SECTION 9
DOING A QUANTITATIVE STUDY

9.1 INTRODUCTION

An essential feature of quantitative research is that it produces numerical data, that is data which are amenable to statistical analysis. In Section 7.11 we outlined the requirements of such data in terms of the concept of levels of measurement. At the very least, quantitative data must consist of frequencies in terms of well-defined and mutually exclusive categories. Higher levels of measurement require those categories to be structured as a scale, either ordinal or cardinal, and this allows more powerful statistical techniques to be used.

Section 7 provided you with an introduction to the logic of quantitative analysis. Necessarily, this was rather selective in terms of the concepts and techniques it introduced, and the sorts of examples discussed. However, we are not able to provide further detailed guidance on analytic techniques here, not least because of the sheer range of these which could be relevant. As a result you may have to do some additional reading about statistical analysis in order to prepare yourself for your research. (Section 10.5 lists some of the sources that you might find useful.)

We want to emphasize, though, that writing a proposal for quantitative research and carrying out some pilot work need not be very demanding in terms of knowledge of statistical methodology. What is required of you in your research proposal is a clear indication of the kind of data which would be collected in the proposed research, some information about how these data would be analysed, and a clear outline of the sorts of conclusions that could be drawn on this basis. You need to specify particular analytic techniques if you can, but most important of all is to explain what the techniques would be used to do. You will not be assessed primarily on the basis of the techniques you propose but rather in terms of your understanding of the logic of statistical analysis and of what conclusions are and are not reasonable on the basis of the kind of data and analysis indicated.

As regards your pilot research, the sort of analysis required depends a great deal on what you plan to do. It may be that even though your research proposal involves a mainiy quantitative approach, your pilot work requires qualitative analysis. And even if you are collecting quantitative data in your pilot research, it is unlikely that these will be susceptible to sophisticated statistical analysis. What will probably be required, above all else, is the presentation of the data in appropriate tabular and/or graphical forms, and the use of simple descriptive statistical techniques, such as percentages, averages, measures of spread, etc.

The amount of data you collect in your pilot work is likely to be quite small, and, given this, the calculations can probably be done with the use of a calculator (one which has statistical facilities would be particularly useful). Alternatively, there is now a wide range of software available for microcomputers which would enable you to carry out statistical analyses, and if you are likely to be dealing with a relatively large amount of data (for example, data relating to more than two

variables for more than 20 cases), it may be worth exploring the possibility of using such a package. There are books that offer guidance in the use of the most popular programs, such as SPSS-x and Minitab, over and above that available in the manuals. You are at liberty to use this software if it is helpful. However, if you are not already a fairiy accomplished user of microcomputers, the time required to familiarize yourself with the program may be prohibitive and it may be more efficient to do the calculations yourself.

Of course, before one can carry out any statistical analysis it is necessary to obtain the data, and this is our main focus in this section. The data required for quantitative analysis can come from a variety of sources. It may be that quantitative data of the kind required for investigating your topic are already available from secondary sources, for example from official statistics such as the Statistics of Education published by the Department for Education and Employment, or from data generated internally by educational organizations, such as SATs scores, examination results, attendance figures, etc. We discuss the analysis of secondary data in Section 9.2 below. More likely, however, the data will have to be produced through research activity. In Sections 9.3, 9.4. and 9.5 we look at three of the main ways in which this is done in quantitative research: survey research, experimentation, and systematic observation.

You will need to be selective in your reading of what follows, concentrating on those parts that are most relevant to your current plans. Of course, in your research proposal you may be intending to combine data from different sources. This is quite acceptable; but do remember the limitations on resources under which the research is to be carried out.

9.2 SECONDARY DATA ANALYSIS

Use of secondary data can have the advantage that the time, effort and costs involved in collecting data are greatly reduced. And, today, there are increasing amounts of quantitative secondary data available. However, the convenience and value of such data can be deceptive, for two closely-related reasons. First, and most obvious, the data necessary to investigate your topic may not be available in secondary sources. Second, even where relevant data are available they may not be presented in the form required for the analysis you want to carry out. Secondary data are usually produced for other purposes than research, and this may well mean that the categories used are not the most appropriate ones for your purposes, that crucial kinds of data are missing, etc. It is also possible that there may be considerable error in the processing and reporting of the data.

Even more than with other kinds of data, then, it is necessary to think carefully about what secondary data can and cannot tell us, and about any likely sources of error built into them. It may be that little can be done about these problems other than being especially cautious in interpreting the results. However, it is sometimes possible to re-process the data in a more suitable way if the raw data are available. For example, the examinations data published by schools may not allow the kind of analysis necessary, but the original examination scores may be on file and it may be possible to re-analyse these so as to meet your needs. However, this can be extremely time consuming: indeed, such work can take more time than the collection of primary data.

In this sub-section we can oniy provide you with a very brief guide to the use of secondary data. You will need to supplement this with some use of the literature referred to in Section 10.5.

Our aims here are twofold:

1 to identify aspects of the use of secondary data which need to be addressed at the stage of initial planning;

2 to outline the sort of pilot work that might be useful in preparing a research proposal involving this source of data.

Developing your research proposal

In framing a research proposal based on secondary data analysis, you need to address the following questions:

1 What variables are the focus of your analysis?

2 What is the quality and accessibility of the data?

You will, of course, also need to outline the sort of analysis proposed and to provide a timetable for the research.

Variables

The first task is to spell out the variables on which your work will focus. The intended product may be a description of some population in terms of these variables or a comparison of two or more populations in these terms; or the proposed research may be concerned with testing an explanatory or theoretical hypothesis. Where the latter is the task, a causal diagram of the kind introduced in Section 7.2 might be of value.

Of course, the variables you will be able to focus on will be determined to a large extent by the data that are available. More than with most research, there may be a tension here between what is of interest and what is feasible. In addition, attention needs to be given to the reliability of the indicators represented in the data. In using examination results, for example, it is important to remember that they measure attainment of particular kinds under particular conditions: they do not cover all aspects of education, nor do they measure underlying ability directly; and they may not be entirely accurate measures even of attainment in the areas of learning they cover, given the conditions under which candidates sit them. There are even more obvious problems with some other kinds of data, for example with using school attendance figures as a basis for measuring levels of truancy. A key aspect of your research proposal, then, will need to be a discussion of the relationship between the variables that are the focus of your interest and the indicators which the data you plan to use can provide.

