Chapter 4   Methodology

In this chapter I discuss the methods I adopted in my research with a view to achieving my aims outlined in chapter two and answering the questions posed there.

The first step was to acquire the raw examination results for past pupils. This was not easy as Sexey's school had not previously made any use of these data and result slips were filed away in cabinets, in case former pupils forgot their grades and enquired as to the results they gained, but not readily available. Having found the actual grades pupils achieved these were then converted into a numerical score for ease of statistical manipulation. The scales used were as follows:

A level: A=10, B=8, C=6, D=4, E=2, N or U =0

GCSE: A=7, B=6, C=5, D=4, E=3, F=2, G=1, U=0

       A*=8 was added later with the introduction of this new grade.

For both examinations I decided to discount absences on the grounds that teachers would not like being held to account for low points totals in their subject area because a pupil chose not to turn up for the examination.
There was a problem with this in that failure to complete coursework at GCSE originally resulted in a candidate being marked as absent. There is now a letter, "Y", to denote coursework not completed so making the distinction between genuine absence from the examination and failure to complete the necessary coursework much clearer. It was also possible that candidates would be encouraged by staff not to turn up to the examination in order to avoid poor departmental statistics, so I chose to record absences but not use them in the calculation of averages etcetera. In this way it would be possible to note high absenteeism whilst discounting these pupils from the statistics achieved by pupils who attended for examination.

The choice of points scales and grade equivalence seemed logical. The

problems with the use of scales have been noted earlier in this thesis and with reference to Gray's comments (Gray, 1986) but there did not appear to be a better scale and it was important that my method should be readily understood by teachers. As the A level points system for university entrance is worked out on an identical scale to my A level scale I did not see teacher comprehension as a great problem.

The results of each pupil were converted to a points total. For A level pupils I used their total A level points as their outcome measure but for GCSE pupils I chose to use the pupil's mean score, the total number of points achieved divided by the number of examinations taken. I chose this particular method for GCSE results because there was a problem in comparing pupils who had come from different schools where different entry policies operated. I deemed it unfair to compare unfavourably a pupil who had obtained, for the sake of argument, seven grade A's giving a points total of 49 points in a school which only allowed pupils to sit seven examinations, with another pupil in another school who had sat ten examinations obtaining C grades and therefore gaining a total of 50 points. The quality of the grades of the first candidate would be lost in comparing points totals with the second candidate who gained a greater total but averaging lower grades. Using the mean grade method the first candidate achieves a mean grade of 7.0, (7 x 7 / 7 = 7.0) equivalent to an average A grade, whereas the second candidate achieves a mean grade of 5.0 (10 x 5 / 10 = 5.0). This system is not perfect for a candidate could sit only one examination and achieve an A grade giving a mean grade of 7.0, the same as our first candidate. The level of achievement is patently not the same so in all printed analyses the number of examinations taken is also made clear to allow users of the information to discriminate between the two candidates with similar means.

At A level too there is a problem related to schools' entry policies. Some schools enter candidates for General Studies and some do not. This is related to

the particular school's policy rather than the ability of the pupils. When making internal comparisons within the school this is not a problem, for in a school that enters candidates for General Studies the more able candidates tend to do the extra A level and therefore deserve the extra points gained from sitting an extra A level. When comparing pupil performance across schools, however, pupils in schools where General Studies was not taken as an extra A level would be at a disadvantage when comparing total points scored. The simplest solution is to omit General Studies results from any inter-school comparisons as is done in most newspaper league tables but not in the Department for Education's performance tables!

Another option is to consider the average points per pupil subject entry. In this way an A level candidate gaining three 'A' grades would average 10 points but so would a candidate who sat one A level and gained an A grade. So, as in the GCSE analysis, for each pupil in my analysis sheets for schools I show the number of examinations taken, the total points achieved and the average points per subject entry.

Two problems encountered very early on in my research work were those of speed of data analysis and accuracy. My initial analysis of two years of A level results from Sexey's School was performed using a calculator, pen and paper. This task alone took in the region of eighty hours simply to calculate indicator averages and outcome scores. When checked later there were a number of errors in the calculations. This was simply not a practical way to proceed. Use of a spreadsheet was not as flexible and with the increasingly large data sample size also became very slow.

