4 APPROACHES TO EDUCATIONAL RESEARCH AND ITS ASSESSMENT

The range of strategies that can be used to pursue educational research is very wide. It includes: laboratory and classroom experiments; large-scale surveys of the behaviour, attitudes, aptitudes, etc., of teachers, pupils and others; secondary analysis of available information; small-scale investigations of particular schools or classrooms, etc. The data may be the product of direct observation on the part of the researcher or it may be produced by others, and can take a variety of forms: ticks in boxes on interview or observational schedules, numbers as recorded in published statistics, text from published or unpublished documents or from field-notes written by the researcher during the course of observations or interviews, audio- or video-recordings and transcripts of these, etc.

As we saw earlier, a common way of conceptualizing this diversity is the distinction between quantitative and qualitative approaches. This is a distinction we shall use in the remainder of this Handbook. It is necessary, however, to emphasize that it is a very crude distinction and one that is potentially misleading. It obscures as much as it reveals.

4.1 DISTINGUISHING BETWEEN QUANTITATIVE AND QUALITATIVE RESEARCH

What do we mean by the terms 'quantitative' and 'qualitative'? The most obvious distinction between the two sorts of research is that the former deals with numbers and usually employs statistical techniques, whereas the latter does not, or does so only to a minor degree. Various other features of the research process tend to be associated with each side of this division. For instance, if one is to count, calculate percentages and use statistical analysis, one must have data that are amenable to these procedures. As a result, quantitative research typically employs what are usually referred to as structured forms of data, consisting of frequency counts or measurements of other kinds.

By contrast, the data that qualitative researchers typically deal with are verbal accounts in natural language, produced by them or by informants, and full or edited transcripts from audio- or video-recordings of events. As they stand, such data cannot be subjected to counting or measurement and therefore are not susceptible to quantitative analysis; though they can subsequently be structured so as to become amenable to such analysis.

Although there are constraints on what methods can be combined, it is easy to exaggerate the degree to which different forms of research design, data collection, and data analysis belong exclusively together. Much educational research combines qualitative and quantitative methods in various ways and to varying degrees; and there is probably even more scope for doing so than is currently exploited. At the same time, there are those who argue that qualitative and quantitative approaches represent contrasting forms of educational research, and that they should not be combined. For example, Smith and Heshusius (1986) claim that the history of educational research, and of social research generally, must be seen as involving the development of these two competing paradigms of research. They further claim that quantitative and qualitative approaches are

fundamentally at odds with one another, since they are based on conflicting assumptions about the nature of the phenomena studied and about how those phenomena can be understood. This contrasts with the view that we have adopted in this Handbook, which is that the difference between quantitative and qualitative educational research is neither so clear-cut nor so deep. (For further discussion of this view see Hammersley, 1992a, chapter 9.)

There is, of course, a wide range of ideas about how educational research should be done, how its products should be assessed, and even about what its purpose is; but these cannot be reduced to two contrasting positions without distortion. Furthermore, much of this diversity stems from the fact that different studies are designed to serve different purposes, and their character is shaped by those purposes. It must also be remembered that research is very much a practical activity carried out in particular contexts under varying constraints.

One of the areas where disagreements about quantitative and qualitative approaches have been most significant is the assessment of research. On the one hand, some have argued that the standards of assessment typically employed in quantitative research – such as internal and external validity, and reliability – should be applied to qualitative work as well. On the other hand, many qualitative researchers argue that distinctive standards are appropriate to qualitative studies. We adopt neither of these approaches. We argue that the same standards should be applied across the qualitative–quantitative divide, but at the same time these are not identical with those conventionally applied to quantitative research.

4.2 STANDARDS OF ASSESSMENT

We suggest that there are just two overarching criteria in terms of which educational research, both quantitative and qualitative, should be assessed: validity and relevance. We shall discuss each in turn.

VALIDITY

By 'validity' we mean truth: the extent to which an account accurately represents the phenomena to which it refers. There are few concepts that have led to more controversy among philosophers than truth. Indeed, there are many educational researchers who reject this concept as inapplicable to the assessment of their work. One reason for this is that they believe that use of the term 'truth' implies the possession of knowledge whose validity is *absolutely certain*, proven beyond all possible doubt; yet knowledge can never be certain in this sense.

These critics point out that in deciding the validity or otherwise of some claim we always rely on presuppositions whose own validity we must take for granted. If we seek to test any of those presuppositions, we shall be forced to rely on further ones. For instance, even in simple measurements of physical objects with a ruler we take for granted certain properties of rigid bodies, in this case the ruler and the object. We assume, for example, that small changes in temperature will not have any significant effect on our measurements. Testing these presuppositions would involve us in measuring temperature, which itself involves presuppositions about the operation of thermometers, and so on.

Now, the second half of this argument is sound, but the first – that use of the concept of truth implies knowledge that is certain beyond all possible doubt – is not. To claim that some statement is true is not incompatible with a recognition that our judgements about it may be wrong. In fact, we can *never* be absolutely certain about the truth of anything, not even in the natural sciences or in our personal lives. There are, however, many things about whose truth we are very confident, and about which we are justifiably confident. In other words, there can be good reasons for believing that some statement is true or false without our ever being absolutely certain about the validity of the claim. We rely on a whole host of presuppositions about the world in our everyday lives and, while many of

them are probably approximately true, we can never be confident beyond all possible doubt about the truth of any of them. Yet, this ever-present uncertainty does not undermine our use of the concept of truth in that context, and there is no reason why it should do so in the research situation either.

A second source of problems with 'truth' arises from beliefs about the nature of human social life. There are those who hold that in the case of social phenomena there is no single reality to which claims made in research reports correspond. It is argued that it is characteristic of human beings that they create multiple social worlds or realities, that all perception and cognition involves the *construction* of phenomena rather than their mere discovery. It may be concluded from this that contradictory views of the 'same' phenomena by different cultural groups are equally 'true' in their own terms. Of course, if we apply this argument to educational research itself, as we should if we are to be consistent, we see the latter as also creating a world (or, given the dissension among social scientists, multiple worlds). This is the sort of extreme relativism that we discussed in Section 2, from which point of view research reports do not provide an account of independent phenomena, but *create* the social reality they purport to describe, for instance through textual strategies of various kinds. Effectively, the argument here is not just that we can never be sure of the truth or falsity of our claims about reality, but that we have no grounds for believing that there are phenomena that are independent of our knowledge of them, since all the knowledge we can ever have is formed by our culture and that culture is only one of many.

While this argument makes an important point about the limits and difficulties of understanding that result from cultural diversity, the conclusion that there are no phenomena independent of researchers for them to document does not follow from it. All social phenomena are human products and are therefore, in some senses, not independent of humanity as a whole. Much social life, however, *is* independent of any particular researcher or group of researchers. Furthermore, we are able to learn other cultures to one degree or another and, thereby, to understand human behaviour that is framed in terms of them. This implies some commonalities among cultures on which knowledge may be built. It is also worth pointing out that to claim that there are other cultures, or even that one's own views represent a culture, itself implies the claim that there is a larger world within which those cultures are to be found and which is not relative. In this way, and others, relativism is self-contradictory.

