

## Epistemology and the theory and practice of program evaluation

Gabriele Lakomski

Department of Education, University of Melbourne, Parkville, Victoria, Australia

### ABSTRACT

In the decades following the education reforms in the sixties in the United States, it became obvious that the practice of evaluation had apparently not produced much useful knowledge. This state of affairs led some social scientists to declare that evaluation research was obsolescent. Others, in a more positive spirit, began the process of critical self-monitoring concentrating on the technical adequacy of evaluations. Finally, a third development took place which is indicated in the emergence of new modes of evaluation commonly characterized as “humanistic” or “naturalistic”. While these expressions of professional concern are a positive development in the debate on the adequacy of “mainstream” evaluation, they fall short in one significant respect. They do not examine explicitly the nature of quantitative evaluation theory and the assumptions upon which it is based. It is argued here that the unsatisfactory state of affairs in contemporary program evaluation cannot be resolved by addressing primarily its technical adequacy. Rather, the problems of program evaluation, as an example of educational research, are of a theoretical, and specifically epistemological, kind. In particular, it is contended that (1) evaluation, whether quantitative or qualitative, suffers from the view that there are secure foundations for our knowledge; (2) that there are no such foundations, and that consequently, evaluation must rid itself of foundationalism on pain of remaining theoretically incoherent and practically irrelevant, and (3) that the solution to these problems is to be found in a non-foundational, materialist-pragmatist theory of knowledge which is theoretically coherent and of practical relevance for evaluation.

### Evaluation and epistemological foundations

In the decades following the immensely expensive and largely unsuccessful education reforms in the sixties in the United States, it became obvious that the practice of evaluation had apparently not produced much useful knowledge and that even when it did it was seldom used to implement or change policies and programs (Alkin, Daillak, and White, 1979; Guttentag, 1973; Guba, 1972). This state of affairs led some social scientists to declare that evaluation research was obsolescent (Dunn, Mitroff, and Dutsch, 1981; Mitroff and Sonoma, 1978). Others, in a more positive spirit, began the process of critical self-monitoring concentrating on the technical adequacy of evaluations (Cook and Gruder, 1978; Glass, McGaw, and Smith, 1981; Cronbach, 1982). Finally, a third development took place which is indicated in the emergence of new modes of evaluation commonly characterized as “humanistic” or “naturalistic”, that is, as proceeding from the field settings in which social actions “naturally” occur (a collection of these modes is found in Hamilton et al., 1977).<sup>1</sup>

While these expressions of professional concern are a positive development in the debate on the adequacy of “mainstream” evaluation (Ross and Cronbach, 1976), they fall short in one significant respect. They do not examine explicitly the nature of quantitative evaluation theory and the assumptions upon which it is based. It is argued here that the unsatisfactory state of affairs in contemporary program evaluation cannot be resolved by addressing primarily its technical adequacy.<sup>2</sup> Rather, the problems of program evaluation, as an example of educational research, are of a theoretical, and specifically epistemological, kind. In particular, it is contended that (1) evaluation, whether quantitative or qualitative, suffers from the view that there are secure foundations for our knowledge; (2) that there are no such foundations, and that consequently, evaluation must rid itself of foundationalism on pain of remaining theoretically incoherent and practically irrelevant, and (3) that the solution to these problems is to be found in a non-foundational, materialist-pragmatist theory of knowledge which is theoretically coherent and of practical relevance for evaluation.<sup>3</sup> How are these claims to be supported and what is the relevance of epistemology for evaluation?

The significance of epistemological issues is clearly accepted by evaluators of different and opposing methodological orientations. It was discontent with the “positivism” of quantitative-experimental evaluation, commonly associated with O. T. Campbell, which first led to the emergence of naturalistic evaluation as an alternative mode (Stake, 1967). R. E. Stake, the originator of Responsive Evaluation, raises the epistemological issue quite clearly when he asks, “How do we know our programs? What counts as evidence that a program is working?” (Stake, 1976, p. 3). At the other end of the methodological spectrum, Campbell, as social psychologist and evaluation theorist, has concerned himself extensively with the theory of knowledge and developed his own descriptive evolutionary epistemology” (Campbell, 1970, 1974a).<sup>4</sup> Since evaluators rarely lay bare their epistemological assumptions, including their political preferences, Stake’s and MacDonald/Walker’s arguments<sup>5</sup> are examined as well as Campbell’s. In addition to the explicit treatment of epistemological issues, a second reason for singling out these evaluators’ contributions is their international prominence in the field of program evaluation and their representativeness of qualitative (naturalistic) and quantitative (mainstream) evaluation respectively. While these comments serve to indicate that the theory of knowledge is recognized by various evaluators as of importance, the reasons for its importance have to be outlined next.

On the naturalistic side, evaluators argue that the laboratory experiment with its researcher-determined variables and measurements is an inappropriate method for the evaluation of social and educational programs. The general argument is that if we are to judge the workings of a program we essentially depend on interpretations and a prior understanding of the program’s and the participants’ reality. The only way to “get at” this reality is to study program and participants in their “natural” state by employing such unobtrusive methods as participant observation. The assumption is that the naturalistic evaluator should intrude as little as possible into the world of the program so that he or she can catch a glimpse of undistorted program life. The way to gain reliable knowledge of “what is going on”, i.e. of a program’s success or failure, is to record as fully as possible what the protagonists say, or describe how they act. The agents’ interpretations, as recorded by the naturalistic evaluator, provide the grounds for knowing: they constitute the evaluation’s epistemological foundations. This does not mean that evaluators in this mould reject out of hand the use of quantitative techniques, they are adamant, however, that educational programs cannot in principle be studied in the way in which a chemist would study amino acids, or an engineer calculate the suspension of a bridge, on pain of trivializing, distorting, or reifying social conditions and hence our knowledge of them. Underlying this assumption is the belief in the familiar division of labour between the social and the physical sciences: the former purportedly deal with the “properly human” world of intentions, motivations, interpretation and understanding, while the latter are in the business of explaining objects in the “natural” world which lack the above listed qualities. It follows for the naturalistic evaluator that the application of the methods of the physical sciences is inadmissible in evaluation since the latter is an example of social/education research.

For the evaluator trained in the Campbell/Stamley school, case studies as the normally preferred modes of naturalistic evaluation may at best contribute useful background information which may enrich the interpretation of quantitative evaluations. But, as Campbell himself emphasizes, case studies cannot stand alone as evidence of program effectiveness, and are hence in no position to replace a good quantitative evaluation (Campbell, 1974b, p. 21). They are thought deficient because of their lack of reliability validity, and above all, replicability. This is so because the key element of good experimental design, internal validity, is absent; and it is this requirement which is of tantamount importance in the interpretation of an experiment or evaluation (Campbell, 1957; Campbell & Stanley, 1966; Cook & Campbell, 1979). The only way to obtain a dependable knowledge base for planning and evaluating social programs in the Campbellian mode is by social experimentation which Riecken & Boruch (1974, p. 3) describe as follows:

one or more treatments (programs) are administered to some set of persons (or other units) drawn at random from a specified population; and that observations {or measurements} are made how or how much some relevant aspect of their behaviour following treatment differs from like behaviour on the part of an untreated group also drawn at random from the same population.