I therefore decided to write my own bespoke software to do the analysis. This had a number of advantages the first of which was flexibility. As my awareness of the statistical techniques required increased so I could add these to my software. Data were input more quickly and could be checked more

easily. The analyses and their printouts could be tailored to my needs and those of the teachers involved.

Statistical techniques had to be learned including the principles of correlation, both Spearman and Pearson, and the creation of regression lines with scattergraphs. The computer programming was fascinating but also time consuming because I was learning many of the techniques as I went along. The resultant package was very flexible and well worth the effort for the speed of analysis encouraged deeper analysis and made research such as this possible for practising teachers.

The input of data and accuracy were both increased when I wrote software to take output from the SIMS (School Information Management System) computer system which schools were beginning to use more and more. The increased speed of input and analysis meant that I was in a position to offer the analysis to other schools and increase the data sample size as well as network with other schools for comparative purposes with the intention of mutual self-help.

Some data are still input by hand but in 1996 the vast majority of schools now submit data on floppy disk from their computer administration systems, this method being more accurate and reliable although checks still need to be made and occasionally schools have results missing from their system data files which need to be input by hand.

Data required

Schools submitting GCSE results for analysis are required to submit the following data: pupil names, pupil gender, prior test scores, examination results including subject, grade, examination board and syllabus.

Pupil names were considered important because in the resultant printouts it was found to be far easier for teachers to relate to pupil names and their

performance figures rather than the figures alone. The names bring with them the image and memories of the pupils concerned, particular circumstances that may have contributed to an individual's performance in the examinations and this is precisely the information that one needs when discussing individual pupil performance with a Head of Subject.

The pupil gender was important so that further analysis of the data could be conducted on a gender basis, addressing some of the issues identified in the section on the aims of this research, namely exploring the data for gender effects, any differences in subject areas or between schools, and whether performance might be linked to the proportion of girls in the year cohort.

Prior test scores for pupils sitting GCSE examinations were mainly Edinburgh Reading Test Stage 4. Sexey's school, in common with all state schools in Somerset, apply this test to their Year 8 pupils in the Spring term during the course of one week so that the baseline data is common and comparable for all Somerset state schools. Sexey's therefore had this prior data for pupils going back a number of years. This was useful but also incomplete. The data were not available for all pupils, many having arrived at Sexey's after the test had been taken and never subsequently tested.

The size of the Year 11 GCSE cohort at Sexey's is small, usually around 45 pupils, so when the number was further reduced because some pupils lacked the prior information it became very difficult to be confident in any statistical findings because of the small numbers of pupils involved. Initial correlations between ERT and mean GCSE scores of individual pupils were sufficiently encouraging (See *Figure 4.1*) to proceed further with the research and seek to involve more schools with more pupils so increasing the size of the overall sample but also looking at other schools' populations.

Figure 4.1

Correlations between ERT and GCSE mean score at Sexey's School

Year	1987	1988	1989
Pearson's r	0.64	0.75	0.71
Sample size	46	51	56

Pearson's product moment correlation formula was used to calculate most of the correlation co-efficients because it is suited to measuring the strength of relationship between two sets of variables on linear scales. Spearman's rank correlation is also used to look at the different rank order positions of pupils according to their indicator scores and examination outcomes, but Spearman's correlation is susceptible to the presence of tied rank positions which inflate the correlation co-efficient. Schools are presented with both correlation figures for their school's data.

Somerset LEA uses the Edinburgh Reading Test (ERT) as a monitoring tool, all pupils taking the test during the space of approximately one week in February during their Year 8 (second year in secondary school), and actually incorporates the standardised age score result in the LEA formula for funding Special Educational needs.

The ERT Stage 4, as used with Somerset secondary schools, was standardised in 1975 against two samples, one from Scotland made up of 2,282 pupils from across the age range to be covered and one from England and Wales made up of 2,216 pupils. The age range covered by stage 4 is 12.0 to 16.0.
The ERT series covers 4 tests, stages 1, 2 and 4 having been produced by the Godfrey Thomson Unit for Academic Assessment at Edinburgh University whilst stage 3 was produced by the Moray House College of Education. The test was designed as an instrument to help teachers assess pupil progress in reading skills.

