

To gain an appreciation of the problems involved in conducting a meaningful analysis of school examination results it was necessary to conduct a review of some of the relevant literature.

Towards the end of the 1960's in Britain the education system was coming under criticism for its "trendy" methods and falling standards. This attack was typified by the publication of a series of papers collectively known as the "Black Papers" (Cox and Dyson, 1969a, 1969b, 1970).

The quality of educational provision was further brought in to question by Jencks in America. His research (Jencks *et al.*, 1972) argued that at least half of the differences in pupil academic outcomes could be attributed to the pupils' social background and prior attainment rather than to the school they attended. Furthermore, he could see no evidence that educationists knew how to go about raising standards,

"We can see no evidence that either school administrators or educational experts know how to raise test scores, even when they have vast resources at their disposal. Certainly we do not know how to do so."
(Jencks *et al.*, 1972).

This was misinterpreted by some as saying that schools did not matter for other things affected pupil performance.

Education continued to be a political issue in the 1970's and was given further impetus by the then Prime Minister, James Callaghan, delivering a speech at Ruskin College, Oxford in 1976. This led to the "Great Debate" on education, conducted by a series of regional conferences. Issues of accountability in education also became a theme of the time: who were the stake holders in the system and to whom were schools accountable (Lello, J., 1979).

It was from the arguments and debate of these times that the academic interest

in school effectiveness grew, in part as a desire to show that schools do make a difference to the outcome of a child's education.

Examination results are by no means the only indicator of a school's success or performance. There are many other indicators such as pupil manners, dress, motivation and behaviour, the popularity of the school, involvement within the local community, success in competitions, sporting or otherwise, to name but a few. Each of us will have our own check-list of criteria against which we could judge a school but the criteria are essentially subjective and not easily quantified.

However, the Office for Standards in Education (OFSTED), the body with the responsibility for inspecting our schools and passing judgement on their success or failure, explicitly stated in their "Framework for the inspection of schools" (OFSTED 1992) that the performance indicators the inspectors would be looking at would be the results of National Curriculum assessments and external examination results. This focus by OFSTED on examination results as a performance indicator serves to emphasise their dominant role in judgements made on the effectiveness of schools.

Examination results remain the currency with which places in Higher Education, employment in the professions or better paid jobs are obtained. Examination results are inextricably tied up with a pupil's prospects in life and therefore one can understand parents' desire for their offspring to do well in their examinations. If a school is perceived, rightly or wrongly, as gaining good examination results, then again it is perfectly understandable if the parents wish their children to attend that school in preference to one with less good examination results. The problems arise with the interpretation of examination results and deciding whether the results are in fact a good performance or not, relative to pupil intake, whether they apply across the range of abilities and whether they reflect the quality of the school, the pupils or indeed the parents.

With a view to making parents better informed about their pupils' schools, the Government for the first time in 1992 published school examination results, school by school with national figures for comparison, (DFE 1992). The Secretary of State for Education, John Patten, wrote the foreword included in the DFE tables from which I quote,

"Parents have many important responsibilities - choosing the right school for their child is one of the most important. Taken together with improved school prospectuses, the new tables will make sure that parents facing that choice are better informed than ever before.

The tables will also ensure that those served by each school - pupils, parents, employers and the wider community- know the standards being achieved and how they compare with others locally and nationally."

There can be no doubt that parents have more information available to them as a consequence of the recent changes regarding reporting of information to parents by schools, the Parents' Charter for example (DFE, 1994, 1995), but more information does not necessarily mean parents are better informed. More facts and figures do not automatically lead to a greater understanding of the factors involved in the attainment of a given set of examination results.

In this chapter, with reference to school effectiveness literature, I intend to illuminate some of the key problem areas associated with establishing whether a school is effective or not in terms of the examination results its pupils obtain.

These key areas are essentially:

- Measures* - defining the instruments one is going to use to quantify effectiveness and improvement;
- Standards* - establishing standards for purposes of comparison and the problems involved;
- Variation* - over time, within and between institutions;
- Gender* - the part gender plays in overall performance and

differences in performance between genders;

Professional problems - interpretation of statistics, the conflict of those requiring more information with those who wish for less, more concise information, and phobias of the teaching profession.

For many schools the catalytic element in focussing attention upon the above mentioned areas has been the production of performance tables by the Government. The narrow focus of these tables has stimulated much discussion in school effectiveness literature as well as in staffrooms and homes throughout the country.

Measures of success

Exactly what measures of examination success should a school report to inform its parents and others? If it is to be the overall pass rate, then what constitutes a pass? Are we to consider the GCSE grades A-C as pass grades or A-G when at A level it is A-E? This issue alone is enough to confuse many parents at the transition point for their children from GCSE to A level.

Schools tend to quote the figures which serve their purpose. What is the figure quoted a percentage of? Is it the percentage of pupils who were entered for the examinations, the percentage of those who turned up for the examination, or the percentage of pupils belonging to a given age group within the school? Each has its merits and faults in terms of expressing the truth about a school's performance but without a clear definition of how the figures are arrived at, stating what is being demonstrated about the school's performance, then these figures can have no validity. The uniformity of the Government's regulations on reporting examination results is helpful in this respect because strict guidelines are laid down as to what should or should not be included in the figures (DFE, 1995). This uniformity of information therefore gives us a degree of reliability when looking at the figures. We know that the figures have been produced in the same way and the figures tell us something about the

levels being achieved. But this is not in itself a measure of the school's effectiveness with its particular pupils, at this point in time or previously. Nor are we told anything of the learning processes that led to such results or what the pupils were like who passed through those processes. It is very difficult to discern whether the particular school has maximised pupils' learning potential, enabling them to gain the highest examination grades they were capable of achieving, or the reverse.

A level indicator measure

In considering A level examination results, Fitz-Gibbon has conducted much innovative work culminating in the formation of an A Level Information Service, or ALIS for short. (See Fitz-Gibbon, 1992). The beginnings of this scheme involved collating the A level results for a number of schools, allocating points for grades according to a fixed scale, and then comparing the points per pupil with the average grade achieved by the same pupils in their O level examinations. It was decided to calculate a mean grade for O level and use this as an indicator of potential for A level, instead of using the total points achieved at O level, because of the fact that some pupils may have sat more examinations than others, quite often because of differing school policies rather than any great difference in ability.

Nuttall and Goldstein (1992a), in a survey for the Guardian newspaper, chose a slightly different method of calculating a GCSE mean grade by counting a pupil's best eight examination results. Their intention was to try and standardise the intake measure as much as possible, disregarding the poorer results of those candidates who sat more than eight GCSEs.

This method does not completely solve the problem of making comparisons with pupils who only sat seven or six GCSEs. Whilst the number of sixth form A level candidates who sat less than eight GCSEs would be small, in my experience there are a number of A level candidates who do sit less than eight GCSEs, particularly those candidates who go on to study subjects such as Art,

Design, Craft or Performing Arts subjects where particular skills are more important than general academic aptitude.

One must also be careful to count only those examination results that were taken in the one academic session. This is for two reasons; firstly it is easier to accumulate GCSE results a few at a time rather than all in one exam season, and secondly all candidates will broadly speaking have endured approximately the same standard of examination difficulty, the criteria being the same for everyone in a given year. (There are problems associated with the comparability of examinations which will be discussed later.)