From these brief characterizations of evaluation modes it is evident that naturalistic evaluation assumes participants' accounts and full description as yielding more valid representations of a program's reality. Quantitative evaluation, on the other hand, puts its trust in scientific experimentation and observations/measurements which are assumed to establish reliable knowledge. Both modes accept implicitly that there are foundations for our knowledge but locate them in different theoretical entities. Such acceptance entails that in the case of competing knowledge claims, one could settle the dispute unequivocally by appealing to those foundations since they are themselves considered immune from doubt and further examination. This would mean, for example, that, if certain measurements of a quantitative evaluation were in dispute, the evaluator might take recourse to "secondary analysis", devise new measurements, or attempt to replicate the study without, however, questioning that measurements and observations are the ultimately valid grounds for the necessary corrections. Similarly, where there are competing interpretations about the efficacy of a program, naturalistic evaluators may try to find yet another interpretation by consulting different groups of people associated with the program, but they will not doubt that (participant) observation and interpretation remain the valid foundations for gathering knowledge. There are substantial difficulties with either view, but before these are discussed, we need to take a closer look at how Stake et al. and Campbell defend their foundations of program evaluation.

### **The Method of Understanding in Naturalistic Evaluation**

The primary justification for the case study as an often preferred method in social enquiry is, according to Stake, the fact that such studies "may be epistemologically in harmony with the reader's experience and thus to that person a natural basis for generalization" (Stake, 1978, p.5). Similarly, MacDonald and Walker point out that many of the questions raised by teachers cannot be answered by experimental methods, "such questions are directed at the experience of the participants, and at the nature and variety of transactions which characterise the learning milieu of the programme. There seems to be a need to find ways of portraying this experience and this milieu so that prospective users of the new programmes can relate them to their own experience, circumstances, concerns, and preferences. The case study is one such way ... (1977, p. 181). The purpose of this kind of study is thus to increase understanding of certain audiences in a way that takes into account their current understandings. Stake maintains that all agents arrive at their practical knowledge through direct and vicarious experience, social scientists and other experts included. He concludes from this that evaluation reports ought to be written in ordinary language in order to approximate our natural experience as much as possible. The case study has found application in a variety of disciplines: medicine, law, engineering, and anthropology. According to

MacDonald/Walker, its appropriateness consists in the fact that it is carried out via an “intimate” social process (1977, p. 184) which leads to a social product which has more direct consequences for the agents involved. This feature separates the case study from experimental method which, the authors claim, “is conceptually asocial”. Since evaluators investigate social processes, they are continuously forced to “find vantage points and roles within a web of human relationships without destroying the fabric. The delicacy and subtlety of his instruments is a precondition of their validity” (MacDonald and Walker, 1977, p. 184). While MacDonald and Walker do not explicitly acknowledge the philosophical roots of their account of the case study, Stake accepts the conception of *verstehen* and the notion of “tacit knowledge” as separate from propositional knowledge. Tacit knowledge, as Stake portrays it, “is that which permits us to recognize faces, to comprehend metaphors, and to know ourselves. Tacit knowledge includes a multitude of unexpressible associations which give rise to new meanings, new ideas, and new applications of the old” (1978, p. 6).<sup>6</sup>

Stake notes the common belief that it is from our ordinary understandings that powerful explanations are constructed, and that “explanation” is the “grandest” of understandings. This common belief is mistaken, he claims, since explanation and understandings are different kinds of things. He accepts a Diltheyan definition of understanding as empathy, as the psychological recreation of an agent’s thoughts, feelings, and emotions, and as involving intentionality.<sup>7</sup> He maintains that Truth in the human sciences is not a matter of propositional statements of lawful relationships, rather, truth is better approximated if we pay attention to perceptions and understanding that come from immersion in and holistic regard for the phenomena (1978, p. 6). The validity of the case study is justified accordingly:

When explanation, propositional knowledge, and the law are the aims of inquiry, the case study will often be at a disadvantage. When the aims are understanding, extension of experience, and increase in conviction in that which is known, the disadvantage disappears (Stake, 1978, p. 6).

Such a view is not incompatible with generalization, according to Stake, if we understand by generalization that which builds upon the knowledge of the particular. This kind of generalization derives from experiential, tacit knowledge which, while it may become propositional knowledge, has not yet passed the empirical and logical tests characteristic of formal generalization. While Stake accepts that the scientific validation of laws is a legitimate enterprise, it is considered an inappropriate one in the human sciences whose aim is, as the practical arts, to “get things done”.

Describing the object of social inquiry, Stake notes that contrary to common belief, a case is not usually a single object, a constituent member of a target population. It is this misconception which is alleged to have led to the devaluing of the case study in that a single case is seen as a poor representation of the population, and hence as an inadequate basis for generalization. Often, he points out, it is far more important to generalize about one particular case, or to compare it to a similar one, and to describe the target case accurately. A “case” can be an institution, a program, a responsibility, a collection, or a population. Its distinctiveness lies in the fact that boundaries of what is, or is not “the case” are kept in focus. Events within those boundaries, rather than hypotheses or predetermined issues constitute the material for study. Case studies can be highly statistical, can be used for testing hypotheses, but Stake maintains that those described in the social science literature have the following features:

descriptions that are complex, holistic, and involving a myriad of not highly isolated variables; data that are likely to be gathered at least partly by personalistic observation; and a writing style that is informal, perhaps narrative, possibly with verbatim quotation, illustration, and even allusion and metaphor. Comparisons are implicit rather than explicit. Themes and hypotheses may be important, but they remain subordinate to the understanding of the case (Stake, 1978, p. 7).

Since holistic and episodic information will remain an important requirements social inquiry. the case study will continue to be popular in Stake’s and MacDonald/Walker’s views.

In the next section, we consider what Campbell has to say in defense of quantitative-experimental methodology.

## Evolutionary Epistemology and Social Experimentation

Campbell's project of a "natural selection epistemology" is his attempt to overcome the definitional operationalism of logical positivism with the aim of building a social science which nevertheless "retains a hardheaded willingness to confront theories with facts", and one which avoids a "would-be direct intuition of social reality" (Campbell, 1970, p. 51). It can be placed in the larger context of the naturalistic tradition which claims the unity of method between the natural and social sciences and considers epistemology as continuous with physical science.

Campbell sees the task of evolutionary epistemology on the one hand "to be descriptive of how people go about it when they think they are acquiring knowledge, or how animals go about perceiving and learning when we think they are acquiring knowledge" (Campbell, 1977, p. 12). On the other hand, is also to "include the theory of how these processes could produce truth useful approximations to it" (Campbell, 1977, p. 12). Descriptive epistemology is considered as consistent with and building upon classical analytic theory of knowledge whose task, Campbell notes, was to counter skepticism but which failed in the attempt. As a consequence of this failure, Campbell states that his epistemology "addresses the problem of knowledge assuming that the biologist's description of man-the-knower, and the physicist's description of the world-to-be-known, are approximately correct, although corrigible" (Campbell, 1974d, p. 141). As he admits, this begs the traditional epistemological question, entailing an "epistemological dualism" and "critical realism". By "dualism" Campbell means the conceptualisation of beliefs ("knowledge") as distinct from the referents of the belief, so that "fit", "nonfit", degrees of "fit" error and correction are conceivable (Campbell, 1974d, p. 141). Epistemological dualism is to be kept separate from ontological-physical dualism, and "critical realism" is to denote a "commitment" to the reality of the external world.

