

Towards checks and balances in educational evaluation

Excerpt¹ from:

Brügelmann, H. (1974): Towards checks and balances in educational evaluation - On the use of social control in research design. In: MacDonald, B./ Walker, R. (eds.) (1974): Innovation, Evaluation, Research and the Problem of Control. Some interim papers. SAFARI project/ Centre for Applied Research in Education/ U.E.A.: Norwich (25-66).

Note: Starting point of the paper and central issue in chapters 1 and 2 are the causes of distortion in processes of gaining information and its evaluation. Basic idea is the distinction of technical precision as a means of enhancing objectivity and validity of these processes and of social control as an alternative (favoured by the author).

[...]

3. TRYING TO CROSS THE QUICKSANDS: A TENTATIVE MAP²

The rules and standards suggested in the following sections are intended to exemplify the 'controlled relativity' paradigm outlined in section 2.3. These rules have experimental status. At the moment they are just hypotheses as there is not sufficient evidence about their practicability and effectiveness. Moreover, I doubt that they can be prescribed generally, anyway: they usually will have to be redefined with respect to the different intentions and circumstances of particular projects.

3.1 ACTION RESEARCH AND CASE-STUDY

There are two research traditions that offer practically relevant suggestions; one is the 'action research', the other the case-study paradigm. I do not want to engage in the tedious debate of definitions, here. Let us look instead at three quotations that illuminate significant features of both approaches:

"Case study research attempts to reach understanding through the detailed study and portrayal of particular facets of individual cases."

(R Walker, 1974 (b), p. 2)

¹ Slightly revised version, September 2006

² Several issues raised in this part have been discussed extensively with Barry MacDonald and Rob Walker. We have not reached agreement on all points (cf. their papers included in this publication). This seems not to be possible, even desirable at the moment. We need, however, a public discussion of and careful experimentation with, ideas of the kind proposed. I hope that some of the suggestions put forward stimulate a reflective exchange of ideas and experiences.

Comparing it with a representative survey, one could say: case-study concentrates on one instance describing as many aspects of it as possible; a survey selects a few variables analysing them on a representative scale. The first is comprehensive while the latter is comparative.

"Action research aims to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework."

(R N Rappaport, quoted in J Elliot/C Adelman 1974)

The FORD Teaching Project tries to integrate the case-study methodology into an action research approach (the last paper quoted is an outline of this approach). In a proposal for a project to study and support the dissemination and implementation of new educational ideas, J Rudduck and myself are taking a similar line:

"The proposed project, then, would have two main functions:

- a research function: it would attempt to build theory out of close observations of particular sets of problems and intentions; it would attempt to understand the nature of these problems in context;
- an action support function: it would cooperate with innovating groups and programmes both by providing guidelines where these could be offered and by providing consultancy where the individuality of the problem eluded a guideline approach

We do not consider these two roles incompatible - indeed, the capacity to fulfil both is the condition of successful consultancy and research."

(J Rudduck/ H Brügelmann 1974,p. 2; cf for this combination also R Rippey, 1973).

In the following sections I shall draw mainly on four projects that roughly represent the range of ideas and experiences relevant here. All of them are further developing the approaches mentioned in practical settings.

Firstly, there is the Teacher's Centre (Pädagogische Arbeitsstelle) St. Gallen/Switzerland where in 1971 E Hengartner has set up a project on primary maths involving teachers from the beginning in an cooperative and continuing effort of curriculum innovation (cf. his

workshop report titled "in service education and action-oriented curriculum development" (together with H M A Weinrebe 1974).

At roughly the same time a project on the "Analysis and Development of a Teacher Orientated Strategy of Curriculum Innovation" was set up at the Centre for Educational Technology (BTZ) in Wiesbaden, West Germany. This Project worked for two years with a few selected schools to explore their problems and to develop tentative solutions cooperatively, and to help with their implementation and evaluation. They have also developed a very sophisticated concept of activating social research (cf. E Müller et al. 1972; P Bonn et al. 1974).

In 1972/73 two projects were set up at the Centre for Applied Research in Education taking up problems that had arisen in the Humanities Curriculum Project in other contexts. The Ford Teaching Project focuses on action-research in the field of "Inquiry and Discovery Learning" involving groups of teachers in the development and exploration of pedagogic as well as research hypotheses and proving a framework of support from the central team (cf. C Adelman et al. 1973; J Elliot/C Adelman 1974).

The Ford SAFARI Project has subscribed to a case study approach in looking at the impact of recent curriculum innovations (especially four major development projects in Britain) and developing a new methodology of educational evaluation - "anthropological design", "democratic model": (cf. B MacDonald 1974; R Walker 1974 (a) and (b)).

The philosophy and actual practice of the four projects show significant but inconsistent overlap (e.g. they do not clearly define the relationship between 'case-study' and 'action research'). They are, however, similar in reacting against distinctive features of the established paradigm of social research. In an internally circulated paper, I summarised these characteristic under four headings (without suggesting that taking the opposite stance on each of these issues would solve our problems):

- the concentration of research on the formulation of general laws and its neglect of significant features of specific instances and their settings;
- the attempt to draw a clear-cut line between values and knowledge in research design;
- the separation of the acquisition of knowledge from its application in terms of methodology, personnel and institutionalisation;
- the clear distinction between the task of the researcher and the role of the practitioner in the research process itself.

All these issues are related to what is usually called the 'theory-practice' problem. It is important, however, to note that the criticism of the established research paradigm and the attempts to develop an alternative approach draw on two quite different strands of argument.

LEGITIMACY AND EFFECTIVENESS: Having decided, for example, the kind of information one is interested in, one is usually looking for instruments and methods that can yield relevant data. The criterion for the selection of certain designs, here, is the effectiveness of such means in achieving one's ends. Thus, criticism is levelled against the statistical nature of statements in social science, against their abstractness and concentration on isolated variables, and more generally against their inadequateness for describing the complexity of social situations and for supporting action within such settings.