In reality there is very little difference in outcome between the methods of Fitz-Gibbon and Nuttall & Goldstein, counting all the O level / GCSE examinations or just the best eight, each method producing similar levels of correlation between indicator score and outcome measure.

Gray *et al.* (1995) with a data set of just under 159,000 candidates for A level in 1994 found the correlation for total GCSE points with total A level points to be 0.62 and for average GCSE grade with total A level points to be 0.70, slightly in favour of the average GCSE grade as the indicator.

Problems with scales

Gray (1986) reminds us that using a scale for equating grades with points has its dangers. In the model for GCSE / O level where an A=7, B=6 and so on down to CSE grade 5 or GCSE grade G being equal to one point, which Gray refers to as the ILEA model, the value that should be attributable to the various grades is a moot point. In Gray's words,

"It embodies an arbitrary system of weighting which is not only unfamiliar outside the research community, but may be anomalous and not reflect the relative values of different grades held in the world at large. For instance, is a CSE grade 5 more valuable than an unclassified 'O' level as this system maintains?" (Gray, 1986)

Although CSE's are now a thing of the past the essential point remains that the use of an ordered scale implies an equally ordered weighting to the value of the different examination grades. If one is to manipulate grades in a statistical manner, however, then this scale seems to be the most straightforward, and has the advantage of needing no change to cope with GCSEs. Mortimore and Byford (1981) were responsible for its usage in the Inner London Education Authority and discuss their selection of an ordered scale more fully in Byford and Mortimore (1980). Provided that one does not consider the resulting numerical score or average in isolation from the actual grades obtained then it can be very helpful and does not preclude deeper analysis of worth.

Neither Fitz-Gibbon (1992) nor Nuttall and Goldstein (1992a) make any mention of the subject composition of the indicator examinations. The problem, more acute today than when Fitz-Gibbon started her work, is that pupils may by careful selection of syllabuses, of their own accord or with the prompting of their school, acquire GCSE passes in several closely associated subject areas. For instance, I have come across pupils with GCSEs in double Science, double award English, CDT Design and Realisation, CDT Technology, Art, French and Maths. This totals nine GCSE passes but to what extent is this total to be judged equivalent to qualifications containing English language, English literature, Maths, Physics, Chemistry, Biology, French, Latin and Geography? Obviously for the purposes of indicating potential examination success at A level it would depend upon what A level choices the pupil made but would the alternative GCSE subject results be equivalent predictors of success at A level if the choice of A level subjects were Maths, Further Maths, Physics and Chemistry?

As a check on the quality of GCSE results obtained by the A level pupils Fitz-Gibbon also administered a test of higher intelligence to the pupils. By demonstrating the correlation between the scores in the test and the points

achieved at A level one has a further indicator measure against which to show any progress made by different pupils in different subject areas and indeed schools. The test scores can then also act as a check on the quality of the GCSE indicator. Clear anomalies between the two indicators would indicate the need for further checking of the reliability of the indicator information for a particular individual.

By calculating averages for the whole sample, represented by the regression line when correlating indicator and outcome information, and plotting individual pupil performance against these Fitz-Gibbon could calculate residuals, basically the difference between what would have been expected of the average candidate and what was actually achieved by a specific candidate. This process was repeated for actual departments and indeed could also be done for schools.

The important element of Fitz-Gibbon's work, as far as schools are concerned, is the feedback showing how they performed against the average for all the schools involved and also against other schools' departments with similar ability pupils. Anonymity is of the utmost importance to avoid the potential for creating league tables and their harmful side effects, the chief of which is misinterpretation of data by the uninformed. The data are, of course, presented in a table, which could be construed as a league table, but only the school concerned can ascertain its position in that table and is at liberty to interpret the information as it will. The threat of possible ridicule is avoided.

Schools are presented with information that makes examination results, essentially summative assessments, useful in a formative sense. Courses can be evaluated and changes made, as and when deemed appropriate, for the benefit of future students as they pass through school.

For changes to be considered, departmental performance must be significantly

different from that which might have been expected and well in excess of measurement error. I should like to see how often variation in departmental performance to such a degree actually occurs and whether many departments show such contrasting results year on year rather than as one offs. This would allay any fears that the variations in the performance of subject departments are too volatile for any meaningful statements about their effectiveness to be made.

Already some problems have been noted, but O level or GCSE mean grades as indicators of likely potential attainment at A level do not tell the whole story, for there are other factors involved. Even where there is a very good correlation co-efficient for GCSE mean grades being compared with points gained at A level, say 0.7 for a sample size of one thousand pupils, then the co-efficient of determination, obtained by squaring the co-efficient of correlation, which gives the amount of variance in A level scores attributable to GCSE mean grades, would only be 0.49. In effect this means that 51% of the differences in A level points scored by the pupils is still unexplained.

Pupil background measures

In an attempt to explain further elements of the variance Fitz-Gibbon has developed other indicators such as scales identifying pupil effort, pupil attitude to school and specific subjects, and socio-economic status of the family background of pupils. Not all of these have been successful in explaining different performance levels of students at different schools. For instance Fitz-Gibbon (1992) found that Socio-economic grouping actually gave a negative correlation with positive attitude to school for A level candidates and was only weakly correlated with achievement in A level work. There is no real explanation offered for this finding, although another finding, that pupils from families where the parents held lower-status jobs were generally more satisfied with their education than those pupils from homes with parents in professional jobs may hold the clue. There is obviously much more research work to be done in this area, but the fact that pupils from families with low-status jobs

may already have undergone a more rigorous process of self-selection than other groups could have some bearing on these results, or possibly they may have developed more self-reliant learning skills.

Jesson *et al.* (1992), in their survey of GCSE performance in Nottinghamshire, used a points scale identical to that used by the DFE (A=7, B=6, C=5, D=4, E=3, F=2, G=1) but sought explanation of performance by the pupils in terms of background information rather than indicators of ability, such as verbal reasoning or general intelligence test scores, information which was not available to the researchers. They therefore considered such factors as gender, pupils in receipt of free school meals, family size, ethnic background, parental occupation, housing type and household lived in (the categories were living: with both parents; with mother only; with father only; and with others). The researchers found that all of these classifications had a bearing upon the achievement of pupils in GCSE examinations but particularly the category of parental occupation with which, to a greater or lesser degree, all the other categories are related:

"The implication of this finding is clear: if schools have differing proportions of pupils from these parental occupational groups then, on the evidence presented here, it is to be expected that their pupils' propensities for securing higher examination scores will also differ. This is not in any sense to set low (or high) expectations for pupils from any given social background, but simply to make the point that pupils' characteristics, obtained from large and representative samples, provide clear and unambiguous indications of the likelihood that examination performance will differ for different groups. It will be higher to the extent that schools take in higher proportions of pupils from occupational groups recorded as 'professional' and lower in so far as they comprise greater proportions of pupils with 'manual' parental occupations."

(Jesson, Gray, Tranmer, 1992).