The data

You will also need to give a clear outline of the data on which the research will rely, and how it will be accessed and used. Faced with any body of secondary statistical data, we must ask a series of basic questions in order to decide what inferences can reasonably be made on the basis of it. First of all, and most obvious, we need to make sure we know quite clearly to what the data relate. What are the units which have been counted, and how have they been defined? For example, it can make a considerable difference whether data about schools in England and Wales relate only to state-maintained schools or also include private schools. Similarly, does the definition of primary school include 8-12 middle schools, and what about 9-13 middle schools? Some investigation may be necessary to ascertain exactly what is and is not included in the figures and under what headings. It is also important to know whether the data are comprehensive or relate onfy to a sample, and if so how big this sample was and how it was selected. You will also need to think about who collected the data and for what purposes, since this will have affected their characteristics and likely validity. Also crucial is the time period to which the data refer; and when comparing data relating to different time periods one needs to take account of differences in the quality of the data collected at different times and of any relevant changes in the classification systems used. The necessary background information about these matters is not always immediately available, even in officially published statistics. Some inquiry may therefore be necessary.

Also necessary in the research proposal is discussion of any preparatory work that is likely to be needed in order to make the data amenable to the kind of analysis proposed. And you will have to consider any problems which may be involved in the analysis of the data, given its likely coverage and quality. Indeed, it might be

necessary to build in checks on data quality, for example through the use of other data sources designed to assess likely threats to validity.

Where data are published, access to them may be relatively unproblematic, though an acknowledgement of the source should of course be made. Where data are not published some negotiation may be necessary to get access to them, and some attention will have to be given to the ethical considerations involved in their use. Will restrictions be placed on the use of the data? And how might these affect the research? Even if no restrictions are imposed, making public data which were not previously available can have effects on the reputations of organizations and of individual people, and this will need to be taken into account.[18]

Summary

Your research proposal, then, should contain the following:

1 a statement of your research problem and its rationale;

2 a discussion of the variables which are your focus and their operationalization;

3 an account of the data set that would be used in the research, of any methodological, access and ethical problems associated with these data, and of any preparatory work that would be required;

4 a description of the kind of analysis proposed;

5 a timetable indicating the time allotted to each stage of the research.

Pilot research

Pilot research can take at least two main forms in relation to secondary data analysis.

1 A small sample of the data may be processed and analysed in order to make an assessment of the time it will take to analyse the whole data set, to test out the procedures for doing this, and to see whether the analysis can produce the kind of results required for investigating the research topic.

2 Data from a different source may be used to check the validity of the secondary data. This may provide a direct check: for example, an observational survey of attendance in lessons could be carried out on a sample of days and the results compared with the attendance figures for those days. Alternatively, the pilot research might consist of an investigation of the process by which the secondary data were produced, with a view to identifying potential sources of error. This could be carried out, for example, through interviews with those involved in compiling the figures.

Either of these options is acceptable, and others are possible.

In your plan for the pilot research (required for STMA 04) you need to indicate:

1 its central purpose;

2 the type of data you plan to use;

3 the problems you anticipate and how you will deal with them;

4 the sort of analysis you will employ.

For a useful general discussion of some of the ethical issues involved in the handling of statistical data see Barnes, 1979; and, specifically in relation to educational research, the articles by Raffe, Bundle and Bibby and by Sammons in Burgess, 1989.

The report of your pilot research (required for STMA 05) will need to make clear:

1 how it worked out in practice;

2 what you have learned from it.

It should include a systematic analysis of the data and presentation of the results.

9.3 SURVEY RESEARCH

One of the most important sources of primary data in educational research is the social survey. In Part 1, Section 4.1, we defined the survey as a case-selection strategy involving the simultaneous selection of a large number of cases for study However, the term is used in a broader sense in most of the methodological literature. There it refers not just to the use of that case-selection strategy, but also to the collection of data by means of self-completion questionnaires or structured interviews. Almost certainly, you will be familiar with these data-collection strategies: you have probably filled in many questionnaires yourself, and may have been interviewed for market research or for some other purpose. Here, we will be using the term 'survey' in this more conventional sense to cover both the survey case-selection strategy and the use of these data-collection techniques.

A central purpose of survey research is to provide comparative data about a relatively large set of cases. This can be done by investigating all of the cases, or it may involve studying a sample of them. The aim is also usually to provide structured data which can be subjected to quantitative analysis. For this reason, the questions included in questionnaires and interview schedules are, for the most part, closed questions: a set of alternative answers is offered to respondents from which choices can be made, though it is good practice to provide space or time for respondents to give their own answers if the answers offered do not exhaust all of the possibilities. Sometimes questionnaires and schedules also involve so-called 'free response' questions to which answers are not prespecified. Where such questions are included, or where respondents have made use of the facility of providing their own answers rather than choosing prespecified ones, answers will need to be coded after the survey has been carried out, so as to produce the kind of structured data required for quantitative analysis.

Below we provide you with a guide to the use of surveys, for the purposes of planning your pilot research and developing a research proposal. You will need to supplement this with some use of the literature on survey research. (See Section 10.5. You might find it useful to start with those parts of one or more of the general introductory books dealing with survey method, and use the more specialized texts if, as, and when necessary.)

It is important to remember that what you are being asked to do here is not a full-scale piece of research, and we are not able to provide you with the necessary support for all aspects of survey research.

Our aims are twofold:

1 to identify the important aspects of a survey which need to be addressed at the stage of initial planning;

2 to outline the sort of pilot work which might be useful in preparing a research proposal involving the survey method.

Developing your research proposal

There are four key issues to be addressed in planning a survey.

1 What are the central variables in which you are interested?

2 Are you going to use questionnaires or interviews?

3 How can questions be formulated to obtain the data needed?

4 What is the population with which the research is concerned? And how is information about it to be obtained?

You will also need to discuss the form of analysis planned, and to provide a timetable for the research.

Central variables

All data collection is selective: some kinds of information will be relevant to the research focus, but much will not be. In survey research the focus has to be decided before the main data-collection phase begins. This is because the aim is usually to collect the same data from every respondent, and (as far as is possible) to standardize the circumstances in which this is done. So, a research proposal needs to identify the variables about which the survey is to provide information. The intended product may be a description of some population in terms of these variables, or a comparison of two or more populations in these terms; or the proposed research may be concerned with testing an explanatory or theoretical hypothesis. Where the latter is the task, a causal diagram of the kind introduced in Section 7.2 might be of value.