From this we have to distinguish the legitimacy of research methods. Here, we are referring to individual and political values, i.e. some sort of 'ethics of social research': From this point of view the 'subject - object' relationship in traditional research and the assumed neutrality of research in relationship to political interests has been heavily criticised (cf. E House, 1973, for instance). Thus, although a certain experiment might be very effective in producing the data we are interested in, we might reject its use on ethical grounds.

3.2 THE PROBLEM OF PROCEDURAL STANDARDS

What we have to do then is to justify standards for research with respect to both criteria mentioned above. Let me give an example for the case-study approach. In the SAFARI Project one of the basic standards is "confidentiality of data". There are two reasons given for the acceptance of this rule. (cf. B MacDonald/R Walker 1974, 7).

Firstly, it helps the researcher to gain access to the field and to relevant data faster as participants will keep control about what they have said until the final stage of the project (where they will be asked to give permission for publication which they can withhold). In this perspective the effectiveness of research can be increased by strictly applying the principle of confidentiality. That the implications are a bit more ambiguous can be seen from the following comment:

"This sharing of control over data with participants does mean that the researcher often has to face the fact that some of his finest data is lost, diluted or permanently consigned to the files. On the other hand his access to knowledge about what are sensitive issues to his informants may guide his research in significant and unexpected ways."

(B MacDonald & R Walker 1974, p. 7)

The rule can draw support, however, from another source too. It can be justified in the context of ethical and political values that in the SAFARI Project have been summarised under the heading "people own the facts of their lives". Here, then we are concerned with the legitimacy of interview techniques:

"Evaluative information is often obtained either without the subject's knowledge that the information is obtained or without his knowledge of the use to which it will be put . . . Some researchers simply spy. The justification for such activities is pragmatic: they are effective . . . (however) the evaluator should consider whether the knowledge gained is jeopardising his own integrity and that of the profession, and if so, whether it is worth the price."

(H. Grobman, 1968, p. 8)

Thus, the researcher clearly has to accept responsibility for the ethical standards governing his work.

THE STATUS OF RULES: Privacy, however, is not the only value relevant to educational evaluation. This makes the task even more difficult. The researcher himself has to consider and to weigh up different interests impinging on his work. He cannot simply subscribe to a prespecified set of rules and rest content with an instrumental role: he has to be prepared to make normative decisions (cf. for a further elaboration of this problem: MacDonald 1974). Let us look at the principle "people own the facts of their lives" from this angle.

Firstly, as any lawyer can tell, there are many facts in your life that you do not own at all. The legal distinction between what is private and where others may intrude has been drawn very carefully. It is not only VIPs who have to accept that "facts of their private life" are made public (as long as such statements do not fall under the law of libel).

Secondly, most facts of one's life are social events, i.e. there are other people participating who may also claim to 'own' these facts. Can we now disentangle the different bits? One could say, perhaps, that people own their interpretations of certain facts, but not the facts themselves. They should be free, then, to decide which thoughts to get publicised and which not (e.g. they should be protected against researchers enticing them into a situation where they cannot watch their own interests, for instance).

Thirdly, people may not even own their interpretations of certain situations. Let us consider, for example, a reform of teacher education which has to be based on data, about actual styles of teaching. To what extent is a teacher free to decline information here? Clearly, a teaching style is shaped by the personality of the individual teacher; on the other hand it is related to the role of a teacher as defined by institutional norms and influenced by formal training. Can you separate, for example, these aspects when looking at the way a teacher behaves in the classroom?

To go one step further: teachers by the very nature of their task intrude into the privacy of the children they are teaching. To justify such intrusion and to improve the education of the teachers, it may be necessary to gain information, e.g. about the style of

communication in the classroom.

Performance might become distorted unless participants do not know what the researcher is looking for.

This is not to say that we should not respect privacy (e.g. to protect the pedagogical relationship with the children). It rather points to the difficulty of establishing a code of conduct for social researchers. To link this to what I have said earlier, I have to take up the distinction between 'technical precision' and 'social control', again. As we have seen, there may be a clash between different considerations of effectiveness or between certain values or between the criteria of effectiveness and legitimacy themselves. We have also seen that we cannot anticipate all constellations to which these standards may apply. Thus, we cannot build up a stable hierarchy of standards accounting for all conflicts possible. Although we can give examples of how certain conflicts might be solved, the system has to remain incomplete. We have to keep it open so that cases not considered before can be solved adequately: We, thus, cannot take the burden of ethical responsibility from the individual researcher by deciding these issues generally and in advance. However, we can and should support him by publicly discussing the problems involved and by providing examples linked to a tentative framework ('sign-posts') that illustrate and systematise relevant considerations. This could help him to become sensitive to the problem and its implications.

SIMULATION AND CONTROL: One final 'preamble' might be helpful before discussing some rules. in more detail. The rules suggested should not presuppose an idealised setting or the common good-will of participants such as 'non'-distorted communication in the proposals of J Habermas and his colleagues. Researchers like everybody else adhere to political values, have career interests at stake, believe in their ideas and look for reinforcement. Rules or standards for research, therefore, should take such divergence of interests into account. They have to be realistic in the same sense as law is. On the other hand one should not regard researchers as psychopaths or potential criminals. Social rules cannot prevent their violation in single cases, anyway. Though they have to be strong enough to avoid abuse in most cases, they should not distort the process.

In this respect one should distinguish the functions of standards for research. Firstly, they can help the researcher to structure his work in such a way that it becomes more productive. In this sense, a code of conduct can be stimulating and supportive. On the other hand, it should help to control research and to protect the interests of other parties concerned (participants, audiences, sponsors). Rules that are more effective in this sense tend to stifle research work and to restrict the researcher's creativeness. Though I do not want to push this distinction too far, I have emphasized the first aspect more in section 3.3, while the second function dominates section 3.4.