Similar research has been conducted by Nuttall *et al.* (1992) for their report to the Association of Metropolitan Authorities, with similar findings. In this case the researchers did have indicators of prior attainment available to them in the form of verbal reasoning test scores and these did prove to be the single most powerful indicator of attainment at GCSE. Ethnicity also proved to have a powerful influence, most notably the Indian, Pakistani and Turkish communities performing better than other ethnic groups when other factors were discounted. Girls tended to perform better than boys generally.

Interestingly, pupils in receipt of free school meals, used quite commonly by researchers as a measure of social class, did not have the expected effect upon examination performance. The researchers highlight the inherent weakness in using this measure namely that many parents who would be entitled to this provision for their children simply do not apply for it. Therefore many families who should appear in the social category as defined by the chosen measure do not do so which rather negates its use as a reliable indicator of social class.

It is worth noting that in producing their paper Nuttall *et al.* (1992) used a multi-level modelling technique, a variant on multiple regression analysis. The essence of the technique is the appreciation that there are many influences having some bearing upon any given set of examination results one is considering. So within the LEA there are individual schools, within the schools there are year groups, in each year group there are forms and / or teaching groups, the base unit being the individual child. Upon all of these, groups and individual units, other effects such as prior learning, social mix, local economy, family resources all play their part. Using this technique it is possible to isolate individual variables and identify any influences they may have, in effect to generate an order of influence. One might therefore be able to look for any effect say ethnicity might have, taking account of factors such as intelligence, social class, gender and so on.

Examination standards

I would now like to consider the interpretation of examination results and what, if anything, they can tell us about a given school. In the very simplest analysis they tell us what the pupils at a particular school in a particular year achieved in the particular examinations which they sat. They are the summative assessment of the pupils' learning throughout their courses of study. As such they are of course subject to the faults of summative assessment, particularly its content validity and also reliability.

Ideally, to have content validity the examination must test every aspect of a course of study to give the full picture of a pupil's understanding and knowledge. The vast majority of examination papers, one could say all GCSE and A level examinations, merely sample areas of knowledge covered by a course because of the breadth of course content. This means that a pupil who is strong in some areas and weak in others could gain an excellent mark if the right questions came up or a very poor mark if they did not. Neither of these circumstances would give a true picture of the pupil's ability. The situation is exacerbated if there is some degree of choice in the pupil selection of questions to answer. This being so then one must accept that the grade achieved in any examination may to some extent be governed by chance yet once the grade is reported it becomes "de facto" an absolute standard set in stone.

Satterly (1989) neatly summarizes some of the problems concerned with examination results derived both from coursework and terminal examination,

"Examination grades tend to enjoy greater respectability in the eyes of the public than course work grades: the latter can too easily be contaminated by favouritism and cheating and be achieved in a variety of conditions which render the principle of comparability inoperative. But whilst the external conditions of an examination (time limit, invigilation, unavailability of sources of reference and so on) remain the same for all examinees and the actual contents of the paper are the same for all, most

examination papers allow choice from a larger set of questions or problems, such that the assumption that a common basis for comparison has been obtained can be illusory. As Rowntree has said, if a student is required to choose four topics out of ten, 210 different choices are available." (Satterly, 1989) .

For each candidate sitting an examination the nature of the examination paper is different in relation to their coverage of the syllabus, their choice of questions and the examiner who eventually marks their examination script. With so many variables in operation in a single examination a true comparison of candidates' ability and knowledge based upon the marks gained for the paper is virtually impossible.

The standardisation of the degree of difficulty of examinations is an area that must give rise to concern. Even in the same subject areas there is the worry that some examination boards may be "softer" or "harder" than others.

Whilst the boards themselves make strenuous efforts to ensure comparability between syllabuses, the figures for the distribution of grades and numbers achieving the various grades would seem to indicate differences in difficulty. To form a valid opinion one way or the other is extremely difficult for we are not privy to the composition of the candidates taking particular syllabuses. For example, if one particular examination syllabus attracted a particular type of candidate or teachers chose to enter weaker candidates for a particular syllabus then it would be unrealistic to expect the distribution of grades awarded to be the same as a syllabus attracting the full range of candidates. Simply looking at the distribution of grades and noting differences in the percentage of candidates who took each syllabus achieving the top grades is not sufficient evidence to pass comment on the relative difficulty of the syllabuses or the achievement of pupils with regard to what might have been expected of them.

Certainly Her Majesty's Inspectorate of schools (HMI) were concerned about the differences in moderation between boards and expressed their concern in their report on standards in the 1992 examinations, "HMI have limited confidence that standards are being maintained; confidence would be more secure if the criteria for awarding were more objective and the procedures used across groups were more consistently rigorous" (DFE, 1992a).

HMI also called for the maintenance of standards by comparing examination papers with those of previous years, "grading is least satisfactory where little use is made of scripts from previous years, where the views of experienced examiners are marginalised, or where there is too little consideration of year-to-year and syllabus-by-syllabus statistical comparisons" (DFE, 1992a).

This criticism was acted upon by the examination boards who undertook to implement procedures to improve standardisation.

A further area of variation in examination standards is in the levels of difficulty of different subject areas. The difficulties of ensuring comparability are discussed by Goldstein (1982) who demonstrates that the number of variables involved in making any assessment of the worth of a pupil's work are so numerous as to make the attempt almost futile. Mortimore and Byford (1981) refer to the Schools' Council form on comparability (1979) which

"concluded that grades should be considered plus or minus the next grade. Thus, a pupil obtaining a grade B result should accept the possibility that his paper might, in fact, have been worth a grade A or C."

The consequences of such a margin of error are potentially compounded, unless the errors operate both up and down, when individual results are combined to give a pupil score, required for entry to University, or further compiled to give an overall school average points score to use for comparing schools.

Torrance (1986) states that the examination boards themselves are not unaware of the problems concerning the validity and reliability of awarding grades. He refers to a paper produced by the Joint Matriculation Board (1983) where they comment upon the 1982 results and show that a minimum of 16 marks out of 196 separated grade D from B in English Literature. They go on to state, "The percentage marks required to take a candidate from grade D to grade B ranged from 2.9 (of a maximum of 340) for Chemistry syllabus A to 9.5 (of a maximum of 200) for Sociology." Clearly the very smallest of errors on behalf of the pupil or marker could result in a very different outcome as judged by the final grade awarded.

More recent work looking at the standards in public examinations has been commissioned by the Schools Curriculum and Assessment Authority (SCAA). Published in 1996 this work set out to look at standards in public examinations at GCSE and A level in English, Mathematics and Chemistry from 1975 to 1995. The report defines its scope in the introduction,

"Standards' in public examinations can be considered in two senses:

- **examination standards** refer to the demands of syllabuses and their assessment arrangements, and the levels of performance required of candidates to gain particular grades;
- **standards of attainment** refer to the knowledge, understanding and skills that candidates demonstrate in examinations.

Examination results depend on standards in both senses. If the examination standards remain unchanged, and candidates' standards of attainment improve from year to year, then examination results will be better. However, results would also seem to improve if standards of attainment do not rise but examination standards fall, or if standards of attainment fall but examination standards fall further.

The comparison of examination results over time can, therefore, prompt questions about examination standards, but cannot in itself provide evidence of a change in them." (SCAA, 1996b)

Here, then, is a dilemma in that consideration of examination results demands consistency of examination standards. To establish consistency of examination standards requires consistency of pupil attainment which in turn can only be established in relation to reliable examination standards. The report goes on to attempt to look at examination standards, considering the demands of syllabuses and examinations and the direct evidence of pupils' work.

