(i)

Arguing Claims for Validity in Case Study Research

Gerald Elsworth
Centre for Program Evaluation, Institute of Education
The University of Melbourne

(DRAFT FOR DISCUSSION ONLY: Please do not cite or quote in this form without the permission of the author)

Paper Presented at the Annual Conference of the Australian Association for Research in Education, Surfers Paradise, November, 1991

Arguing Claims for Validity in Case Study Research

Gerald Elsworth
Centre for Program Evaluation, Institute of Education
The University of Melbourne

Introduction

For perhaps twenty years1 the so-called "paradigm wars" (Gage, 1989) have immersed educational research and evaluation in a frequently vitriolic, mostly self-defeating debate about competing epistemologies and appropriate research strategies. Even within paradigms, protagonists have found significant areas of disagreement, particularly about the history and structure of the discourse on their preferred position (see, for example, the recent debate around the notion of "traditions" of qualitative research in Jacob, 1987; Atkinson, Delamont, and Hammersley, 1988; Buchmann and Floden, 1989; Jacob, 1989; and Lincoln, 1989). Yet it seems apparent that a resolution of the conflict cannot easily be found in the kind of methodological pragmatism advocated by Gage and, amongst others, Cronbach and Associates (1980, Ch. 4), Miles and Huberman (1984) and Salomon (1991). The continuing debate in such journals as the "Educational Researcher" and the forceful statements in early chapters of books such as that by Guba and Lincoln (1989) suggest that a strategic retreat by either side is unlikely; the contestation and polarization represented by Gage's military metaphor is too entrenched, the competing positions too firmly held, and much too much work has gone into the critique and counter-critique that have lead to their formulation.

There is, at the least, an analogy here with Lakatos' account of the structure of an on-going scientific research program, in which the highly resistant "hard core" propositions of the program are distinguished from the "protective belt" of auxiliary assumptions, ad hoc hypotheses and positively interpreted empirical results (Phillips, 1987). In an evolving program, elements of the protective belt may be modified or discarded in order to preserve the core. The paradigm wars have framed the development of rationalities for social research in which contrasting methodologies are given explicit and detailed ontological and epistemological justification. These justifications might be seen as the core of a methodological program in which the particular research strategies constitute auxiliary elements; the refutable yet protective belt. From this viewpoint, shifts between research strategies within a methodological paradigm might be accommodated relatively easily. But, given that competing paradigms are truly distinctive and have conflicting yet supportable rationalities, the pragmatic selection and mixing of strategies from each might involve ontological and epistemological commitments that are much closer to the core and which thus might be considerably harder to dislodge or ignore.2

In addition to the contestants, the paradigm debate has had a number of important commentators and analysts, the front-line correspondents (albeit sometimes quite partisan), of Gage's wars. Phillips (1987), for example, has presented the conflicting issues, properly viewed, as a contrast between a new conception of "naturalistic" social science on the one hand and, on the other, various anti-naturalist arguments and commitments to specific methodologies. As a group, the latter are characterized as The

Position.3 There were three general factors, Phillips (1987) proposed, which led to the recent critiques by those who hold the Position. They were:

(a) a growing awareness by social scientists of the work being done in the new philosophy (and sociology) of science and the drawing of a "relativistic moral" from these contributions;

(b) a strengthening interest in European modes of thought in the humanities, in particular phenomenology and hermeneutics; and,

(c) the discovery of new practical difficulties within the social sciences themselves which had led to a growing uncertainty that the received methods of the physical sciences would yield appropriate answers.

Phillips pointed out, however, that a frequently unrecognized "Catch 22" lurked in the arguments of many of the critics. He suggested that

if the "new philosophy of science" is on the right track - and, as indicated, many of them (the critics) believed that by and large the new points are sound - then, in this new and liberal sense of "science", there would seem to be little bar to the social sciences being regarded as scientific. On the old and now discredited view of science it no doubt was quixotic to suppose that the social sciences could be naturalistic; but of course not even the physical sciences were naturalistic in this sense (for this account of naturalism was inadequate)! (Phillips, 1987, p. 49)

In a similar manner, recent counter-critiques of the rationality of the Position have emphasised that much has been built around an attack on the "straw paradigm" of "positivism". Many aspects of positivism have been rejected in the natural sciences, in particular, various tenets of logical positivism such as the verifiability principle of meaning. It is claimed in contrast that there is much in their rejection of the possibility of theoretical explanation in the social sciences, the primacy given to direct sensory observation and the importance accorded tacit knowledge which brings supporters of the Position quite close to the viewpoints held at certain times by the logical positivists (Phillips, 1983, 1987). Reform was under way in the philosophy of the natural sciences and was influencing many areas of research in the social sciences well before the constructivists had their critique ready for a full launching; reform which seems to have been almost completely ignored in many of the arguments which have framed the Position. No better examples of this attack on a paradigm that, itself, seems to have been "constructed" to facilitate the rhetoric of the contest can be found than in the opening pages of "Fourth Generation Evaluation" (Guba and Lincoln, 1989 pp. 8, 12). Here, for example, the authors proclaim that

it is our intention to define an emergent but mature approach to evaluation that moves beyond mere science - just getting the facts - to include the myriad human, political, social, cultural, and contextual elements that are involved.

