast, the capacity of the experimenter to assign units (usually people) to treatment and control groups neutralizes nearly all subject-centered threats to validity. Experimental control over the “how much” of x assures adequate variation in the independent variables. Control over the timing (the *when*) of exposure implies a solution to the endogeneity problem (since values of the independent variable can occur independently of the dependent variable).

The logic of experimental research provides guidelines in ex post facto settings. The random assignment technique directs us to find ways to control extraneous variables, for example, by using stratified designs that reduce variation in confounding variables or by building in variation and doing partial correlation and regression analysis. The manipulation procedure translates into the scheduling of units and observations so as to assure variation on the independent variable. Ex post facto research is the “continuation of experimental logic through other means.” On this important philosophical point, I do not think there are differences with KKV. Why, then, do they reject quasi-experimental analysis?

They say “We reject the concept, or at least the word, ‘quasi-experiment’” (p. 7n.). They further state that “investigator control over observations and values of the key causal variables” is the determining factor in deciding whether something is an experiment. Two points need to be made. First, researcher control over values of the independent variables is not enough to define experimentation. The ability to assign randomly is also crucial as is experimental isolation (a lab). In a pure experiment, the three properties go together. Without manipulating the independent variables, we cannot be sure that hypothesized effects will have a chance to occur. Without random assignment and laboratory isolation, we cannot be sure we would detect such effects even if they did occur.

The second point is more nuanced. If KKV mean that quasi-experimental designs do not represent a logically distinct category, I agree. However, the numerous designs pioneered by Campbell and Stanley (1963) were possible because they “unpacked” three properties that merge in pure experiments (manipulation of the independent variable, random assignment, and lab setting). These properties were then combined in various ways to produce various hybrid designs (see Achen 1986; Cook and Campbell 1979). For example, a field experiment allows for some ability to manipulate the independent variable but no control over random assignment and setting. Other designs allow random assignment (e.g., of court cases to different processing procedures) but no ability to affect the independent variables.

Quasi-experimental designs strive for three things: (1) natural settings in which abrupt variation occurs in independent variables, (2) some natural controls as one is likely to find when adjacent units of analysis experience different “treatments,” and (3) a checklist of concepts and techniques to use to address internal and external validity. The interrupted time-series design and multiple control-series design provide valuable controls in many situations of interest to the political science researcher. Indeed, Campbell’s (1969) “Reforms as Experiments” is a model for the policy analyst attempting to unravel complex interactions in the policy process. Because these designs are...
meant for a world of biased selection, differential exposure to threats to validity, measurement error, and researcher wish fulfillment, I find them helpful. Finally, I find Campbell and Stanley’s distinction between internal and external validity helpful—indeed almost necessary—to assessment of the value of any design. Internal validity is fundamental; it is the starting point: “Did x have an impact on y?” External validity asks the question, “To what other groups, units, populations can these results be extended?” The former question concerns us all, but the second is often reduced to a sampling instability issue, which it is not. It is a contextual (hence theoretical) issue: “If it is true in the Bronx, will it also hold in Cook County?,” rather than “How are elements 1–30, sampled randomly from a population, different from 31–60, sampled from the same population?” The distinction is all the more helpful in that it maps nicely onto main effects versus interaction effects (main effects are threats to internal validity, interactions, to external validity). Thus, to take one example, if selection biases operate independently of one’s hypothesized causal variable, it is a threat to internal validity; if these same selection factors interact with the causal variable, it is a threat to external validity. In the former case, x (causal variable) had no impact. In the second case, x did have an impact, but only because of its conjunction with selection factors.

Let me bring my comments back to the starting point. I regard this as a fine book that will be widely used in methodology courses. I am persuaded that much of what goes under the label of qualitative research is concerned with explanation and causality and must therefore be attentive to the main arguments of this book. By outlining a research strategy applicable in both descriptive and causal settings and relevant to qualitative and quantitative research, KKV hold the promise of unifying previously fragmented parts of our discipline. At the very least, Designing Social Inquiry encourages us to talk to one another and to learn more precisely where our differences lie.

Notes

I am grateful for the comments of Anthony Gill.

1. I believe this phrase, or something like it, is attributable to Sir Arthur Eddington.

2. Prediction, unlike forecast, requires a theoretical structure. The logical structure of a prediction is, “If x occurs, then y will occur.” By contrast, a forecast simply asserts that y will occur in the future as a result of extrapolation (“casting forth”) of y.

3. It does to some degree. Random sampling helps to eliminate chance as a factor explaining an association. However, if many variables are correlated (confounded) in the population, random sampling will only provide a more accurate assessment of this confounding. It will not control these variables in the sense of neutralizing their influence.

4. By subject-centered threats to validity, I mean those differences among groups which are the result of differences among the individuals that compose the groups. Random assignment does not control for differences in the environment of the groups (differences irrelevant to the treatment) or variation that the experimenter may introduce by treating the two groups differently.