Questionnaires versus interviews

Questionnaires and survey interviews share much in common in terms of the sorts of questions that are asked. In both cases, given that the aim is usually to collect the same information from all respondents, and in such a way as to avoid the behaviour of the researcher introducing bias, the ordering and wording of the questions are usually standard across respondents. Thus, interviewers are generally required to follow the outline of a schedule whose character is very similar to that of a questionnaire. (Indeed, the term 'questionnaire' is often used to refer both to the questionnaires which respondents fill in themselves and to the interview schedules used to conduct survey interviews.)

However, there are some important methodological implications of the choice between postal questionnaires and survey interviews. It is, of course, generally much cheaper to cover the same size of sample by means of a questionnaire than by interviews. At the same time, questionnaires make literacy demands which some respondents may find it difficult to meet. As a result, they may fail to respond to the questionnaire, or their answers may be difficult to understand or misleading. There are other reasons, too, why the response rates to postal questionnaires could be lower than those for interview surveys. Faced with an interviewer who has made an appointment to carry out the interview, a respondent will usually put other activities aside and answer the questions (though perhaps not all of them!), whereas postal questionnaires are often forgotten or mislaid even by those who intended to complete them.

The *quality* of the data may also be superior in the case of interviews, since clarification by the interviewer can reduce misunderstandings. But it is important to remember that there is also the possibility of interviewer bias: the social and personal characteristics of the interviewer can affect the answers which respondents give (Sudman and Bradburn, 1974). Age, sex, social class and 'race'/ethnicity are important factors here. Indeed, some people may be prepared to answer a questionnaire who would not agree to an interview; in fact, an interviewer can sometimes constitute a major barrier to honest response. As we pointed out in

Section 4.4, doing research in an institution in which one is known or plays a key role may affect the character of the data collected. Thus, a headteacher asking members of her staff about school morale might receive very different responses using these two strategies, at least where the questionnaire does not require the identity of the respondent to be recorded.

When questionnaires are administered to a group of people face-to-face (as is often done with school classes), the response rate is usually much higher than in the case of postal questionnaires. And in this situation, like that of an interview, the researcher may be able to clarify anything which respondents do not understand, and perhaps even to help those who have difficulty in completing the questionnaire, for whatever reason; though this may introduce bias, to the extent that respondents are given differential help.

Another approach which is sometimes used, especially where a sample is widely distributed geographically, is the telephone survey. Here a structured interview is carried out over the telephone. This preserves some of the advantages of the face-to-face interview, yet also facilitates coverage of a larger sample than would normally be possible by face-to-face contact. However, it also involves some disadvantages. First, not everyone is comiortable talking on the telephone, and this may affect the quality of the data produced. Indeed, some people who would have filled in a questionnaire may refuse a telephone interview. Furthermore, unlike face-to-face interviews, telephone interviews cannot yield direct observational data about the circumstances in which people live or work, and these may be of value; nor do they give access to nonverbal behaviour, which can often be useful as a basis for judging the answers which respondents give.

These are some of the considerations which need to be borne in mind in choosing between the use of questionnaires and interviews, and different means of administering them. What they reveal is that the choices made have consequences both for the quantity of responses which will be obtained and perhaps also for their quality. However, to talk of 'quality' of data is slightly misleading. What we perhaps should say is that the responses obtained by different strategies may be divergent. Whether this makes the data better or worse depends to some degree on the purposes of the research.

Constructing questionnaires and schedules

Whatever the variables which are the focus of the proposed survey, these will need to be operationalized. In other words you will need to decide how they can be measured effectively. What sorts of questions will need to be asked in order to produce data that will enable the researcher to come to conclusions of the kind specified in your research problem? And, looking at it from the other end, in constructing a questionnaire or schedule you will need to think carefully about how you are going to process and analyse the data produced so as to illuminate the research topic One of the functions of pilot work in survey research is to test whether the data-production process is well designed to serve adequately the proposed data analysis; and whether that analysis serves the overall purposes of the research.

The information which questionnaires and interviews are used to collect usually falls into two categories. First, there is often background information which is needed to provide the basis for comparison of respondents: age, sex, number of years spent working in a particular educational institution, whether employment is part-time or full-time, etc Then there is the type of information that is directly related to the research focus. This may be of a kind to which respondents have access and which they can supply directly. Alternatively, the purpose could be to ask questions which will elicit the attitudes of respondents towards various issues. Here, the aim may be to collect the respondents' relatively conscious, carefully-considered views about some issue; or interest may lie in their gut reactions or unconscious attitudes, for example because it is believed that the latter are likely to be more closely related to their actual behaviour.

Which sort of information the research is concerned with will make a difference to the sorts of questions that must be asked. If the focus is information known to the informant or their consciously formulated views, the questions will need to make as clear as possible exactly what information is wanted. If the aim is to collect data about unconscious attitudes, the usual strategy employed by survey researchers is to ask respondents to answer a battery of questions which are designed to tap attitudes without respondents being aware of their purpose. (Note that ethical considerations are relevant here.)

Whatever types of questions are involved, though, it is necessary to ensure that they are understandable to the various kinds of respondent belonging to your population, and are as unambiguous as possible. Much of the discussion in the literature covers common faults in questionnaire design, and great care needs to be taken here (see Oppenheim, 1986).

You will also need to consider whether the questions are to be of the closed or the free-response kind. This has' implications for the nature of the information likely to be produced, and also for the practicalities of running the survey: completion of a questionnaire or interview takes longer the more free-response items there are. And, even more important, free-response questions increase considerably the time involved in processing the data, since they have to be coded. On the other hand, such questions will sometimes produce more valid data.

Where closed questions are involved careful thought needs to be given to the range of possible answers identified. Sometimes these will represent a scale: for example, asking respondents whether they agree strongly, agree, have no view, disagree, or disagree strongly. There are technical considerations involved in the construction of answer frames, and you may need to draw on the literature to address these.