This proved very difficult as examination boards were unable to provide many candidate scripts, if any in some cases, from a number of years previously for comparison with scripts of candidates in more recent examinations.

Also, the nature of the syllabuses has changed over the years, both in content and approach. Different forms of assessment are used now with less reliance on terminal assessment. Coursework and open-book examinations are much more prevalent as are modular examinations at A level in particular.

With such huge changes as these it was virtually impossible to make any meaningful comment on the changes in examination standards beyond general statements on the changed emphasis and content of the examinations. General recommendations were made that the examination boards should set up an archive of examination scripts so that these may be referred to in future.

Without being able to establish a stable baseline, consistent examination standards, against which to measure pupil examination attainment over time then reports such as SCAA (1996) and SCAA (1996a) looking at GCSE and GCE results and trends over time can show no more than that - the results obtained and the prevailing trends in those results. They certainly cannot provide definitive proof of rising standards of pupil attainment. The annual controversy in the media over this question will no doubt continue to be aired.

That the validity and reliability of examination results can be questioned would be taken by some as evidence of the futility of using examination results for

any sort of comparative measure. Yet we live in the real world and whilst our measures may not be perfect we do need measures in order to differentiate. That there are doubts about the validity and reliability should make us more cautious in the use of those measures but not cause us to dismiss them entirely.

There are consequences for this research. The use of average grades, rather than relying upon a single examination subject grade, is likely to help avoid the extreme effects of margins of error. The importance attached to any decisions based upon differences in examination outcome should be tempered by the knowledge of the flaws in the system; the more important the decision the more pressing the need for corroboration from other evidence. Where subject results are going to be compared it is more useful to compare a school department's results with other departments using the same syllabus and examination board, not merely the same subject area, where possible. Maintaining some concept of improvement over time will be difficult when the measuring instrument, the standard of work required for the examination grades awarded, may well be changing.

Improvement over time

Gray *et al.* (1996) sought to look at changes and improvements in school effectiveness over a five year period, 1990 - 1994. Using prior reading test information on pupils from over 30 schools in one Local Education Authority they considered the examination results of five cohorts of pupils. Over the five year period 14,965 pupils were included in the sample having both indicator data and GCSE examination results. The prior indicator data mean score for the sample remained largely the same over the length of the study as did the proportion of males and females indicating a relatively stable population. Other measures mentioned by Gray *et al.* were percentage of pupils on free school meals and percentage of pupils in receipt of educational statements. The figures for these two measures both increased over the period 1991 - 1994, figures for 1990 not being available.

As Gray *et al.* point out, this survey covers pupils' educational experience over almost a decade, the first cohort of pupils having entered secondary school in 1985 and the last cohort taking their GCSEs in 1994. For this reason I consider the figures on free school meals and pupils with statements to be very suspect measures of pupil background. Over this period of time the changes in regulations regarding Family Income supplement and those entitled to Free school meals changed so that variation in the numbers claiming this benefit are unlikely to reflect truly any changes in the nature of the pupil sample. Also, as has been pointed out by Nuttall, Thomas and Goldstein (1992), free school meals as a measure of social deprivation is a flawed measure in that those most in need of free school meals do not always claim them and hence do not appear in the figures.

Thomas and Mortimore (1996) also find the free school meals indicator potentially misleading when fuller information on pupil backgrounds, such as parental education and occupation, is available, "When rich and wide ranging pupil level data (for example prior attainment measures in different areas) are available and taken into account in the analysis, school context factors (such as percentage of pupils entitled to free school meals) are not significant in predicting pupil outcomes."

Aware of the faults of free school meals as an indicator contributing to the study of school effectiveness, in particular its potential for misleading conclusions because of incomplete data, the limited extra information it could contribute when there was a good prior attainment indicator available, and mindful of the extra workload producing the information would impose on participating schools I chose to avoid this sort of measure in my research data.

The percentage of the school year cohort with statements of educational need is also a suspect measure in that it is very dependent upon the schools' special

needs department and its efficiency, the school policy on special needs and the LEA policy with regard to those pupils requiring statements, all of which are likely to have changed over a ten year period. Parents too are now more aware of their rights and their children's educational entitlement (DFEE, 1994 and 1995). They are far more ready to pursue their rights and seek statementing than ten years ago. Increasing financial constraints have meant that some authorities have dragged their feet in statementing pupils because of the legal obligation to provide the resourcing these pupils require. There is also the point that pupils may be statemented because they are exceptionally bright rather than less able so the picture of a school painted by this particular figure is not clear.

Thomas and Mortimore (1996) make the point that once prior attainment and other disadvantage factors had been taken into account they found no difference in the performance of pupils with statements from those without.

That this measure too seems flawed in its contribution to the explanation of differential pupil performance in school examinations, particularly when a measure of prior attainment, which in itself would tend to reflect those 'disadvantage factors' operating upon a child's chances of success, seemed yet another good reason to maintain the simplicity of my research data and restrict intake measures to evidence of prior attainment.

Gray *et al.* make the point that schools identified in their research as improving more than most, in terms of average examination points per pupil, had also increased their examination entries per pupil. The initial average examination entry per pupil figure was 7.4 in 1990 and rose to 8.6 in 1994 but the majority of this increase was made in 1991. What is not made clear is whether this increase came about because schools actively sought to improve their examination statistics, particularly with the advent of national performance tables, or were increasing their subjects per pupil entry figures because of the National Curriculum requirements. Rather than the increase in entries

accounting for better improvement rates in some schools, might the increase not be more indicative of schools which have made a conscious decision to change and seek improvement, the rise in examination entries per pupil being indicative of schools seeking to improve?

A problem for these and all researchers in trying to discover what makes some schools more effective than others is the rarity of schools whose performance is significantly better than average. Even in quite large samples, such as that of Gray *et al.* , the majority of schools are performing largely in line with expectations and only a very few are performing significantly above expectations. When there are so few it is very difficult to have statistical confidence in any deductions made from the data. Because these schools are exceptional it is also difficult to ascertain why they perform so well. Linear studies of schools' performance over time are also in short supply in the United Kingdom, Gray's research being a notable exception.

It is clearly important that examination data for schools should not be looked at in isolation from other schools or from other years. My research would have to consider a number of schools over as long a period as possible in order to ascertain any trends in performance that might be apparent. Not only that, but judgements on effectiveness or otherwise of schools must be considered in the light of their performance being significantly better or worse than might be expected from the average school.

Effective schools or effective parts?

Even if one were to restrict one's comparisons to apparently similar schools there are so many variables to take into account - the size of the school, the age of the school and any tradition for excellence, the stability and qualifications of the teaching staff, the general state of repair of the buildings and fabric of the school, the resourcing and general finances of the school, the social and racial background of the catchment area and so on.