Of course, even if we accept that validity is a feasible and legitimate criterion in terms of which to assess educational research, the question remains: on what basis can assessment of it be carried out, given that there is no foundation of evidence whose validity is absolutely certain and by which researchers can validate their accounts? The only basis available, it seems to us, is judgements of the likelihood of error. From this point of view, there are three steps in assessing the validity of research claims:

1 The first question that we must ask about a claim is how 'plausible' it is: that is, whether we judge it as very likely to be true given our existing knowledge. In the case of some claims, they will be so plausible that we can reasonably accept them at face value without needing to know anything about how researchers came to formulate them or what evidence is offered in support of them. The first test, then, is plausibility.

2 A second question we may need to ask is whether it seems likely that the researcher's judgement of matters relating to the claim is accurate, given the nature of the phenomena concerned, the circumstances of the research, the characteristics of the researcher, etc. We shall call this 'credibility'. In assessing credibility we make a judgement about the likely threats to validity involved in the production of a claim, and the likely size and direction of their effect. As with plausibility, there are claims

whose credibility is such that we can reasonably accept them without further ado (albeit always recognizing that we *could* be mistaken).

3 Of course, where we conclude that a claim is neither sufficiently plausible nor sufficiently credible to be accepted at face value, to be convinced of its validity we shall require evidence. When we examine the evidence, however, we shall have to employ much the same means to assess its validity as we applied to the claim itself: we will judge *its* plausibility and credibility. Of course, in turn we may require further evidence to support the evidence, which we shall again judge in terms of plausibility and credibility.

In many respects, this seems to us to be the sort of basis on which we judge claims, our own and those of others, in everyday life. It is important to point out, though, that we do not do it on a purely individual basis. As co-participants in various communal activities we compare our judgements with those of others; and where there are disagreements these may need to be resolved. In a research situation, moreover, there is an obligation to try to resolve disagreements and to do so through rational discussion. Such discussion may reveal to us that what we have been disposed to accept as adequately plausible or credible should not be accepted (or it may reveal errors in the presuppositions on which others' judgements were made). It is the function of the research community to act as a corrective to the beliefs of individual researchers in this way.

Needless to say, plausibility and credibility are a relatively weak basis for judging the validity of claims, compared to the idea that we can assess claims directly according to their correspondence with reality, or by relying on some body of evidence whose validity is absolutely certain. The approach we are suggesting provides no guarantee that our judgements will be correct, nor any way of knowing for certain whether they are correct. Neither will judgements always be consensual, since there are very likely to be different views about what is plausible and credible; and rational discussion carries no guarantee of reaching agreement. In our view, however, this is the only viable basis for assessing the truth of knowledge claims that we have.

Truth or validity is the first criterion, then, in terms of which we believe that research accounts should be judged. Also important is 'relevance'.

RELEVANCE

To be of value, research findings must not only be valid, but must also be relevant to issues of public concern. This second criterion is curiously neglected in many treatments of educational research methodology. It is rarely mentioned explicitly in discussions of standards of assessment, whether those appealing to the quantitative tradition or those offering distinctive qualitative criteria.

Our interest in facts, in everyday life as much as in educational research, is selective. All descriptions are for some purpose and the nature of the purpose will crucially shape the character of the description. If, say, we are describing a classroom lesson, it will make a considerable difference whether our interest is in the extent to which there is balanced participation between girls and boys, in whether the teacher seems to discriminate against black students, in how the knowledge that forms part of the syllabus is presented, or in something else. The descriptions produced on the basis of these various interests may overlap, but equally they may be so different as to be not recognizably referring to the same lesson (although if we believe that they are all true they should not contradict one another). Just as there is always a large number of points of view from which we could *describe* a phenomenon, so, too, with *explanations*. These will not only be concerned with accounting for some aspects of a phenomenon rather than others, but will also involve the selection of explanatory factors partly according to the purpose that the explanation is to serve.

It is also of significance in this context that research reports are communications addressed to an audience. When we communicate with people they assume that we are telling them something that is likely to be of significance to them. If it turns out that we merely communicate facts, any facts, we shall soon find that we have few listeners. In the case of most educational research, the aim is to communicate with a relatively large audience (otherwise, what would be the point of publication?). It follows that what is communicated should be relevant in some way to such an audience. The obvious questions that follow from this are: who are the appropriate audiences and what sort of relevance should educational research reports have for them?

Intended audiences for research reports vary of course. Sometimes reports are directed primarily towards other researchers; sometimes they are addressed to a particular practitioner audience; sometimes they are directed at a general audience. Whatever the intended audience, though, there are two aspects of relevance that need to be distinguished:

- *The importance of the topic*. The research topic must relate (however remotely) to an issue of importance to the intended audience.
- *The contribution of the conclusions to existing knowledge*. The research findings must add something to our knowledge of the issue to which they relate. Research that merely confirms what is already beyond reasonable doubt makes no contribution.

In these terms research findings may connect with an important topic, but still not be relevant since they do not tell us anything new about it. Conversely, research may add new knowledge, but this may relate to no topic of any significance and so still lack relevance. Importance and contribution are necessary and jointly sufficient conditions for relevance, and we must assess any set of findings in terms of both; though, of course, there may be disagreement about these matters.

Later in this section we shall look at what is involved in applying the standard of validity to examples of educational research.[9] Before we can do this, though, we must look at the sort of reading of research reports that is necessary to lay the basis for such assessment.

4.3 READING FOR UNDERSTANDING

The first task in reading a research report is to try to understand what it focuses on and the arguments it contains. This means that we need to look for information of the following kinds: about the research focus; about the cases studied; about the methods of data collection and analysis; about the major claims and the evidence provided for them; and, finally, about the conclusions drawn concerning the research focus. While research reports are not always organized in these terms, information relevant to all of these matters can usually be found and it is useful to organize one's notes in terms of these headings. We shall spell out in a little more detail what we mean by each of them.

THE RESEARCH FOCUS

We use the term 'research focus' to refer to the most general set of phenomena (one or more) about which a study draws conclusions, and the aspects of those phenomena that are of concern. Equally, we can think of the research focus as the general set of questions that the study addresses. There are, of course, different sorts of question that can be addressed. The research may be primarily designed to answer descriptive, explanatory, predictive, evaluative or prescriptive questions. The distinctions among these types of question are not always made by researchers and, thus, may not be clearly marked in research texts. These different

[9] Applying the standard of relevance is something that only individual readers can do, since judgements of the importance of the topic, and perhaps even of whether the conclusions of the research contribute to existing knowledge, are likely to vary considerably across audiences.

types of question, however, place different requirements on the researcher, in terms of the sort of evidence that is required to support answers to them. They therefore have implications for what evidence the reader should expect.

