**Issues related to correlation data**

Fundamental to this research, indeed to the whole field of school effectiveness research, is the issue of correlation - establishing a link between prior indicator data on pupil ability and examination outcome measures. If no link can be established then the prior information is not an indicator and cannot be used as a baseline against which to measure school or pupil performance. With no common baseline the effectiveness of different schools cannot be established. Correlation then is a *sine qua non* as far as school effectiveness research using quantitative data is concerned.

This is not to say that high correlation between indicator data and outcome data is synonymous with high effectiveness or low correlation with low effectiveness. A subject department which obtains for some pupils higher grades than expected, or predicted by correlating the results of a large sample of pupils against their indicator scores, the rest of its pupils performing in line with expectations, will have a lower correlation than a department where all the pupils perform as expected. It is the first department that is more effective, because its pupils gained some exceptionally good results, even though its correlation coefficient is lower. What defines a particularly effective department are the outcomes it achieves with the pupil material it has, not the correlation coefficient for its results.

In considering the relationship between pupil Edinburgh Reading Test results and GCSE mean grades at the level of individual schools I looked at eighteen different schools in 1996. These schools are, bar one, LEA controlled Comprehensive Co-educational Day schools, the exception being a Grant Maintained Comprehensive State Boarding School. All of these schools apart from one in Cumbria are in Somerset or the new authority of North Somerset. Other schools are involved in the analysis I conduct each year but these tend to use different tests, such as the NFER Cognitive Abilities Test for example, and

so I exclude these from my discussions of correlation here, there not being enough of them with a common baseline test.

I have examination data for schools from 1992 till 1996. Nine schools supplied both GCSE and Edinburgh Reading Test information in 1992 and 1993, twelve schools in 1994 and 1995, eighteen schools in 1996. Six schools supplied data for every year from 1992 to 1996. The correlation coefficients for all these schools and the years they were involved are given in *Appendix A pages 1 & 2*. The highest correlation produced using Pearson's Product Moment method was 0.82 but this was a school where the pupils had been tested using Edinburgh Reading Test within twelve months prior to their taking their GCSEs and one would therefore expect the correlation to be higher.

In 1996 of the eighteen schools using Edinburgh Reading Test, the highest correlation coefficient was 0.79 and the figure for the combined sample of eighteen schools was 0.73 . The lowest correlation was 0.45 for Sexey's School. This low correlation can be explained by the fact that two pupils were known to have misleading ERT scores and the small sample size of only 44 pupils. The smallness of the sample size means that these two pupils represent a much greater proportion of the school sample than two pupils in a school with a year cohort of 200. One boy was known to be much more able than his score indicated and one girl had improved consistently since her ERT. She was expected to do better in her examinations than her very low test score would have indicated. If these two candidates were excluded then the correlation coefficient rose to 0.60 . I am not proposing that these candidates' examination results be ignored by mentioning the much improved correlation if their results were excluded but simply pointing out how exceptional their achievements were. Only two other schools had correlation coefficients below 0.70 and these were 0.68 and 0.66 .

Spearman rank correlation was also applied to the data, as part of the procedure

for producing particular printouts of pupil rankings for schools, and this technique gave equally high correlation coefficients. There was never more than 0.05 difference between the Pearson and Spearman correlations and therefore no particular benefit in using one technique rather than the other to calculate the correlation coefficients.

The figures for the combined samples of all schools with ERT information, produced by correlating the ERT scores and GCSE average grades for all pupils from all the schools, over the four years from 1992 to 1996 are as in *Figure 7.1*.

**Figure 7.1**

## Correlation figures for Combined Schools

| *Year* | *Pearson* | *Spearman* | *Sample size* | *Standard error* |
|---|---|---|---|---|
| 1996 | 0.73 | 0.74 | 2834 | 1.04 |
| 1995 | 0.73 | 0.76 | 1630 | 1.03 |
| 1994 | 0.71 | 0.73 | 1489 | 1.08 |
| 1993 | 0.72 | 0.73 | 990 | 1.06 |
| 1992 | 0.71 | 0.71 | 1092 | 1.14 |

In total the results and test scores of some 6,035 pupils were analysed. Correlations were found to be typically in excess of 0.70 with standard errors of estimation of about one grade.

I also obtained the correlation figures for the whole of Somerset secondary school population ( *Figure 7.2* ), which subsumes some of the schools that submitted data directly to me. The calculations are done from my software used under licence by Somerset LEA.

**Figure 7.2**

## Somerset Local Education Authority

| *Year* | *Pearson's r* | *Sample size* | *Standard error* |
|---|---|---|---|
| 1996 | 0.73 | 4159 | 1.11 |
| 1995 | 0.73 | 4108 | 1.14 |
| 1994 | 0.74 | 3839 | 1.07 |

These correlation figures are remarkably similar to those I obtained with my smaller samples and individual schools. The degree of correlation seems very stable.