Jesson *et al.* (1992) looked at the performance of schools in Nottinghamshire having collected considerable social data on the pupils but lacking any prior test data. His team found that school performance was broadly in line with the social makeup of the pupil intake. Nuttall *et al.* (1992) commented upon the fact that using multi-level modelling information on social background can contribute to a greater understanding of examination results, particularly if prior test information is unavailable or LEAs without a common test can share information likely to affect the performance of pupils in their schools. More recently Thomas and Mortimore (1996) looked at the 1993 GCSE results of pupils in schools in Lancashire using multilevel analysis to include the effects of measures of prior attainment and school background factors taken from census data.

Importantly this study by Thomas and Mortimore found evidence of some schools being more or less effective with pupils of different ability groups. Within schools too, they found there was evidence that some departments were more effective than others. Their evidence for this were the low correlations within schools between the grades obtained by the same pupils in different subject departments and the disparate value-added scores for different departments. This was not gone into in great detail. I would have liked to have seen some more detail, particularly on subject area performance between schools comparing like with like, subject areas and pupil ability.

This whole area of variable performance by different subject groups within schools needs greater discussion, particularly as schools tend to be compared on measures of overall effectiveness rather than sub-group performance in Government performance tables, media league tables and public awareness,

"Overall, these findings show that on its own an overall measure of school effectiveness may be of limited value in some schools and may mask substantial differences between departments, groups of pupils or

both." Thomas and Mortimore, 1996

Fitz-Gibbon (1990) argues strongly that all-encompassing figures such as percentage pass rates are misleading when used as performance indicators, and points clearly to the inherent problem areas,

"No account is taken of the kinds of pupils entered...

... Barely passing is counted in the same way as getting a high grade...

... Adding up passes in all subjects loses information...

... A distorting effect on educational practice may occur...

... The percentage pass rate ignores the curriculum balance...

... Percentage of what?"

Fitz-Gibbon's first point is that any judgement about the performance of pupils and / or school should be based upon what could be reasonably expected of pupils of a given ability and a percentage pass rate figure alone does not give any information regarding performance in relation to expectations.

Accepting this, I would suggest that percentage pass rates being presented along with an indication of pupil group ability, a prior test score for example, would improve the information conveyed considerably, particularly if percentage pass rates were quoted for different ability bandings.

Fitz-Gibbons's second point reflects the problem of all "hurdle type" measures in that only the fact that a pupil cleared the hurdle or did not is recorded, not the quality of the clearance. Just "failing" a GCSE examination by gaining a D grade is counted in the same way as obtaining an F or G grade: the focus is on the pass or fail rather than the quality of the grade obtained. I would agree that simply adding all the grades to produce a global figure for the school, Fitz-Gibbon's third point, does conceal the performance of individual departments, and this distinction is vital if a school is to identify areas of strength and weakness in order to plan effectively.

Measures of overall school performance are temptingly easy to compare, however, and are likely to remain. For closer scrutiny it is essential to have access to both the aggregated and disaggregated data if one is to look at the performance of individual children or particular groupings of children. The potential, pointed out by Fitz-Gibbon, for distortion of the educational processes if schools were to pursue certain strategies for improving their examination statistics is becoming a reality. If schools were to avoid entering all but the most able pupils for what were perceived to be the most difficult subjects there would develop a system of curricular provision for the elite and another for the rest. The range of subject options available to pupils, already curtailed by the imposition of the National Curriculum which was intended to ensure a minimum entitlement to curricular provision, would be restricted. Only those pupils who were considered racing certainties to gain high grades in examinations would be allowed to enter any but the most compulsory subject areas.

In Somerset, since the introduction of performance tables based on percentages of pupils achieving five or more GCSEs at grade C or above, Physics, Biology and Chemistry GCSE syllabuses have ceased to be taken by pupils in state schools. All pupils are now entered for either the Double Science syllabus for which two grades are awarded or the Single Science syllabus. The three separate sciences were regarded by teachers as too difficult and taking too much time to teach in a curriculum that offered too little time. It was therefore a combination of National Curriculum time pressures, the fact that teachers considered Physics, Chemistry and Biology GCSEs more difficult than other science GCSEs and the push to gain higher grade GCSEs at the expense of subject content.

Fitz-Gibbon makes the further point that only allowing those pupils to sit the examination whom the school thinks are sure to pass is a self-defeating mechanism for,

"In what are to a large extent norm-referenced examinations this tactic could not succeed if adopted by everyone and is therefore a tactic to be discouraged. At A-levels for example, about 30% are predictably going to 'fail' so it is reasonable to expect most institutions with average intakes to take on a sacrificial 30%. If percentage pass rates are used, then A-level institutions with higher than a 70% pass rate should be asked if they are pulling their weight in this respect, not praised!" (Fitz-Gibbon 1990).

There has of course been much debate about whether A level examinations are norm-referenced in terms of calculating the numbers of candidates who will pass and at what grades. The examination boards deny that there is a significant norm-referenced element at A level and GCSE examinations are predominantly criterion referenced, all be it to criteria judged to be 'normal' for 16 year olds.

The Audit Commission (1993) report, commissioned by the Government inspection body, OFSTED, would certainly take issue with Fitz-Gibbon's comments about taking on a "sacrificial 30%" of students who are going to fail. The report is most concerned that schools and colleges consider the financial cost of allowing students to undertake courses for which they are not suited and may not even complete, never mind pass.

The percentage pass rate for GCSE examinations has increased each year since its introduction. In 1992 the results had scarcely been published when the Secretary of State for Education pronounced that he believed examinations were getting too easy and that he would therefore instigate an immediate inquiry into standards. It is a sad comment on our society that we lament the fact that so many of our pupils apparently fail examinations and yet, when the numbers of those passing increase, we deny them the praise they deserve for meeting our published criteria and instead claim that standards must be falling. In order to make judgements on standards it is necessary to have a reliable

measure and know from which point you are starting to measure. To ascertain performance one needs to establish a reliable baseline measurement and this will be a crucial element in my research.

In claiming that percentage pass rates ignore the curriculum balance, taking no account of how many "easy" or "difficult" subjects were attempted, Fitz-Gibbon refers to evidence, although she does not say what, which suggests that a D in A-level mathematics is about as academically difficult to get as a B in A-level English, "A school or college which, by attracting many candidates into mathematics, helps to ameliorate the national problem of a shortage of numerate persons might thus find itself at a disadvantage if percentage pass rates were examined." (Fitz-Gibbon, 1990).

Subject comparability, highlighted here, is an area that must be considered in looking at school performance indicators and overall school effectiveness. Percentage pass rates on their own do not help in illuminating this issue which I will return to later in my research.

By asking "Percentage of what?" Fitz-Gibbon makes a plea for clarity of information so that it is perfectly obvious exactly which candidates' results are included and which are not. To a large extent this concern has been addressed by the DFE regulations concerning the reporting of examination results.

Fitz-Gibbon's comments on percentage pass rates echo many of my own views and experience prior to starting this research, and these, added to the value-added issue, highlight difficulties of interpretation. Percentage pass rates as indicators of examination performance fail to show any measure of progress made by the individuals and this makes them of limited value for addressing many school effectiveness issues. There is no indication of the level from which a candidate started and so it is perfectly possible for the greatest achievement and improvement to have been made by a candidate who just "failed". No account is taken of the kinds of pupil who were entered for

examination, their intellectual, social or even financial background, all of which factors have an influence upon their potential performance in examinations.