3.3 STAGES AND STANDARDS

In his stimulating paper on "Problems of Inference and Proof in Participant Observation", H S Becker has suggested to structure of the process of research in the form of distinct

'stages', and presenting its conclusions as a "description of the natural history of our conclusions" (1958, 653, 660). Becker has distinguished three stages of field analysis, i.e.

- selection and definition of problems, concepts, and indices;
- checking the frequency and distribution of phenomena, and
- construction of social systems models;

He adds a fourth stage after the field-work has been completed which comprises the final analysis and the presentation of results (op. cit., p. 653).

The number of stages, presumably, will vary according to the nature and length of a project. There are, however, certain principles that have to be considered generally.

PRE-RESEARCH STAGE: Firstly, I think it useful to add a stage before the actual field work starts. Its purpose would be to provide time for the researcher to clarify the interests that have led him to undertake the study, to determine the values that should guide it, and to outline the concepts and the methodology that he will use (at least initially) in the field. As G Myrdal (1970, p. 55) points out:

"A 'disinterested' social science has never existed, and for logical reasons, can never exist . . . The only way in which we can strive for 'objectivity' in theoretical analysis is to expose the valuations to full light, make them conscious, specific, and explicit, and permit them to determine the theoretical research."

Of course, we cannot completely expose our belief system and our "cognitive map". Nor should Myrdal's proposal imply that we are not allowed to change our attitudes and concepts during the research. The term 'determine' in his statement, therefore, is misleading. Such an exercise and the publication of its results can be helpful, however, in several respects.

The researcher, himself may become more aware of his own bias and of the interests of his reference groups (sponsors, for instance). He can offer his outline to future participants in the field to help them in making an informed decision about their participation in the project. Such an account also will help audiences of the final report to understand the research better and to check its basis. Thus, an account of this kind is a necessary condition for research becoming a 'shared' or 'vicarious' experience (cf. R E Stake, 1972).

There is a catch in asking for a statement of this kind, however. The researcher may become identified with certain interpretations at too early a stage. Thus, funding bodies may interpret shifts in this framework as 'failure'. Or, in defending the merit of his proposal, the researcher himself may become committed to it in a way that prevents him from pursuing discrepant information. I do not see any alternative to this suggestion, however. What we have to do, therefore, is this - the researcher has make clear that his

propositions are tentative; that they are conjectures, the value of which should not be assessed in terms of stability, i.e. survival of as many tests as possible, but only with

respect to their fertility in stimulating the exploration of problems (the refutation of a hypothesis can yield more information than its corroboration under trivial conditions). Secondly, expectations towards research have to be changed.(and it is in the interest of all researchers to contribute to this change). Audiences have to understand that research is a learning process for all the people involved. Learning, as K R Popper (1963) puts it, depends on bold conjectures and their continuous refutation or refinement.

The situation is similar to that in curriculum development:

"A. curriculum without shortcomings has no prospect of improvement and has therefore been insufficiently ambitious. What we ask of a curriculum offering is not that it should be right or good but that it should be intelligent or penetrating. Its dilemmas should be important dilemmas, its shortcomings should reflect real and important difficulties."

(L Stenhouse, 1974)

INTERIM REPORTS: It also seems to be useful to distinguish different stages in the field work itself. I would suggest provisional summaries of observations at the end of each, to establish a common language and to check interpretations. Such interim reports, therefore, should be made available to participants immediately. They could also be used as a basis for the "natural history of conclusions" that H S Becker has suggested.

Moreover, this procedure contributes to achieving more 'transparency' in research. This standard is particularly relevant to the relationship between project and audiences: it enables audiences to control the research process (legitimacy) and to use its findings more intelligently (effectiveness).

"In terms of research design the point is not simply that vertical studies are circulated for approval, but for validation. The responses to the study perhaps being added to the study itself may reveal significant new interpretations. Like works of art, case studies are never finished, only left."

(R Walker, 1974 (a) 29)

Because at the beginning, reports have to remain cursory and tentative, anyway, and will pose questions rather than systematise and defend interpretations, participants may feel encouraged to contribute. It is also (psychologically) easier for the researcher to accept comments at that stage and to build them into his framework. The sequence of reports, then, should reveal the changing perspective of the researcher as his role is redefined, from being a detached observer ('stranger') to becoming familiar with the peculiarities of

the 'subculture' under study. Audiences can benefit from both perspectives and their specific advantages, and the effect of "immersing into a culture" can be made explicit.

Reports should also help participants to understand what the researcher is after (not in terms of judgements, but in terms of problems and aspects he is interested in). Moreover, the way the researcher is reporting could be a means of encouraging them to continue the research: i.e. they themselves can test how he is using evidence, if he is impartial, if he is keeping informants' names anonymous when they have requested this, etc. Participants can calculate the risk of providing information. This procedure, then, will help to gradually build up mutual understanding and trust. As the researcher cannot easily cheat (usually he would have to report conflicting points of view), such 'tests' might help him to gain permission for the publication of data at the end which he would not have got otherwise.

Again, I think one should not underestimate the problems of such a procedure. Reports will often bring conflicts out into the open. They may contain comments of participants about each other. Thus, such reports carry a considerable disturbance potential. On the other hand, one might argue that it is better for all participants to explore the degree of frankness that can be tolerated gradually and internally instead of being confronted with a big clash at the end. The assumption then is that conflicts cannot be avoided, anyway, and that the researcher has some obligation to face the consequences of his interference rather than leaving participants with a well-equipped arsenal of weapons on the battlefield at the end. I. shall come back to this action-research aspect later.

PUBLICATION: The separation of a final stage for the selection and presentation of data has several advantages. The main point is that we can employ an independent person for this task. The longer the researcher stays in the field, the more he will become a participant with a defined status and personal affiliations - however much he strives to be impartial. Moreover, he has had a chance to present his perspective in the interim reports. It may be useful to have somebody more 'innocent and ignorant' being responsible for the selection and presentation of evidence (cf. the summary of the evaluation reports of the Humanities Curriculum Project by D Hamingson (1973) for example).