With reference to my initial aims in this thesis, the relationship between Edinburgh Reading Test results of pupils in Year 8 of their secondary school education and their GCSE results in Year 11 was established by the strong correlation figures I had found. With sample sizes in excess of 1000, as large as 2834 in the most recent combined schools' sample for 1996, ( in excess of 4000 for the Somerset LEA data ) these correlation values are very highly significant but must also be looked at in terms of their usefulness as indicators of the predictive validity of the ERT.

In seeking to establish the predictive validity of using ERT results as indicators of likely success at GCSE the correlation coefficient becomes, to quote Gronlund (1976), "a validity co-efficient", specifically a coefficient of 'predictive validity' with the criterion for success being the GCSE mean grades obtained by the pupils. The higher the validity coefficient (correlation) the more confident schools and their teachers can be that the ERT scores are indicative of future success at GCSE level and monitor pupil progress in accordance with expectations.

Some teachers would say there is the danger of teacher expectations of pupil performance, based on the previous performance of pupils with similar indicator scores, becoming self-fulfilling prophecies and possibly limiting the aspirations of pupils. The key to this problem is how it is approached.
Any examination results predicted from a point on a regression line represent a mean of the performances of pupils with that particular indicator score and therefore are subject to a standard error of prediction. If the standard error of prediction were one grade and a pupil were predicted a 'C' grade then there is 68% certainty that the actual score obtained will be within the range of a 'D'

grade to a 'B' grade and some pupils may exceed the predicted range, above or below.

The expression of the prediction as a range of grades makes the challenge less threatening and gives the pupil / teacher something to aim for. If pupils live up to teachers' expectations rather than down, as long as those expectations are not excessive, this is a far better course of action than allowing pupils who are more able than their current work would indicate to under-perform badly in the actual GCSE examinations in relation to other pupils with similar indicator scores.

School or departments need to consider their performance relative to a larger sample, such as the results of a number of schools combined, and check that the predicted range of grades from the larger sample is not too dissimilar from that of their own pupils at the level of the department and school. It would be wrong for a school / department that was generally more effective than the larger sample to lower their expectations. In such a scenario the school / department would be better advised to set targets based on its own data.

With correlation coefficients at the level of the school being typically around 0.70, squaring this value and multiplying by 100 to give the coefficient of determination as a percentage one arrives at a figure in the region of 50%. That is to say, 50% of the variance in the GCSE mean grades of the pupils could be attributed to variation in the pupils' ERT scores of some three years earlier. In 1996 only three schools had correlations lower than 0.70 and most were in excess of this value so more than 50% of the variance can be attributed to ERT scores.

With standard errors of estimation for each school population and for the larger combined schools' samples of around 1.0 this means that statistically one can expect 68% of the pupils to have GCSE mean grades within plus or minus one

grade of that predicted by their ERT score. When one considers that this range of error, plus or minus one GCSE grade, is not per examination subject but in the calculated mean taking into account all the GCSEs the pupils sat, this would seem to be a fairly accurate indicator of general attainment in GCSE examinations.

The predictive validity of ERT results is good but how good is a moot point and the importance of such a question depends upon the intended use of the predictor information. If we, as teachers, wanted to know exactly what our pupils are going to achieve overall in their GCSEs then we would need a perfect correlation of 1.00 and anything less would not fulfil our need, but this is not the intended usage; we want to gain some *indication* of pupil potential which can guide our efforts as teachers of young human beings and human beings, thankfully, are not 100% predictable.

Since a substantial correlation between Edinburgh Reading Test (ERT) results and GCSE examination mean grades has been established over the whole population, then a further aim of my research was to look at the possible link between ERT and individual subject areas with a view to understanding any link and using the information formatively. Again correlation would be used to test the strength of any link and whether the predictive validity of such a link could be established.

My first step was to look at the correlation for each subject area using the combined sample of pupils from all schools with ERT data. For example, in 1996 there were 2767 pupils in my sample of schools who sat GCSE English language. The correlation coefficient for ERT and English language GCSE grades, using the same points equivalent scale as before (A*=8, A=7, B=6, C=5, D=4, E=3, F=2, G=1), was 0.70 . The correlation coefficients for the various subjects in 1996, 1995 and 1994 are shown ( *Figure 7.3* ) along with the number of pupils taking the subjects and the standard error of estimation.