Variable performance within the school unit

Fitz-Gibbon goes to great lengths to emphasise the importance of the subject department within the school for this is one level at which action can be taken to improve performance in examination results. Most schools have departments whose results are a credit to their schools and other departments whose results are less so. I myself, in analysing the 1992 GCSE results of Somerset schools, can say that without exception each school had departments whose results for the pupils in their care were outstanding, when in some cases the overall statistics for the school were not. Nor is it a simple matter of comparing the English department's results with the English department in another school, for the ability of the groups within subject areas within schools varied quite dramatically between schools. I repeat, to aggregate the results for a school is to lose the distinction necessary to offer praise where it is due and help where necessary.

Schools are the sum of their parts and these are constantly changing year on year, not only the pupils but the staff as well. Even the staff who remain in a school for a length of time are not equally effective with each examination class they oversee. The consistency necessary to establish standards and measure improvement, as mentioned earlier in this chapter, is very difficult to obtain. The pupil intake each year varies in ability, social mix, general attitude and number which in turn affects resourcing. Peer pressure, so influential on the young for good or bad, cannot be easily quantified. As teachers are well aware, some years simply do not gel and fail to reach the standard they are perceived as being capable of. The "collective properties of a pupil body have an effect on pupil achievement over and above the effects of individual pupil characteristics", (Willms, 1985, quoted in Gray, Jesson and Jones, 1986.)

Blakey and Heath (1992), reporting preliminary findings from the Oxford University School Effectiveness Project, take the point a little further. Whereas research into individual pupils' social grouping is now prevalent in considering pupil performance in examinations, Blakey and Heath stress the importance of the social mix of the school as a whole, governed largely by its catchment area, and its influence on the individual. Using questionnaire information on parental occupation and frequency of attending school functions such as parents' meetings, Blakey and Heath looked at the social context of the whole year cohort as well as that of the individual pupil in relation to examination expectations. They found that the social context of the year group helped explain more of the variance in examination results between schools than pupil prior ability and individual pupil background alone.

"In explaining a pupil's performance, then, we need to take account both of his or her own social class background and of the background of the other children in the school. That is to say, we need both individual - level and school - level measures of social background." (Blakey and Heath, 1992)

No league table or report on school examination results takes this into account.

Some factors causing variation in performance

Fitz-Gibbon (1992) talks about the futility of searching for school effectiveness measures when she finds little stability if any in the performance of schools at A level from year to year. Willms and Raudenbush (1989) emphasise the need for longitudinal study of school effects noting how unreliable performance indicators can be because of sampling variability, "If we are to use performance indicators for monitoring school effectiveness, we need to distinguish between variability amongst schools in their stable components of performance, variability in their true changes in performance, and variability associated with sampling error."

They go on to discuss the changes in a school's performance attributable to changes in the local economy of the catchment area of the school such as rising unemployment which might change, as it has done recently in Britain, the pupils' attitude to staying on at school. "Changes in school performance from year to year can also be attributable to the effects of specific school policies or to changes in a school's social class and ability composition."

We might also add the effects of Government initiatives to this list of variables. Into the farrago of factors affecting a school's ability to help its pupils obtain the best possible examination results are thrown new assessment techniques, changes to syllabus provision, changes to curricular provision and examination entry policy and, perhaps most important of all of these, the ability of the school to cope with change.

Finance too must play its part. A school with inadequate funding cannot provide the breadth of resources necessary to compensate for the shortfall in educational resources in less well endowed homes. Official statistics on LEA spending per capita of school population are very misleading for it would appear that some authorities, such as North Yorkshire, achieve good results on a very low outlay whereas others, such as Knowsley, have the opposite situation (Times Educational Supplement, March 31, 1995). Without comparable figures on the social make-up of the LEAs these spending figures are almost meaningless.

Nor can there be savings made by entering pupils for fewer examinations for Blakey and Heath (1992) have found that schools with generous entry policies do tend to achieve better results than schools with comparable pupils but a more restrictive entry policy. There was no evidence of pupils gaining higher grades by doing fewer subjects. It is scarcely surprising that Blakey and Heath came to this conclusion when the measure they were using for academic success was the average examination points per pupil in the various schools.

As was noted earlier in this thesis looking at improvement over time, Gray *et al.* (1996) also found pupils were gaining higher average points scores in schools with more generous entry policies, but again it is hardly surprising that where pupils sit more examinations they will tend to gain higher average points per pupil scores for the school. Another measure which could have been used would have been average points per entry, then the two measures, average points per pupil and average points per examination entry could have been compared. Greater detail is needed to see whether the more able pupils are being entered for more examinations to maximise their impact upon the statistics or whether the increased entry policy applies to all candidates regardless of ability.

In my research figures are given for average entries per candidate, average points per entry and average points per candidate.

Gender differences

One area that is not included within the performance tables is the relative performance of the different genders in school examination results. Given similar abilities the question of whether girls and boys perform better in single sex or co-educational schools, for example, cannot be addressed from the tables. The information provided by the obligatory publication of school examination results in the DFEE performance tables does not allow such issues to be addressed, merely providing bald figures which limit interpretation and are subject to misinterpretation by the lay person / parent.

The proponents of the Girls' Schools' Association would certainly put the case for their sector of education but Jane Steedman (1984) showed that there was very little, if any, difference in outcome between mixed and single-sex schools. What differences there were could generally be attributable to different intake standards, although there was a slight overall advantage for girls in girls' schools, particularly bright girls, over boys in boys schools or girls in mixed schools. There were some differences in the performance of girls and boys in

particular subject areas, for instance girls tended to do better in French and English whereas boys tended to gain higher grades in Mathematics and sciences such as Physics and Chemistry.

So, even where the results were apparently higher for girls and particularly girls in girls' only schools the reality was far more complicated and less clear cut. This is another example of the potentially misleading nature of indicators without skilled interpretation.

Steedman does not give the exact size of the sample for the data she was using but as her information came from the National Child Development Study data, information being collected about all people born in Britain during a particular week in 1958, it was large and comprehensive. She was also fortunate in having access to extensive information about educational tests, applied at ages 7 and 11 as well as O level and CSE grade 1 results, parental occupations and types of school attended.

The research, although dated in terms of current performances of boys and girls, indicates the complexity of the simple question of relative gender performance, for example taking account not only of the prior ability of the students, the social background of the families but attempting to make allowance for different school types. However the actual results must be considered somewhat dated now, for although the research was published in 1984 it was actually concerned with examination results gained in 1974 since when there appears to have been a marked change in the performance of the two genders.

That the data allowed a longitudinal study of pupil progress on such a large scale is impressive but is limited in that the study could only look at one year's examination results. The data were not available to repeat the exercise with successive cohorts of pupils. Nevertheless, Steedman's work does emphasise the complexity of an issue which to many outside the research community

would seem misleadingly straightforward.

The work of Hedger and Raleigh (1990) as reported in a paper concerning GCSE results in Shropshire particularly highlights the differential attainment by gender. With the exception of Mathematics, Hedger and Raleigh found girls to be out-performing boys in every National Curriculum subject area. In English in particular the average results for girls throughout the LEA were almost a complete grade higher than the boys. Furthermore, 53% of girls achieved a GCSE grade A-C in English compared to 32% of boys.