And that

in the past, the methodology employed in evaluations has been almost entirely scientific, grounded ontologically in the positivist assumption that there exists an objective reality driven by natural laws, and epistemologically in the counterpart assumption of a duality between observer and observed that makes it possible for the observer to stand outside the arena of the observed, neither influencing it nor being influenced by it. (italics in original, underlining is my emphasis)

Science (and by implication, the social science of those who hold "positivist" points of view) is characterized as being atheoretical, yet holding to a belief in immutable natural laws; as having no interest in human, political, social, cultural or contextual matters; and as being based on a duality which results in observers being uninfluenced by their observations. What curious people the evaluators who work within Guba and Lincoln's "positivist" paradigm are - researchers who ignore the very stuff their work deals with; practitioners of an applied science who none-the-less believe in a static empiricism which will reveal natural laws in the absence of theory; and detached observers who, being uninfluenced by their observations, can make no modifications at all to their a priori expectations, neither supporting nor refuting them!

In this sense then, the paradigm wars might be seen as illusory. Far from being entrenched behind its "straw paradigm" of "positivism", mainstream educational research and evaluation itself has been actively experimenting with new paradigms, incorporating many insights from the new philosophy and sociology of science, and embracing a much more liberal conception of the social sciences. An understanding of this paradigm shift makes redundant much of the critique of those who hold the Position and, because it was based largely on the "negative strategy" of the critique, renders vulnerable much of their epistemology. Hence, the wars, themselves, turn out to have been a one-sided and unnecessary construction, an attempt to legitimate privileged ownership of ethnographic, hermeneutic and phenomenological thought and method in educational research. This paper, then, starts from the proposition that the constructivist arguments for these methods were falsely contrived, requiring those who practice interpretive educational research (be it qualitative or, indeed, quantitative) to accept an unwieldy and restrictive epistemology or to abandon altogether their attempts to address new research problems with appropriate, yet rigourous, methods.

Scientific Realism and the "New Social Science"

Postpositivist social science has been evolving for many years, in interaction with more general developments in the philosophy and sociology of science. As House (1991, p. 2) has recently pointed out (see also Bhaskar, 1989), there are various versions of the new "realist" conception of science and a considerable array of contributors, including those who have taken an "antimonistic" view, emphasizing instead the social character of science (for example Popper, Lakatos, Feyerabend, Sellars and Kuhn) and those who have developed an "antideductivist" approach to explanation placing particular emphasis on the role played by models, metaphors and analogies (Scriven, Polyani, Toulmin, Hesse, Harre and Bhaskar). Some of the relevant work of Popper, Polyani and Kuhn is well known to educational researchers (although it seems frequently to have been misinterpreted; Phillips, 1987), less so, however, many other important contributions, including the earlier writing of Scriven. Following House (1991) Bhaskar's account of scientific realism is taken as exemplary of these developments.

Bhaskar (1978) argues that events must be seen as occurring independently of the experiences through which they come to be known and, further, as independent of the "structures and mechanisms" that generate them. Reality is stratified into three distinct levels; experiences, events and generative structures, and the existence of phenomena at each level is independent of the existence of phenomena at other levels. Both positivism and its combatants incorrectly equated events and structures with experiences and thus concluded that reality could only be known in the domain of immediate experience (the empirical). Thus a fundamental tenet of Bhaskar's realism is that, in addition to experiences, these events, and structures and mechanisms, are real. Scientific explanation, furthermore, consists of descriptions of the manner in which structures act and interact to generate events and experiences. Therefore,

the construction of an explanation for some identified phenomenon will involve the building of a model, making use of cognitive materials and operating under the control of something like a logic of analogy and metaphor, of a mechanism, which if it were to exist and act in the postulated way would account for the phenomenon in question....And so we have in science a three-phase schema of development, in which in a continuing dialectic, science identifies a phenomenon (or range of phenomena), constructs explanations for it and empirically tests its explanations, leading to the identification of the generative mechanism at work, which now becomes the phenomenon to be explained, and so on. On this view of science its essence lies in the move at any one level from manifest phenomena to the structures that generate them. (Bhaskar, 1989, pp. 68-69; italics in original, underlining my emphasis).