Campbell accepts that the nature of the world and the possibility of knowing it are intimately connected, and that "if one is expanding knowledge beyond what one knows, one has no choice but to explore without the benefit of wisdom (gropingly, blindly, stupidly, haphazardly)" (Campbell, 1974d, p. 142). He calls his theory the theory of "blind" variation and selective retention in order to demarcate it from metaphysical conceptions of evolution such as, for example, the "argument from design" prevalent in Darwin's time (see Gould, 1977). It is in keeping with these caveats that Campbell describes his epistemology as physicalist, reductionist and materialist. Its validity turns on the assertion that evolution, even in its biological aspects, is a knowledge process and that the natural selection model can be generalized to other epistemic activities such as learning, thought and science.<sup>8</sup> The latter is of particular interest for our purposes.

Science, as but one aspect of sociocultural evolution, is characterized in that the "selective system which weeds out among the variety of conjectures involves deliberate contact with the environment through experiment and quantified prediction, designed so that outcomes quite independent of the preferences of the investigator are possible. It is preeminently this feature which gives science its greater objectivity and claim to a cumulative increase in the accuracy with which it describes the world" (Campbell, 1974b, p. 434). He admits that the application of the selective retention model is problematic because of the different trial and error processes within science,<sup>9</sup> and yet Campbell maintains that it, like biological evolution, is opportunistic in relation to both solutions and problems; that is, a researcher may change the research problem to one which is solvable following the encounter with a new phenomenon.

In conclusion, then, Campbell believes that traditional epistemology (in particular classical empiricism), while admirable in its attempts to counter skepticism, has nevertheless failed in its valiant endeavour to justify its foundations for knowledge. As a result, he believes that we can never be sure that we know, or when we know.<sup>10</sup> Consequently, Campbell believes he is justified in abandoning the business of traditional epistemology to seek whatever we can know about the natural world in those enterprises whose job it is to tell us: the physical sciences, and especially evolutionary biology.<sup>11</sup> As we saw, natural science generally progresses in the same manner as does natural selection, with experiment and prediction being the "selectors" which "weed out" among



the various conjectures in the light of a presumably “better fit” of some hypotheses rather than others. Given his physicalist assumptions which rightly do not allow for a substantive and methodological difference between the sciences, Campbell applies his model of natural science to the social realm and the evaluation of educational programs. Campbellian social scientific experimentation is seen as allowing for inferences of superior dependability about cause and effect, as making possible comparisons between the effectiveness of two programs, and generally as providing a dependable base for policy decision-making.

### The Problem With Foundations

We learnt in the previous section that naturalistic evaluators argue for the epistemological priority of holistic, complex program descriptions based on various participants’ understandings. These descriptions are in part gathered by “personalistic observation”. Furthermore, it is presumed that people’s motivations and intentions somehow provide more valid or “deeper” access to the social reality of the program, access denied the “hard” sciences. It is obvious that foundations are sought in a different philosophical tradition, phenomenology, which considers itself as the alternative to “positivist” natural science. Campbell, in contrast, remains in the natural science tradition whose foundations, he thinks, cannot be justified, but whose experimentalist methodology is still the best available.

The idea that our knowledge is in need of foundations is, as Richard Rorty so well described, a relatively recent invention of philosophers (Rorty, 1980, Ch. 3). Without going into its history here, the core assumptions of (the various kinds of) foundationalism<sup>12</sup> (1) are that certainty or unchallengeable provability is a necessary feature of any piece of knowledge and (2) that it was believed, as Walker (1985, p. 56) explains, that

if such certainty was achievable it was in virtue of secure epistemic foundations, or epistemic items which need no justification and which supported all other genuine epistemic items. Claims to knowledge, then, as opposed to mere belief or opinion, were justified if they could be shown to follow by some procedure of rational inference from the epistemic foundations or to be foundational items themselves.

Classical empiricists, for example, believed that their secure epistemic foundations could be found in the “evidence of the senses”, or in more contemporary language, in “sense data”. Classical rationalists, in turn, put their trust in “innate ideas” which were so obviously true as to be beyond doubt (see Walker and Evers, 1982).

First and foremost, it has proven very difficult to determine any candidate which might serve as an epistemic foundation, and one which does not give rise to criticism and doubt as well as demands to justify itself and to provide its own foundation (Walker and Evers, 1982; Evers and Walker, 1984, Ch. 1). Consider, for example, the classical empiricists’s claim that we gain knowledge from, roughly speaking, the evidence of our senses, particularly through perception. This assumption, while intuitively appealing, turns out to be erroneous on closer inspection. It has been known for some time that our senses, while generally trustworthy, may mislead us subject to the constitution of our perceptual apparatus.<sup>13</sup> Hence it is possible for humans to see things which are not there, such as “Fata Morganas”, or not to see things which do exist, such as molecules and gases. It is this awareness which provided the skeptic with an entering wedge to The empiricist’s argument about the certainty of sense perception, and one which is acknowledged and shared by Campbell.<sup>14</sup> While the physical causes of the images on our retinas are partly responsible for what we see, what we can perceive is also constituted by the way in which the information is stored in our brains. Our brains, in turn, are shaped by the way in which we have been socialized and by our expectations (cf. Chalmers, 1978). To expand this point, talk of “brains”, “socialization” and “sensory experience”, say, implies that we already know what kind of creature we are and what our cognitive capacities. That is to say, a theory which explains this is, at the very least, presumed; it is neither self-evident, nor

simply given. When Campbell claims that the task of traditional analytic theory of knowledge has been completed with a negative answer, i.e. that “knowledge claims going beyond immediate experience cannot be proven with either deductive or inductive certainty. Even were a belief to be correct, we would not know this for certain” (Campbell, 1977, p. 4), we can see where his skeptical approach shares the classical empiricist’s foundations (see Williams, 1977, Ch. 1). While Campbell denies the possibility of justification for his own evolutionary epistemology, he accepts, if not the provability of “immediate experience” its validity as obvious. Let us take up this notion of the “obvious” and let us call the empiricist’s claim about the epistemological status of sensory experience, *p*. We can now ask: how do we know that *p*? Unfortunately, this question has not been satisfactorily answered in classical empiricism precisely because *p* was taken as obviously true. But principle *p*, considered from a psychological perspective, is quite shaky because we have to learn about our own sensory experience first, and usually on the basis of an already acquired language of ordinary objects (Quine, 1964). The epistemic notion of “obvious”, then, presupposes a whole network of theories including accounts of human development and language learning. Hence, the theory of evolution is no more true than, say, creationism, just by dint of describing the development of the species itself. Campbell states that he cannot provide us with a justification because he implicitly accepts the classical empiricist’s restricted foundations which, under his skeptical gaze, do not hold up in terms of certainty or unchallengeable provability. This is so because skeptical doubts, as Quine (1975) points out, are scientific doubts. While doubt prompted the theory of knowledge, it was also knowledge which first prompted doubt:

The basis for scepticism is the awareness of illusion, the discovery that we must not always believe our eyes ... Illusions are illusions only relative to a prior acceptance of genuine bodies with which to contrast them ... The sceptic’s example of the seemingly bent stick owes its force to our knowledge that sticks do not bend by immersion (Quine, 1975, pp. 67-68).

The long and the short of this is that Campbell presumes a certain model of science and its growth: in his case an inductivist model rooted in the “immediate experience ... of sense perception. While induction is characteristic of biological science in its infancy, science as we have developed it, aided by our innate similarity standards and their development through species learning, has become a “ponderous linguistic structure, fabricated of theoretical terms linked by fabricated hypotheses and keyed to observable events here and there ... (Science) is no longer simple induction. It is the hypothetico-deductive model” (Quine, 1975, pp. 71-72).