Besides the focus itself, we must also look out for any rationale that the author offers as to why the focus is important and thus why the conclusions might be of interest to us. Most research reports provide some explicit statement of their focus and of its assumed significance, but this is not always the case. Here, for example, is the opening paragraph of an article on 'Gender imbalances in the primary classroom':

(French and French, 1984, p. 127)

Activity 2

On the basis of this brief extract, what do you take to be the focus of French and French's research?

French and French give a clear account here of the focus of the research and of its relationship to other literature. The aim is to provide an explanation for gender imbalances in the distribution of teachers' attention and of pupils' participation in classroom interactions. They give no indication, however, of why they believe this focus to be important; nor do they provide this elsewhere in the article. Where a rationale for the focus is absent, we need to think about what the rationale might be.

Activity 3

Why do you think the authors believed this topic to be important?

There are various possibilities, but it seems to us that French and French probably regarded gender imbalance in classroom interaction as important because it might have effects on the relative achievement levels of girls and boys later in their school careers, which in turn might affect their job prospects.

Where there is a discussion of the focus of the study, and of a rationale for this, this usually occurs at the beginning of the account, though relevant material may be found in any part of it. In book-length studies, the rationale for the research focus may amount to one or more chapters, including reviews of the literature that identify the gap that the study is intended to fill.

THE CASES INVESTIGATED

There is a need for care in distinguishing the focus of the research from the case or cases studied. We defined the focus as the most general set of phenomena about which the study makes claims. Usually, this will be a *type* of phenomenon or some feature of a large *aggregate* of cases. Examples include 'school bullying', 'learning algebra', 'the pedagogical styles of secondary-school teachers in the United Kingdom', 'the curricular organization of primary schools in England and Wales', etc. By 'the case studied', on the other hand, we mean the phenomena specifically located in place and time to which the data analysed relate.

Educational research can vary considerably in the number of cases investigated. Quantitative research typically, but not always, studies quite large numbers: fifty or more, perhaps even thousands of cases. Qualitative research, on the other hand, usually deals with a small number of cases, perhaps even just one.

Sometimes the focus of the research is restricted to the cases studied, where these are of general interest in themselves. Sometimes the stated focus is the case studied even when this does not seem to be of obvious general interest. For example, in a study of classroom interaction, Hammersley explicitly limits his conclusions to what occurred in the particular inner-city secondary modern school he studied at the time he studied it (Hammersley, 1974). Such a cautious limitation of focus is surely misleading, however. After all, why would this study warrant publication if it did not have implications for some more widely relevant focus? Where this happens, we should probably conclude that the focus is different from the case studied, but is left implicit; and we must try to identify what seems likely to have been the intended focus.

Most accounts provide at least some information about the cases selected and the reasons for their selection. Of course, the information that could be provided is virtually endless. What we need is information that allows us to judge how confidently we can draw conclusions about the focus of the research from findings about the cases studied. This is necessarily a matter of judgement. Here, for instance, is what French and French provide in their article on gender imbalances in classroom interaction:

(French and French, 1984, p. 128)

In addition to information about the cases studied, we must also look out for any indication of why the cases are believed to be of relevance to the research focus, or at least the basis on which they were selected. This may involve the claim that the cases are typical of some category or population of cases. In the case of quantitative research the cases studied may be a systematically selected sample from the aggregate that is the focus of the research. Indeed, sometimes the selection will have been made on a random basis, in such a way as to maximize the chances that the cases studied are representative. Often, though, there will be an element of convenience built into the sampling strategy, and this sometimes makes generalization problematic. For example, in a well-known study of teaching styles and pupils' performance in primary schools, Bennett studied a large sample of teachers, but they all worked in schools in the North West of

England, within easy access of Bennett's base at the University of Lancaster (Bennett, 1976). This means that, in some respects at least, these teachers were probably unrepresentative of larger populations, such as primary teachers in England and Wales. In the case of qualitative research, where a relatively small number of cases is usually studied, generalization to a larger population is often even more problematic. Here, though, as with Bennett's work, judgements of the wider significance of the findings can usually be made on the basis of information about the sample itself, on the one hand, and whatever information is available about the wider population to which generalization is to be made, on the other.

IDENTIFYING THE MAIN ARGUMENTS AND EVIDENCE

The most central task in understanding a study is to identify its main findings, and the evidence presented in support of them. Above, we drew a distinction between the research focus and the case or cases studied. Corresponding to this, we shall also make a distinction between the conclusions of a study, which relate to its focus, and the major claims, which relate to the cases. In looking for the major claims of a study we are therefore looking for those findings that refer specifically to the cases investigated, leaving any that refer beyond them for later consideration as conclusions. Of course, there will often be considerable overlap in content between the major claims and the conclusions; that which is found in the cases studied frequently being generalized to other cases. The distinction between claims and conclusions is important because assessing the validity of such generalizations is an extra task over and above assessment of the claims themselves.

Within any research report there will be a host of claims made and the task of identifying the main ones involves detecting the argumentative structure of the account so as to see which are superordinate and which subordinate. Usually this is not too difficult: authors will indicate to one degree or another which are the most important points they are making, and how the others relate to these. Sometimes, of course, authors provide summaries of their claims. These may occur in the conclusions at the end of an article, or in concluding chapters of books, or they may come earlier. Here, for example, is the author's summary, provided quite near the beginning of his article, of an account of the effects of variations in teachers' expectations on pupils' performance in the kindergarten and early grades of an inner-city school in a predominantly black urban community in the United States (we have introduced paragraphs into this extract for the purpose of clarification).

> The argument may be succinctly stated in five propositions.
>
> First, the kindergarten teacher possessed a roughly constructed 'ideal type' as to what characteristics were necessary for any given student to achieve 'success' both in the public school and in the larger society. These characteristics appeared to be, in significant part, related to social class criteria.
>
> Secondly, upon first meeting her students at the beginning of the school year subjective evaluations were made of the students as to possession or absence of the desired traits necessary for anticipated 'success'. On the basis of the evaluation, the class was divided into groups expected to succeed (termed by the teacher 'fast learners') and those anticipated to fail (termed 'slow learners').
>
> Third, differential treatment was accorded to the two groups in the classroom, with the group designated as 'fast learners' receiving the majority of the teaching time, reward-directed behaviour, and attention from the teacher. Those designated as 'slow-learners' were taught infrequently, subjected to more frequent control-oriented behaviour, and received little if any supportive behaviour from the teacher.

> Fourth, the interactional patterns between the teacher and the various groups in her class became rigidified, taking on caste-like characteristics during the course of the school year, with the gap in completion of academic material between the two groups widening as the school year progressed.
>
> Fifth, a similar process occurred in later years of schooling, but the teachers no longer relied on subjectively interpreted data as the basis for ascertaining differences in students. Rather, they were able to utilize a variety of informational sources related to past performance as the basis for classroom grouping.
>
> (Rist, 1970, pp. 413–4)

Such summaries are extremely useful in giving us a sense of the overall argument of a study, but we should not rely on them entirely. Sometimes, we shall find that, in our own judgement at least, they are not entirely accurate or complete. For example, *our* summary of Rist's main claims would include one that he omitted: that the pupils themselves responded differentially to the teacher's behaviour towards them, in ways that reinforced the teacher's expectations. Also, where Rist refers to two groups of children in the summary (fast and slow learners), the teachers he studied divided their classes into three groups, though still ranging from 'fast' to 'slow'.