Bhaskar (1989) has also dealt with the question of whether this account of realism in the natural sciences is appropriate to the domain of the social sciences (i.e. the problem of naturalism). He asserts that it is, but under certain, limiting, conditions. These are that:

(i) social structures do not exist independently of the activities they govern;

(ii) social structures do not exist independently of the agent's conceptions of what they are doing in their activity; and

(iii) social structures may be only relatively enduring (Bhaskar, 1989, p. 79).

In addition to these limits, Bhaskar (1989, pp. 82-83) also argues that, by virtue of the fact that social objects are manifest in "open systems" where there are no invariant regularities, predictive theories are not possible; social science theories are therefore restricted to being "exclusively explanatory". But, crucially, these limits do not affect the form of causal laws that might be developed, only the criteria for accepting them and, possibly, the confidence with which they may be held.

Shifting to the question of an appropriate methodology for the social sciences, Bhaskar (1989, p. 83) proposes that, because the "methodological directives" generated by orthodox philosophies of science presuppose closed systems, they are "totally inapplicable" to the social sciences. He thus draws the radical conclusion that

Humean theories of causality and law, deductive-nomological and statistical models of explanation, inductivist theories of scientific development and criteria of confirmation, and Popperian theories of scientific rationality and criteria of falsification, together with the hermeneutical contrasts parasitic on them, must all be totally discarded. (Bhaskar, 1989, p. 83).4

This conclusion, however, does not impair the possibility of independently validated knowledge in the social sciences, only the expectation that experimentation (along with the logical and statistical procedures developed in the natural sciences which determine the design of critical experiments and the appropriate interpretation of their results) can lead to this knowledge. Bhaskar argues that, while attempts to define the reality of forms of social life will, in general, precede attempts to generate causal hypotheses, both can be empirically validated by the explanatory power of the hypotheses derived from them. Rigourous procedures for the validation of descriptive and causal claims in the social sciences, then, are still required. And while their logical structure might follow that of validation procedures in the natural sciences, their content necessarily will be different, being concerned with the force of explanatory, rather than predictive, conclusions.

Parallel Developments Within "Mainstream" Social Science

From within social science, and more particularly educational and social program evaluation, the development of this new scientific realism has been paralleled by the work of a number of individual authors. Of particular interest is that of Campbell and his colleagues (see, particularly, Brewer and Collins, 1981; Cook, 1987; Campbell, 1988). While rarely recognized, this aspect of Campbell's writing has opened a large "window of opportunity" for mainstream educational researchers to utilize explanatory and interpretative methods.

Having expressed a cautious interest in qualitative methods and case study designs in earlier papers (collected under the section heading "Interpretative Social Science" in Campbell, 1988), Campbell (1986) made a particularly important move in a paper titled "Relabling internal and external validity for applied social scientists". The new label that he proposed for internal validity was "local molar causal validity", with the additional qualifications of "pragmatic" and "atheoretical". Far from simply being a rhetorical or pedagogical device, the relabling represented an important clarification in the evolution of Campbell's "critical realist" epistemology.

The criterion of internal validity, Campbell (1986, p. 68) argued, had come to be misunderstood as indicating similarity to "pure treatment (rule-of-one-variable), fully controlled, laboratory experiments", a meaning that was never intended. The new label was designed to emphasise that this form of validity was concerned specifically with the truth of the claim that a treatment as a complex package, implemented in a particular setting, had indeed made a simple real difference to the desired outcome in the direction intended. Campbell's notion of threats to validity can be viewed as a dialectic of refutation, a particular case of Bhaskar's dialectic which underpins the development of theory.5 A threat to validity is a plausible alternative hypothesis that can be put forward to support the claim that something other than the treatment package was responsible for the observed change. This alternative hypothesis can, in its turn, be refuted by a strong design and/or a "patch-up" post-hoc analysis of further data (Cordray, 1986). Procedures for establishing local molar causal validity therefore consist of design elements and/or post-hoc analyses that strengthen the claim that the program package worked as intended. A study designed to establish local molar causal validity is seen as a step, in an evolving scientific program, that is intermediary between exploratory quantitative and/or qualitative studies that suggest plausible constituents of the package, and subsequent validity studies which partition and refine these constituents. In Campbell's words

the molar approach assumes that clinical practice, participant observation, and epidemiological studies already have accumulated some wisdom, suggesting treatments that are worth further testing as molar packages. If these packages turn out to have striking molar efficacy we will, of course, be interested in further studies, both clinically and theoretically guided, that help us to determine which of several conjectured major components is most responsible for the effect. These later studies in turn will still be using complex packages, rather than testing theoretically pure variables in isolation or in experimentally controlled higher-order interaction. Pure-variable science can, of course, be a source of treatment packages (as in brain metabolite therapy for children diagnosed as potentially schizophrenic due to metabolite abnormality), but in preventive intervention these, too, will inevitably be part of a complex system of diagnosis and delivery. (Campbell, 1987, p. 69).