These figures and those for previous years are given in *Appendix B.*

**Figure 7.3**

**1996 GCSE and ERT score Correlations by Subject Area**

| | *1996* | | | *1995* | | | *1994* | | |
|---|---|---|---|---|---|---|---|---|---|
| ***Subject*** | ***r*** | ***n*** | **se** | **r** | **n** | **se** | **r** | **n** | **se** |
| Art | 0.35 | 1156 | 1.37 | 0.43 | 861 | 1.46 | 0.36 | 674 | 1.45 |
| Biology | | | | 0.58 | 36 | 1.02 | 0.34 | 94 | 1.15 |
| Business Stds. | 0.41 | 234 | 1.45 | 0.69 | 142 | 1.27 | | | |
| Chemistry | | | | 0.50 | 36 | 0.80 | 0.30 | 93 | 1.15 |
| Child Studies | 0.57 | 165 | 1.19 | 0.42 | 53 | 1.41 | 0.67 | 118 | 1.32 |
| Drama | 0.45 | 709 | 1.12 | 0.47 | 426 | 1.17 | 0.48 | 307 | 1.15 |
| Design Techn. | 0.48 | 827 | 1.70 | 0.36 | 720 | 1.36 | 0.42 | 254 | 1.86 |
| Electronics | 0.53 | 79 | 1.48 | | | | | | |
| English Lang. | 0.70 | 2767 | 1.03 | 0.69 | 1593 | 1.06 | 0.68 | 1457 | 1.14 |
| English Lit. | 0.61 | 2178 | 1.16 | 0.64 | 1381 | 1.11 | 0.64 | 1255 | 1.19 |
| Food | 0.52 | 530 | 1.22 | 0.52 | 336 | 1.36 | 0.43 | 326 | 1.48 |
| French | 0.63 | 1555 | 1.38 | 0.69 | 1089 | 1.35 | 0.72 | 954 | 1.28 |
| Geography | 0.66 | 1189 | 1.33 | 0.66 | 793 | 1.40 | 0.65 | 728 | 1.37 |
| German | 0.58 | 745 | 1.36 | 0.69 | 428 | 1.39 | 0.60 | 413 | 1.28 |
| History | 0.63 | 874 | 1.42 | 0.64 | 613 | 1.45 | 0.65 | 664 | 1.38 |
| Humanities | 0.66 | 536 | 1.41 | 0.71 | 306 | 1.25 | 0.63 | 146 | 1.26 |
| Info. Tech. | 0.29 | 299 | 1.29 | 0.44 | 72 | 1.16 | 0.78 | 33 | 1.22 |
| Italian | 0.73 | 30 | 0.99 | | | | | | |
| Maths | 0.69 | 2746 | 1.27 | 0.71 | 1567 | 1.27 | 0.69 | 1450 | 1.26 |
| Media Studies | 0.58 | 160 | 1.17 | 0.36 | 34 | 1.08 | 0.58 | 89 | 1.38 |
| Music | 0.41 | 276 | 1.53 | 0.54 | 182 | 1.47 | 0.55 | 127 | 1.57 |
| Physical Educ. | 0.59 | 346 | 1.20 | 0.40 | 195 | 1.39 | 0.59 | 54 | 1.34 |
| Physics | | | | 0.62 | 36 | 1.06 | 0.39 | 93 | 1.34 |
| Religion | 0.67 | 264 | 1.44 | 0.51 | 137 | 1.49 | 0.68 | 59 | 1.18 |
| Science Dbl | 0.68 | 2236 | 1.17 | 0.71 | 1279 | 1.11 | 0.64 | 1441 | 1.10 |
| Science Single | 0.65 | 465 | 1.16 | 0.75 | 273 | 1.04 | | | |
| Sociology | 0.60 | 129 | 1.18 | | | | 0.68 | 51 | 0.94 |
| Spanish | 0.62 | 164 | 1.46 | | | | | | |
| Statistics | 0.48 | 53 | 1.08 | | | | | | |

For many subjects, particularly those with large numbers of pupils taking them, the correlations are very significant. There is some variation in the exact correlation coefficients from year to year. The correlations for English Language and English Literature are always high, as might be expected when using a prior "reading" test, but correlation coefficients are also high in Maths, Double Science, Humanities, Geography and French to name but a few. Patently the Edinburgh Reading Test is a good indicator of academic success in more subjects than English and English Literature.

In subjects with smaller numbers being entered there tends to be more variation

in the correlation coefficient from year to year. Examples of this would be Media Studies and Information Technology as shown in *Figure 7.4*.