The Stage 4 test itself incorporates five sub-tests:

a)   Skimming - 30 items designed to test pupil ability to extract the required information

b)   Vocabulary - 35 items to test pupil understanding of the meaning of various words

c)   Reading for facts - 30 items to test ability to decide what a passage says or does not say

d)   Points of view - 35 items to test pupil ability to perceive the consistency of various points of view

e)   Comprehension - 25 items to test pupil ability to draw conclusions of an inferential nature about various passages.

The total testing time for stage 4 is exactly one hour making it relatively swift and simple to administer within a school environment. Kuder-Richardson reliabilities (K.R. 20) were calculated during the development of the test on the basis of the whole sample and produced an overall internal consistency reliability for the test of 0.956. The internal consistency of various individual sub-tests, again using K.R. 20 were A=0.858, B=0.911, C=0.776, D=0.844, E=0.732. Tables are given in the test documentation allowing the conversion of raw scores into age adjusted quotient scores operating on a scale of 70 to 130 being standardised around a mean of 100 with a standard deviation of 15. The stated error margin for the quotient scores is $+/-3$.

That the test was produced and standardised as long ago as 1975, and has not been revised or updated since, particularly when the vocabulary commonly used by young people of this age range is likely to have changed quite markedly since then, must be an area of concern. Indeed, within Somerset some Headteachers would claim that there has been a perceived drop in the general scores by successive year groups, although I have not seen figures to substantiate this claim and it would be very difficult for individual schools to disentangle any change in average quotient figures for their schools from any

changes in the nature of their intake. However, it remains a reliable and useful indicator of pupils' academic potential.

That all the secondary schools in Somerset had this information for their pupils, particularly those submitting their GCSE examination results for analysis, was particularly fortunate because it enabled measures of performance to be judged against a common baseline or starting point. This starting point was to be the Standardised Age Score, or quotient score, with no weighting for gender differences. As no account was taken of gender in the standardisation of the ERT scores, this could mean that some of the differences in performance at GCSE by the respective genders may be masked by the girls already having progressed at a greater rate than the boys and so gaining higher ERT scores. This cannot be checked with the data available and so any differences that become apparent between the genders at GCSE can only be attributed to progress made since the ERT was taken in Year 8. It is quite likely, of course, that individual pupils of whatever gender are progressing at their own individual rates in any case.

The larger numbers involved when looking at pupils from a number of schools meant that I could give greater statistical weight to my findings than when looking at the very small numbers involved at Sexey's School alone.

However, I came to use the Edinburgh Reading Test scores as a baseline for measuring pupil performance rather as a secondary issue, for I was first concerned with looking at Advanced level performance, and then rather by default as it was the only data I happened to have on the majority of our pupils. I was also extremely sceptical, as a practising teacher, that a supposed reading test could tell me anything about a pupil's ability in Maths, Physics, French or Latin. I will return to the details of how I was proved wrong later in this thesis.

Further data collected included the individual grades achieved by pupils in each subject they sat, including the full subject name, examination board and syllabus. Not all schools submitted the last two pieces of information. These data were necessary to allow a full analysis of the results for each subject department in a school, pupil by pupil as well as for the subject as a whole, and the part they played in contributing to the school's overall results. Information on examination board and syllabus would be useful information for teachers to consider if large differences appeared in the performance of similar ability groups taking the same subject in roughly similar schools. By roughly similar I mean schools of approximately the same size serving similar communities and in similar areas.

On running the data through my software detailed analyses were produced and printed out to be returned to the schools involved. These analyses would initially be only the schools' own data but would then be followed by analyses based on the combined data of all the schools involved. Examples of the various printouts are included in the next chapter with an explanation of what they show and how the figures are derived. Findings from the analyses and highlighted by the printouts will be discussed in the following chapters.

The detail of my methodology outlined so far may seem heavily biased in favour of quantifiable measures, figures based on pupil outcomes rather than the processes by which these outcomes are achieved. As such it could be claimed that there is more emphasis on the "school effectiveness" side of my research than the "school improvement", for if schools are to improve then they need to know both what and how to improve rather than simply being told they are inferior in comparison to other schools. McCormick and James (1983) argue that the processes operating in schools have merit in themselves,

> "First, it is possible to regard curricular processes as having intrinsic
> value. For this reason, they are worthy of attention in their own right i.e.
> without reference to objectives or outcomes."