It would be worthwhile to submit this outsider version to the researcher and the participants for further comments. An editor should decide, therefore, which (if any) of these comments are illuminating for readers and should be included.

One should bear in mind that a case-study has two quite different functions: one is to describe a specific instance as accurately as possible ('verisimilitude'); the second is to stimulate other people's perception and judgement ('richness'). In this sense a 'finished' case-study can be the stimulus and instrument for further research (cf. R Walker 1974 (a), p. 29). A report, thus, can serve as a 'foil' for others analysing their situation. This leads me to the second standard I want to consider in this context.

RESEARCH AS A LEARNING EXPERIENCE FOR ALL PARTICIPANTS: The value behind this is the idea of 'democratisation of knowledge' (cf. B MacDonald 1974, esp, p. 10). This implies two things the researcher should strive for increasing the relevance of knowledge;

and he should ensure its accessability to all parties interested. As R E Stake (1972, pp. 5 and 1) has pointed out, case-study is the paradigm that in both respects is most 'responsive' to the needs of audiences:

"The responsive approach tries to respond to the natural ways in which people assimilate information and arrive at understanding... The researcher arranges for various persons to observe the program, with their help he prepares brief narratives, portrayals, product displays, graphs, etc. He finds out what of that is of value to his audience. He gathers expressions of worth from various points of view. Of course, he checks the quality of his records. He gets program personnel to react to the accuracy of his portrayals. He gets authority figures to react to the importance of various findings. He gets audience members to react to the relevance of his findings ... The promise of gain is two-fold. The readers will comprehend and the complexity of the program is not lost."

Another facet of this is the exchange of information between researcher and participants.

From the researcher's point of view, 'learning' implies a broadening of design, especially in not limiting it to the control of pre-specified criteria (cf. the notion of a goal-free evaluation in M Scriven (1973)). This also points to the need to focus attention not only on students, but on other participants as well; not only on persons, but also on institutions; not only on short-term, but also on long-term effects; not only on intended outcomes, but also on side-effects; not only on effects, but also on process. and context (cf. J M Atkin, 1963, E W Eisner, 1967; B MacDonald/ M Parlett, 1973, for further elaboration). As R E Stake (1972, p.1) put it, evaluation should be 'responsive' rather than 'preordinate'.

If participants' needs of information are taken into account, the researcher has to depart from the role of a detached observer who collects data, takes them away, processes them on an abstract level, and, after a long time, feeds them back in a format that makes it difficult for participants to recognise their own problems in the publication. Instead, we need speedy feedback of useful information that has to be gathered and processed by less formal methods (cf. W Klafki 1973, p. 514).

Sometimes it is necessary to make the research and the relationship between researcher and participants itself a focus of attention. This has led the curriculum project at the Centre for Educational Technology in Wiesbaden to emphasise the role of 'practice-oriented discourse'. Thus, they document the communication between different parties in

the project setting and discuss the documents together as soon as possible (cf. P Bonn et al 1974, p. 36; also cf. C Adelman, 1973).

The immediate feedback of information in a way that is useful for participants in solving their actual problems also makes the researcher more attractive for practitioners. Thus, it increases the effectiveness of research by giving him access to the field. (for difficulties related to this, cf. below 3.4: "Evaluation - consultancy - support - policy").

"Learning experiences for all", however, means more than feedback of data. As I have argued earlier, social situations are specific. They differ from one point in time to another, as well as from one setting in space to another. E.Hengartner/H M A Weinrebe have stressed the need to organise research in a way (1) that researchers can learn from teachers (whom they regard as 'experts of classroom practice' (1974, p. 10), and (2) that teachers can learn about procedures of curriculum development and evaluation as well as about general concepts that can help them to analyse their experiences from another angle (cf. for a similar concept: F M Connelly, 1972). This perspective led to a long-term cooperation between researchers and teachers, and to the establishment of a network of teacher groups that can itinerate the process of development and evaluation and, thus, provide a learning experience for an increasing number of teachers with the support from more experienced colleagues. (cf. for a similar model: J Elliott/C Adelman 1974) .

In this way, the results of research should become more readily disseminated and intelligently implemented. Moreover, the competence of practitioners is enhanced and a social nucleus for reflective practice built up:

"Another facet of the democratic mode, however, is that the methods of research should, through the process of the research, become available to those being studied, so that when the project terminates they will have, not just a copy of the report, but access to skills which allow them to continue to research unaided."

(R Walker, 1974, p. 15)

ACCOUNTING FOR DISCREPANCIES: Another standard will be the range of divergent views presented. The researcher should not only be interested in regularities, but also in inconsistencies. Moreover, he should not try to eliminate discrepancies, but to account for them. As I put it in an earlier part of this paper (cf. 2.3) he should aim at increasing the degree of interrelatedness of data and their sources. This, again, has several implications.

Firstly, both the legitimacy and the effectiveness of research demand that the researcher should accept the mode of thinking and communication in the specific setting which he is studying: he should accept and strive to understand the 'Lebenswelt' of the participants in the project (cf. H Dräger et al. 1973, p. 46). Thus, he is trying to explore and to reconstruct the concepts underlying their practice. It is assumed that perceptions

will vary according to biography and role (interests, reference groups) and, thus, have to be represented.

This does not necessarily mean that the researcher should refrain from giving his own interpretation of events. As it was stated in the Humanities Curriculum Project, the divergence of views available and the range of interpretations taken into consideration

seem to be a useful means to increase understanding. (cf. The Humanities Curriculum Project: An Introduction, 1970, p. 20). Now, the researcher is a person who can introduce discrepancy when things are taken for granted. This technique is well-known from the theatre scene (cf. B Brecht, for example) and has deliberately been used in research, too.

It was one of the aspects of the 'researcher as stranger' in the FORD Teaching Project (cf. J Elliott/C Adelman, 1974) to introduce an additional perspective (cf. also E Müller et al. 1972 for the Wiesbaden Project). J Rudduck/H Brügelmann (1974) have suggested using international cooperation as a means to confront and complement experiences from different settings in the interpretation of common problems.