**Figure 7.4**

**Correlation coefficients for small subjects**

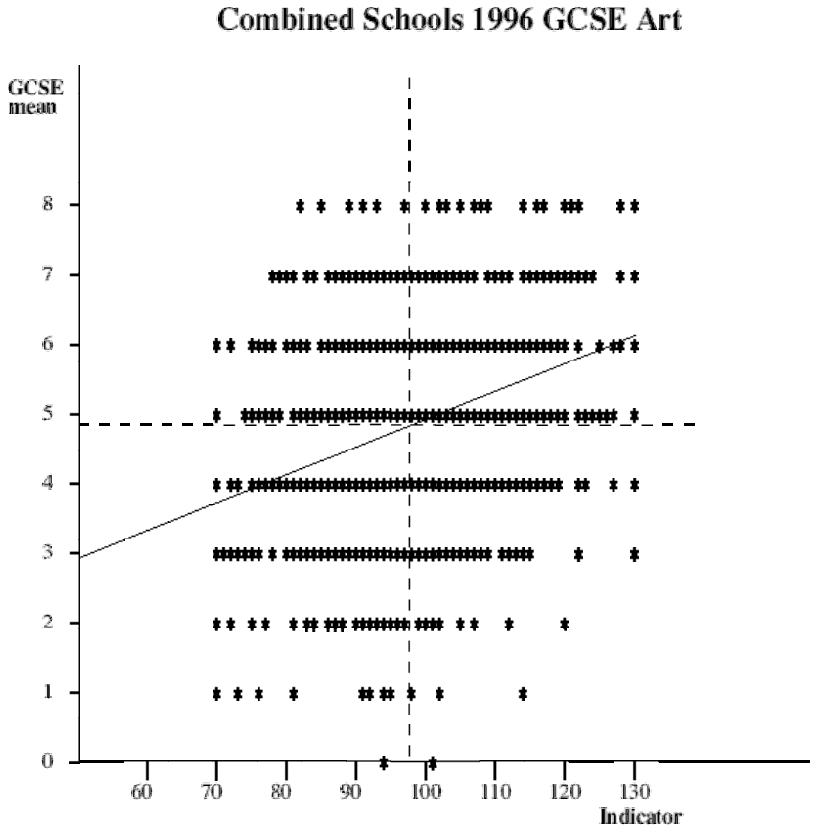
| | | *1996* | *1995* | *1994* |
|---|---|---|---|---|
| Media Studies | r | 0.58 | 0.36 | 0.58 |
| | n | 160 | 34 | 89 |
| Info Tech. | r | 0.29 | 0.44 | 0.78 |
| | n | 299 | 72 | 33 |

Some subjects, such as Drama, Design Technology and Music, consistently have relatively low correlations. This is likely to be because of the way these subjects are assessed, with a high degree of practical work and less dependence upon the written word, reading comprehension or sequencing skills. Reading comprehension and skimming through a passage of text for information are particular skills which the Edinburgh Reading Test (ERT) is designed to quantify and are important skills in subjects such as English, Maths, Science and Foreign Languages, not least because the medium of the examination is by way of a question paper which has to be read and understood before the specific subject skills can be assessed.

Looking at the results of individual pupils, some who have low ERT scores and in other subjects fail to achieve high grades can and do score the highest grades in subjects such as Art, Drama, Design Technology and Music because, according to their teachers, of their enjoyment of the subject, enthusiasm, special talents and so on. A typical example is the regression line and scatter graph for pupils who sat GCSE Art in 1996 shown in *Figure 7.5*.
As can be seen from the scattergraph, pupils with ERT scores of 70 were gaining GCSE Art grades ranging from G to B (1-6) and pupils with ERT scores of 130, at the opposite end of the range, were gaining grades from E to A* (3-8). A* grades were achieved by pupils with ERT scores ranging from

just over 80 to 130. This is reflected in the circular cloud like scatter, the low covariance and correlation coefficient.

**Figure 7.5**




| | | | |
|---|---|---|---|
| Number of pupils in the sample | 1156 | | |
| Mean for X is | 97.63 | Mean for Y is | 4.85 |
| Standard deviation for X | 12.71 | Standard deviation for Y | 1.46 |
| Covariance is | | 6.44 | |
| Coefficient of correlation | | 0.35 | |
| Coefficient of determination | | 12.02% | |
| Standard error of estimation for Y upon X | | 1.37 | |

The coefficient of determination would indicate that only just over 12% of the variation in GCSE grades is attributable to the variation in ERT scores. The standard error of estimation means that any grade predicted from ERT scores must be considered plus or minus 1.37 grades at the 68% certainty level. Pupils with low ERT scores and high grades, or vice versa, in subjects such as Art, Drama, Design Technology and Music, lower the correlation coefficient for these subjects. I must emphasise that the low correlation is not a reflection

upon the quality of the results or the teaching the candidates have received. In these subjects the Edinburgh Reading Test is not a good indicator of likely outcomes at GCSE level.

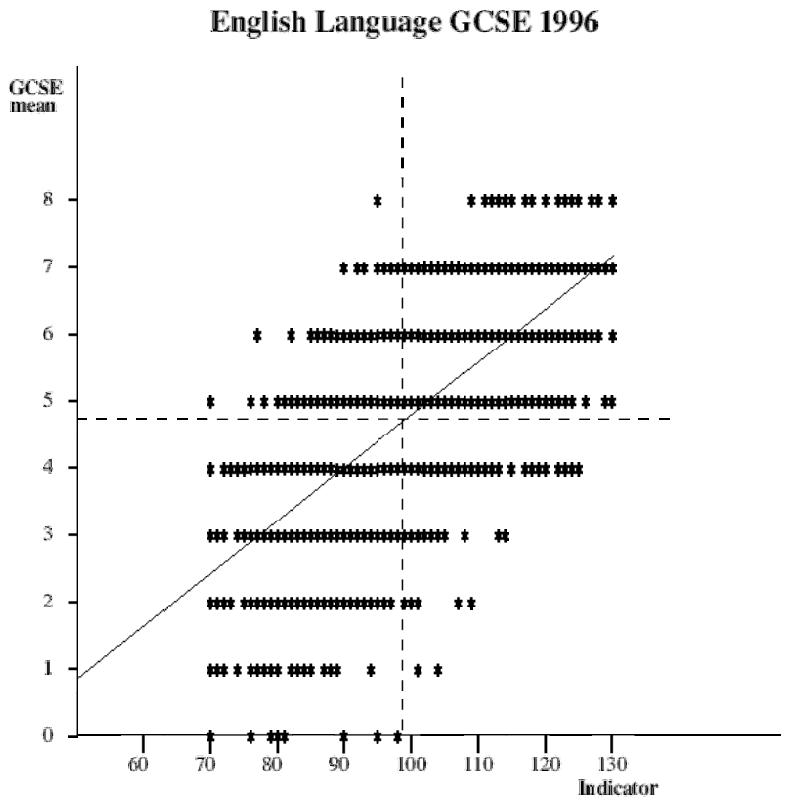
The standard errors of estimation for individual subject areas in general are slightly higher than for pupil combined subject mean scores and tend to be within the range of 1.1 to 1.4 . This in itself poses problems when attempting to predict a single subject grade for an individual pupil with a particular ERT score because the scale of grades at GCSE is not a continuous scale.