If for all its detailed description, Campbell’s evolutionary epistemology is indicative of an inadequate model of natural science, could he still claim to gain knowledge as the result of employing a reliable method, that of experimentation? The answer is, not surprisingly, “no”.

Recall that experimentation was to control and sort out plausible rival hypotheses with the aim of providing “reassurance, (that) if one can intentionally and successfully produce (or prevent) a phenomenon, one has a surer sort of knowledge about its causation than can be obtained in any other way” (Riecker and Boruch, 1974, p. 10). Unfortunately, such reassurance has not been confirmed by the history of science in relation to “crucial” experiments (see Lakatos, 1977; Hanson, 1959; Kuhn, 1970; Feyerabend, 1975). “Crucial” experiments, particularly in physics, were supposed to falsify unambiguously particular singular hypotheses on the one hand, and on the other, were simultaneously to confirm singular alternatives.<sup>15</sup> As the physicist and philosopher Pierre Duhem showed (1906) “crucial” experiments could do neither of these things. Duhem was the first to provide systematic proof of what has come to be accepted as the theory-dependence thesis in philosophy of science. Simply put, the thesis maintains that the scientist’s observations are never independent of theory and hence of other hypotheses (that is, hypotheses other than those explicitly examined). This is so because in using laboratory equipment to observe or measure temperature, for example, we presume the theoretical context which gives readings and apparatus their empirical significance in the first place. It follows that we never appraise just a single hypothesis but a corporate body of statements (see Quine’s description of science above). It should be clear then that observation does not unambiguously falsify particular singular hypotheses. If we

encounter a recalcitrant observation, all we know is that one of the hypotheses is wrong, but we do not thus know which one. As for the second point, Duhem showed that it, too, is mistaken. “Crucial” experiments do not confirm a single hypothesis but an “indefinite disjunction of hypotheses” (Walker and Evers, 1983, p. 398). More strongly put, there is an indefinite number of hypotheses which can be added to the disjunction. The upshot of Duhem’s arguments are two methodologically important points:

- a. experimental situations always drastically underdetermine any conclusions they are alleged to imply, and
- b. the slack between experiment and conclusion is taken up by theory to such an extent that different theoretical perspectives can warrant different conclusions from identical experiments (Walker and Evers, 1983, p. 398).

In the light of the theory-dependence of observation with its implications for scientific experimentation (whether “social” or “natural”), we have to conclude that quantitative-experimental evaluation, as described, with its emphasis on “hard” data and certain cause-effect relationships, makes extravagant claims which cannot be supported even in principle. Put another way, its claims far outrun its theoretical resources to justify them. This is so because Campbell presupposes an inadequate model of physical science, one which is grounded in the assumption of “immediate experience” in form of sense perception. Thus, the evaluation model which prides itself as the most scientific turns out to be quite the opposite, namely arbitrary. What this discussion also brings to light is a fact of perhaps even greater significance. Quantitative evaluation with its heavy emphasis on “internal validity” necessarily depends on a wider theoretical context for its measures, statistics and results to have any significance at all, a context which, while unacknowledged, is also implicitly tested in any quantitative evaluation. This is the explicit result of Duhem’s methodological points which seriously undercuts the claim of scientificity in quantitative evaluation. This issue will be taken up again later.

What remains to be done is to show that the foundations assumed by naturalistic evaluation are a variant of the doctrine of “immediate experience” which, epistemologically speaking, makes both evaluation approaches members of the same empiricist family. Hence, it will not come as a surprise if the remainder of the epistemological story sounds familiar. Central to naturalistic evaluation is the claim that understanding is a superior method for creating valid evaluation knowledge because the method gets at the epistemologically prior “tacit” knowledge apparently characteristic of all humans. Tacit knowledge, as Stake describes it, is immediately available to us while propositional knowledge is not. The two are quite distinct and separate. The tacit/propositional knowledge dichotomy is, however, merely a variant of a rather old philosophical theme: *the doctrine of the given* with its corresponding “two components view” (Williams, 1977, Ch. 2). Following Williams’ (1977) exposition, this view maintains that there are two elements in cognitive experience: sense data immediately available to the mind, and a form or interpretation which represents the activity of thought. It is not hard to understand why the distinction was drawn like this. The major worry was that if there were no distinction between given sense data and their interpretation, then knowledge would be quite arbitrary since there would be nothing to match it up against to see whether it is true or false. Indeed, it would become meaningless to talk about truth, falsity, or error. Hence, on this account, “bruta data” have to be presupposed in order for us to have inferential knowledge at all. We see that we have struck another form of empiricism which, not surprisingly, runs into problems we have already encountered. If the given is to provide the final check upon empirical knowledge, it must be some kind of knowledge itself in order to avoid a regress. But as we saw tacit knowledge is fundamentally different from propositional knowledge (although the former can become the latter).<sup>16</sup> Since this is so, the requirement is that the mind apprehend tacit knowledge without the help of conceptual mediation (“immersion” in the phenomena). Tacit knowledge, then, is, in Williams’ term (1977, p. 29) “ineffable”. And ineffable knowledge is in no position to provide us with a check on anything, “let alone the entire edifice of



empirical knowledge" (Williams, 1977, p. 29). To demonstrate the dilemma more clearly, consider Stake's claim that tacit knowledge allows us to recognize faces. Now it is all very well for everyday purposes to insist that we know our spouse's, child's, or the mailman's face when we see it. But this is an entirely different thing from knowing that the face we see is that of our spouse and not the mailman's. How do we know that what we see before us is a face (and not just a pink colour-patch) and that the face belongs to the mailman? Surely, guessing would not do, we must be able to tell one from the other - and get it right.<sup>17</sup> But in order to do so, we cannot rely on tacit knowledge, for the reasons we discussed earlier. What is given in tacit knowledge are things or phenomena rather than facts.<sup>18</sup> This must not be confused. If the core of tacit knowledge is the givenness of things/phenomena it is easy to see that the appeal to it apparently involves no judgment. Hence, naturalistic evaluators such as MacDonald/Walker and Stake can claim that they merely ... "collect" program definitions and present portrayals, activities which do not commit them to any factual claims whatsoever! The moral is quite clear as William points out, "...the sense in which the given is immune from error is without epistemological significance. There is nothing to be wrong about, but nothing to be right about either" (1977, p. 36). As a consequence of this epistemological dilemma, naturalistic evaluators cannot sort valid from invalid program interpretations or descriptions, a state of affairs reflected in the conception of "value pluralism" defended by both Stake and MacDonald/Walker.<sup>19</sup> It follows, then, that contradictory findings of the same program are implicitly accepted as of equal value. Given the self-understanding of naturalistic evaluation as more client-centred, holistic, and above all, democratic, the incapacity to judge is more than an acute embarrassment. It makes this approach epistemologically as arbitrary as the one it sought to supersede. This conclusion is unavoidable because the conception of tacit knowledge is a confused attempt to merge two conflicting notions of the given: that things or phenomena are given, and that our relation to them must be some kind of non-inferential knowledge. While the notion turned out to be incoherent on its own account, it is ultimately untenable because, as was argued earlier, all observation is theory-dependent, and none of our experiences are given simpliciter. In conclusion, judged on their epistemological merits, both evaluation modes outrun their theoretical resources to justify their claims to knowledge. They are incoherent, where "coherence" in a general sense is understood as a theory's possessing more virtues of system (Evers, 1984, p. 24, Note 3). On this account, a theory is judged superior if it is more coherent than a rival theory. While lack of coherence is a deficiency of theory in the sense of providing the lower threshold of a rational theory, such deficiency has direct practical consequences. What these consequences are in terms of the practices of both approaches is considered in the following.