Another consequence is the way interpretations are selected and presented at the end. Not only participants, but also audiences will benefit from studying discrepancies. To take a concrete example, if nine people agree on a certain interpretation (A) and only one person opts for (B), it might be more useful to report both (A) and (B.) and to relate these interpretations to background data than to jump to the conclusion that "90 per cent of this group answered (A)". Both types of information are important and complement each other. But the majority principle is an insufficient criterion for judging the adequateness of interpretations. Consensus is often rooted in convention, while minorities may be the avant garde of a better understanding.

The advantage of case-studies is that they can interrelate conflicting points of view and provisionally 'reconcile' them on a higher level. Such reconciliation, however, always will depend on the point of view of the interpreter. Thus, the model could lead to a process of 'indefinite triangulation' (A V Cicourel 1973, p. 124)

In section 2.3 I have mentioned three pragmatic criteria by which to determine the adequate degree of 'interrelatedness'. I want to recall them here for convenience: firstly, the resources available; secondly, the increase of information gained; thirdly, the feasibility of the representation for researcher and audiences. Let me give a concrete example of this, again.

The FORD Teaching Project is interested in what they call 'teacher development', i.e.

" the process of helping teachers to become more aware of what they actually do in a way that changes their perceptions of possibilities in the classroom situation and consequently affects the degree of reflectiveness expressed in their decisions."

(J Elliot, 1973, p. 2).

To achieve this and to develop a medium for communicating findings to other people, the project prepared case-studies. One of these is, "Three points of view in the classroom", which contains the following (cf. J Elliott/D Partington 1973):

- a transcript from a tape-recording of a lesson;
- the observer's notes taken during the lesson;
- the observer's notes on the transcript written independently from looking at any other data 10 weeks after the lesson;
- an interview between the observer and the teacher after the lesson;
- an interview between the observer and the students after the lesson;
- a joint discussion between students and teacher (after he had listened to the tape) chaired by the observer.

The approach shows interesting similarities to the procedure adopted by R D Laing/A Esterson (1970) in their studies of "families of schizophrenics" where they repeatedly interviewed members of a family in different constellations to build up a network of interrelated interpretations

EVALUATION OF EVIDENCE: This brings me to a question glossed over on the previous pages: by what means can we assess the usefulness of evidence?

In his summary of different "models and methods in evaluation research", M Guttentag refers to M Levine's (1968 - 1973) proposal to adopt a 'legal model' in evaluation. According to this summary Levine argues:

"... that the scientific enterprise as a whole can be conceived of as following an adversary model in which there are claims and counterclaims, arguments and counter-arguments, and each side advanced by an advocate who attempts to make the best possible case for this position. Issues of evidence, the inferences which can be drawn from evidence, and methods for evaluating arguments and evidence constitute the process of ascertaining truth in the legal system. In the legal system there are rules which govern the presentation and evaluation of evidence."

(M Guttentag, 1971, p. 86)

Apparently, the 'adversary model' is a useful social mechanism to control the process of research as it is protecting divergence of views. The device must not, however, be misunderstood as a technical procedure 'ascertaining truth' as Guttentag put it. We have to be careful not to make the same mistake that many lawyers have made in using the legal methodology as if it were a set of technical rules and not a system of social roles (cf. for this distinction chap. 2 above). We should also note that in recent years lawyers becoming

aware of the deficiencies of their own system are increasingly looking for alternative models in the field of the social sciences. By simply adopting their paradigm we might fall out of the frying-pan into the fire.

Finally, we have to be aware of a crucial difference between the legal system and social research, respectively. Law is about settling conflicts within a limited amount of time. The

legal procedure, therefore, is decision-oriented. Its outcome has to be one interpretation. Given that two points of view conflict, we have to falsify one of them. Transferring this to social research, we come close to I Lakatos' position that research is not about falsifying theories by confronting them with facts but about resolving inconsistencies between conflicting theories. From what I have said earlier (cf. section 2.3) it should be clear that I do not agree fully with this position.

Research is about increasing understanding, i.e. it has to increase complexity rather than reduce it. The latter is a significant feature of decision-oriented procedures. Understanding should feed decision-making, it should not pre-empt it. That is to say, research should help to make decisions more adequate rather than make them easier.

Moreover, actions are always specific. They are located in a unique setting in time and space. Understanding strives for generalisation. To guide action it has to be reinterpreted within specific contexts. We would lose information if we tried to translate it into solutions on this general level. Indeed, this is what happens with legal rules, and I think it is appropriate as we need some general guidelines that are susceptible to as unambiguous an interpretation as possible. That is to say, legal rules have to exclude ambiguities and to refer to more easily identifiable indicators: to minimise the risk of arbitrary decisions in single cases we accept some arbitrariness in the general rules, already. We put reliability over validity. The case is oversimplified here, but I think one can agree on this: in solving the conflict between material justice and formal consistency, law more than research, tends to the latter. (This comparison, perhaps, may help to understand better some of the deficiencies of the classical research model and its plea for 'standardisation' too.)

In science we can afford ambiguities as long as the 'consumer' himself makes the decisions. As soon as somebody is making decisions in authority we have to look for procedures similar to the ones in law. It should be clear, then, that no decision whatsoever can override opposing points of view solely with reference to the 'scientific' nature of the evidence it is based upon.

Thus, I am strongly pleading for a division of labour between research (and the standards governing it) and policy-making (and its procedures and criteria). I am not saying, by the way, that there is no overlap between the 'context of valuation' and the 'context of justification' in research (cf. for this distinction, E R House, 1932, p. 416), or that 'knowledge' and 'values' are two independent factors. However, for practical purposes it is useful to devise different procedures for different aspects of social reality in both of

which cognition and evaluation play a significant role (cf. for further implications and some difficulties of this position, section 3.4 'research - consultancy - support - policy').