As always with Campbell's writing, there is much embedded and implied in this short passage. He has long held the view that to be able to establish with some certainty that a complex treatment package had a desired effect within a particular context is all that reasonably can be hoped for in applied social science at the present time. This view is clearly evident in the concern with the "molar" efficacy of treatments. Also evident is Campbell's relative subordination of the role of theory and cautious support for qualitative methods and clinical judgement. In addition, Campbell's continued concern with the conditions under which "molar" program impact can reasonably be claimed echoes his commitment to a pragmatic conception of causality which rests on the observation of contingent events in approximately closed systems. It is acknowledged that social systems are open and hence cannot be subject to laboratory-like single-variable control, but social science none-the-less must aim for the fullest possible specification of contingent relationships. While "molar causal relationships in society may not be reproduced easily", causation never-the-less

implies that by varying one factor I can usually make another vary. For many valid causal laws we may not in practice be able to manipulate the putative cause at will, if at all. This has grave consequences for our ability to test the law, but it does not negate its truthfulness. However, it does decrease the immediate practical importance of the law, for it suggests that the causal powers implicit in the law cannot be easily used to make desirable changes in persons or environments. (Cook and Campbell, 1979, pp. 35-36).

It is not easy to read Campbell correctly on this point, but it seems reasonable to conclude that, though he initiated a new epistemology in social research and evaluation, he has not fully embraced its methodological implications. As argued by House, "scientific realism" involves a new role for the concept of causality. "The actual world is a world of incompletely described and incompletely known causal agents" (House, 1991, p. 5), and scientific explanation is not an account of the regularities to be observed in contiguous events, but rather an account of the interaction between structures which may, under certain conditions, result in observable events. Campbell would appear to agree, but, through his conception of "local molar causal validity" he seems to reaffirm the primacy of a contrived experimental or quasi-experimental set-up in producing the conjunction between events (intervention and impact) which, he believes, is the best way to arrive at a causal understanding of social programs at the present time. The "relabling" is seminal none-the-less as, by shifting from the generic notion of "internal" validity to a qualified form of "causal" validity Campbell has (deliberately?) left the way clear for speculation about other forms of causal validity and the investigation of the relative strengths of alternative research strategies in relation to them.

Scientific Realism and The Case Study in Educational Research

The realist view of the nature of scientific explanation is, at the least, sceptical of the role of experiment in the social sciences. Experiments are artificially contrived to bring about a constant conjunction of events such that the proposition of causal efficacy can be held open to refutation. To the extent that constant conjunctions of events are rare in the real world and that causal explanation from the realist perspective, crucially, involves an account of the manner in which generative structures are related to events and experiences in open systems, experiments are unlikely to be able to give a valid account of real world phenomena. As House (1991) suggests, educational programs must be seen as both events that are themselves produced by causal structures, and as structures which might be expected to cause events, only some of which might be manifest in a particular implementation. Multiple interactions between events and structures, and the possibility of complex feed-back relationships within and across levels must also be considered (Cronbach and associates, 1980). It might be added that, in-so-far as prediction is not possible in open systems, it is not logically possible to construe which events might be the outcomes of any particular implementation of a program; hence the a priori selection of outcomes to monitor in any experimental implementation is hazardous.6 The realist analysis of social science therefore suggests that alternative methods which directly address the problem of explanation and which can "stand alone" in providing valid accounts of the causal relationships between generative structures and events in the social world need to be explicated.

If the extent of formal publication is taken as an index, it might be concluded that non-experimental "explanatory" research in education has been dominated in recent years by large-scale cross-sectional and longitudinal surveys and associated structural and causal modelling techniques for the analysis of data. In contrast to these "extensive" research procedures, "intensive" research (Stoecker, 1991), especially the case study, has had a more modest "public" profile. But, frequently in the form of "multi-site" evaluations of the implementation of federal programs, case studies have flourished in the field of policy-related commissioned research. In the United States, a marked swing in Federal evaluation funding away from large-scale surveys seems to have occurred in the late 1970s, and by 1982 twenty-five separate studies funded at over one million dollars each were reported to be underway (Herriot and Firestone, 1983). Similarly, in Australia, the large-scale evaluations of the Transition Education, and Participation and Equity programs were based on multi-site case study methods (Kemmis et. al., 1983; Elsworth and Hartley, 1988; Owen and Hartley, undated; Hartley and Owen, undated).