### **Practical Implications of Foundationalist Evaluation**

The present epistemological inquiry leads to the conclusion that both evaluation modes determine prior to any investigation being carried out what is to count as valid knowledge without, however, being able to further justify their presumptions. Consequently, their claims to knowledge cannot be made good. They are epistemologically deficient and hence arbitrary. This result does not sit well with the assumed purposes of either mode since they are to provide knowledge and information for policy-makers, teachers, administrators, and sometimes even parents and children. Evaluations are supposed to contribute to change and improve educational practices and procedures. They are to provide a rational base for decision-making in the light of scarce resources (see Campbell, 1969). Take, for example, the case where a quantitative evaluation comes up with the result of "no significant difference" between experimental and control group. The conclusion that therefore the innovative program should be discontinued would seem plausible. Without going into the statistical-technical details here (see Meehl, 1967; Lykken, 1968; Deming, 1975), let us note that such a conclusion is not at all warranted. This should be obvious now in the light of Duhem's two points discussed earlier. The quiet conflation of statistical with social significance is a good example of Duhem's second point that the slack between experiment and conclusion is taken up by theory. Had the recommendation been a positive one, the point still remains valid. It is not possible to claim in

this instance that any causal links have been established in the evaluation as the result of which the program should be discontinued. No such “fact of the matter” has been found which could serve as the basis for the decision. Still, quantitative evaluation continues to maintain that it is the only scientific approach to evaluation, and its results continue to be accepted as objective and scientific. To be sure, secondary evaluations are carried out because of conflicting evidence in the primary evaluation, or for political reasons, but such re-analyses can best be described as “fine-tuning” of an instrument which is basically accepted as sound. This is to say that even when the results of a quantitative evaluation are challenged, the challenge does not ordinarily press the issue of epistemological justification of results because that is either riot seen as the problem, or the criticism itself presumes unwittingly the same epistemological foundations as the criticized mode. As a result, recommendations are made which have direct impact on people working in, or supposedly benefitting from, a program. These recommendations cannot be justified by arguing that they are based on knowledge as opposed to mere opinion, belief, or “informed guess”. Drawing out the implications, it becomes clear that decisions made about the continuance or discontinuance of programs -- to the extent that they are either implicitly or explicitly based on foundationalist views of knowledge -- are rooted in the personal preferences of policy-makers, power groups, or individuals, none of which become accessible to public scrutiny and debate. While it cannot be assumed a priori that any group or individual has suspect motives, the point is that their preferences, which are also theories about the way education or social welfare, for instance, ought to be organized, do not become explicit and hence cannot be assessed on their merits. These options are prohibited by evaluations which presume before the event what is to count as valid (and admissible) knowledge. In this way, the range of possible evaluations of particular programs is limited in advance, and so are possible solutions or ways of action.

While naturalistic evaluators are only too ready to recognize the political nature of, particularly, quantitative evaluation (which is taken to be conservative), they believe they evade if not politics altogether then at least undesirable politics. They tend to refrain from giving recommendations at the end of an evaluation in order to escape the “hand-maiden” role of evaluation as a paid service.<sup>20</sup> This is done on the assumption that the giving of recommendations equals unjustifiable meddling of professionals in the affairs of practitioners, an interference which might accord the evaluator qua exert more influence than is warranted. The naturalistic evaluation report as in Stake’s model, for example) normally offers a collection of case descriptions which are considered as of equal information value to everyone involved in the evaluation. The “right to know”, fundamental to MacDonald and Walker’s model, is seen as applying across the board in equal measure. While giving different groups and individuals a voice in the case study is a step in the right direction, it is by no means sufficient because, as we saw, such choice is rooted in the preferences of the individual evaluator. In addition, as long as naturalistic evaluators assume that everyone’s knowledge, interpretation, or observation is of equal value, their approach while political can certainly not be called “democratic” in any substantive sense. The upshot of such “value-pluralism” is what Petrie (1984) calls a “noncommittal subjectivism” which, in the absence of rational criteria, leaves decision-making where it was located earlier: in the non-public preferences, opinions, and ideologies of whatever body is responsible for the decision. While these preferences remain unexamined, and indeed unexaminable, they simultaneously provide the rationale for decision-making in educational and social research generally. Implied in this is an additional problem which needs to be examined more closely.

The problem referred to is that any evaluation (or educational research) takes place as an interaction between groups and individuals, and it is this interaction which first makes the research possible. What kinds of social relations and social control are involved in evaluation is nicely illustrated by Cronbach (1982, p. 36) in his report of a comment made by Stake:

The scientist (quantitative evaluator, G.L.) sees the rank and file as suppliers of data and as an audience awaiting what he distils from the data; the humanist (naturalistic evaluator, G.L.) is a teacher who turns the rank and file into investigators. They are to refine their perceptions without

the aid of an intermediary or authority. Evaluation and use of evaluation thus become a communal process. This approach minimizes the role of power and expertise.

The relations between experimentalist and subjects are characterized by the former assuming total control both epistemological and social, and by the latter not possessing either.<sup>21</sup> Epistemological control is exerted in that the experimentalist is by definition the legitimate definer of knowledge and of what is knowable. What is knowable is in turn delimited by the quantitative method (controlled by the researcher). While both experimenter and subjects find themselves in the same social situation, the evaluator controls the subjects' behaviour in that situation. Indeed, it is a prerequisite of the experimental method that subjects be controlled. They must conform to the strictures of the scientific method, that is, answer certain questions, demonstrate only specific behaviours, etc., since the canons of internal validity demand that external variables be kept out. The quantitative evaluator is not a variable in the program evaluation and must not be one since the replicability requirement necessitates that any experimentalist be capable of carrying out the procedure. This is achieved by standardization through professional training and is controlled by the certificated and accredited members of the profession. This means that professional training attained at the appropriate institutions puts any researcher in a position of epistemic privilege in relation to persons not trained in like fashion. Practically then, quantitative evaluators can only be professionals on the assumption and continuation of authoritarian social relations whose very nature must remain hidden. Hence, criticisms made of quantitative evaluations by un-certificated persons, or people subscribing to a different methodology, can be disqualified as coming from people who do not understand the purposes of experimentation (they are not "properly trained"). Quantitative-experimental evaluation becomes ideological when its proponents use their professional privilege to exclude lay-people from sharing in the production of evaluation knowledge by denying them epistemic power at the outset. If this argument is correct, then we must conclude that in main-stream evaluation as in educational research generally, "the most worthwhile and well-founded forms of knowledge are contingent upon the protection of the epistemic privilege of the professionals, such that to violate that privilege is to violate the internal validity of any relevant experiment" (Walker and Evers, 1983, p. 402).

Turning to the social relations obtaining between naturalistic evaluators and their subjects, it is apparent that they are more conducive to a democratic research practice. Since naturalistic evaluation does not require experimental control subject to its epistemology, there is no epistemic requirement demanding control over subjects. On the other hand, naturalistic evaluators, too, are accredited and "properly trained" professionals working in university departments or research institutes and are, in this sense, as privileged as their quantitative colleagues. In the concluding section of this paper, the practical advantages of naturalistic evaluation in particular are discussed within the context of a different epistemology, materialist pragmatism.