This excursus should have strengthened our assumption that in research we are not concerned with judging the adequacy of interpretations. We have, however, to make explicit the conditions under which information is gained. Here again, I find useful H S Becker's (1958, p. 645) inventory of criteria such as:

- Did participants act or did they talk in accordance with a certain interpretation or did they both? Did they show consistency under changing conditions and in different constellations?
- Was there agreement between different participants or not?
- Did participants volunteer to give a statement, did they answer a direct question, or did they agree with a suggested interpretation?
- Did the researcher record a certain bit of information accidentally or was he looking for evidence related to a certain hypothesis?

All this should not be taken to distinguish between 'better' or 'less valid' information. It may help us, however, to account for discrepancies. Thus, 'volunteering' in one case may mean that somebody has given evidence without being influenced; in another case it may be that somebody deliberately attempts to 'guide' the researcher in a certain direction. Therefore, it is very important to accumulate information from different sources and by different procedures to cross-check data and to build up a densely interrelated 'social system' to accomplish

„ ...greater accuracy by successively refining the model to take account of evidence which does not fit his previous formulation, by searching for negative cases (items of evidence which run counter to the relationships hypothesised in the model) which might force such revision; and by searching intensively for the interconnectedness in vivo of the various elements he has conceptualised from his data."

(H S Becker, 1958, p. 658)

TRANSFERABILITY: The "Construction of Social System Models" (loc. cit.) leads us to the final stage and the problem of how to present information. There are two kinds of audiences that have to be distinguished. Firstly, the participants themselves are interested in an account of the situation they are living in. Secondly, outsiders are interested in data that are useful for them in a situation perhaps quite different from the one where the data were collected. Both groups, however, need information that is of

relevance to more than just one situation. Our standard, therefore, has to be the generalisability, or, as I would prefer to call it, the transferability, of information.

Now, we have reached a critical point, as it is usually assumed that at least in this respect the classical model of representative surveys is superior to the description of specific instances. Underlying this judgement, however, is an assumption that I hope to have attacked successfully in the course of this paper: the problem of generalisation is interpreted as a technical one. The effect of this definition, however, is this:

"In its current state of development, educational research cannot bridge the gap between 'we know from statistical evidence that this innovation works' and 'you can make this work in your class!'"

(L C, Taylor, 1971, p. 229; quoted from H Simons, 1971, p. 119).

R Walker talks about a 'paradox' within which educational research has been operating:

„... it aims for understandings which have been generalised from specific and localised information, always aiming for high levels of predictive reliability. Yet it also aims to inform practitioners who operate in the context of problems and decisions which are essentially idiosyncratic and unique. The problem they face is not how to generalise from, but how to generalise to. In this context quantitative research offers the practitioner trends and patterns, but never judgement in particular cases."

(R Walker, 1974 (b), p. 2)

This, I want to argue, is a problem of communication: it poses the question if we can present information and convey experience in a form that helps other people to learn from it. In answering this question, we have to take into account the nature of the situations in which a practitioner has to solve problems:

"...relationships among variables in education are highly interactive. To put the matter in slightly different language, the relationships which educational researchers investigate are highly context dependent... If physical laws were as limited in generality as the laws so far discovered by social scientists, we would hesitantly creep out of bed each morning not knowing whether we would float to the ceiling or crash to the floor."

(G V Glass, 1972, p. 11).

The standard we have to adopt, then, is openness'. As I have put it in the context, of curriculum development, generalisations have to remain non-prescriptive, specifications can only be tentative and have to be adaptable, and the design as a whole must be perceptive of specificities. That is to say, the presentation of research (as well as of development

work) has to be in the form of hypotheses and should provide support for the translation of these hypotheses into different settings (H Brügelmann, 1974, p. 1).

In this sense, case-studies are not only a means of research, but also a medium of communication (cf. p. 9 above; I come back to this issue in section 3.4. below: "confidentiality vs. identifiability"). The presentation of a case-study should not only describe how more general hypotheses were developed from specific situations, it should include also some guidance for how to translate such hypotheses into a different setting.

Thus, it has to report an instance of tentative concretisation under specified conditions as well. It should make explicit the difficulties of this process which has to be conceived of as an 'exploration' of problems rather than as a mere 'application' of authoritative laws.

This procedure of generalizing from one specific situation to another (which, however, as I believe can be facilitated by middle range 'theories') is very close to everyday practice. I sometimes refer to it as the 'Miss Marple paradigm'. Like Hercule Poirot, the other main character in A Christie's detective stories, she adheres to the 'triangulation' principle (cf. section 2.3 above) in building up an image of the world the crime happened in. Unlike Poirot, however, she always draws on the very elaborated model of the neighbourhood in the small village where she normally lives. She is using her reflected experience of that sub-culture as an heuristic framework or foil to interpret new situations.

The task, then, is to show how case-studies can be used as a research instrument. In this sense, I think, we can say that they 'respond' to the needs and abilities of people working in the field: they can be written in a language that is theirs, and they are comprehensive, i.e. they can provide some sort of 'vicarious experience' (cf. chap. 2 and p. 7 above). This is not to say that there are no problems with the transfer of knowledge through case-studies (the most serious ones will be discussed in the following section). Especially, one should not forget the need to help practitioners to understand the different nature of this kind of research evidence from the 'cargo cult' into which they have been initiated by classical research. This is a tremendous task as experiences in the field of curriculum innovation show (cf. for the problem to be overcome the ambiguous attitude of practitioners to 'experts': Hengartner/H M A Weinrebe 1974, p. 17). Moreover, the skills and techniques necessary for an adequate use of this kind of research are not common. This points to the need of considerable investment into in-service education and the establishment of support structures. We have to gradually build up a network of competent and confident practitioners and to provide for more exchange of ideas and experience between them.