Intensive research methods including the case study appear to be ideally suited to generate the kind of explanatory causal inferences demanded of social science by the realist perspective. Certainly, a strong case can be made for the case study over extensive cross-sectional survey research in this regard (Stoecker, 1991). It can be shown by the path analysis of survey data, for example, that being brought up in a family of non-English speaking background increases the likelihood of a young person making the transition from secondary school to higher education (Elsworth, Day, Hurworth and Andrews, 1982). But even the most detailed attention to the analysis of mediating effects can only offer fairly gross indications of the causal mechanisms involved. To gain a clearer view of the actual causal processes at work within the families, and to develop and test satisfactory theories about these processes, the more detailed and intensive data available through case studies are necessary. (It might be added, of course, that this is just one of the vast number of problems in contemporary social research where experimental studies are logically impossible.) As a final stage in the analysis, the causal hypotheses developed from the case studies might be validated by matching to the mediating effects discovered from the analysis of survey data (Elsworth, Day, Hurworth and Andrews, 1982) or by new (qualitative?) forms of meta-analysis, a methodology which House (1991) argues is also congruent with a realist conception of causation. But the "yield" of usable knowledge from this kind of "two-stage" procedure would be greatly enhanced by the development of appropriately rigourous procedures for the validation of the first-level inferences derived from the intensive strategies themselves.

Arguing the Validity of Inferences from Case Studies

While there has been considerable debate about the validity of the causal inferences that are frequently drawn from cross-sectional and longitudinal studies, little formal attention has been given to validity issues associated with the case study and related "intensive" methods. The problem has frequently been handled by relegating the case study to a secondary role in hypothesis development or theory confirmation, roles in which the method is not required to stand alone to support causal inferences. Campbell, for example, has discussed these two possible roles for case studies. In the paragraph quoted in a previous section, he suggested that "clinical practice" and "participant observation" (both of which might be regarded as informal case studies) are useful in suggesting the detailed strategies which might be included in intervention packages, and earlier, (Campbell, 1975) he elaborated a role for the "single culture case study" in theory confirmation (or more crucially disconfirmation). To the extent, however, that validity issues in case study research have been discussed in the literature, they are either:

(i) addressed in terms of analogies with the problem of the validity of experimental designs, as, for example in Campbell's (1975) argument that the multiple observations carried out in case study work and the multiple implications for theory disconfirmation might be regarded as comparable to the multiple replications in an experiment and the associated degrees of freedom of a statistical analysis; or

(ii) consist of suggestions for specific strategies,which, in some unexplicated general way, are proposed to enhance validity (see, for example the "parallel criteria" of trustworthiness proposed by Lilcoln and Guba (1986) and Stoecker's (1991, p.106) suggestion that checking back with respondents or participants in the research is "perhaps the best validity check" .

It is useful, here, to examine further the idea that a claim for the validity of an inference in social science is part of a dialectic of refutation. Drawing on a conceptual framework proposed by Toulmin (1958), Dunn (1981) has elaborated this proposal into a formal structure in which a threat to the validity of an inference is viewed as a rebuttal in a socially transacted argument. There are six elements in a transactional model of argument as opposed to three in the classical syllogism (see Figure 1). They are the data (D), claim (C), warrant (W), backing (B), rebuttal (R) and qualifier (Q) (Dunn, 1982, p. 306). The data element is equivalent to the minor premise in the syllogism, the warrant the major premise and the claim, the conclusion. In Toulmin's model of substantive social argument these elements are augmented by the second triad. The backing consists of additional data, claims or arguments that support the status of the warrant if it is in doubt, the rebuttal consists of conditions under which the relevance of a claim can be challenged, and the qualifier expresses the extent to which the force of a claim needs to be modified in the face of the rebuttals. When applied to the problem of the validity of a knowledge claim, the rebuttals of the transactional model take the form of threats to validity. When these are known, as in the classical set of threats to local molar causal validity developed by Campbell and Stanley (1963) and Cook and Campbell (1979) the rebuttals can either be used to: (a) evaluate a prospective design for a research study by suggesting plausible alternative explanations for an anticipated positive effect, or (b) judge the worth of a conclusion that has already been drawn on the basis of an implemented design. Dunn (1982, p. 307) illustrated the second case using the well known quazi-experimental analysis of the impact of the British "breathalyser" legislation (Ross, Campbell and Glass, 1967).

The primary purpose of Dunn's paper, however, was to demonstrate how the transactional model might be applied at a "second-order" level to evaluate more general threats to "usable knowledge" of which threats to "knowledge adequacy" (including threats to validity) are one subset. Irrespective of the generality of the objective, the key to the proposed strategy is the identification of threats as rebuttals in a socially transacted argument about the value of a knowledge claim. When applied to the issue of validity, application of the model also presupposes the formal isolation of threats as a necessary precondition for the development or application of research strategies which will test the causal proposition while, concurrently, addressing the threats (i.e. threats are plausible rival hypotheses to a causal explanation). The transactional model is therefore a general strategy, independent of the particular nature of the research design that generated the knowledge claim, and equally applicable to explanatory as well as predictive research and to qualitative as well as quantitative data. As Dunn (1982, p. 308) pointed out, the

transactional model can also assist in transforming the empirico-analytic and hermeneutic sciences into critical ones....since the model forces the inspection of causal and ethical assumptions, as well as their underlying warrants, as part of a social process of interpreting qualitative and quantitative data.