In practical terms it is not much use saying that 68% of pupils will be within 1.03 grades either side of the English Language mean grade for any particular ERT score when GCSE grades equate to whole units, despite teachers' fondness for predicting A/B rather than A or B. Pupils will be awarded one of the grades A*-G, not B plus ½.

The following graphic (*Figure 7.6*) shows how the scattergraph for an individual subject differs from one showing the average grade achieved across a basket of subjects. Note how although the ERT scale is continuous the GCSE scale is clearly banded because pupils cannot score half grades. Any statistical prediction must be converted to the nearest actual grade.

For the majority of subject areas when looking at the combined schools' population there is a strong correlation between GCSE and ERT although in a few subjects, as mentioned above, the correlation is considerably weaker. In subjects where there is a strong correlation then using ERT scores to indicate likely GCSE grades that will be achieved has predictive validity. Using ERT to predict GCSE grades in subjects such as Art is a much less valid exercise.

**Figure 7.6**




At the level of individual schools and their subject departments the correlations found at the combined school level persist but the generally much smaller numbers, particularly in subject areas with small entry numbers nationally anyway, do mean that the correlations are much more subject to the vagaries of individual pupil performances (See correlations for individual school subject departments in *Appendix B*) and any special conditions pertaining within the teaching of a particular subject. For instance, the degree of correlation, in subjects where the combined schools' sample indicates that there is a strong relationship between ERT and GCSE grade, can be affected:-

• Where the subject department within a school has so few candidates that the aberrant performance of just one or two candidates has a disproportionate effect upon the correlation figure. An example of this is highlighted in the case study on the French Department in Chapter 6 of this thesis.

• Where the subject is only taught to a restricted ability range. With the introduction of the Double Science and Single Science syllabuses at GCSE level increasingly few candidates are entered by schools for the Biology, Physics and Chemistry syllabuses. In 1996 no candidates were entered for these three subjects by any of the schools submitting examination results to me. Those candidates that were entered for these subjects tended to be selected on ability and therefore the ERT range of scores was restricted to the higher scores. Most of these pupils were likely to gain grades A, B or C and therefore the outcome range, the grades actually obtained at GCSE in these subject areas, was also restricted resulting in poor correlations.

The figures for the years 1993 - 1995 can be seen in *Figure 7.7* showing the percentage of the candidates who had ERT scores in excess of 100 and the percentage of candidates who achieved a grade in the range A-C.

(The A* grade was only introduced in 1995 and represents the top 50% of the A grade range.)

**Figure 7.7**

**Restricted intake and outcome ranges for Biology, Chemistry & Physics '93 -'95**

| | | Sample size | %ERT >100 | %GCSE A-C |
|---|---|---|---|---|
| Biology | 1995 | 36 | 74.0 | 80.6 |
| | 1994 | 93 | 80.6 | 68.8 |
| | 1993 | 191 | 71.1 | 51.3 |
| Chemistry | 1995 | 36 | 75.0 | 66.6 |
| | 1994 | 93 | 80.6 | 68.8 |
| | 1993 | 208 | 61.5 | 43.3 |
| Physics | 1995 | 36 | 75.0 | 63.9 |
| | 1994 | 93 | 78.5 | 59.2 |
| | 1993 | 202 | 56.4 | 51.5 |

At one time within the South Somerset area this situation also applied to German GCSE but, as it became common practice in these schools to offer German as an alternative to French as the mandatory foreign language to be studied under the National Curriculum, then the range of ability of the candidates broadened as did the range of their results and consequently the correlation increased.

• Where one looks at the correlation for teaching groups selected on ability

within subject areas in a school. Here the results of the candidates within the subject department as a whole may well correlate strongly with their ERT scores but when looking at the correlation for the pupils within any particular set it is likely that the correlation will be lower because of the restricted indicator and outcome ranges.