In classical research it was more expensive to gain knowledge than to disseminate it. In the proposed paradigm it may be the other way round. However, the distinction between an 'acquisition' phase and an 'application' phase may become obsolete, altogether. It is hoped that the 'use' of research evidence in practical situations can increasingly be conceived of as part of some sort of continuous evaluation and revision of hypotheses. In this sense we may talk about inductive accumulation of knowledge. However many tests we

use, though, we can never assume we have reached safe ground. Therefore, we need some additional checks and balances to increase the credibility of research evidence. Some aspects of this problem will be discussed in the following section.

3.4 ROLES AND RESPONSIBILITIES

H Simons (1971, p. 122) has suggested that we could increase the credibility of case-studies if the emerging reports were authentic, detailed, accurate, impartial, fair and inconclusive. On the other hand, as she herself points out, authenticity of evidence and an increase of contextual data may make it more difficult to publish a report. In this section, I want to explore some of the potential conflicts inherent in the case-study and action research approaches.

Confidentiality vs Identifiability

L Stenhouse (1974 (a), pp. 20, 16) has made an interesting point in comparing the status of evidence gathered by standardised measurement and by case-studies, respectively:

"Field study data are strong data which are difficult to organise. Test data are weak data which are susceptible to organisation . . . The field situation in which the action takes place is unique. No attempt to replicate it can succeed . . . Accordingly, the research must aspire to situational verifiability. That is, the findings must be so presented that a teacher is invited not to accept them but to test them by mounting a verification procedure in his own situation . . . Data drawn from such case-studies in schools are in a useful sense hard data. They are rooted in real situations and have a high degree of verisimilitude."

This, I think, is a helpful clarification of the use of case-study data. However, even if results of case-studies have to be regarded as hypotheses, people will be more interested in data the acquisition of which they can check. There are two meanings of evidence being 'rooted in real situations' which we have to distinguish here.

Firstly, there are data that are identifiable: e.g. names of people, institutions, and places are given so that one can check if the information presented is authentic. In this sense, evidence will serve as a basis from which we can infer statements. Its task is to support interpretations and to make their basis explicit, i.e. accessible to scrutiny.

On the other hand, case-studies can be a means of illustration. Here we use examples to communicate interpretations, not to justify them. In this respect we could also use fictional instances because what we want to achieve is some sort of more concrete and lively presentation (in this sense we can talk about the evaluator as an artist).

The standards that apply to both these functions of case-study data are different: in the first instance we are interested in authenticity, while in the second we are striving for richness. Both aspects are important. They should not be mixed up, however.

In terms of social control we are interested in identifiable data, of course. On the other hand, the principle of confidentiality may make it difficult to publish interesting, i.e. complete, accounts in that way. There are two compromises possible as far as I can see.

The first would be to postpone the identification of informants and to publish a case-study "by employing fictitious names and by not pinpointing its geographical location", (H Simons, 1971, p. 122). At the same time one could offer access to the files where the original documents are kept on request. Thus, if anyone should doubt the authenticity of certain data ('a chief education officer can't have said that") one could let him listen to the tape.

With the permission of informants one could also reveal their identity (although this is a difficult problem as the name of one person often will enable outsiders to reconstruct the whole setting). This procedure has been adopted in the evaluation of the Humanities Curriculum Project and no difficulties seem to have arisen up to now.

On the whole it might be more suitable to use the following arrangement. Researchers and participants select several confidential persons who do not belong to either group but will have access to the files. Audiences, then, could submit their requests to these arbiters for checking the accuracy of information. This mechanism perhaps would be most adequate to protect the interests of participants and audiences equally.

It would also help the researcher who could then adopt a three stage procedure for the publication of data: (1) publication with reference to the identity of informants; (2) publication with reference to the roles of informants using unidentifiable descriptions; (3) use as background information without explicit reference (hearsay). In terms of supportive power there is an increase from (3) to (1), of course. In terms of illustrative verisimilitude, the distinction does not matter, however.

In a teacher's centre case-study I suggested to participants that they should:

"... adopt the following principle: data should not be published with reference to the sources unless the interviewees themselves have agreed to it; the report should not include reference to the historical and geographical setting of the centre under study unless the warden and the LEA have agreed. If agreement should not be reached, however, I am entitled to use data and circulate a report anonymously with an additional note giving the reasons why participants withheld their permission. Such a report will have less credibility, but it still might be useful in an heuristic sense."

(Letter to a Teacher's Centre Warden, 28 September 1974).

An arrangement of this kind gives participants enough power over their information to protect their privacy (without the researcher becoming entirely dependent on their willingness to accept certain interpretations).

The principle of confidentiality poses additional problems in the relationship between participants and researcher. Apparently, this relationship is an asymmetric one, i.e. when giving information to the researcher, the informant has to trust him. If the researcher should abuse data (e.g. publish them without permission), the informant has not got any means to defend himself. The only thing he can do is to try to make this known publically afterwards, or limit the amount of information he is putting at risk from the beginning. It seems, therefore, that two things are needed:

Firstly, an ethic as outlined here has to be endorsed by a professional association. Thus, a member of such an association might find it easier to get access to the field by referring to the code of such an association and the risk he is running in not adhering to it. The problem at the moment is that no such tradition exists and that it takes time until a public image is built up that could make this a reliable mechanism. Moreover, an organisation of this kind would represent researchers only. Though it would not be interested in keeping black sheep among its members, it may also be tempted to deny any breach of the rules altogether in order to protect the image of the profession. Thus, this mechanism would seem to be useful but not sufficient.

Therefore, an additional arbitrary body to which participants (and researchers) could appeal would be needed. On this body the same people could be elected who act as confidential persons to audiences. Usually, the procedure would be that a version of the report to be published would go back to participants for agreement. They then could appeal to the arbitrators.

Whatever form such a mechanism may take, its mere existence would be an effective means to prevent most violations. Thus, it is important to establish it before the actual research begins. It would help the researcher to gain the trust of participants more easily and it would support the legitimacy of research as well (without distorting it). An ethical code is fine. Effective sanctions seem to be better.