There is also the problem of routeing respondents or interviewers through the questionnaire or schedule. Very often you will want to ask further questions of people who respond in a particular way to one of your questions. This means that respondents will have to follow different routes through the questionnaire or schedule depending on how they answer particular questions. This routeing needs to be structured in a way which meets your needs for information, but also allows the respondent or interviewer to understand exactly what is required.

In the light of all this, you will need to think carefully about the length of the questionnaire or interview schedule, and the time it will take to complete. There is a temptation to include as many questions as possible, sometimes ones which are not directly relevant to the research focus but which might produce interesting information. In general, this temptation should be resisted.

Sampling

A requirement in any survey is clarity about the population which is of interest. This is not always as straightforward as it might seem. Perhaps one's primary interest is in the staff of a particular school or college, but at the same time one hopes that the findings will be applicable to other schools or colleges of the same type within a particular geographical region, or even more generally. In such a situation one must make a decision and state clearly which population one is focusing on: very different strategies will be necessary according to which of these populations is selected. (Of course, this does not prevent one from suggesting that the results may be of wider relevance.)

Sometimes a whole population can be surveyed. However, usually the aim will be to study a representative sample. There are several different types of sampling strategy, but a broad distinction can be drawn between probability and non-probability sampling. Which of these is chosen will depend partly on technical and partly on practical considerations. Generally, the aim is to produce a sample which represents the target population in relevant respects. Ideally, this should be done in

such a way that statistical sampling theory can be used to assess the likelihood that the sample selected is representative. This requires probability sampling. Here, all members of the population have a known and non-zero chance of being selected. The most straightforward kind of probability sampling is the simple random sample, where respondents are chosen completely at random, using a random number table or random numbers generated by a computer. However, survey researchers do not always adopt this strategy; indeed, national surveys rarely do so because it would be extremely expensive. More common is some form of stratified or cluster sampling. This involves random sampling from within relatively homogeneous strata (for example social classes) or from natural clusters (schools and colleges constitute natural clusters of both teachers and students).

However, not all surveys can meet the requirements of probability sampling. One of these is that there is a sampling frame: a list of all respondents in the population, or within the relevant clusters. Where probability sampling is not feasible, some other sampling strategy will have to be employed. There are various possibilities here. One is what is sometimes referred to as strategic sampling. Here, people may be selected who are known to vary sharply in relevant ways. Sometimes, however, it is possible to do no more than collect data in the form of what is called an 'opportunity' or 'convenience' sample: those people are surveyed whom it is possible to contact within the constraints of very limited time and other resources. It may be possible for this to be done in such a way as to represent particular categories of respondent in proportions which reflect their occurrence within the larger population, thus stratifying the sample in terms of some key dimensions of heterogeneity such as age or sex. This is referred to as quota sampling. With non-probability sampling of these kinds we *cannot* use statistical sampling theory to tell us the likelihood that our sample is representative. However, it may be possible to assess the likely representativeness of the sample in relevant respects by comparing its characteristics with what is known about the target population.

Clearly then, some discussion of the advantages and disadvantages of the sampling strategy proposed and its implications for the data to be produced will be required in your research proposal.

Another problem which will need to be considered is how to deal with non-response. The sample from which the data to be analysed actually come may be very different from that intended, because a substantial proportion of respondents do not supply data, for one reason or another. This is a problem because non-responders may differ in relevant characteristics from those who do respond, so that the actual sample may be unrepresentative of the intended population even though the sampling strategy promised to produce a representative sample.

Thus, there needs to be some indication given in your proposal of how many times interviewers should seek to contact respondents when they are not initially available, and how they should respond to reluctance or refusal. In the case of a postal questionnaire, there is the issue of how many reminders should be sent out. These issues have practical consequences in terms of cost and time, as well as technical implications for the quality of the sample. Sometimes, a brief non-response questionnaire is sent to non-responders to assess the extent to which they differ in their characteristics from responders; though, of course, some proportion will not respond even to this!

Another key issue is the size of the sample you need. This is not a straightforward matter. Statistical theory allows us to calculate the size of probability sample required in order to achieve results of a given likely representativeness. However, there is no means of calculating this for non-probabilistic samples. Almost whatever the size of the population, the sample should be at least 30 cases; and the larger and more heterogeneous the population is, the bigger the sample will need to be in order to facilitate its representativeness. Another crucial consideration is to try to ensure that the numbers in the sub-categories you will be comparing are sufficiently large to allow reasonable conclusions to be drawn. Reading other studies dealing with the relevant population and conducting a pilot study may be helpful in making decisions about the size of sample required.

Thoughout this discussion of sampling we have assumed that your unit of analysis is the individual respondent. This may not be the case, however. It could be schools or colleges, local authorities, etc. In this case you need to consider both the issue of what is an appropriate sample and of who would be appropriate respondents. All the considerations we have already discussed still apply However, in addition there is the task of deciding who would have the information you require and would be prepared to supply it, and how you can contact them effectively.[19]

Summary

In summary, then, at the design stage you need to be clear about the overall aim of your research and to have an idea about the *types* of statement that you will be hoping to make in your conclusions. Also, you need to think about what is likely to be possible, given the limitations on resources and other practical constraints within which the research will have to be done. Having some conception of the whole project will help you to make decisions about the amount and level of data which are desirable and what sort of analysis will be appropriate.

Your research proposal should contain:

1 a statement of your research problem and its rationale;

2 a discussion of the survey which is planned, in terms of:

(a) the specification and operationalization of the variables;

(b) the choice between questionnaires and interviews;

(c) the construction of the questionnaire or schedule;

(d) the sampling strategy to be employed and the kind of analysis proposed.

Under each heading there should be consideration of problems likely to arise and how they would be dealt with.

3 an indication of the analytic techniques which would be employed;

4 a timetable indicating the time allotted to each stage of the research.

Pilot research

It is common in survey research to use pilot studies to develop questionnaire or schedule items, and/or to check the viability of these. This is not a phase of research to be hurried through. The aim is to anticipate, as far as possible, what could go wrong in the main phase of data collection.. This is essential because problems which arise then can be very costly to put right and may even undermine the whole value of the survey. Oppenheim reports that 'it is not unusual for the pilot work to take many months, to use up hundreds of subjects, and for any given instrument to require half-a-dozen or more revisions' (Oppenheim, 1979, p. 53). We do not expect you to carry out pilot research on this scale, but the function of your work will be analogous to that which Oppenheimer describes.