Although summaries of the main claims provided by the author are useful, then, they are no substitute for one's own careful reading of the whole text. At the very least, one may find differences of emphasis between the summary and the actual account, and sometimes there can be important discrepancies. Finally, it is worth noting that authors may mix what we have distinguished as claims and conclusions.

Once we have identified the main claims, we need to think about what *sorts* of claims they are. There are several types of claim to be found in educational research reports, and it is worth distinguishing among these because they require different sorts of evidence. They can be listed under three headings, as follows:

1. Definitions
2. Factual claims
   - (a) descriptions
   - (b) explanations
   - (c) predictions
3. Value claims
   - (a) evaluations
   - (b) prescriptions

Definitions

Definitions tell us how a particular term is being used in an account. For instance, in introducing a distinction between teaching and survival strategies on the part of secondary school teachers, Woods provides the following definition of 'survival':

> It is, in short, a *survival* problem. What is at risk is not only [the teacher's] physical, mental and nervous safety and well-being, but also his continuance in professional life, his future prospects, his professional identity, his way of life, his status, his self-esteem ... .
>
> (Woods, 1979, p. 145)

Woods is spelling out here what he means when he claims that teachers are concerned with 'survival'. This is essential information if we are to be able to assess his argument that survival was a predominant concern among the secondary-school teachers he studied.

It is unlikely that the central claims of a study would be definitional. As in the example of Woods' work, however, definitions may well form an important part of the substructure of the argument.

Factual claims

By factual claims we mean those that document features of the cases studied by the researcher, and their causes and consequences. These claims may take several forms: descriptions, explanations, or predictions.

Descriptions Descriptions are one of the most important sorts of argument to be found in research texts, since all of the other types, apart from definitions, depend on them. We cannot explain, predict, evaluate, or prescribe about something without describing it, or at least assuming some description of it.

What we mean by a description is a verbal or numerical representation of some feature of an object: for example, of a person or a situation. Often, descriptions will be quite complex, relating to multiple features of many objects. Here is some of the descriptive information about the participation of boys and girls that French and French provide in the research on gender imbalance in primary classrooms referred to earlier:

(French and French, 1984, pp. 127–8)

In this extract we have a mixture of numerically and verbally presented information that describes some aspects of the pattern of interactions between the teacher and pupils in the lesson studied, in particular as regards gender differences.

It is worth noting that descriptions are always selective, they never include all the information that could, in principle, have been provided about the phenomena studied. Thus, French and French do not tell us about the amount of time the children spent on and off a task, what their relative levels of ability were, what proportions came from working-class and middle-class homes, etc. What is selected should, of course, relate to the focus of the research, as indeed it does in this case. And all of the information relevant to that focus needs to be presented.

Explanations Where descriptions provide an account of features of some phenomenon, explanations are concerned with *why* one or more of those features occurred; and seek to show that they are the product of particular factors, operating directly or indirectly. We can illustrate this in relation to Rist's study of teacher-pupil interactions in the early years of schooling. It is the central claim of his account that the different levels of achievement of the children are largely a product of the teachers' expectations of them, based on social class. Therefore, besides providing descriptions of the teachers' procedures for allocating children to groups, and of their interactions with members of these groups, Rist also makes

an explanatory claim. Indeed, the descriptive claims are subordinated to the task of explanation.

Explanations, like descriptions, are selective. For any phenomenon, we can identify a potentially infinite number of causes. There may be one or more immediate causes and each of these can in turn be explained in terms of other more remote causes – and so on, *ad infinitum*. For instance, we may explain the differential performance of the children that Rist studied in terms of differences in academic ability, or in their attitude to school. Each of these variables, along with Rist's own preferred explanatory factor (the teachers' expectations), could be explained on the basis of more remote factors, for example differences in the children's home background. Furthermore, home background could, in turn, be explained by appeal to family histories, or to the nature of the local communities from which the children come. These, again, may be accounted for on the basis of various features of the character and history of American society, notably its racial and social-class structure (as indeed Rist seeks to do later in his article). The implication of this is that where explanations are presented we must look out for any indication of the criteria that have guided the selection of explanatory factors.

Predictions By 'predictions' we mean claims that if particular conditions are met certain other events will occur. Predictions are symmetrical with explanations. In explaining something we start with what happens and look for significant causes; in predicting we start with some actual or possible state of affairs and pick out likely effects that are significant. Once again, relevance to the focus underlies the process of analysis here. There is an infinite number of potential consequences, immediate and more remote, that may follow from any given conditions. Not all will be judged relevant.

Predictions do not often form part of the major claims in educational research. They do occur as part of the substructure of the argument, however. Indeed, as we shall see in Section 6, they play an important part in much quantitative analysis, where what actually happened is compared with what would have been expected to happen given particular theoretical assumptions. As with explanations, when we identify predictive claims we must look out for the theoretical claims implicit in them, and for the relevances that have structured them.

Value claims

By value claims we mean those which not only describe, explain and/or make predictions about a feature but also express some view about it in terms of one or more values. There are good arguments for suggesting that research reports should not include value claims, since the primary task of research is to produce factual knowledge and, as we have seen, value judgements cannot be based solely on such knowledge. However, research reports do sometimes include such judgements.

Evaluations Evaluations involve descriptions of phenomena and perhaps explanation of them. In addition, they also give some indication of whether the things described or explained are good or bad, or in what respects they are good or bad. Some educational research openly shuns evaluative intent, being exclusively concerned with describing or explaining what *is* rather than what *ought to be*. As we saw in Section 2, however, there is a considerable amount of such research that is explicitly devoted to evaluation. Moreover, it is not uncommon to find evaluative claims embedded in many other educational research reports, too. For example, in Peter Woods' description of the process of choosing a course at the secondary school he studied, he contests the teachers' claim that 'the advice and guidance offered is given in the best interests of the pupil' (Woods, 1979, p. 51). Here he is moving beyond description and explanation to an evaluation of the teachers' actions (Hammersley, 1990a).

Of course, evaluations necessarily imply some set of values and we need to look out for information about what these are.

Prescriptions Occasionally, on the basis of their research, researchers outline some recommendations about what changes ought to be made in the phenomena that they have studied. In the conclusion to an account of the coping strategies that teachers use to deal with the constraints imposed upon them by the British educational system, Andy Hargreaves offers the following:

> The crucial axis which might provide the possibility for radical alteration and humanisation of our educational and social structures would seem to be that which connects teacher 'experience' to structural constraints. Change of such a magnitude demands the active involvement of teachers in particular, and men and women in general, in the collective criticism of existing practices, structural arrangements and institutional goals. Furthermore, the possibility of change is contingent upon the provision of institutional conditions under which such collective criticism could take place and be reflexively integrated with ongoing practice. Paradoxically, this requires the fulfilment of 'gradualist' policies such as small class-sizes and the creation of more 'free-time' so that a meaningful integration between theory and practice might arise and thus produce a reconstruction of teacher 'experience' on radical lines, infusing it with the power of transformation.
>
> (Hargreaves, 1978, pp. 81–3)

What is being suggested here is not very clearly specified and it is worth noting that Hargreaves does not explicitly recommend the possible 'radical alteration and humanisation' or the gradualist policies he mentions, but it seems obvious that he is nonetheless to be read as prescribing them.