Research - Consultancy - Support - Policy

The role of the researcher in the SAFARI Project has sometimes been referred to as an "honest broker". It is his task, therefore, to facilitate the exchange of judgements. A responsibility relevant to this role is impartiality.

"The purpose of the case-study is to make the experience of innovation accessible to public and professional judgement, and not provide a vehicle for the biases or personal judgements of the evaluator."

(H Simons, 1971, p. 122)

I think it is important to emphasise that the researcher should strive for impartiality; this does not imply that impartiality can be used as a demarcation criterion. Selection of data and their interpretation inevitably are value-laden; they are not, however, determined by the values the researcher holds. Thus although somebody may believe in the potential of compensatory education, this does not mean that he cannot detect that hereditary factors play a significant part in the development of children. Impartiality, therefore, has to be seen as a regulative criterion or role expectation. E R House (1972, p. 410) has used 'fairness' and 'honesty' as standards that describe a similar stance:

"We measured the program at perhaps its point of greatest strength - that's what I mean by being fair. . . . Where the program did turn out badly we reported it. Through, familiarity with the program we also knew where the weakest points lay . . . That's what I mean by being honest."

Another problem we have to face is this: should the researcher try to restrict his task to just presenting judgements or should he give his own interpretation of the situation under study as well. I think both avenues are possible. There may be situations where both participants and audiences can learn most from a mere collection of interpretations focussing on some key issues:

"Case-study presentations should be basically inconclusive accounts of what happens in a particular school and should contain accurate reports of the judgements, convergent or divergent, of those involved in the events. Misinterpretation can be largely avoided by . . . seeking the agreement of those involved to the fairness of the report. Of course, there is no escaping the judgement of the evaluator in terms of what to include and what to leave out of the account . . ."

(H Simons, 1971, p 122)

As the selection of evidence will inevitable follow a certain strategy, it might be useful to make these assumptions explicit. This, then, comes close to the attempt to build up a 'model' of the situation under study, especially, when audiences need contextual data providing some sort of background against which controversial interpretations become meaningful. Moreover, the perspective of a 'stranger' will throw additional light on the events reported. His perspective will be particularly helpful to audiences who are in a similar position as he was in the beginning: they are 'innocent' and 'ignorant' of the sub-culture they want to understand. It is important, however, that the researcher is aware of his image as an authority and does not attempt to impose his interpretations. Secondly, he has to make explicit the basis of his judgements and to provide 'excess information' so

that audiences can check his interpretations. The obligation to submit any report to participants for agreement of accuracy and fairness and the separation of the role of an 'editor' from that of the researcher are additional safeguards (cf. 3.3 above).

Now, the standard of impartiality throws up particular problems in the relationship between researcher and participants. The question is to what extent is the role of the researcher passive (i.e. restricted to collecting evidence) and to what extent should and can he accept responsibility for becoming actively involved in the situation under study? From what I have said earlier, it should be clear that the researcher cannot claim any particular authority on which to base a certain policy. It seems difficult, then, to justify a change-agent approach aiming at the implementation of a particular solution. Moreover, this would endanger the possibility of doing research in the sense described here, at all.

On the other hand, a researcher staying in the field for some time may be asked by participants to give his opinion on certain issues, he may be consulted when difficulties arise, or he may even feel obliged to support participants instead of taking a detached stance of scientific interest when they run into difficulties. Thus, ethical considerations as well as the effectiveness of research (exchange of information and mutual support may make the researcher more attractive to participants) point to the need of adopting an action-research. approach.

Moreover, a consultancy or support role enables the researcher to gradually refine his hypotheses by offering them to participants in situations where they can test them. In this way, the researcher would provide instruments rather than impose ends. In accepting a support role he also could help to build up a research capacity in the field that makes practitioners less dependent on outside authorities.

While the SAFARI Team accepts the interventionist impact of research as a side-effect, only, I would try to use it as an instrument of research (cf. J Rudduck/ H Brügelmann, 1974; C Adelman et al. 1973). The kind and degree of support, then would have to be determined by its fertility in increasing understanding both in participants and audiences. A second criterion would be compensation to participants for the difficulties caused by the research itself.

The use of the notion 'participants', however, may still blur the problem. Different participants will have different interests. Giving information to somebody in need of it may strengthen his value position instead of that of somebody else.

The SAFARI Project has attempted to overcome this difficulty by defining its stance as follows:

"Democratic evaluation is an information service to the whole community about the characteristics of an educational programme... The democratic evaluator recognises value pluralism and seeks to represent a range of interests in his issue formulation. The basic value is an informed citizenry,

and the evaluator acts as broker in exchanges of information between groups who want knowledge of each other... The key justificatory concept is 'the right to know'."

(B MacDonald, 1974, p. 11)

The basic assumptions of this role definition are these:

- the researcher cannot claim any particular authority on which to base interventions;
- the democratisation of knowledge is the most effective means to destroy monopolies of information;
- the accumulation of knowledge depends on public criticism of conflicting points of view.

This is not to say that this position is value-free. The values governing it have been clearly stated. Now, we have to explore their potential and implications in practice. I do see problems associated with this position, e.g. unequal opportunities for different groups to use knowledge because of the unequal distribution of power in society. On the other hand, there seems to be no viable alternative available.

The practicability of the position suggested rests on the effectiveness of the checks and balances outlined: participants are controlling each other through the involvement of all significant parties; researchers and participants are controlling each other by submitting evidence from different points of view and negotiating the publication of reports; researchers are controlling each other because of the separation of responsibilities; audiences can control participants and researchers by means of an independent arbitrary mechanism.

Such distribution of power and responsibilities can draw on models from social psychology and law, in particular (such as in-group/out-group relationships, for example). As I mentioned earlier, concrete suggestions put forward in this paper have to be regarded as hypotheses. Their adequacy should be explored in the course of projects in different areas. We are in need of such experiments to develop a framework of guidelines. It would be helpful if critical-comments on this paper could be based on such experience.