There are at least two main possibilities for your pilot work.

1 You may wish to carry out open-ended interviews, analyse the data, and develop a section of the proposed questionnaire or schedule on the basis of this. Here, your primary concern would be to develop appropriate questions and pre-specified answers.

[19] There are ethical issues involved in all aspects of survey research, and these will need to be borne in mind in constructing your research proposal, and in your pilot research. For useful discussions, see Barnes, 1979, and (specifically in relation to education) the articles by Raffe, Bundle and Bibby and by Sammons in Burgess, 1989.

2 You may want to construct the proposed questionnaire or schedule on the basis of your existing knowledge or to model it on one used in another study. It could then be administered to a small sample of respondents, and some simple statistical analysis of the results could be carried out. Here, the main aim would be to check the draft questionnaire or schedule for intelligibility, ambiguity and usefulness across the range of types of people who make up the population. You might also want to employ a number of open-ended questions in your pilot questionnaire and develop a coding frame for analysing them.

Either of these options is acceptable, and there are other possibilities.

For the purposes of the pilot study, then, you need to collect a small amount of data only. This may amount to around two or three open-ended interviews, in the region of five or six structured interviews, or ten to twelve self-completed questionnaires. It is important that your respondents are drawn from your target population. However, it would make little sense to select these at random: the sample is too small for it to be representative in statistical terms. A better strategy is to select respondents in such a way as to reflect the known heterogeneity of the target population. So, the idea would be to select people who differ from one another in ways that you know are likely to represent significant differences within the population relevant to your research topic.

In the plan for your pilot research (required for STMA 04) you need to indicate:

1 its central purpose;

2 the type of data you plan to collect and by what means;

3 the problems you anticipate and how you will deal with them;

4 the sort of analysis you will employ.

The report of your pilot research (required for STMA 05) will need to make clear:

1 how it worked out in practice;

2 what you have learned from it.

Some of what you have learned will have arisen from the practical experience itself: discovering that a question you thought would be easily understandable or was unambiguous is not; or that the range of possible answers to a question you had identified omitted something important. However, in addition, while the data produced by your pilot research are unlikely to be amenable to sophisticated statistical analysis, they will require some analysis, qualitative or quantitative, if you are to gain the most from your research. This analysis and its results should be reported.

9.4 EXPERIMENTS

In Section 4.1 we defined the experiment as a case-selection strategy in which cases are created in order to facilitate the exploration of a research problem or the testing of a hypothesis. In fact, most experimental research is concerned with hypothesis testing, that is with testing a predicted relationship between an independent (or causal) and a dependent (or effect) variable. (These terms were introduced in Section 7.2.) This is done by varying the independent (or treatment) variable, and controlling relevant extraneous variables (in other words, other variables which might also be expected to have an effect on the dependent variable).

Experimental designs can be quite complex, but at the core of all of them is a contrast between a set of subjects who have been treated to a high level of the independent variable and another set who have been subjected to a low or zero level of it. This is often formulated as a contrast between the treatment group and the control group. What is significant about this contrast is that it provides a basis for assessing the effects of the independent variable; assuming that the effects of other factors, such as initial differences between the members of the two groups, can be eliminated or allowed for.

In some experiments the same subjects constitute both the treatment and control groups, on different occasions. This is often referred to as a 'within subjects' design. Here, initial differences between the groups have effectively been controlled, but allowance has to be made for the fact that the order in which they experience the experimental and control situations may affect their responses. Those who have already experienced the experimental situation may behave differently in the control situation to those who have not. This is usually dealt with by allocating equal subsets of the subjects to the two situations in different orders, thereby minimising the danger of spurious conclusions arising from what are referred to as 'order effects'.

When different subjects are used in the treatment and control groups (referred to as a 'between subjects' design), elimination of the effects of irrelevant differences between members of the two groups is usually attempted in one of two ways: by randomiy assigning people to the experimental and control groups, or by setting out to match the members of each group on what are taken to be key extraneous variables.

Whether the experiment involves the same subjects or different subjects experiencing experimental and control situations, it is important that those two situations are kept as similar as possible in all respects other than the level of the independent variable. Otherwise differences in the results for the two situations may arise from extraneous differences between them, rather than from variation in the independent variable.

This manipulation of experimental and extraneous variables is usually referred to as 'physical control'. It allows more effective testing of causal hypotheses than is possible in other kinds of research. However, in addition, experimental researchers usually employ statistical analysis to assess the extent to which any differences discovered between the treatment and control situations were the product of chance. This is done in much the same way that Ball sought to assess whether the differences in numbers of middle-class and working-class students assigned to top and bottom streams was a product of chance or of systematic bias in the allocation procedure. Significance tests, such as chi squared, are used for this purpose (see Section 7.6).

Experiments vary considerably in their complexity. While the contrast between treatment and control is the basis of much experimental method, many experiments involve a comparison between three or more conditions. This has implications not just for the number of subjects required, and for the practical organization of the experiments, but also for the kinds of statistical analysis which are appropriate.

It is also conventional to distinguish between kinds of experiments involving different *levels* of physical control. The pure type of experiment is carried out in a specially designed laboratory, usually employing subjects who have been recruited (and often paid) for the purpose. However, there are also field and quasi-experiments. In the former, the experiment takes place 'in the field', that is outside the laboratory. Here the researcher has some control over the explanatory variable, but less control over extraneous variables. In quasi-experiments, the researcher has even less control. Here, what is involved is the study of some naturally occurring change and its impact.[20]

As we argued in Section 4.1, it is best to see the distinction between the sort of case creation which is characteristic of experiments and the selection of natural cases by survey researchers and qualitative researchers as actually representing a dimension. We noted that experiments vary considerably in terms of how far they are carried out in a specially prepared environment, and in the degree to which the experimenter controls and manipulates relevant variables, especially extraneous ones. Physical control may be reduced because it is not possible or is ethically unacceptable randomly to assign subjects to (or to match subjects across) treatment

[20] For the classic discussion of different types of experimental design and the threats to validity relating to these, see Campbell and Stanley, 1963.

and control groups. It may also be reduced because the actual situation in which the experiment is carried out may be by no means completely under the control of the researcher, so that differences in the situations experienced by treatment and control groups may occur. Similarly, some independent variables are simply not open to manipulation, for physical or ethical reasons: sex is an obvious example.