I do not disagree with this idea but in schools today it is the apparently objective measures of examination results by which the performance of schools is judged and held to account. Indeed the external pressures upon schools are shaping the delivery of the curriculum within schools with regard to subjects taught and to whom they are taught as well as examination entry policy.

Using quantitative data, examination results, allows schools to use the Objectives model for evaluation as expounded by Shipman (1979) who believed that setting out one's working objectives before starting an evaluation was the only way to proceed. He claims that the tasks of teaching and learning must be defined if they are to be operationalized and that the definitions should include some indication of what will be considered success or failure. These ideas are considered by McCormick and James (1983) who go on to state that using this method the indicators thus defined become criteria and standards against which curricula may be systematically judged.

Stenhouse (1975) traces the origins of the Objectives model back to Bobbit (1918 and 1924), who sought to bring some scientific objectivity to the understanding of the purpose of schooling, his work being further developed by Tyler. I quote from Stenhouse (1975, 53),

> "Tyler starts from the question: what educational purposes should the school seek to attain? and he equates a purpose with an objective or goal. Education is a means towards ends."

Tyler was clearly attempting to cast light upon the role of education,

> "Since the real purpose of education is not to have the instructor perform certain activities but to bring about significant changes in the students' patterns of behaviour, it becomes important to recognise that any statement of the objectives of the school should be a statement of changes to take place in students." Tyler (1949, 44)

It was in attempting to quantify the role of education that Tyler developed his ideas on educational objectives, "One can define an objective with sufficient clarity if he can describe or illustrate the kind of behaviour the student is expected to acquire so that one could recognize such behaviour if he saw it." (Tyler 1949, 59-60). Here one can see the generally accepted definition of "Objectives" as referring to 'Behaviours". Elsewhere behavioural objectives are synonymous with "intended learning outcomes" or "goals", the general theory being that having defined one's goals when evaluating the curriculum it is possible to ascertain whether they have been attained, and therefore the curriculum as delivered is successful, or not. The problem, obviously appreciated by Tyler, is one of definition -- making the Goal sufficiently clear that it is obvious whether it has been attained or not.

Tyler had eight steps towards curriculum evaluation:

1. Establish goals / objectives

2. Categorize them

3. Define them in behavioural terms

4. Establish situations in which they can be demonstrated

5. People must know about them in advance

6. Devise appropriate measuring techniques

7. Collect performance data

8. Compare the data with the objectives

The apparent advantages of this methodology are quite clear. By establishing goals initially there is a focus for all involved; a strong force for organization allowing many disparate areas of educational research to make a contribution.

> "Just as Taba argues that objectives 'provide a common, consistent focus for the multifarious activities we call the curriculum', so the objectives model provides a logical pattern of co-operative action and intellectual synthesis for those engaged in educational research and the academic study of education." Stenhouse (1975).

81

The process of categorizing the goals makes the goals themselves that much clearer and may be a useful process in conjunction with point 3 in widening the perception of those involved with the evaluation. By this group I mean not only the evaluators but the subjects of the evaluation and the recipients. An understanding of what the purpose of an evaluation is and an identification of the desired outcomes should ease the course of the evaluation and help bring about the intended changes. Evaluation must be concerned with change else it becomes an assessment mechanism leading to a report.

McCormick and James (1983) refer to Bolam (1975) when they say, "a school's organizational structure, the role of its head, and the values and attitudes of its teachers come to be regarded as crucial to the survival of any curriculum project." It is therefore very important that these people understand the importance and relevance of any evaluation project that is to be undertaken in their school.

The use of measuring techniques, collection of performance data, and the comparisons made give actual physical results which can be examined, displayed and debated. In terms of providing data for those to whom a school is accountable the process used is very clear. The method is heavily geared towards the collation of quantitative data such as examination or test results. Indeed, our GCSE syllabuses explicitly use stated objectives and aims to identify those behaviours which are expected of candidates attaining prescribed grades.

Hopkins (1989) describes the objectives model under another guise, the Research and Development or Factory model, and summarizes it as follows:
1. Translate agreed aims into specific, measurable behavioural objectives
2. Devise appropriate learning experiences
3. Devise tests to assess student performances

4. Administer tests with a sample of classes using the new programme

5. Process results to yield useful information to the team which is producing

    the new programme, or to the sponsors and potential users of the project.