This trend for girls to out-perform boys in GCSE examinations has a distorting effect upon the overall results for mixed schools, particularly comprehensives, when compared to those of single sex schools.

Hedger and Raleigh's findings would suggest that the more girls in a particular year cohort the more likely it is that the school's results will be higher than if boys predominated in the year. Single sex schools tend to be selective in their intake and this would tend to ameliorate the situation for boys' grammar schools, their results being high because of the high ability of their intake.

This view is supported by the work of O'Donoghue *et al.* (1997) in their large scale study of 1996 A level results using a multi-level modelling approach,

"When both types of entry policy (gender and selective / non-selective) are included together in the fixed part, in Model 12, a different story emerges. Progress by students in single sex schools are now no longer significantly different from students' progress in mixed schools: differential progress is dependent on whether students are in selective schools or not. The single sex school factor therefore seems to be accounted for by admissions policy; and most single sex schools are selective."

Other variables such as the relative abilities of the boys and girls in co-educational comprehensive schools and the distribution of pupil ability within

the gender groups will also play their part in the school's overall performance in terms of GCSE results.

I intend to explore these issues further in this thesis.

The School Curriculum and Assessment Authority (SCAA) in their analysis of the 1995 GCSE results (SCAA 1996) found a large difference in the performance of girls and boys which has been increasing since the introduction of GCSE examinations in 1988, "In 1995, there was a difference in the grade profiles for boys and girls, with just over 57% of girls being awarded A*-C grades, compared with just under 49% of boys..."

The SCAA report goes on to consider the changes in the proportion of the Year 11 cohort in England achieving five or more A*-C grades from 1975 to 1995, by gender,

"The increase since 1988 (the year of the introduction of GCSE examinations) is greater for girls than it is for boys. In the final year of GCE 'O' level and CSE (1987), the difference in favour of girls was 1.6%. By 1990 it had increased to 7.6%, and by 1995 to 9.1%."

At A level the difference in overall performance between girls and boys is also marked. In 1989 the percentage of girls in the 17 year old cohort achieving two or more A level passes overtook the figure for boys for the first time (14.9% compared to 14.8%). This difference in the performance of the two genders reached a peak in 1992 when 21.8% of girls in the 17 year old cohort achieved two or more A level passes compared to 16.8% of boys. Since then there has been approximately a 2% difference in favour of girls with the figures for 1995 being 21% for girls and 18.9% for boys (SCAA 1996a).

A natural conclusion would be that at GCSE and A level the more girls in your school the better your overall examination statistics are going to be and yet the Government tables on performance do not warn parents of this fact nor do they

break the statistics down by gender so that this differential performance is readily apparent.

Who is responsible?

In reporting examination results, be they successes or failures, there is the question of whose responsibility they are - pupils, parents, teachers or schools? Mortimore and Byford (1981) reflect upon exactly whose attainment is being discussed when one considers school examination results,

"Some commentators on examination results, such as university admission tutors and personnel managers, may see the attainment, or lack of it, as being the direct responsibility of the candidates. Others, including LEA administrators and inspectors, may see the attainment as being the direct responsibility of the school. Yet another group, for example Her Majesty's Inspectors and educational journalists, may see the attainment as principally reflecting the work of the LEA."

Mortimore and Byford tend towards the view that the examination results must be a product of all three contributions.

This is an important point and should not be dismissed lightly when looking at school improvement. It is to the pupils that examination results and certificates are given and it is they who carry forward their achievements to the next stage of their lives. Yet, rightly or wrongly, teachers, schools and LEAs are judged on the performance of the pupils. This begs the question as to whether researchers into school effectiveness should not rather be looking for schools which allow their pupils to be more effective in gaining their desired outcomes, which may not be examination results at all.

Effective schools are those with effective pupils.

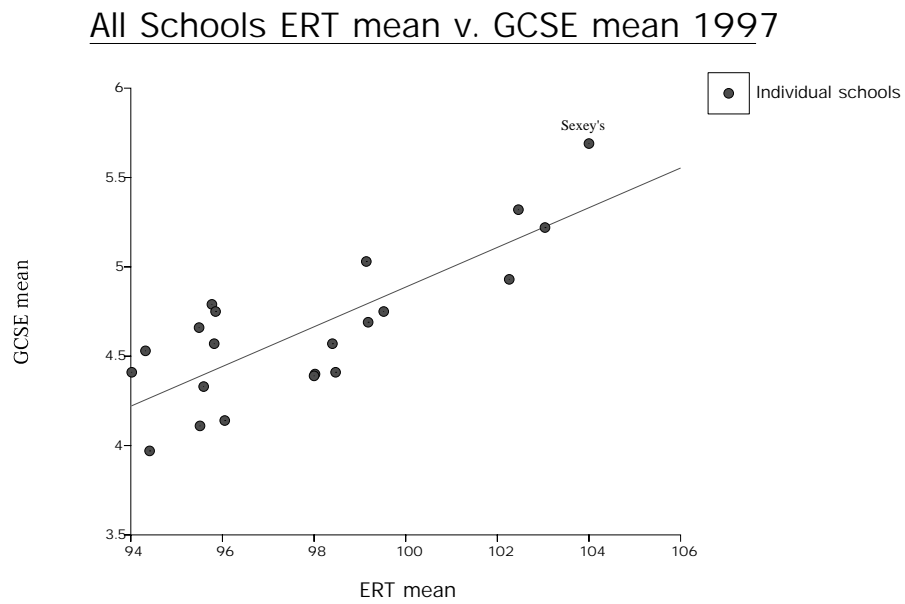
Problems with statistical analysis

Much of the work that has been done on analysing examination results depends upon the use of regression analysis, a technique not without its problems as Gray *et al.* (1986) point out. They comment that researchers into school effectiveness using quantitative data on examination results have tended to rely upon mean values for individual schools and regressed these figures on some aggregate measure of the school's intake. Comparisons are then made between schools but these can only be on the most general level for the fine detail is lost in the process of aggregation. Statements on school effectiveness must then be restricted to comments upon the average school effect rather than that for any given individual.

There is then also the problem that any regression analysis residual value can only be relative to the performance of other schools in the group and therefore subject to the influence of one or two possibly untypical schools.

By way of illustration I include a regression line graph (Figure 3.1) produced from the data for 21 schools in 1997. Sexey's school is indicated on the graph and exerts an influence on the slope of the line of best fit equal to any other of the schools indicated. This is despite the fact that Sexey's had but 25% of the pupil numbers of many of the schools also shown. Its influence, measured in terms of pupil numbers ought to be far less but the graph takes no account of pupil numbers merely comparing average ERT scores and average GCSE grade per pupil for each institution.

Figure 3.1



Gray's final comment upon what he calls the "mean - on - mean" approach is that it assumes that schools are equally effective across the ability range, an assumption which at best is questionable. As will be shown later, in my own work on the results of Somerset schools I have found considerable variation in the attainment of pupils of differing abilities both within and between schools. Some schools are very effective with less able pupils but less so with the more able, others are the reverse, but the situation is not often stable year to year and is very susceptible to changes in intake and numbers in relation to the school year population as a whole. There may be a case for plotting different regression graphs for different ability bandings within the schools.