### **Evaluation as Open-Ended Social Experimentation**

In the preceding text it was argued that foundationalist modes of evaluation have direct practical and political consequences. It was suggested that these consequences are the result of the degree of social control presumed and required in either evaluation mode. As a result of the discussion, naturalistic evaluation, while deficient on epistemological grounds, is nevertheless accorded a working priority over quantitative-experimental evaluation because it does acknowledge "the web of human relationships" (MacDonald/Walker) and does not impose experimental controls. Its social relations are more democratic in that participants, qua their knowledge of the program situation, are of central significance as informants. Hence, naturalistic evaluators have to subject themselves to the participants' circumstances and to their willingness to co-operate. It is possible in this mode of research to minimize the expert/dominance position for example, by signing over ownership of an evaluation report to the program personnel and participants. While this is a minority position even within naturalistic evaluation (MacDonald being the main proponent), such a stance would be

impossible for the experimentalist given the epistemological and professional premises of the model. Unlike Cronbach (1980, p. 7) who suggests that evaluators should not yet declare allegiance to either mode given the present state of the art, it is argued here that it makes good scientific sense to opt for naturalistic evaluation, particularly participant observation. As already noted, it is the preferred mode because it accepts the epistemic requirement of co-operative as opposed to hierarchical social relations between evaluator and subjects. Prioritising naturalistic evaluation in this way offers a more flexible framework since interpretations of the “rank and file” are not devalued *ex cathedra* on professional grounds. To avoid misunderstanding, however, the singling out of the procedures of naturalistic evaluation is not to be confused with endorsing the mode as currently understood in the field. For in order to justify the claims made by naturalistic evaluation, and in order to make the study of evaluation more coherent, we need a non-foundational epistemology with quite specific characteristics. While just such an epistemology was implied in this text, it must now be introduced more fully.

The epistemology offered here differs significantly from its traditional rivals in that it does not give rise to such dualisms as mind/body; fact/value; and professional/nonprofessional. Unlike its competitors, it accepts that all observation is theory-dependent. If, as we saw in the discussion of Duhem’s methodological points, theory always precedes observation, then our epistemology denies that there are privileged foundations for knowledge in the sense in which “immediate experience” in form of sense perception was privileged. It acknowledges as its smallest epistemic units whole theories as opposed to “sense data” or “innate ideas”. On this view, as implied in the preceding text, scientific theories enjoy the same epistemic status as commonsense ones. In fact, all theories are epistemologically on a par so that the theory of knowledge is continuous with natural science. Note that this is quite different from asserting, as Campbell does, that epistemology is replaced by natural science on the assumption that traditional analytic theory of knowledge has run its course and could not deliver what it promised. While the assessment is correct, the conclusion to be drawn is not that epistemology has become irrelevant (Campbell’s solution) but that traditional analytic epistemology, as Quine showed (1961) has to be done away with since it asked the question of knowledge incorrectly. In addition to being non-foundational, then, our epistemology is naturalistic in Quine’s sense, a characteristic it shares with Campbell’s project.

While this theory of knowledge does not accept privileged foundations, it is self-referential like all epistemologies. But unlike Campbell’s descriptive epistemology, or the naturalistic evaluators’ phenomenology for whom circularity becomes vicious, it escapes their fate because it recognizes equivalence between the principles of assessment of knowledge claims and the claims themselves. Simply put, “proof” and “claim” are epistemic equals. This makes it a very powerful theory indeed because it is immensely resourceful. Since it does not lay down in advance what is to count as valid knowledge, it is free to draw on whatever theory appears useful in the evaluation of an educational or social program. Evaluation knowledge, which is knowledge of why particular programs (solutions) do or do not “work”, is the result of ongoing inquiry, and the epistemology’s power lies precisely in its facilitation of such open-ended, continuous inquiry. In this sense, it is pragmatic. To put the point differently, its pragmatism is methodological in that neither participants’ understanding nor statistical-mathematical techniques are ruled out of court prior to considering the program under review. Denying the relevance of a particular research tool in advance is tantamount to denying systematic, open-ended inquiry. Continuous, systematic inquiry is, in turn, the hallmark of science. This shift to the methods<sup>22</sup> of dealing with evaluation has been made possible since secure foundations, as here understood, have proven to be non-existent. Since this is so, we follow Dewey in that any activity, including program evaluation, administration, or educational research generally, can be scientific provided it possesses “systematic methods of inquiry, which, when they are brought to bear on a range of facts, enable us to control them more intelligently, less haphazardly and with less routine” (1929, pp. 8-9).<sup>23</sup>

There is another important criterion which has to be mentioned, and that is the question of this epistemology’s own standard of validity. Since it is the very characteristic of a non-foundational

theory of knowledge that it cannot withdraw to secure foundations to stop the regress of reasons, it subscribes instead to a coherence theory of evidence (Quine, 1969; Quine and Ullian, 1978). Accepting a coherence theory of evidence that theory is true which provides the best, that is, internally coherent, evidence in the explanation of a phenomenon. Translated into education evaluation, we can say that that evaluation is true which is internally coherent in the explanation of a specific program's performance. Within the context of the coherence theory of evidence, Campbell's methodological notion of triangulation (Campbell and Fiske, 1959; Campbell, 1970 b) would find a secure place. For the coherence theory of evidence the usual criteria for the quality of a theory also apply. These are simplicity, elegance, explanatory power, and comprehensiveness. Another of this epistemology's defining characteristics is then that it is coherentist.

Finally, a word of explanation in regard to the term "materialist" in the materialist-pragmatist theory of knowledge. Materialism is here understood in the sense in which the term "matter" is used in the natural sciences, especially contemporary biology and biochemistry. These sciences demonstrate that the vital functions of all living organisms can be explained by chemical, and ultimately physical, processes. Hence, it is more precise to describe the epistemology as "physicalist." In this respect, there is no disagreement between Campbell and the present position, but considerable disagreement with the "mentalism" (or dualism) of naturalistic evaluation. It follows from the physicalist assumption that we can give a causal account of the genesis of the species, and see it as of a kind with all other living organisms. However, being thus part of nature does not mean that humans are like amoebas (despite the fact that we are their "cousins" as Campbell is fond of saying), or that human behaviour can be explained by reducing behaviour X directly to gene Y on a one-to-one basis. Materialism, then, does not entail or mean traditional reductionist behaviourism of the Skinnerian kind. There are significant differences between us and other organisms which are primarily the result of the development of our large brains (see Gould, 1981). As a consequence, we have become self-reflective. And precisely because we can know what we do know from the present perspective of the developed sciences, our theory of knowledge is historical. Entailed in this definition is an acceptance that educational and social problems generally are rooted in the division of labour with its hierarchical organization of professionalization. It follows that those problems can only be overcome practically through the actions of those whose problems they are. This distinction is an important one to keep in mind. As was noted earlier, the truth of any theory is, on the present account, dependent on the coherence of evidence it musters. This validates it as theory. In practice, however, the theory is validated through the actions of the agents concerned in relation to the solving of their problems. This means for program evaluation that participants and program personnel carry out their own evaluation of their program's performance, and adjust or discontinue it depending on their assessment. Outside evaluators, to the extent that their services are mandated, as well as potential and actual beneficiaries of the program, are also legitimate participants in the evaluation. On the present materialist-pragmatist account of evaluation, each group or individual is an epistemological equal entitled to offer their theories of a program's or practice's effectiveness. This also includes the accounts of students and parents since they, too, have their own theories about the educational practices they are involved in, whether directly or indirectly. Given the epistemological argument presented here, it can no longer be maintained that (a) quantitative evaluators *qua professionals* and possessors of "scientific" methodology are in a privileged position to evaluate, nor (b) that program participants *qua better local* knowledge have privileged access to the program's reality, as assumed by naturalistic evaluators. Knowledge about a specific program or educational procedure or set of practices can only be obtained by pitting the various theories offered against each other, and by working out how coherent or incoherent these accounts are, and where the contradictions occur (if any). The process of working through inconsistent or contradictory accounts is what program evaluation is. The best, most valid, or scientific evaluation is the one in which these conflicting frameworks have been rendered coherent.<sup>24</sup> While the achievement of coherence is by no means guaranteed, what is guaranteed by the acceptance of a materialist-pragmatist theory of knowledge is the process of



continuous inquiry into a program's effectiveness. While better evaluation research does not automatically improve educational practice, it makes it a whole lot more likely.