The use of field experiments and quasi-experiments sometimes results from physical or ethical restrictions, then. However, non-laboratory experimental designs are also sometimes selected because the situations in which they are carried out are less 'artificial' than the laboratory, so that there is less chance that subjects' behaviour is a product of the experimental situation and more chance that it is generalizable to 'natural' conditions (in other words, it is assumed that, other things being equal, non-laboratory experiments increase ecological validity). As with research design generally, there is a trade-off here between advantages and disadvantages, and it is rarely if ever possible to maximize all advantages and minimize all disadvantages, even in principle. What is important is that researchers are aware of the threats to validity which may be operating on the data they produce and that they assess the effects of these in their analyses.

Below, we provide an outline of the main considerations involved in experimental research, to assist you in planning your pilot research and developing your research proposal. However, there is a limit to the support we can provide here; you will also need to draw on the relevant methodological literature. (See Section 10.5. Many of the general introductions cover experimental research, but you will probably also need to look at more specialist texts.)

Do remember that you are not being asked to do a full-scale piece of experimental research, but rather to develop a research proposal and to carry out some pilot research.

Our aims in the discussion which follows are twofold:

1 to identify the most important aspects of experiments which need to be addressed at the stage of initial planning;

2 to outline the sort of pilot work which might be useful in preparing a research proposal involving experiments.

Developing your research proposal

There are three key issues to be addressed in developing your research proposal:

1 specification of the hypothesis to be tested;

2 operationalization of the independent and dependent variables;

3 formulation of the experimental design.

The form of analysis to be used will also have to be decided, and you will need to provide a timetable for the research.

Specification of the hypothesis

The first requirement in planning an experiment is to make explicit the hypothesis which is to be tested. This usually has to be done before any experiment can be carried out, though the hypothesis may often be developed or even changed dramatically after one or more experiments have taken place. The hypothesis will be causal, predicting a relationship between at least one independent and one dependent variable. You will need to specify what those variables are and what the expected nature of the relationship is between them. For instance, you might be interested in whether segregation of the sexes improves the quality of students' learning. Here, sex segregation is the independent variable, the quality of students' learning is the dependent variable, and the relationship between them is taken to be positive (the greater the degree of segregation, the greater the quality of learning). The nature of the relationship - whether it is positive or negative - may or may not

be specified in the hypothesis. Where it is, the hypothesis is described as uni-directional; where it is not, as bi-directional. (Whether or not direction of effect is specified has implications for the kind of statistical test which is used.)

Operationalization

A second requirement is that the independent and dependent variables are operationalized. What will be experimentally manipulated is not the independent variable itself, but some exemplification of it. Similarly, what will be measured is not the dependent variable directly but some behaviour which is taken to be an indicator of it. In the example already discussed, segregation of the sexes might be operationalized in terms of single-sex learning groups (high level of segregation) versus mixed-sex learning groups (low level of sex segregation). And the dependent variable could be operationalized in terms of a test of some specific form of learning relevant to the material that the learning groups are working on.

It needs to be remembered that the relationship between concepts and indicators is complex. We can often subsume any indicator under a number of conceptual headings; and, conversely, we can operationalize a concept in different ways. As a result, the process of operationalization introduces some scope for ambiguity and error. In the case of our example, it might be that while the experiment measures the effect of single-sex/mixed-sex grouping on a particular form of learning, it does not measure very well the effect of sex segregation on quality of learning in general. The most obvious problems arise with the measurement of the dependent variable. The test may measure only superficial not depth learning, or it may measure learning in relation to one sort of material which is not representative of others. Sometimes multiple measures may be required, or even parallel experiments dealing with different operationalizations of the variables. What is most important, however, is an explicit awareness of the problems which operationalization may involve.

Formulating the experimental design

A third element in the process of planning an experiment is to determine its design. An important consideration here is the identification of the extraneous variables which need to be controlled if the experimental hypothesis is to be tested effectively. In our example these would include any differences in relevant abilities of the male and female students who are participating in the experiment. Another might be differences in teaching style adopted by the teacher in relation to single-sex and mixed-sex groups.

An experimental design will often involve, as a minimum, pre-testing (that is measuring the dependent variable) in relation to both the treatment and control groups at the start of the experiment, and testing each group again afterwards. However, much more complex designs are possible, involving multiple conditions and therefore multiple groups. Thus, even in the example we have been using, there would probably need to be at least three groups: two single sex and one mixed. Furthermore, as we noted earlier, experiments are usually run many times, with the experimental procedure being refined or changed over the course of the work. However, we suggest that you adopt a relatively simple design for your research proposal, and especially for your pilot research.

Another issue which you need to consider is what you are going to tell the respondents about the experiment. There is a dilemma here. On the one hand, it is a normal rule of research that people participate in it on the basis of informed consent. On the other hand, if you tell people a great deal about the purpose of your experiments there is a strong danger that this will affect their behaviour. So, you need to think about what initial information it is ethically and practically appropriate to give, and also what debriefing may be necessary subsequently. You must remember that people are often intensely interested in the results of an experiment in which they have been involved, and they may regard it as some sort

of test of their own abilities, personalities, etc. This is perhaps especially true of children, and if your experiments involve children you need to think about these issues with particular care.[21] Debriefing can also be of value in assessing the success with which extraneous variables have been controlled in the experiment.

Summary

Your research proposal, then, should contain:

1 a clear statement of the hypothesis which is to be tested and its importance;
2 some discussion of the key variables and how they are to be operationalized;
3 a description of the experimental design proposed;
4 an indication of the analytic techniques which would be employed;
5 a timetable showing the time allotted to each stage of the research.

There should also be a consideration of problems which are likely to arise and how they would be dealt with.

Pilot research

The pilot research must be on a much smaller scale than what you plan in your research proposal. Just one or two trials of an experimental procedure would be quite sufficient. However, in doing this you will need to consider most of the issues we have discussed in relation to developing a research proposal.

For the plan of your pilot research (required for STMA 04) you need to indicate:

1 what the independent and dependent variables are, and how you intend to operationalize them;
2 where the experiment will take place, exactly what it will involve, the practicalities of recruiting, briefing and debriefing subjects, running the experiment, etc.;
3 what sort of analysis you intend to employ.