The reader needs to watch carefully for evaluations and prescriptions that seem to take the form of descriptions, explanations or predictions. As with evaluations, in trying to understand prescriptions we must try to identify the values underlying them.

IDENTIFYING CONCLUSIONS

Earlier, we drew a distinction between major claims and conclusions, the former applying to the case or cases studied, the latter going beyond these to deal with the focus of the study. Therefore, as the final step in trying to understand a research report, we need to identify the conclusions of the report. Sometimes there will be a section labelled 'Conclusions', but as we noted earlier authors do not always distinguish between major claims and conclusions. It is therefore not unusual to find major claims as well as conclusions summarized in closing sections. As an example, here is the final section from the article by French and French:

(French and French, 1984, pp. 133–4)

Here, on the basis of their claim that there was an imbalance in participation in the lesson they studied in favour of the boys, and that this was produced by attention-seeking strategies used by a small number of them, French and French conclude that such strategies may also be responsible for gender imbalance in other contexts. On this basis they recommend that teachers need to be aware of this if they are to avoid distributing their attention unequally.

As this example indicates, conclusions, like main claims, can be both factual and value-based. We will focus here on factual conclusions. There are two ways in which researchers may draw such a conclusion. One is by means of what we shall call 'theoretical inference' and the other is by generalizing the findings in the case studied to a larger number of cases, which we shall refer to as 'empirical generalization'.

Theoretical inference

In the sense in which we are using the term here, theories are concerned with why one type of phenomenon tends to produce another, other things being equal, *wherever* instances of that first type occur. While studies concerned with drawing theoretical conclusions do, of course, have an interest in particular phenomena occurring in particular places at particular times, that interest is limited to the relevance of those phenomena for developing and testing theoretical claims.

Most experimental research relies on theoretical inference in drawing conclusions. For example, in the case of Piaget's research on children's cognition, mentioned in Section 2, he sought to create situations that would provide evidence from which he could draw conclusions about the validity or otherwise of his theoretical ideas. His critics have done the same. Thus, Donaldson and others have carried out experiments designed to test his theory against competing hypotheses, such

as that the children in his experiments simply did not understand properly what they were being asked to do.

Non-experimental research may also rely on theoretical inferences, but often the distinction between theoretical inference and empirical generalization is not drawn explicitly, so that sometimes it is not easy to identify the basis on which conclusions have been reached.

Empirical generalization

Instead of using the cases studied as a basis for theoretical conclusions, researchers may alternatively seek to generalize from them to a finite aggregate of cases that is of general interest. For example, Woods claims that the secondary school he studied was:

> ... ultra-typical in a sense. Pressure was put on the teachers to prosecute their professional task with extra zeal; both the task, and the strategies which supported or cushioned it, were, I believe, highlighted in consequence. In turn, the pressures on the pupils being greater, their resources in coping were stretched to great limits and appeared in sharper relief. Thus, though the school could be said to be going through a transitional phase, it was one in which, I believe, typical processes and interrelationships were revealed, often in particularly vivid form.
>
> (Woods, 1979, pp. 8–9)

Woods does not tell us which larger population of schools he believes this school to be typical of, but we can guess that it was probably secondary schools in England and Wales in, say, the 1970s and 1980s.

In examining the conclusions drawn in research studies, we therefore need to consider whether theoretical inference or empirical generalization, or both, are involved. As far as theoretical conclusions are concerned, we need to be clear about the theory that the cases have been used to develop or test, and about why those case studies are believed to provide the basis for theoretical inference. Where empirical generalization is involved, we must look out for indications of the larger whole about which conclusions are being drawn, and for the reasons why such generalization is believed to be sound.

4.4 READING FOR ASSESSMENT

Understanding the argument of a research report is usually only the first task in reading it. Often, we also wish to assess how well the conclusions are supported by the evidence. In Section 4.2 we introduced the concept of validity as a standard to be used in assessing research reports and explained how we thought the validity of claims and conclusions could be assessed. We suggested that this assessment had to rely on judgements of plausibility and credibility.

It is rare for the major claims in research reports to be so plausible that they need no evidential support. It is unlikely that any such claims would be judged to have much relevance. Faced with a claim that is not sufficiently plausible to be accepted, the second step is to assess its credibility. Here the task is to decide whether the claim is of such a kind that a researcher, given what is known of the circumstances of the research, could have made a judgement about the matter with a reasonably low chance of error. Here, we must use what knowledge we have, or what we can reasonably assume, about how the research was carried out. For instance, we must look at whether the research involved the researcher's own observations or reliance on the accounts of others, or both; and whether the claims are of a kind that would seem unlikely to be subject to misinterpretation or bias. Again, it is rare for major claims to be sufficiently credible to be accepted at face value.

As an illustration of judgements about credibility, let us consider again the research of Ray Rist into the effects of teachers' expectations of pupils' school performance (Rist, 1970). In the course of his study, Rist makes claims about which pupils were allocated to which classroom groups by the kindergarten teacher he studied. It seems to us that we can conclude that his judgement of this distribution is unlikely to be wrong, given that it involves a relatively simple matter of observation, that he observed the class regularly over a relatively long period, and that on this issue he seems unlikely to have been affected by bias. However, Rist also makes the claim that the three groups of children received differential treatment by the teacher. In our view, the validity of this second claim should not be accepted simply on the basis of his presence as an observer in the situation. This is because multiple and uncertain judgements are involved: for example, judgements about amounts and types of attention given to pupils by the teacher over a lengthy period of time. Thus, while we might reasonably accept Rist's first claim as credible on the basis of what we know about his research, we should not accept his second claim on the same basis.

If we find a claim very plausible or highly credible, then we should be prepared to accept it without evidence. If we judge a claim to be neither sufficiently plausible nor credible, however, then we must look to see whether the author has provided any evidence to support it. If not, then our conclusion should be that judgement must be suspended. If evidence *is* provided we must assess the validity of that evidence in terms of *its* plausibility and credibility. Where that evidence is itself supported by further evidence, we may need to assess the latter too.

As we saw earlier, claims can be of several types, and different sorts of evidence are appropriate to each. Let us look at each type of claim and the sort of evidence required to support it.

DEFINITIONS

Definitions are not empirical claims about the world, but statements about how an author is going to use a term, about what meaning is to be associated with it. As such, they are not open to assessment in the same manner as factual claims, but this does not mean that they are open to no assessment at all.

One obvious assessment we can make of a definition is whether it has sufficient clarity for the purposes being pursued. Where there is a standard usage of a term that is clear enough for the purposes at hand, no definition is required. Many concepts used in educational research, however, are ambiguous or uncertain in meaning and yet they are often used without definition. A notorious example is 'social class', which can have very different definitions, based on discrepant theoretical assumptions; usage is often vague. Many other concepts raise similar problems.