Hopkins admits the popularity of this approach but goes on to criticize the model stating that this is "due to the fact that it is apparently simple to use and appeals to the ends - means, cause - effect, linear and bureaucratic view of the world commonly held by policy makers."

Eisner as quoted in McCormick and James (1983) makes a similar but perhaps more cogent criticism,

> "The outcomes of educational programmes are not completely
>
> predictable, and hence to evaluate only for those goals one has intended
>
> can lead one to neglect equally important, and at times even more
>
> important, outcomes that were unintended. Using objectives to serve as
>
> criteria for evaluating educational programmes has an important function
>
> to perform, but a conception of evaluation that limits itself to what has
>
> been pre-planned in terms of goals or objectives is likely to be
>
> educationally thin." Eisner (1979, p.174)

To summarize some of the criticism of the objectives model; it relies far too much on the outcomes of the educational experience, assuming a distinct relationship between these and the initial intentions. To quantify the benefits of a programme it is necessary to establish a baseline from which to measure, yet Tyler's method does not explicitly make any mention of pre-testing and post-testing to calculate any gain score. In theory then it is possible that a school may already be meeting its stated objectives and the evaluation would serve no other purpose than to provide a metaphorical pat on the back, for an objectives model is able to give little information about *how* the educational objectives are achieved. Similarly, if the objectives are not met, there is no information to say why not.

Any feedback the objectives model does offer is not available until the end of the process. In a large evaluation project taking considerable amounts of time it would be helpful to receive some progress report. This would be helpful for the morale of those being evaluated and could be used to encourage some degree of ownership of the end report. It would also be helpful to the evaluators in that they could gain some idea of whether they were on the right lines and that the programme they were actually studying was not actually changing as they conducted their enquiries.

The objectives model's concentration on quantitative data, whilst appealing to those who demand hard evidence, ignores much of a subjective nature which in itself can be said to have validity. A teacher's "gut feeling" or intuition, built on years of experience, is incapable of being quantified in any meaningful way but should certainly not be ignored. In my own research I have found that the experienced teacher is often extremely accurate in making predictions, sometimes in contradiction of their own mark books!

> "In short, in order to evaluate one must understand. It can be argued that conventional objective-type evaluations do not address themselves to understanding the educational process. They deal in terms of success or failure. But a programme is always a mixture of both and a mixture which varies from setting to setting." Stenhouse (1975 p.109)

Relying purely on quantifiable examination result data is insufficient to inform a school how well it is educating its pupils. Use of prior attainment data adds another dimension in that it helps to explain the variation in attainment of different pupils, along with such data as gender, social background and circumstances operating within the school. Setting targets and benchmarking can be useful in aiding schools focus on self-improvement but they must not ignore the human element.

Illuminative evaluation offers a very different approach to the evaluation of school effectiveness, one with a more holistic approach. Stake advocates a more responsive approach to evaluation. I quote from Stenhouse (1975 p. 114),

"An educational evaluation is a 'responsive evaluation' if it orients more directly to programme activities than to programme intents, if it responds to audience requirements for information, and if the different value-perspectives present are referred to in reporting the success of the programme" Stake (1972).

The essence of what Stake is saying is that the programme of evaluation should look more at what is actually happening during the programme itself and should take full account of the intended audience for the report. There may be several groupings, each with their own views of what is required in terms of information, who will receive the end report and must be made aware of the other parties' concerns.

MacDonald, as quoted in Stenhouse (1975 p.112), takes this point further, "Evaluation is the process of conceiving, obtaining and communicating information for the guidance of educational decision making with regard to a specified programme." MacDonald (1973, 1-2).

It was in 1972 that Parlett and Hamilton argued that evaluation should be of a more descriptive and interpretive nature than was possible using the objectives model. Rather than measurement and prediction they proposed a 'social - anthropological' approach calling this their "illuminative" model. Fundamental to this approach was an understanding of a school's "instructional system", the arrangements for the formal provision of the curriculum, and the "learning milieu", the complicated social-psychological and material environment in which the pupils work. The method is very much defined by the problem in hand, and as such will vary from situation to situation, but the general strategy is as follows,

1. Observation --- become familiar with the situation

2. Further inquiry --- select data, hypothesise, focus on particular areas

3. Seek to explain --- seek general principles, spot patterns of cause and effect.

The process thus informs itself continually and is capable of highlighting unexpected issues which may be critical to the project and yet would have been disregarded by an Objectives model. "Beginning with an extensive database there is progressive focusing on the emerging issues which allows unique and unpredicted phenomena to be given due weight." Hopkins (1989, p.24).