## Notes

1. The meaning of "naturalistic" in evaluation is not to be confused with the philosophical doctrine of naturalism which holds to the unity of method between the natural and social sciences. Naturalistic evaluation is not "naturalistic" in the slightest since it advocates an essential difference between the methods and substance of the sciences.
2. While incorrect and careless application of statistical technique are also considered problematic in quantitative evaluation, this is not the cause of the contemporary discontent with program evaluation.
3. The epistemology proposed here is the central part of a research program carried out by C.W. Evers and J. C. Walker. A comprehensive introduction is provided in their *Epistemology, Semantics and Educational Theory*. Occasional Papers No.16. Department of Education, The University of Sydney, Australia, 1984.
4. This aspect of Campbell's work is not well known in the evaluation community. The recent volume entitled *Scientific Inquiry and the Social Sciences*, edited by M.B. Brewer and B.E. Collins (San Francisco: Jossey-Bass) 1981, is the first comprehensive attempt at discussing Campbell's theoretical-epistemological contributions to evaluation research.
5. Barry MacDonald and Rob Walker, as British evaluators, have done much to publicize the case study method in the European context and are hence included here. Of particular importance is their essay "Case Study and the Social Philosophy of Educational Research". In D. Hamilton et al. *Beyond the Numbers Game* (Berkeley, CA: McCutchan) 1977. MacDonald is the more overtly "political" evaluator, advocating "Democratic Evaluation". See his "Democracy and Evaluation". Centre for Applied Research in Education, University of East Anglia, 1978; and "Evaluation and the Control of Education". In D. Tawney (Ed.) *Curriculum Evaluation Today: Trends and Implications* (Aylesbury, Bucks.: MacMillan) 1976.
6. Stake derives the notion of "tacit" knowledge from M. Polanyi, *Knowing and Being*. (Chicago: The University of Chicago Press) 1969.
7. See W. Dilthey "The Rise of Hermeneutics" for extended discussion of the concept of understanding. In P. Connerton (Ed.) *Critical Sociology*. (Harmondsworth, Middlesex: Penguin) 1976, pp. 104-116.
8. Campbell has been taken to task over his account of both biological and social evolution in a special issue of the *American Psychologist*. May 1976.
9. Replying to an imaginary critic, Campbell asks, "Do I really, with a straight face, want to advocate that the discovery of new scientific theories is through an 'unjustified' or 'blind' variation of theories and a selective retention process? Yes, I do, implausible as it may seem". In his "Unjustified Variation and Selective Retention in Scientific Discovery". In F. Ayala and T. Dobzhansky (Eds.) *Studies in the Philosophy of Biology*. (London: MacMillan) 1974 (d), p.153.
10. An excellent discussion of the skeptical argument and its problems is M. Williams *Groundless Belief*. (Oxford: Basil Blackwell) 1977.
11. Campbell, who often cites Quine approvingly in his work, appears here to accept Quine's "Epistemology Naturalized" thesis which argues that there is continuity between epistemology and natural science. In Campbell's *William James Lectures* (Harvard 1977) he agrees with Quine on the nature of natural knowledge, the "indeterminacy of translation", and he even accepts the latter's criticism in relation to skepticism. However, his "agreement" misses Quine's most central epistemological point. Campbell writes that Quine, among others, recognizes the circularity of assuming knowledge before knowledge, and even grants that the latter provides "subtle arguments that such circularity need not be vacuous or vicious (Campbell, 1977, p. 17), only to go on to say, "However that may be, my preferred vision is that no nonpresumptive knowledge is possible, neither in ordinary knowing nor in our knowing about knowing" (ibid.). He continues to accept explicitly the analytic-synthetic distinction whose refutation was Quine's final blow to analytic epistemology.

12. D. M. Armstrong provides a list of various types in his *Belief, Truth and Knowledge*. (Cambridge: Cambridge University Press) 1973, p.160.
13. For a discussion of the philosophical problems with perception see A. F. Chalmers *What is This Thing Called Science?* (St. Lucia, Queensland: The University of Queensland Press) 1978 Second Printing.
14. He has himself carried out experiments which demonstrate the social influence on visual perception. But from the psychological fact that people can be mistaken about what they see, Campbell goes on to claim that visual perception, while fallible, can be rendered more objective by way of constructing increasingly accurate measuring devices. He considers the whole course of science as indicating this very trend. In "Natural Selection as an Epistemological Model". In R. Narroll and R. Cohen (Eds.) *A Handbook of Method in Cultural Anthropology*. (New York: Natural History Press/Doubleday) 1970.
15. This account is indebted to J.C. Walker and C. W. Evers "Professionalisation and Epistemic Privilege in the Politics of Educational Research". In *Educational Research for National Development: Policy, Planning and Politics*. Collected Papers, Australian Association for Research in Education, Annual Conference, 1983, Canberra, Australian National University. Research in Education, Australian National University, Canberra, November 1983.
16. This is as should be for the empiricist. Since tacit knowledge immediate in the sense of our having it subject to being conscious, concepts consequently have to be learnt, often by "ostensive" definition.
17. Since we normally do without any apparent problem, and instantly, it appears plausible to attribute recognition to tacit knowledge as causal. But this would be drawing an incorrect inference: from the fact that recognition seems instant, we may not conclude that it is subject to an assumed epistemologically privileged kind of knowledge. Analogously, just because program participants have better local knowledge of a program, one may not thus conclude that they are in an epistemologically superior position to judge "what is going on".
18. That is to say, the distinction between phenomena and facts is one which is internal to the doctrine of the given. The argument in this paper denies the validity of this distinction, as well as the doctrine.
19. For a critical appraisal of this issue in context of MacDonald's evaluation approach see G. Lakomski "Ways of Knowing and Ways of Evaluating: Or, How Democratic is 'Democratic Evaluation?'" *Journal of Curriculum Studies*. 1983 (3): 265-276.
20. MacDonald, in particular, argues that while evaluation data should contribute to better decision-making in the light of administrators' priorities, circumstances and values, such information does not necessarily only benefit that group since the (non-recommendatory) evaluation report leads administrators to make the "right" decisions to give their rhetoric some backing, at least some of the time. See B. MacDonald *The Experience of Innovation*, Vol. 2 of the revised edition of the publications of the Humanities Curriculum Project Evaluation Unit (Occasional Publication No. 6) Centre for Applied Research in Education, University of East Anglia, 1978.
21. For a careful and detailed discussion see J.C. Walker and C. W. Evers "Professionalisation and Epistemic Privilege in the Politics of Educational Research" op cit.
22. More precisely, the shift to a wider notion of "methods" than is currently employed in the field of program evaluation which includes the justification of using particular methods within a whole theoretical context.
23. This is a much more generous account of science and evaluation than Campbell or Stake et al. may wish to endorse. But as was argued in this paper, it is a superior one to the empiricism entertained by both representatives.
24. Such a coherence model has been developed specifically for educational research by J.C. Walker and P. Moran in their "Theory, Practice and Discipline - Problems and Solutions" *Radical Education Dossier*, 20, pp. 20-23 Sydney, 1983.