Faced with uncertainty about the meaning of key concepts, whether or not definitions are provided, we must give attention to two aspects of that meaning: intension (the concept's relationship to other concepts) and extension (its relationship to instances).

To clarify the intension of a concept we must identify other elements of the network to which it belongs. Concepts get some of their meaning by forming part of a set of distinctions that is hierarchically organized. We can illustrate this by looking at Woods' contrast between teaching and survival strategies on the part of secondary school teachers, mentioned earlier (Woods, 1979, chapter 7). We can note that despite their differences, these are sub-types of a higher-level category – teachers' strategies. This opens up the question of what other types of action teachers use in the classroom, besides strategies. Similarly, at the other end of this conceptual network, Woods himself identifies a variety of different sorts of survival strategy. Putting these two points together, we can see how the distinction between teaching and survival strategies forms part of a larger

conceptual structure which can be represented diagrammatically as shown in Figure 1.

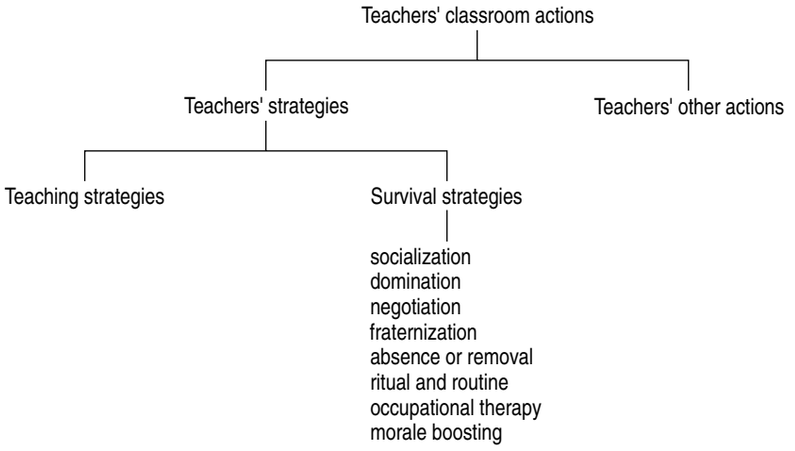

Figure 1

By mapping out conceptual networks of this kind we may be able to see weaknesses in the formulation of key terms. In the case of Woods' study, it seems that the distinction between strategies and other forms of classroom action on the part of teachers might need clarification.

The second aspect of meaning, extension, concerns what would and would not count as instances of a category. Sometimes the problem of identifying instances may be quite difficult. Staying with the example from Woods, he argues that survival strategies:

> ... expand into teaching and around it, like some parasitic plant, and eventually in some cases the host might be completely killed off. However, like parasites, if they kill off the host, they are a failure and they must die too, for they stand starkly revealed for what they are. The best strategies are those that allow a modicum of education to seep through. Alternatively, they will appear as teaching, their survival value having a higher premium than their educational value.
>
> (Woods, 1979, pp. 146–7)

Although the definition of survival is reasonably clear in its intension, its extension is problematic. Given that the concern with survival may masquerade as teaching and that survival strategies may have educational value, how are we to distinguish instances of the two? This issue may be of considerable importance if we are to assess the validity of Woods' claims.

Another basis on which we may criticize definitions is that they fail to make distinctions that we believe are important, given the goal of the research. An example arises in Jean Anyon's report of a study of teacher–pupil relations in five elementary schools in the United States. Discussing her findings in two of the schools, catering primarily for working-class pupils, she writes the following:

> A dominant theme that emerged in these two schools was student *resistance*. Although some amount of resistance appeared in every school in this study, in the working-class schools it was a dominant characteristic of student–teacher interaction. In the fifth grades there was both active and passive resistance to teachers' attempts to impose the curriculum. Active sabotage sometimes took place: someone put a bug in one student's desk; boys fell out of their chairs; they misplaced books; or forgot them; they engaged in minor theft from each other; sometimes they rudely interrupted the teacher. When I asked the children during interviews why they did these things they said, 'To get

> the teacher mad'; 'Because he don't teach us nothin'; 'They give us too many punishments'. When I asked them what the teachers *should* do, they said, 'Teach us some more'; 'Take us alone and help us'; 'Help us learn'.
>
> The children also engaged in a good deal of resistance that was more passive. They often resisted by withholding their enthusiasm or attention on occasions when the teacher attempted to do something special. ... Passive resistance can also be seen on some occasions when the children do not respond to the teacher's questions. For example, they sit just staring at the board or the teacher while the teacher tries to get them to say the answer, any answer. One such occasion, the teacher shouted sarcastically across the room to me, 'Just *look* at the motivation on their faces'. On occasions when teachers finally explode with impatience because nobody 'knows' the answer, one can see fleeting smiles flicker across some of the students' faces: they are pleased to see the teacher get angry, upset.
>
> (Anyon, 1981, pp. 11–12)

It has been argued, with some justification, that the concept of resistance used by Anyon is insufficiently discriminating. As Hargreaves comments: 'the mistake Anyon makes is to assume that acts of overt social protest are of the same nature as more minor transgressions, pranks and absences of enthusiasm ... . Almost all pupils' actions that fall short of absolute and willing compliance to teachers' demands are counted as resistance by her' (Open University, 1984, pp. 31–2). He contrasts this usage with the more restricted one found among other writers on pupils' orientations to school. They often distinguish a wide range of pupils' adaptations and recognize that not all of pupils' actions in the classroom are oriented primarily towards the teacher (see for example Furlong, 1976; Woods, 1979; Hammersley and Turner, 1980).

Definitions may be an important part of research accounts and, even where they are absent, they may need to be reconstructed by the reader (as far as that is possible). While they cannot be judged in empirical terms, we can assess their clarity; and whether they make what seem to be necessary distinctions given the purposes of the research.

DESCRIPTIONS

There are two main considerations we must bear in mind in looking at evidence for descriptions. First, how plausible and how credible are the evidential claims themselves? Secondly, how convincing is the relationship between them and the descriptive claim they have been presented to support.

Validity of evidential claims

Assessment of the validity of evidence must proceed in much the same fashion as we recommended in assessing the validity of major claims. To start with, we must assess its plausibility in terms of our existing knowledge. If it is very plausible, then we may simply accept it at face value. If not, we must assess its credibility.

In assessing credibility, we must take account of the process by which the evidential claims have been produced. The two basic sorts of evidence to be found in educational research reports are extracts from observational reports by the researcher and information provided by others, whether via interviews, responses to postal questionnaires, documents (including published statistics), etc. Let us look at the sorts of threat to validity associated with each of these sources of evidence.

The researcher's own observations These may take the form of tallies of the responses of experimental subjects, or of answers by interviewees to survey

questions, or they may consist of field-notes or transcriptions of audio- or video-recordings, etc.

There are three general sources of error in observational reports that we need to consider.