• Ceiling effects may also come into play with the indicator data, the outcome measure or both. That is to say, the most able pupils may well be capable of gaining higher indicator / outcome marks if the scale allowed for it but instead must be content with the highest mark available which is not necessarily indicative of their true ability. The same limitations operate at the bottom end of the scales as well. The discrimination between candidates at the extreme ends of the scales can be lost, limiting any correlation.

That the correlation, or coefficient of predictive validity, is often lower in these circumstances does not mean that the ERT score information is any less useful to us. For the sake of example, if we know that pupils with ERT scores in excess of 110 are going to gain A, B or C grades in Chemistry then that information is useful even though the correlation for such a selected group may be low and therefore the validity of using the ERT scores to discriminate between candidates within that particular group is flawed.

Mehrens and Lehmann (1984) summarise the problem well,

> "A paradox exists with respect to validity. In evaluating a test to determine whether it will assist in decision making, we want the test to have high validity coefficients on unselected groups. However, if we then use the test data to help make wise decisions, the validity coefficient among the selected individuals may be quite small. The more successful the test is as a selection device, the smaller will be the validity coefficient within the selected group provided that the proportion being selected is small."

They go on to say that if good and valid use is made of the test instrument for selection purposes, effectively weeding out those who are not going to succeed, then one also successfully reduces the validity coefficient for those remaining in the selected group. The original selection procedure remains a valid use of the indicator information.

Correlation is a useful tool in establishing the strength of a relationship between two variables, in this case the Edinburgh Reading Test and GCSE mean grade or, at the single subject level, the actual grade obtained by a pupil. In using the prior indicator information (ERT score) for predicting GCSE results the correlation coefficient can be used as a coefficient of predictive validity, the stronger the correlation the greater the validity of the exercise.

However, in the search for effective schools and subject departments it must be remembered that the search for high correlations is not the main aim of the exercise and correlation has its flaws.

In a small sample, such as a subject department, where some pupils have achieved results significantly above what would have been expected from their prior test scores and the remaining pupils have performed in accordance with expectations, based on a number of schools' results or a number of years' results in the same school, the correlation will be low and yet this would be a very effective department because all pupils achieved at least what was expected or better.

It is quite possible for a subject department's results to correlate very highly with the prior test scores of the pupils yet their results to be significantly below those that would have been expected in other departments or the average for that subject in the combined sample of schools. The same applies at the level of the individual school. The correlation coefficient is not a measure of success

or failure. It is a measure of the strength of the relationship between two sets of data.

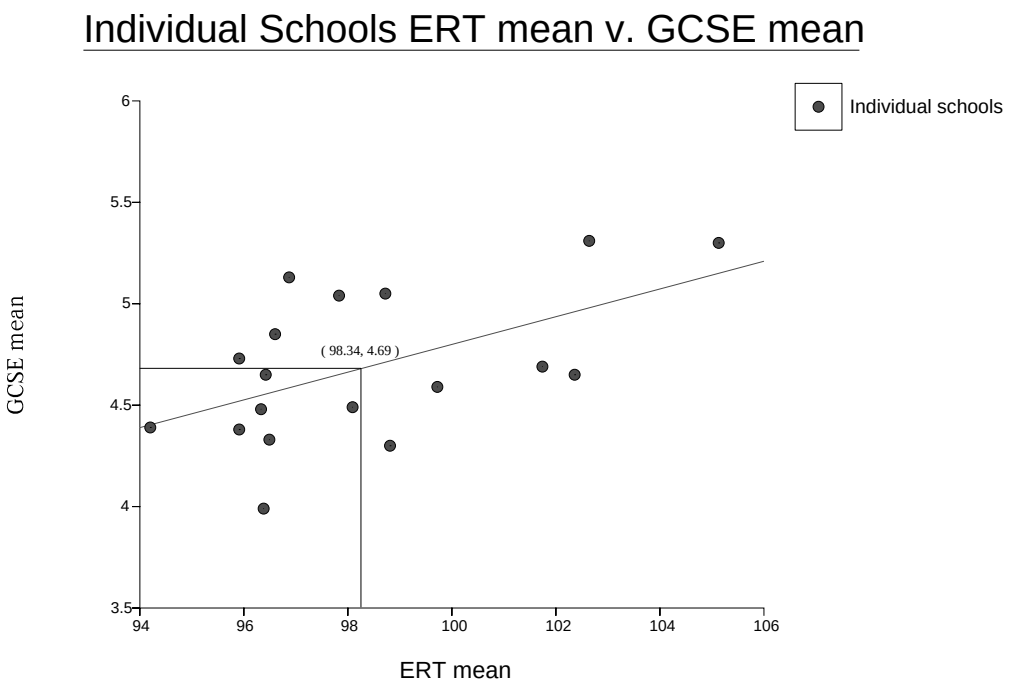
**Distribution of pupil ability**

In considering school effectiveness within a group of schools, such as an LEA, the published research has tended to rely on using 'means' to report the relative ability of schools' populations where prior attainment information was available. More recently there has been an acknowledgement that schools may be differentially effective with different ability groups within their schools, Blakey and Heath (1992) and Thomas and Mortimore (1996) giving examples of this.