## References

- ALKIN, M. C., DAILLAK, R. and WHITE, P. *Using Evaluations: Does Evaluation Make a Difference?* (Beverly Hills, CA: Sage) 1979.
- CAMPBELL, D. T. "Factors Relevant to the Validity of Experiments in Social Settings". *Psychological Bulletin*, 1957, (54): 297-312.
- CAMPBELL, D. T. and FISKE, D. W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix". *Psychological Bulletin*, 1959 (56): 81-105.
- CAMPBELL, D. T. and STANLEY, J. C. *Experimental and Quasi-Experimental Designs for Research*. (Chicago: Rand McNally 1966.
- CAMPBELL, D. T. "Reforms as Experiments". *American Psychologist*, 1969 (24): 409-429.
- CAMPBELL, D. T. "Evolutionary Epistemology". In P. A. Schilpp (Ed.) *The Philosophy of Karl Popper*. (La Salle, Ill: Open Court) 1974 (a).
- CAMPBELL, D. T. "Qualitative Knowing in Action Research". Kurt Lewin Award Address. Society for the Psychological Study of Social Issues. Presented at 82nd. Annual Meeting of the American Psychological Association, New Orleans, September 1, 1974 (b).
- CAMPBELL, D. T. "Descriptive Epistemology: Psychological, Sociological, and Evolutionary (Preliminary Draft of the William James Lectures, Harvard University) Spring 1977.
- COOK, Th.D. and GRUDER, Ch.L. "Metaevaluation Research". *Evaluation Quarterly*, February 1978, 2(1): 5-51.
- COOK, Th.D. and CAMPBELL, D. T. *Quasi-Experimentation: Design and Analysis Issues for field Settings*. Chicago: Rand McNally 1979.
- CRONBACH, L. J. and Others. *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*. San Francisco: Jossey-Bass 1980.
- CRONBACH, L. J. *Designing Evaluations of Educational and Social Programs* (San Francisco, CA: Jossey-Bass 1982.
- DEMING, W. E. "The Logic of Evaluation". In E. L. Struening and M. Guttentag (Ed.s) *Handbook of Evaluation Research* Vol. 1, (Beverly Hills, CA: Sage) 1975.
- DEWEY, J. *The Sources of a Science of Education*. (New York: Horace Liveright) 1929.
- DUHEM, P. *The Aim and Structure of Physical Theory*. English edition published by Princeton University Press, 1954.
- DUNN, W. N., MITROFF, I.I., and DEUTSCH, S. J. "The Obsolescence of Evaluation Research". *Evaluation and Program Planning*, 1981, 4: 207-218.
- EVERS, C.W. and WALKER, J.C. *Epistemology, Semantics and Educational Theory*. Occasional Paper No. 16. Department of Education, The University of Sydney, Australia, 1984.
- EVERS, C. W. "Epistemology and Justification: From Classical Foundationalism to Quinean Coherentism and Materialist Pragmatism". In C. W. Evers and J.C. Walker, *Epistemology, Semantics and Educational Theory*, op. cit.
- FEYERABEND, P.K. *Against Method: Outline of an Anarchistic Theory of Knowledge*. (London: New Left Books) 1975.
- GLASS, G.V., MCGAW, B., SMITH, M.L. *Meta-Analysis in Social Research*. (Beverly Hills, CA: Sage) 1981.
- GOULD, S.J. *Ever Since Darwin*. (New York: W.W. Norton) 1977.
- GOULD, S.J. *The Mismeasure of Man*. (New York: W.W. Norton) 1981.
- GUBA, E. "The Failure of Educational Evaluation". In C.H. Weiss (Ed.). *Evaluating Action Programs* (Boston: Allyn and Bacon) 1972.
- GUTTENTAG, M. "Evaluation of Social Intervention Programs". *Annals of the New York Academy of Sciences*, 1973, 218-3-13.
- HAMILTON, D. et al. *Beyond the Numbers Game*. (Berkeley, CA: McCutchan), 1977.
- HANSON, N.R. *Patterns of Discovery*. (Cambridge: Cambridge University Press) 1958.
- KUHN, Th. S. *The Structure of Scientific Revolutions*. (Chicago: Chicago University Press) 1970.
- LAKATOS, I. "Falsification and the Methodology of Scientific Research Programmes". In I. Lakatos and A. Musgrave (Eds.) *Criticism and the Growth of Knowledge*. (London: Cambridge University Press) 1977.
- LYKKEN, D. T. "Statistical Significance in Psychological Research", *Psychological Bulletin*, 1968, 70 (3): 151-519.
- MacDONALD, B. and WALKER, R. "Case Study and the Social Philosophy of Educational Research". In Hamilton, D. et al. *Beyond the Numbers Game*. (Berkeley, CA McCutchan) 1977.
- MEEHL, P.A. "Theory Testing in Psychology and Physics: A Methodological Paradox". *Philosophy of Science*, 1967 (34): 103-115.
- MITROFF, I. and BONOMA, Th.V. "Psychological Assumptions, Experimentation and Real World Problems -- A Critique and an Alternate Approach to Evaluation". *Evaluation Quarterly*, May 1978. 235-257.

- PETRIE, H.G. "Theory into Practice: Educational Policy Analysis and the Cultural Perspective". In T. J. Sergiovanni and J. E. Corbally (Eds.) *Leadership and Organizational Culture*. (Urbana: University of Illinois Press) 1984.
- QUINE, W. V. O. "Speaking of Objects". In J. A. Fodor and J. J. Katz. *The Structure of Language*. (Englewood Cliffs, N.J.: Prentice Hall)
- QUINE, W. V. O. "Two Dogmas of Empiricism". In his book *Ontological Relativity and Other Essays* (New York: Columbia University Press) 1969.
- QUINE, W. V. O. "The Nature of Natural Knowledge". In S. Guttenplan (Ed.). *Mind and Language*. (Oxford: Clarendon Press) 1975.
- QUINE, W. V. O. and ULLIAN, J. S. *The Web of Belief* (New York: Random House) 1978.
- RIECHKEN, H. W. and BORUCH, R. D. (Eds.) *Social Experimentation - A Method for Planning and Evaluation Social Intervention* (New York: Academic Press) 1974.
- RORTY, R. *Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press) 1980.
- ROSS, L. and CRONBACH, L.J. "Review Essay: Evaluating the Handbook of Evaluation Research". In G.V. Glass (Ed.) *Evaluation Studies Review Annual*, Vol.1 (Beverly Hills, CA: Sage) 1976.
- STAKE, R. E. "The Countenance of Educational Evaluation". *Teachers College Record*, 1967, 68: 523-540.
- STAKE, R. E. *Evaluating Educational Programmes: The Need and the Response* (Washington, DC: OECD Publications Center 1976.
- STAKE, R. E. "The Case Study Method in Social Enquiry". *Educational Researcher*. 1978, 7: 5-8.
- WALKER, J.C. and EVERS, C.W. "Epistemology and Justifying and Curriculum of Educational Studies". *British Journal of Educational Studies*. 1982, 30(2).
- WALKER, J. C. "Materialist Pragmatism and Sociology of Education". *British Journal of Sociology of Education*. 1985. 6(1).