Later in the same paper Gray and his fellow researchers highlight the problems with errors of estimates. Much of the inter-school variation in school effectiveness studies is very small in terms of residuals either side of a regression line. In fact much of the difference between schools lies within the

95% confidence intervals for the error of estimation and therefore it is unreliable to attribute any weight to the relative positions of the various schools. Far better then to restrict any comment about the relative effectiveness of schools to general comment regarding those which are close to the regression line showing what one would expect the average performance to be, either above or below, and those schools showing extremes of performance, again above or below the line.

More versus less information

There is an essential conflict in the analysis of examination results and that is, on the one hand, the need to provide ever more detail and information in order to understand better the meaning of those results and, on the other hand, the need to keep the analysis simple with key indicators so that parents, governors and staff can gain some grasp of what is going on. Broad strokes are needed to paint a picture that is not going to be caught out putting too much weight on small figures within error bandings, and yet a fine brush if the information essential for improvement is not to be lost.

In our search for indicators to explain outcomes, we must be careful not to equate such measures as intelligence with other correlates of examination success such as social class.

There has been much research published looking at examination results in the light of social factors; Gray *et al.* (1990) who consider parental social class, housing tenure and the number of siblings in their multilevel model; Sammons *et al.* (1993) where the fathers' occupations, entitlement to free school meals and racial background are the social measures taken account of; Thomas and Mortimore (1996) who used census data related to the pupils' home area in order to consider social status in relation to school examination results. These are but a few of the studies which have considered social status and its contribution to examination success. In all cases, where it has been available, information on prior academic achievement has been the strongest correlate

with academic success, social correlates offering a degree of fine tuning when used in a multilevel model.

That both intelligence and social class correlate with examination success does not imply any direct relationship between class and intelligence. Any such statement is ill advised both because of the lack of evidence and, within the school environment, because of the damage it does to pupil, parent and teacher morale. The key point arising from research, such as that just mentioned, is not any link between intelligence and social class but the potential for examination success which may or may not be realised because of social advantage or disadvantage.

For a school seeking indicators of pupil potential and a means of monitoring progress in relation to potential, the standardised test score offering a measure of prior ability, which will in itself already reflect some degree of social advantage or disadvantage, seems to offer a more straightforward method which I will explore in this thesis.

Some professional phobias

I hope I have been able to show that there are a great many factors exerting an influence upon examination results achieved by the individual pupils within schools and corporately by the schools themselves. The reporting of any single statistic purporting to show performance is a highly dubious practice. Even when using the mean GCSE result for individuals, as an indicator for potential at A level as do Fitz-Gibbon (1992), Gray *et al.* (1995) and DFE (1995a), or pupils' GCSE points as a measure of attainment as do Hedger and Raleigh (1990), Gray *et al.* (1996), Thomas and Mortimore (1996) amongst others, it is dangerous to neglect the constituent individual subject results which make up that average score. There is a conflict between the need to aggregate data in order to give a general indicator of potential and the need to consider the fine detail of the constituent parts of that general indicator in order to gain more

information on candidate suitability for particular courses or areas of weakness.

Other factors, such as home background, parental education, family income / deprivation, play their part in explaining a pupil's eventual examination results but the very number of factors that need to be considered is in danger of overwhelming the purpose of the task, a case of not being able to see the wood for the trees.

This very problem plays into the hands of those who would have us give up the search for indicators of performance in examination results.

Sammons *et al.* (1996) felt compelled to defend their approach to school effectiveness research against comments made by Hamilton (1996), "Hamilton claims that school effectiveness research is 'ethnocentric' and unconcerned with democracy, equal opportunities or social justice, which suggests that it ignores the powerful impact of socio-economic factors, gender and race." (Sammons *et al.* 1996).

Other comments on school effectiveness research attributed to Hamilton by Sammons *et al.* include, "social Darwinist eugenic rationale" and "ethnocentric pseudoscience".

Hamilton is not alone in being critical of school effectiveness research. Elliott (1996) is concerned that much of the research and its findings is fed, uncritically, to politicians and policy makers who then use it to suit their own ends. He sees the findings of school effectiveness research as, "a structure of coercive control, inasmuch as they appear to leave little room for the exercise of self-directed and autonomous thinking on the part of pupils within the learning process." He also regrets the lack of attention in school effectiveness research given to the efforts of individual teachers in their classrooms and feels unable to have any sympathy with the educational values which he believes the school effectiveness movement embodies.

Clearly there are those who harbour severe doubts about the research into school effectiveness and its motives. These doubts are not restricted to researchers but in my experience are held by some teachers too. We must get away from the various phobias people have about examination results; staff who feel it is they who are being judged rather than the pupils; that there are too many variables to make any objective judgement and, far from being the hard evidence one was seeking, examination results lead to subjective judgements; explaining away examination performance in terms of pre-test scores, social background, ethnicity and so on is to set limited expectations or create self-fulfilling prophecies; in effect one is searching for the Holy Grail, an essentially futile thing to do.

Whilst one can sympathise and understand the reasoning behind these views this attitude simply will not do. Staff will be judged on their pupils' examination results by their managers, their peers, their pupils, governors and parents. It is naive not to understand this and therefore how much more important it is that we seek ways of clarifying our performance as teachers in relation to that of our pupils.

As teachers we have a fear of being held accountable for events which are not fully under our control. Pupils are individuals and not the compliant "imbibers of knowledge" we might wish them to be. They too have pressures impinging upon them which affect their ability to learn, about which teachers can do very little. In presenting school effectiveness data one must be careful to acknowledge this and emphasise the role the data has to play in highlighting the factors which can and cannot be changed by teachers.

By searching for reasons behind pupil performance we can examine our practices, consider our techniques, skills and their appropriateness for the pupils in our care. There are many factors involved but far better to

acknowledge that than to be under some misapprehension. As professionals those of us involved in education must strive to identify the factors, make them widely known and therefore give people the opportunity to alter them, where possible, for the benefit of future generations.

There is always a degree of subjectivity involved in any judgement even though it is based upon purportedly hard quantitative evidence. If 50% of the milk has been removed from the bottle, is it half - full or half - empty? To seek to understand, even if that understanding seems far out of reach, is never a futile act. The process is helpful and purposeful, the alternative is apathy and despair.

The reading of the research literature reviewed in this chapter was essential to gain an understanding of the issues and complexity of both the data and its interpretation necessary for research in the area of school performance. This in turn has strengthened my view that the data I present should be at a level which is accessible to teachers and school managers but within the context of current school effectiveness research.

My methodology, discussed in the next chapter, must be clear to all its audiences, researcher and practitioner alike. Better use of the performance data available to schools, with a better understanding of the issues involved, should lead to a better educational experience for the pupils in our schools.

To conclude this chapter I include two quotes from Nuttall (1990) concerning indicators and indicator systems which I believe to be apposite and relevant to this research,

"To be an indicator, an educational statistic must also have a reference point against which it can be judged",

"Indicator systems must also produce information useful to the policy

community if they are to survive as publicly supported endeavours".

In order to judge performance we must know from which point our children's educational progress is starting and where we would ordinarily expect it to be by a given stage in the process. For any benefit to come of a system which purports to judge educational progress it must produce information that can be used by those in charge of the educational process to benefit those who are going through it, the pupils themselves.