Before one looks at the differential effectiveness of schools with different ability groups one really ought to consider the existence and relative size of such groups within schools. The use of an ability 'mean' on its own is widely reported (Mortimore and Byford, 1981; Maughan *et al.*, 1990, Thomas and Mortimore, 1996) but does not give sufficient information on the ability of the school cohort to make accurate comment upon the examination performance of the cohort for a "mean" gives little information on the distribution of ability in the particular sample.

A typical use of the "mean on mean" approach would be to produce a scattergraph and regression line for the performance of schools as below ( *Figure 7.8* ). The mean GCSE outcome for the 18 schools is 4.69 and the mean ERT is 98.34. The standard deviation for GCSE mean is 0.36 and the correlation coefficient is 0.54 .

**Figure 7.8**




If all the schools had populations showing a normal distribution of pupil ability in each school then the mean indicator score in each school could be claimed to be representative of the ability of each school's year cohort . The problem with this is that not all the school populations are normally distributed and, even if they were, unless the standard deviations for each school were similar the distribution of the ability within the schools would also be different.

Of the 18 schools shown in the graph above, 8 schools have mean ERT scores of between 95.91 and 96.87, less than a score of 1 between the lowest and the highest, and therefore remarkably similar in terms of the average ability of their pupils considering the disparity in the GCSE means. Of these 8 schools the highest GCSE mean was 5.31 , just over a C grade average, and the lowest was 3.99, just under a D grade average. This difference is equivalent to a grade difference in every GCSE the pupils took.

However, as the information in the following table ( *Figure 7.9* ) shows, the schools are not so close when one looks at the percentage of pupils in their year cohorts with ERT scores in excess of 100. The rank order of the schools on this second measure has changed quite markedly.

**Figure 7.9**

| *School* | *ERT mean* | *Pos'n* | *%>100 ERT* | *Pos'n* | *Mean GCSE* | *Pos'n* |
|---|---|---|---|---|---|---|
| A | 95.91 | 8 | 33.96 | 5 | 4.73 | 3 |
| B | 95.91 | 7 | 31.86 | 8 | 4.38 | 6 |
| C | 96.33 | 6 | 33.33 | 6 | 4.48 | 5 |
| D | 96.38 | 5 | 35.14 | 3 | 3.99 | 8 |
| E | 96.42 | 4 | 41.22 | 1 | 4.65 | 4 |
| F | 96.49 | 3 | 38.12 | 2 | 4.33 | 7 |
| G | 96.60 | 2 | 31.87 | 7 | 4.85 | 2 |
| H | 96.87 | 1 | 34.00 | 4 | 5.13 | 1 |
| Mean | 96.36 | | 34.93 | | 4.56 | |
| Std. dev. | 0.30 | | 3.01 | | 0.32 | |

Schools A, B and H have ERT means which differ from the group mean by more than a standard deviation, but none differ by as much as two standard deviations.

Schools B, F and G have percentages of the year cohort with more than ERT scores of 100 which differ from the mean for the group by more than one standard deviation whereas school E differs by more than two standard deviations.

Schools D and H have GCSE means which differ from the average of the means for the group by more than one standard deviation but none differ by as much as two standard deviations.

A single ERT point difference between top and bottom schools in their ERT mean score has become an almost 9.4% difference in the proportion of pupils with ERT scores in excess of 100. In terms of GCSE mean grades the gap between top and bottom schools is greater than a grade difference across all the subjects the average pupil would sit.

The school which had the highest ERT mean has achieved the highest GCSE mean but that is coincidental to my main argument here which is that the mean ERT score is potentially misleading if one doesn't take account of the distribution of pupil ability within a school. It is not sufficient to assume a

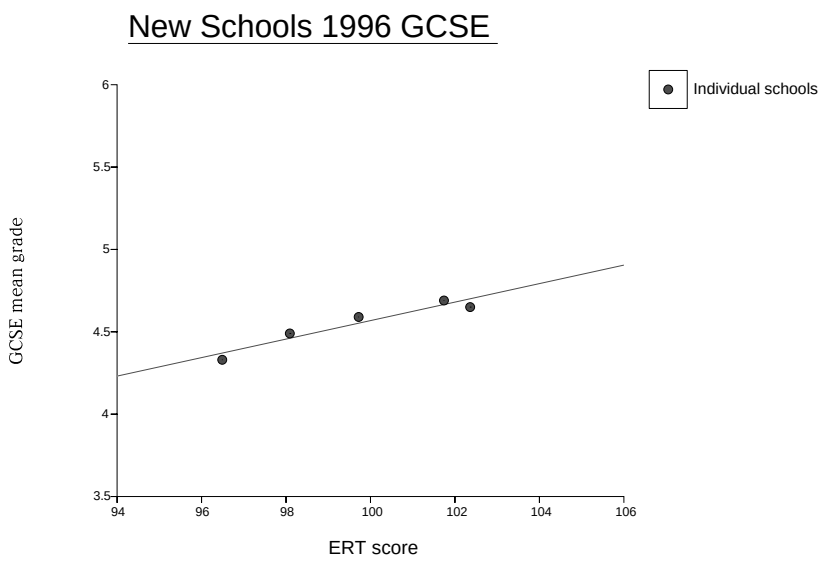
normal distribution or comparable variances. This point was borne out in the case study of School X where the distribution of pupil ability varied year on year and particularly between genders.