Some Possible Threats to the Validity of Case Study Research

The arguments developed in earlier sections of this paper suggest that case studies are a plausible methodology for generating the explanatory causal inferences required by a realist perspective on the social sciences and sought with increasing frequency by practicing researchers, especially, perhaps, those working in the more applied fields such as evaluation. The challenge now is to develop rigourous procedures for their validation. Following Dunn's arguments, these procedures seem to require the a priori recognition of threats to validity . Given that explanatory research necessitates the identification of causal inferences subsequent to data collection, potential threats must be generated as general classes of plausible rival hypotheses from the accumulated experience of theorists and practicioners in the field. Once the range of potential threats is established, those that represent specific rival hypotheses for a particular causal claim might be identified through more or less formal applications of the transactional model of argument (the more formal applications following, possibly, some form of quazi-judicial process). This suggests a cyclic or iterating approach to explanatory research through which causal hypotheses are generated from early observations, subject to a process of "threat identification", and retested against data from the same case that are subsequently gathered specifically to test the plausibility of the threat. Those hypotheses from a particular case that withstand the test against plausible rivals might then form the basis for further studies where the cases are purposefully sampled to test the generality and robustness of the claims, or, in order to avoid the potential problem of the "decay" of the generalization over time, "pattern matched" (Campbell, 1966), through an appropriate meta-analysis, to hypotheses developed from further case studies undertaken as part of the same multi-site study.

Table 1 has been constructed from the results of a preliminary review of an idiosyncratic selection of literature in program evaluation and sociological method. It contains suggestions for six possible categories of threat to the validity of explanatory inference in case study research and evaluation. The categories are ordered roughly in terms of the focus of the group of threats on observation, classification, causal inference and valuing in case study research. The list is not meant to be definitive in any way; simply as an illustration of the form that threats to the validity of explanatory inferences might take, and as an heuristic for further research and debate.

TABLE 1

Six Possible Categories of Threat to the Validity of Observation, Interpretation and Evaluation in Case-studies

1. Selection (of events and settings). Analogous to Campbell and Stanley's (1963) threat to causal inference of "selection" of individuals for comparison groups. This group of threats includes various limitations of attention (such as looking in the wrong places, looking at the "ground" instead of the "figure") and the effect of efforts by "gatekeepers" in the field to direct the fieldworker's attention to certain events and settings rather than others. (Campbell's metaphorical treatment of visual illusions is pertinent here.)

2. Premature classification/Mis-classification. Threats in this category occur when an observation is noted to be a member of a particular class when it is better regarded as being in another. Examples are widespread and may often result from the use of psychological (eg. Freudian) or sociological (eg. Marxian) typologies to label an observation or event prematurely, with the consequent "stickiness" of the label. At a more "macro" level, threats of classification have been discussed in relation to a perceived tendency towards compulsive and premature theorizing in the social sciences.

3. Simple inductive triangulation. Use of the (frequently recommended) strategy of drawing a firm conclusion about a class of events from the observation or report of a small number of members of the class. There are at least two general problems with this strategy: (i) the logical problem of justifying the inference when all members of the class have not been observed, and (ii) the likelihood the observations "stored up" for later use in developing a theory about a class of events will "degenerate" or "decay" as other factors come into play (Phillips, 1987). A related issue is the potential "rubberiness" of the criteria that an observer uses to include an observation as a member of the class of events being "stored" and to disregard others. Should be carefully distinguished from "multiple method" triangulation where an active attempt is made to refute a generalization about the behaviour of a class of events by showing that it doesn't hold when an alternative method of observation or measurement is used.

4. Misjudged salience. Judging and subsequently interpreting an observation, event etc. as salient to the case in the context of the aims of the study when in fact it is not. Often discussed as a threat associated with survey methods (eg. the coercive effect of previously developed or readily available questionnaire items and scales on the research focus) but is equally applicable to qualitative observations and interpretations. In evaluation, this threat is frequently manifest in conflicting claims for the salience of problems, observations, data etc. by competing groups of "stakeholders". In critical theory, cognisance of this threat led to the pejorative application of the label "Methodological Empiricism" to quantitative sociological research on status attainment, the implication of the label being that researchers in this tradition were seduced by their methodological expertise and interests into misjudging the salience of certain problems in sociology that were particularly amenable

to their method of analysis.

5. Spurious causal attribution. The attribution of a prior and correlated event as the "cause" of a later event . Threats in this category include but are by no means exhausted by Campbell's classical threats to "local molar causal validity" (history, selection, maturation, mortality, statistical regression, ambiguity about the direction of a causal relationship etc.). Importantly, this group of threats also includes the interpretation (for eg. on the basis of persistent informant claims or the use of "time-ordered", "site-dynamics" or "process-outcome" matrices in data analysis) of a reported or observed correlate of an outcome as a "cause" of the outcome (whatever the language used, eg. the event that "led to", the reason "why" an event occurred).