First, we must think about the potential effect of the research process and of the researcher on the behaviour observed. This is often referred to as the problem of reactivity. Thus where people know that they are being observed they may change their behaviour: for instance, they may act in the way they believe they are supposed to, rather than the way they usually do. Equally important are the possible effects of the personal and social characteristics of the researcher on the behaviour observed. Thus, the age, gender, social class, 'race', professional identity, etc., of the researcher, or the participants' perceptions of them, may affect what people do and say when being observed.

Clearly, reactivity can be a significant source of error in researchers' observations. It is important, however, to remember what is at issue here. It is not whether the research process or the characteristics of the researcher have affected the behaviour that was observed, but rather whether they have affected it in respects that are relevant to the claims made (and to a significant degree). Often, reactive effects may be judged likely to have occurred, but unlikely to have had a significant effect on the validity of the findings.

A second source of error lies in the nature of what is being observed. Some features (e.g. which pupils are allocated to which group) are less likely to be misperceived than others (e.g. the similarities or differences in treatment of groups). This is not a matter of some features being directly observable in a way that guarantees the validity of observational reports, while other features are merely inferred. All observations involve inferences, but some are much less likely to be erroneous than others.

Thirdly, and equally important, we must consider features of the researcher and of the circumstances in which the research was carried out, in so far as these might have affected the validity of the researcher's observational reports. We need to think about the sorts of constraints under which observation occurred, and the resulting danger of misperceiving what took place (especially when simultaneously trying to record it).

There are two main strategies that researchers may use to record their observations of events, often referred to as 'structured' (or 'systematic') and 'unstructured' observation. The former involves the recording of events of predefined types occurring at particular points in time, or within particular intervals. A very simple example is the categorization of a pupil's behaviour as 'working' or 'not working' on a task, as judged, say, every twenty-five seconds. The aim of this would be to produce a summary of the proportion of time spent by the pupil on a task. As this example illustrates, 'structured' observation typically produces quantitative data (information about the frequency of different sorts of event or of the proportion of time spent on different types of activity). 'Unstructured' observation, by contrast, does not produce data that are immediately amenable to quantitative analysis. This form of observation involves the researcher in writing field-notes, using whatever natural language terms seem appropriate to capture what is observed, the aim being to provide a description that is relatively concrete in character, minimizing the role of potentially controversial interpretations.

These two forms of observation typically involve different threats to validity. Among the dangers with structured observation is that the predefined categories used will turn out to be not clearly enough distinguished, so that there is uncertainty in particular instances about which category is appropriate. There may also be relevant events that do not seem to fit into any of the categories. Unstructured observation generally avoids these problems because the language available for use in the description is open-ended. This, however, is only gained

at the cost of the information being collected on different cases or at different times often not being comparable.

Increasingly, observational research uses audio- or video-recording, which usually provides a more accurate and detailed record than either 'structured' or 'unstructured' observation. These techniques, however, still do not record everything: for instance, audio-recordings omit non-verbal behaviour that can be very significant in understanding what is happening. Furthermore, audio- and video-records need to be transcribed and errors can be introduced here. Even transcription involves inference.

Besides features of the research process, we must also take account of what we know about the researcher, and the resulting potential for bias of various kinds. For example, we must bear in mind that Rist may have been unconsciously looking out for examples of what he took to be differential treatment of the 'slow learners', interpreting these as negative, and neglecting the respects in which all the children were treated the same. This is not a matter of dismissing what is claimed on the grounds that the researcher is probably biased, but rather of taking the possibility of bias into account in our assessment.

Information from others Information from others may take the form of responses to interview questions or to a postal questionnaire, documents of various kinds and even accounts given by one participant to another that were overheard by the researcher. They may be handled by the researcher in raw form, extracts being quoted in the research report, or they may be processed in various ways: for example, by means of descriptive statistics.

All three sources of error that we identified as operating on observers' reports must also be considered in relation to information from others. Thus, where they are reports from witnesses of events, we must consider the possible effects of the witness's presence and role on what was observed. Second, we must assess the nature of the phenomenon being described and the implications of this for the likelihood of error. The third source of error, the reporting process itself, is more complex in the case of information from others. We must note whether the account is a first-hand report, or a report of what others have told someone about what *they* saw or heard. Evidence of the latter kind is especially problematic, since we will not know what distortions may have occurred in the passage of information from one person to another. It should also be borne in mind that those who supply information are likely to rely on memory rather than field-notes or audio-recording, so that there is more scope for error.

In addition to assessing the threats to validity operating on the information available to the people supplying the information, we must also consider those that relate to the transmission of information from them to the researcher. For example, in the case of data obtained by interview we must assess the effects of the context in which the interview took place: for what audience, in response to what stimulus, with what purposes in mind, under what constraints, etc., did the person provide information to the researcher? Also, what threats to validity may there have been to the researcher's recording and interpretation of the information? Of course, it must be remembered that what are made available in research reports are selections from or summaries of the information provided, not the full body of information.

These, then, are the sorts of consideration we need to bear in mind when assessing the validity of evidence. Equally important, though, is the question of the strength of the inferences from the evidence to the main descriptive claim.

The relationship between evidence and claim

Evidence may seem quite plausible or credible in itself, and yet the support it offers for the claim can be questionable. Evidence sometimes gives only partial support, at best, for the set of claims that the author is presenting. Sometimes, too, one finds that there are other plausible interpretations of the evidence that

would not support the claim. For example, French and French provide information about the number of turns taken by girls and boys in the lesson they studied. When they draw conclusions from this about the differential distribution of the teacher's attention, we might reasonably ask whether the inference from evidence to claim is sound. Does the number of turns at talking provide a good measure of the distribution of the teacher's attention? If it does, does the fact that the researchers were only concerned with discussions involving the whole class create any problems? Could it be that if they had taken account of informal contacts between the teacher and pupils their conclusions would have been different? The answer is that it is difficult to know on the evidence provided, and we probably should suspend judgement about their conclusions as a result (Hammersley, 1990b).

In this section we have looked at the assessment of descriptions, suggesting that this requires examination of the plausibility and credibility of the claims and of any evidence provided in their support. In the case of evidence, we must look at both the likely validity of the evidential claims themselves and of the inferences made on the basis of them. We have spent quite a lot of time looking at the assessment of descriptions because these are the foundation of almost all research. We shall deal with the other sorts of claim more briefly.

EXPLANATIONS AND PREDICTIONS

As we noted earlier, all types of claim (except definitions) include a descriptive component. Given this, the first step in assessing the validity of explanatory, and predictive, claims is to identify and assess their component descriptions, explicit or implicit. This is done in precisely the same way as one assesses any other description. Over and above this, though, we must look at how well the evidence supports the specifically explanatory or predictive element of the claim. There are two steps in this process. First, all explanations and predictions involve theoretical assumptions, and it is necessary to assess the validity of these. Second, it must be shown that the explanation or prediction fits the case at least as well as any competing alternative.