This necessary information is acquired through as wide a range of methods as possible, including observation, interviews, questionnaires, test data, documentary and background information. The "Aim" is to "Illuminate" the complexities of the various "realities" for a varied audience rather than concentrate upon decision making. In turn this will enable a sharpening of discussion, the complexities to be disentangled, significant information to be isolated from the trivial, and the raising of the level of sophistication of the debate.

With this model each evaluation is unique to the particular situation,using the techniques judged to be most useful in the particular circumstances. This is both a strength and a failing of the model: a strength in that time and effort are not wasted using inappropriate techniques, but a failing in that it can never be replicated and so is open to the criticism of having a non-specific methodology with no history by which it could be judged.

> "Without an adequately formulated body of theory or methods, the
> illuminators have been, and will be, unable to progress and generate a
> coherent, cumulative research tradition. They cannot transcend the short-
> term practicalities of any given programme of curriculum innovation.
> They merely substitute one variety of atheoretical 'findings' - based
> mainly on observation and interview - for another - based mainly on test

scores." Atkinson and Delamont, 1986:252 as quoted in Hopkins (1989 p.25).

Hopkins also refers to criticisms made by Lawton (1980:117-118) suggesting that illuminative evaluation has a controversial and non-specific methodology, tending towards subjectivity, role conflict and esoteric language. This does mean that the skill and experience of the evaluator is absolutely vital to the success of the project. S/he must be able to identify correctly those areas of information that merit more focus, be able to formulate theories, maintain a sensible level of detachment and yet present information in a useful way for possibly disparate groupings of interested parties. Good inter-personal skills are obviously essential.

Validity is also a problem that must be addressed with the illuminative model of evaluation. With the objectives model there is a clearly defined method, agreed beforehand, and often quantitative measures that are available for inspection. This is not so for the illuminative model and therefore a greater need to keep a log / diary of the actions taken which may be looked at by others. Triangulation, wherever possible, should be employed, so that more than one view of a situation can be called upon to support or refute a deduction.

Therefore, where a school has a good idea of exactly what it wants to check and possibly alter following an evaluation, particularly if they are striving to demonstrate their accountability, then an Objectives model would seem to satisfy their needs. This model offers reassurance in that a school is able to "quantify", "awareness raise" or even "train" its staff by working through the various stages of the model. Its well defined methodology leaves little scope for criticism by those who would hold the school to account.

However, if those who desire an evaluation are unsure of their aims, except that they wish to find out the merits or faults of a given project, then the

Illuminative method seems to have much to offer, particularly if the management structure of the school is open and adaptable enough to take on board the issues such an approach could throw up. The illuminative model is better able to cope with the unknown and, by virtue of more regular feedback, is far more able to change course in the middle of an evaluation should the need arise. It is far less dependent upon quantitative data and therefore able to look at qualitative issues, such as pupil - teacher or teacher - teacher interaction. This method even holds out the possibility of throwing some light upon the very essence of the education experience in a given institution.

The nature of the data provided to me by schools lends itself to analysis which is primarily quantitative, using examination results and pupil data. These data, analysed properly, allow for the setting of objectives and the measurement of performance against these objectives but in a value-added sense where prior measures of attainment are available. The use of this more accurate analysis should advise the process being used with current pupils in a formative sense. Quantitative data do not preclude illuminative analysis of situations pertaining in schools but can add considerably to the understanding of human factors. I would hope to capture some elements of this in the production of case studies based upon particular scenarios (see chapter 6) in schools and therefore relate statistical findings to actual experience.

> "Social Scientists have come to abandon the spurious choice between qualitative and quantitative data: they are concerned rather with that combination of both which makes use of the most valuable features of each. The problem becomes one of determining at which points they should adopt the one, and at which the other, approach." (Merton and Kendall, 1946).