6. Spurious/coerced positive (or negative) interpretation or valuing. Involves the problem of coercion of the case-study worker to the point of view (criterion) held by program participants or managers through the perceived imperative to repay the "gift" of access to their program. (May be generalized to the problem of repayment for the gift of access to any culture or case-study setting; a qualitative researcher cannot "hide behind the numbers"!) Also involves the use of a criterion (criteria) to allocate value to an event (product, program, process, outcome etc.) when another criterion is more appropriate (e.g. normative vs rational models of valuing). Related, in this sense, to the problem of misjudged salience.

REFERENCES

Atkinson, P., Delamont, S., and Hammersley, M. (1988). Qualitative research traditions: a British response to Jacob. Review of Educational Research. 58, 231-250.

Buchmann, M., and Floden, R. E. (1989). Research traditions: diversity and progress. Review of Educational Research, 59, 241-248.

Brewer, M. B., and Collins, B. E. (Eds.) (1981). Scientific Inquiry and the Social Sciences. San Francisco: Jossey-Bass.

Campbell, D. T. (1966). Pattern matching as an essential in distal knowing. In K. R. Hammond (Ed.). The Psychology of Egon Brunswick. New York: Holt, Rinehart and Winston.

Campbell, D. T. (1986). Relabling internal and external validity for applied social scientists. In W. M. K. Trochim (ed.) Advances in Quasi-Experimental Design and Analysis. New Directions for Program Evaluation, no 31. San Francisco: Jossey-Bass.

Campbell, D. T. Methodology and Epistemology for Social Science. (1988 - E. S. Overman, ed.). Chicago and London: The University of Chicago Press.

Campbell, D. T., and Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (ed.) Handbook of Research on Teaching. Chicago: Rand McNally.

Cook, T. D. (1987). Postpositivist critical multiplism. In W. R. Shadish and C. S. Reichardt (Eds.) Evaluation Studies Review Annual, Volume 12. Beverly Hills: Sage.

Cook, T. D., and Campbell, D. T. (1979). Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

Cordray, D. S. (1986). Quasi-experimental analysis: A mixture of methods and judgement. In W. M. K. Trochim (ed.) Advances in Quasi-Experimental Design and Analysis. New Directions for Program Evaluation, no 31. San Francisco: Jossey-Bass.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., and Weiner, S. S. (1980). Toward Reform of Program Evaluation. San Francisco: Jossey-Bass.

Dunn, W. N. (1982). Reforms as arguments. Knowledge: Creation, Diffusion, Utilization, 3, 293-326.

Elsworth, G. R., Day, N .A., Hurworth, R., and Andrews, J. (1982). From School to Tertiary Study: Transition to College and University in Victoria. ACER Research Monograph No 14. Hawthorn. Australian Council for Educational Research.

Elsworth, G R, and Hartley, R. (1987). The Implementation and Impact of PEP. Final Report of the National Evaluation of the Participation and Equity Program. Melbourne: Centre for Program Evaluation, Melbourne College of Advanced Education.

Guba, E. G., and Lincoln, Y. S. (1981). Effective Evaluation. San Francisco: Jossey-Bass.

Guba, E. E., and Lincoln, Y. S. (1989). Fourth Generation Evaluation. Newbury Park, CA: Sage.

Herriot, R. E., and Firestone, W. A. (1983). Multisite qualitative policy research: Optimizing description and generalizability. Educational Researcher, 12, 3, 14-19.

Hartley, R., and Owen, J. M. (undated). Case Studies in Seven Schools. A Report of the National Evaluation of the Participation and Equity Program. Canberra: Department of Employment, Education and Training.

House, E. R. (1980). Evaluating with Validity. Beverly Hills: Sage.

House, E. R. (1991). Realism in research. Educational Researcher, 20, 6, 2-9.

House, E. R., Mathison, S., and McTaggert, R. (1989). Validity and teacher inference. Educational Researcher, 18, 7, 11-15.

Jacob, E. (1987). Qualitative research traditions: a review. Review of Educational Research, 57, 1-50.

Jacob, E. (1989). Qualitative research: a defense of traditions. Review of Educational Research, 59, 229-235.

Kemmis, S., Dawkins, S., Brown, L., Cramer, B., and Reilly, T. (1983). Transition and Reform in the Victorian Transition Education Program. The Final Report of the Transition Education Case Study Project. Melbourne: Transition Education Advisory Committee.

Kutchinsky, B. (1973). The effect of easy availability of pornography on the incidence of sex crimes: the Danish experience. Journal of Social Issues, 29, 163-181.

Lincoln, Y, S. (1989). Qualitative research: a response to Atkinson, Delamont and Hammersley. Review of Educational Research, 59, 237-239.