In the report of your pilot research (required for STMA 05) you need to provide:

1 an account of what problems arose in carrying out your plan and how you dealt with them;
2 the results of your analysis of the data, with an indication of what conclusions can and cannot reasonably be drawn from this.

9.5 STRUCTURED OBSERVATION

Observation is a common method of data production used by educational researchers. However, observation, like interviewing, can take different forms. The most important distinction here is between 'unstructured' and 'structured' observation, sometimes referred to as 'ethnographic' and 'systematic' observation, respectively.[22]

[21] For a useful discussion of the ethical issues surrounding experimental research, see Diener and Crandell, 1978.

[22] Neither of these terminological contrasts is entirely appropriate. Ethnographic observation is not unsystematic; and neither is it entirely unstructured. However, we will use these labels here since they have become a matter of convention in the literature.

Like ethnographic observational work, structured observation takes place in 'natural' rather the experimental contexts; but it is distinctive because it involves the coding of behaviour in terms of a relatively small set of categories which has been explicitly defined. Figure 20, for example, shows one of the most frequently used observational schedules in educational research, Flanders' Interaction Analysis Categories.

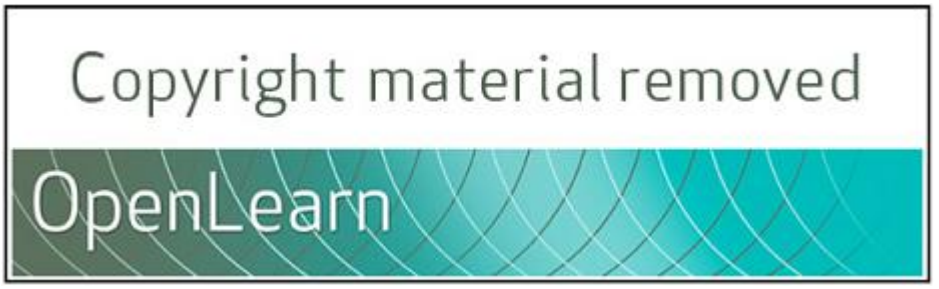

Figure 20 Flanders' Interaction Analysis Categories (FIAC)
(Flanders, 1970, p. 34, Table 2-1)

This schedule focuses primarily on the behaviour of the teacher, and the interest is in the extent to which he or she engages in various types of behaviour which relate to 'directiveness' of teaching style. The task of the observer is to code the teacher's behaviour every three seconds in terms of these mutually exclusive categories. This produces a set of frequencies indicating the extent to which the teacher engaged in these different types of behaviour within the observational period. Different teachers (or the same teacher on different occasions) can be compared in these terms. Modified versions of this category system have been used for somewhat different purposes to those for which Flanders originally developed it. Thus, Green (1983) used it to assess the differential frequency of these types of teacher action in relation to children from different ethnic groups.

Flanders' schedule is oniy one of a very large number of schedules which have been developed, for a variety of purposes, and on the basis of a range of different theoretical assumptions (see Simon and Boyer, 1974; Gallon, 1978). Moreover, while most systematic observation research in education has been classroom focused, there is no reason why this approach cannot be applied in other settings, for example in meetings or in selection interviews, though obviously the kinds of categories involved would need to be different.

More than with ethnographic observation, systematic observation requires that the researcher be quite clear early on in the research process about what is and is not relevant to the research focus. And the implication of this is that the observational schedule must be carefully designed to serve the particular purposes of the researcher. There is a close correspondence here with what is involved in developing a questionnaire or interview schedule. Furthermore, just as the latter can include free-response questions, so too some observation schedules may allow observers space to describe relevant aspects of the scene in an open-ended way, or to include information which is not covered by the categories but which they believe may be important. We can see, therefore, that the distinction between systematic and ethnographic observation is not as clear-cut as it might seem. There is some scope for a blending of these two approaches; though this is restricted by the practical constraints involved.

In this sub-section we will provide you with a guide to the use of structured observation, for the purposes of planning your pilot research and developing your research proposal. You may need to supplement this with some use of the methodological literature (see Section 10.5 for further information).

Our aims here, then, are twofold:

1 to identify the important factors that need to be addressed in the planning of a structured observation study;

2 to outline the sort of pilot work which might be useful in this work.

Developing your research proposal

There are four key issues to be addressed in planning a structured observation study.

1 What are the central variables in which you are interested?

2 How can an observational schedule be constructed to measure these?

3 How is the observation to be carried out? Who is to be observed, when, and where?

In your research proposal you will also need to outline the analytic techniques to be used, and to provide a timetable for the research.

Central variables

As we have already noted, in structured observation the research focus has to be closely defined before the main data-collection phase begins. This is because the aim is to collect the same data from every case observed. An important requirement in a research proposal, then, is to identify the variables about which the observation is to provide information. At the same time, this has to be done in the light of what is likely to be observable with low levels of likely error. Also important are practical limits to the range of variables about which data can be collected in a single study.

Structured observation studies are sometimes descriptive in goal. Here, the aim is to provide information about the frequency of particular types of behaviour in a population, or to document the character of a particular set of lessons or meetings in terms of some set of types of interaction, for example relating to teaching or leadership style. However, structured observation may also be used for the purposes of testing an explanation or theory. An influential tradition of systematic observation studies is referred to as 'process-product research'. Here, the method is

used to investigate the extent to which different kinds of behaviour, for example different teaching styles, produce different outcomes, for instance different levels or types of student learning.

Constructing a schedule

Constructing an observation schedule is one of the central tasks in this kind of research. Here the question which must be asked is: what sort of categories will be necessary in order to produce data that will enable the researcher to come to conclusions of the kind specified in the research focus? This is the problem of operationalization. It involves translating relatively abstract concepts into categories referring to behaviour which is identifiable with a minimum of likely error. In the history of systematic observation research there has been a progression from an initial reliance on rating scales which depended on considerable judgment on the part of the researcher to the use of much more concrete categories. Indeed, even the categories included in Flanders' schedule have come to be criticized for being open to multiple interpretations. As a result, many of the more recent systematic observation studies have engaged in detailed specification of what does and does not count as an instance of each category.[23] As with survey research, structured observation requires some training and monitoring of the people who will be collecting the data, whether this is the researcher him or herself or others specially recruited for the purpose. This is necessary to try to ensure that data are collected in as standardized a way as possible across all the cases studied, so that they are comparable.