As an illustration, let us return again to Rist's study of early schooling. He argues that the differential achievement of the children after three years of schooling is significantly affected by the differential expectations of the teachers. The first question we must ask, then, is whether the theoretical idea he is relying on is plausible. The idea that teachers' expectations can affect children's learning has stimulated a great deal of research and certainly seems plausible at face value, though the evidence is mixed (see Rogers, 1982). We have no reason to rule it out of account.

The next question is whether Rist successfully shows that this factor is the most plausible cause in the cases he studied. The answer to this, in our view, is that he does not. This is because he does not deal effectively with the possibility that the differences were produced by other factors: for example, by differences in ability among the children before they entered school.

A similar sort of approach is required in assessing predictions. It is necessary, of course, to begin by assessing the description of the situation or causal factor from which the predicted event is believed likely to stem. If the period over which the prediction should have been fulfilled has expired, we must also examine any account of what actually occurred and its relationship to what was predicted. The next concern, as with explanations, is the validity of the theoretical assumptions on which the prediction is based. Can we accept them as plausible? Finally, we must consider the possibility that the predicted event would have occurred even if the factor that is claimed to have produced it had not been present, or that some other factor explains both of them. Here, again, we must rely on thought experiments to assess the likelihood of different outcomes under varying conditions; though we may subsequently be able to test out different interpretations in our own research.

The assessment of explanations and predictions, then, involves all the considerations that we outlined in discussing descriptions, plus distinctive issues concerning the specifically explanatory or predictive element. The latter, like the former, requires us to make judgements about what is plausible and credible, judgements that can be reasonable or unreasonable but whose validity or invalidity we can never know with absolute certainty.

EVALUATIONS AND PRESCRIPTIONS

In assessing evidence for value claims, once again we begin with their descriptive components. Is the phenomenon evaluated, or the situation to be rectified by the prescribed policy, accurately represented? In addition to this, in the case of prescriptions a predictive assumption is involved: that if such and such a course of action was to be taken a particular type of situation would result. The validity of these various assumptions must be assessed.

The distinctively evaluative or prescriptive element of claims concerns whether the phenomenon described, or the situation the policy prescribes, is good or bad. We must decide whether evidence is required to support this component (that is whether it is insufficiently plausible) and, if so, what evidence is necessary. What we are looking for here are arguments that appeal in a convincing way to generally accepted value and factual assumptions. On this basis, we must consider both what values have and have not been, should and should not have been, taken into account in the value judgement. We must also consider how those values have been interpreted in their applications to the particular phenomena concerned.

ASSESSING CONCLUSIONS

When it comes to assessing the conclusions of a study we have to look at the relationship between the information provided about the cases studied and what is claimed about the focus of the research on the basis of this evidence. Earlier, we identified two strategies by which researchers seek to draw such inferences: theoretical inference and empirical generalization. We need to look at how to assess examples of each of these types of inference.

Theoretical inference

Theoretical claims are distinctive in that they are universal in scope. They refer to a range of possible cases, those where the conditions mentioned in the theory are met, rather than specifically to a finite set of actual cases. What we must assess in the case of theoretical conclusions is the extent to which evidence about the case or cases studied can provide a basis for such universal claims.

We should begin by recognizing that there is one sense in which no basis for universal claims can be made: evidence about a finite number of particular cases can never allow us to draw conclusions on a strictly logical basis (i.e. with complete certainty) about a universal claim. This is known as the problem of induction. Various attempts have been made to find some logical basis for induction, but it is widely agreed that none of these has been successful (Popper, 1959; Newton-Smith, 1981). However, once we abandon the idea that a claim to validity must be certain beyond all possible doubt before we can call it knowledge, and accept that we can distinguish between claims that are more or less likely to be true, the problem of induction becomes less severe; though it is still not easy to deal with.

What we need to ask is: does the evidence from the cases studied provide strong enough support for the theory proposed to be accepted? Some cases will by their nature provide stronger evidence than others. Take the example of testing the theory that streaming and banding in schools produces a polarization in attitude on the part of students, with those in top streams or bands becoming more

pro-school and those in bottom streams or bands becoming more anti-school. Hargreaves (1967) investigated this in a secondary modern school, where we might expect to find a polarization in attitude towards school resulting from the effects of different types of home background and school experience. Lacey (1970) tested the theory in a grammar school where most, if not all, of the students had been successful and probably pro-school in attitude before they entered the school. The fact that Lacey discovered polarization at Hightown Grammar constitutes much stronger evidence than any that Hargreaves' study could provide, in this respect, because in the case he investigated a key alternative factor (differences in attitude among pupils before they entered secondary school) had been controlled.

Empirical generalization

The other way in which researchers may seek to draw conclusions about their research focus from their findings about the cases studied is through empirical generalization. Here the aim is to generalize from the cases studied to some larger whole or aggregate of cases of which they form a part. A first requirement in assessing such generalizations is to know the identity of the larger aggregate. The second step is to make a judgement about whether that generalization seems likely to be sound. We can be reasonably confident about such judgements where statistical sampling techniques have been used effectively. Indeed, in such cases we can make precise predictions about the likelihood of the generalization being false. It is important, however, not to see statistical sampling techniques as the only basis for empirical generalizations. It may be possible to draw on relevant information in published statistics about the aggregate to which generalization is being made, and to compare this with what we know of the cases investigated. This at least may tell us if those cases are atypical in some key respect.

4.5 CONCLUSION

In this section we looked initially at the relationship between qualitative and quantitative approaches to educational research, and then at how we should set about assessing educational research reports. As is probably very clear from what we have written, our view is that 'quantitative' and 'qualitative' are simply labels that are useful in making sense of the variety of strategies used in educational research. They do not mark a fundamental divide in approaches to educational research.

To adopt this position is not to deny that there is considerable diversity in the philosophical and political assumptions that motivate or are implicit in educational research, as well as in the techniques of research design, data collection, and analysis that are used by educational researchers. Quite the reverse, we have tried to stress this diversity. What we deny is that it can be reduced to just two contrasting approaches.

We also question how much a commitment to particular philosophical and political assumptions determines what educational researchers do or should do, or how research reports should be evaluated. There is undoubtedly some influence in this direction, but it is less than might be imagined. In particular, it is very important to remember that research is a practical activity and, like other practical activities, it is heavily influenced by the particular purposes being pursued, by the context in which it has to be carried out (including the resources available), and by the particular audiences to be addressed. In other words, the approach one adopts depends at least as much on the sort of research in which one is engaged as on one's political or philosophical assumptions. The latter will shape how one goes about research, but will not determine it. Hence, one finds people doing research in similar ways whose philosophical and political assumptions are quite different. Equally, those who share the same assumptions often engage in very different kinds of research. In short, just as teaching is a practical activity that involves judgement and the taking into account of local circumstances, so too

research is not simply the application of methodological principles. Those principles are important, but they are not all-important. This is essential to remember whether one is assessing research or engaged in it oneself.

Up to now we have only provided rather superficial information about the methods used by educational researchers to obtain and analyse data. In the next two sections of this Part we shall look in more detail at some of the principles and techniques involved in both qualitative and quantitative research.