Lincoln, Y. S., and Guba, E. G. (1986). But is it rigorous? Trustworthiness and Authenticity in Naturalistic Evaluation. In D. D. Williams (Ed.) Naturalistic Evaluation. San Francisco: Jossey-Bass.

Lincoln, Y. S., and Guba, E. G. (1985). Naturalistic Inquiry. Beverly Hills: Sage.

Miles, M. B., and Huberman, A. M. (1984). Qualitative Data Analysis: A Source-book of New Methods. Beverly Hills: Sage.

Owen, J. M., and Hartley, R. (undated). Case Studies in Five TAFE Colleges. A Report of the National Evaluation of the Participation and Equity Program. Canberra: Department of Employment, Education and Training.

Phillips, D. C. (1983). After the wake: Postpositivist educational thought. Educational Researcher, 12, 5, 4-12.

Phillips, D. C. (1987). Philosophy, Science, and Social Inquiry. Oxford: Pergamon Press.

Parlett, M (1972) Evaluating innovations in teaching. In H.J. Butcher and E. Rudd (Eds.). Contemporary Problems in Higher Education. London: McGraw Hill.

Parlett, M., and Hamilton, D. (1985, unpublished manuscript; 1986).

Evaluation as illumination: A new approach to the study of innovatory programs. In G.V. Glass (ed). Evaluation Studies Review Annual, Volume 1. Beverly Hills: Sage.

Ross, H. L., Campbell, D. T., and Glass, G. V. Determining the social effects of a legal reform: the British "breathalyser" crack-down of 1967. American Behavioural Scientist, 3, 493-509.

Rossi, P., Berk, R., and Lenihan, K. (1980). Money, Work and Crime. New York: Academic Press.

Salomon, G. (1991). Transcending the qualitative - quantitative debate: The analytic and systemic approaches to educational research. Educational Researcher, 20, 6, 10-18.

Stoecker, R. (1991). Evaluating and rethinking the case study. The Sociological Review, 39, 89-112.

1 I have dated the opening salvo of the "wars", at least in-so-far as evaluation is concerned, to the (deliberately provocative?) use of the phrase "agricultural-botany paradigm" by Parlett (1972) and Parlett and Hamilton (1975) in their development of the "alternative" approach of illuminative evaluation.

2 See, for example, the remark by Guba and Lincoln that "we have come to appreciate that the central feature of our paradigm is its ontological assumption that realities, certainly social/behavioural realities, are mental constructions." (1989, p19 - footnote). They appear here to be claiming the "constructivist" ontology as the "hard core" of their methodological program.

3 It should be pointed out that Phillips uses the term "naturalism' in the sense initiated by Popper and other philosophers of science to characterise the methods of the natural sciences; the social sciences are, in this sense, naturalistic if "they have no features that offer insuperable obstacles to the adoption of the methods of the natural sciences" (Phillips, 1987, p. 204). In contrast, Williams, Guba and Lincoln (prior to their latest work) and others used "naturalism" to refer to research which was interpretive and based on qualitative methods. Those who hold Phillips' Position have derived their arguments from earlier work in hermeneutics, phenomenology, ethnography and, perhaps to a lesser extent, critical theory and see various qualitative methods as most appropriate for their epistemological program. I will use Guba and Lincoln's term "constructivist" in addition to Phillips' notion of the Position to characterize their arguments.

4 Taken in isolation, and in the absence of the second-last phrase, this quotation would be "sweet water" indeed to constructivists. Later in the same chapter, however, Bhaskar (1989, p. 86) identifies briefly two difficulties with hermeneutics. Deriving from hermeneutics' "continuing belief in empirical realism" they are that (i) the conditions for social phenomena (experiences) can, themselves, be real and (ii) the phenomena

(experiences) may be false, or inadequate. Given that, if experiments are disallowed in the social sciences, an important source of data must be reported experiences, these observations strengthen the need for rigourous procedures for the development of explanatory conclusions and criteria for their validation.

5 Dunn (1982) has proposed a similar characterization of a threat to validity as a rebuttal to a socially constructed knowledge claim. Dunn's contribution will be elaborated in a later section of this paper.

6 It is interesting to note in this context that much of the weight of inference in a number of "classical" examples of experiments and quazi-experiments in social program evaluation has fallen on the "patch up" procedures and/or post hoc argument followed, rather than the clear demonstration of a "molar" contiguous relationship between intervention and impact. This may be seen, for example, in the account of the TARP experiment by Rossi, Berk and Lenihan (1980), Kutchinsky's (1973) time-series study of the impact of liberalization of Danish pornography laws on the incidence of sex crimes, and the well known time-series study of the British breathalyser legislation by Ross, Campbell and Glass (1970). The latter two studies, Kutchinsky's especially, are also largely explanatory rather than predictive.