If an observational schedule is to produce data which can be subjected to quantitative analysis, the categories must have certain characteristics. Most important of all, they will need to be mutually exclusive. In other words, at each point of observation, for each variable, the researcher must be faced with a choice between one or other of a fixed set of categories. Any ambiguities about in which of the categories a particular instance of behaviour belongs is a potential source of error, since as a result different observers (or the same observer on different occasions) may code what is effectively the same behaviour differently. One of the functions of pilot work in this kind of research, then, is to test the specification of the categories to discover whether they cover the whole range of relevant behaviour, and whether there are instances which are ambiguous in terms of that specification.

Another issue which needs consideration concerns the sampling of behaviour. In live coding (as opposed to the coding of audio- or video-recordings or transcriptions of these) it is not possible to record every item of relevant behaviour, unless the behaviour focused on is fairly rare. If a range of frequent types of behaviour is to be recorded, some sampling system will be required. There are various options here. Some schedules involve recording whether particular types of behaviour occurred within a particular time interval. Perhaps the most common system, however, is point sampling. Here, the behaviour of the target person(s) is recorded at regular points in time. (With Flanders' system, as we noted, this is every three seconds; with the ORACLE schedules it is every 25 seconds.) These different forms of behaviour sampling produce somewhat different kinds of information, and some thought is required about which is most appropriate for the purposes of the proposed study. Again, pilot research can serve a useful function here.

As we noted earlier, schedules can take a variety of forms. They vary in the number of variables they cover, and in the number of observational categories they include; and they may also involve scope for open-ended description on the part of the

[23] The article by Scarth and Hammersley, 'Questioning ORACLE', in Reader 2, discusses examples from some of the training materials developed by one of the major systematic observation studies carried out in Britain in the past few decades, the ORACLE project. This article also provides a critical analysis of the operationalization of teaching styles in this research.

researcher. However, in constructing a schedule one must not onfy be concerned with how effectively it will measure the variables concerned, but also with how feasible the observation process is likely to be. Obviously, the more variables and categories involved the more difficult the task of observation becomes; and this has implications for the kind of behaviour sampling which can be employed: for example, how widely spaced the behaviour sampling points need to be. The implications of the complexity of the observational process for the likely quality of the data must also be assessed. In other words, what is important here is not just whether it is physically possible to fill in the schedule in the time available, but also whether time pressures are likely to affect the quality of the data produced. Furthermore, it is necessary to remember that while an observational process may be feasible over a relatively short period, and may produce reasonable quality data, observer fatigue can affect the data when the observation takes place over relatively long stretches of time. For this reason, the observation might have to be carried out in short bursts.

The inclusion of scope for open-ended descriptions in a schedule increases the time which will need to be allowed for processing data, since such descriptions will require coding. Moreover, this may not be at all straightforward since the nature of these data will vary across observers. At the same time, data of this kind may provide important information, or insights, that would otherwise not have been available. Here, as elsewhere, judgments have to be made balancing both methodological and practical considerations.

Who, when and where?

Very often, structured observation is used to produce quantitative data about a large number of cases which can provide the basis for conclusions about the frequency of different types of behaviour within a population, or differences in frequencies among two or more populations. Here, researchers using systematic observation are faced with a trade-off between the number of cases they can study and the amount of time which they can spend observing any particular case. Such dilemmas are characteristic of all forms of research, but they can have important consequences for the quality of the data and the soundness of the findings. For this and other reasons it may be necessary for systematic observation research to study a sample of the cases making up the target population. Here, much the same considerations apply as in the case of survey research, and it would be worthwhile reading the section on sampling in our earlier discussion of survey method (Section 9.3).

Systematic observation research is also sometimes used to provide information where the focus is a relatively small number of people in a specific context. Here sampling across cases may not be involved, but there will still be sampling over time. Decisions will need to be made about the time periods when the observation will take place. This can be of some significance given that people's behaviour varies over time: there may be weekly cycles; different lessons taught by the same teacher may involve different behaviour on his or her part in some respects; meetings dealing with different agenda items may bring about very different behaviour on the part of members of a senior management team, etc. These sampling issues, and the implications of the strategies adopted, will also need to be discussed in the research proposal.

Summary

Your research proposal, then, should contain:

1 a statement of your research problem and its rationale;

2 a discussion of the variables which are your focus and of how they are to be operationalized in terms of an observational schedule;

3 an indication of the target population and information about how this population is to be observed: who will be observed, where, over what period, using what behaviour sampling strategy, etc.;

4 a discussion of any problems that are likely to arise and how they would be dealt with;

5 an account of the analytic techniques which it is proposed to use;

6 a timetable indicating the time allotted to each stage of the research.

Pilot research

As we have noted, it is common in structured observation research to use pilot studies to develop, test and refine schedules. This can provide information about the extent to which the categories are exhaustive of the relevant behaviour and sufficiently unambiguous. Pilot work can also test the feasibility of the observation task. As with survey research, an important aim is to anticipate, as far as possible, what could go wrong in the main phase of data collection.

There are at least two main possibilities for pilot work here:

1 unstructured observation, perhaps supplemented by interviews with relevant actors, leading to the development of an observation schedule;

2 the development of an observation schedule on the basis of existing knowledge, and/or the modification of an existing schedule, and its application to a small sample of cases with a view to identifying and correcting problems in its use, and assessing its value for the purposes of the research.

Either of these options is acceptable, and there are probably others.

For the purpose of the pilot study you only need to collect a small amount of data. This may amount to one or two periods of unstructured observation, or around four or five periods of systematic observation.

In your plan for the pilot research (required for STMA 04) you need to indicate:

1 its central purpose;

2 what type of data you intend to collect and by what means, and what sort of analysis you will employ.

You should also consider any problems you anticipate and how you intend to deal with them.

The report of your pilot research (required for STMA 05) will need to discuss:

1 how it worked out in practice;

2 what you have learned from it.

Some of what you have learned will have arisen from the practical experience itself. However, in addition, while the data produced by your pilot research are unlikely to be amenable to sophisticated statistical analysis, they will require some analysis, qualitative and/or quantitative, if you are to gain the most from your research. This analysis should be clearly presented in your report.