Some of the variation in the rank order of the schools in each column of *Figure 7.9* will be because of "natural" variation year on year and one could expect the order of these same schools ranked according to ERT mean to be quite different in successive years.

Other variation is caused by peculiarities in the population of the schools. For example, the distribution of pupil abilities in school E is unusual in that at 41.22% of pupils with scores in excess of 100 this figure is more than two standard deviations above the mean for the eight schools. In order to pass comment upon whether this school ought to have achieved a higher GCSE mean one would have to consider the spread of abilities within the remainder of its pupil population, the relative numbers of girls and boys and their ability distributions. It would also be advisable to look to trends over a number of years before any judgements were made on the effectiveness of the school. The comparison of the performance of these schools is further complicated in that there exist two distinct populations within the larger sample shown above.

A smaller population of schools, shown in the graph below ( *Figure 7.10* ), only joined the consortium of schools looking at value-added in 1996 and so have not had time to develop their responses to the value-added data in the way that those schools have which have been involved in the consortium for much longer. When isolated out, the small sample of new schools from a fairly restricted geographical area have a very high correlation between ERT and their schools' GCSE mean. The correlation coefficient was 0.96, mean ERT was 99.68 and the mean GCSE grade was 4.55 with a standard error of estimation at 0.04 .

**Figure 7.10**

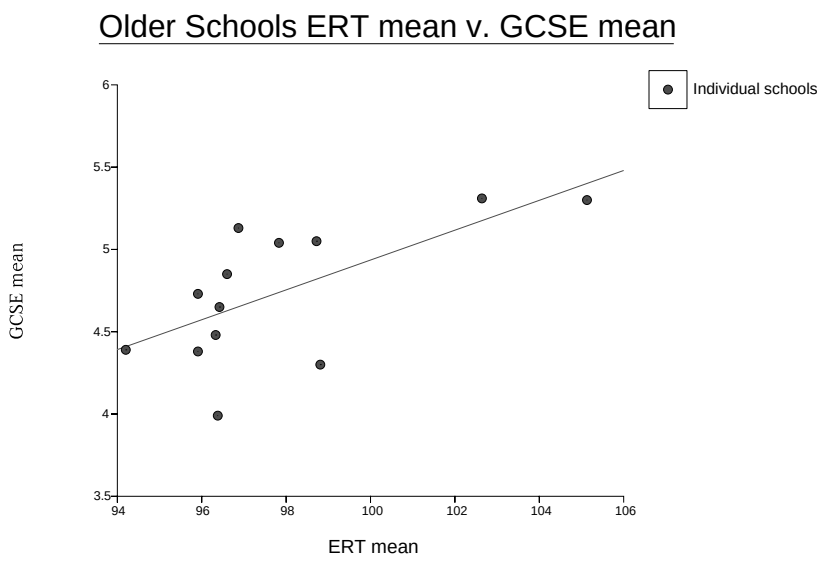



It will take time for these schools to train their staff in the use and interpretation of indicator data and value-added concepts. Systems will have to be put in place to make information available to staff, such as pupil indicator test information, progress information, reporting mechanisms may have to be amended to reflect progress in relation to expectations. Staff, pupils and parents will have to become familiar with the systems and learn the limits of the information provided by such systems. In some schools these changes will involve only minor alterations to existing procedures but in others a culture change may be necessary to encompass the linking of pupil progress to expectations based upon quantitative data.

That the performance of these "new" schools correlates so very highly with the average ERT score for the schools indicates that in 1996 the major influence upon their differential performance at GCSE level was the ability of their pupil intake as indicated by the ERT.

The thirteen other schools, shown in the graph below (*Figure 7.11*), had been involved in the value-added analysis for much longer but from a much wider geographical area, and also had a greater correlation than the whole sample when isolated out as a discreet population.

**Figure 7.11**




The correlation coefficient was 0.65 with a mean ERT of 97.83, mean GCSE of 4.74 and a standard error of estimation at 0.30 .

It will be interesting to see how the examination performance of the small group of new schools changes as they adjust their school development plans in response to the value-added data. My experience with other schools would suggest that their performance will improve at a rate above the national improvement in examination performance if they make use of the value-added information. The information must be disseminated to all the teaching staff and not just filed in the management's filing cabinets. School improvement over time will be discussed later in this chapter.

Continuing the theme of distribution of pupil ability, the following distribution graphs ( *Figures 7.12 a-d* ) show how different the spread of ability can be within schools with approximately the same average (mean) ability. If the different schools were equally effective with the full range of abilities, which they were not, then these distributions would have led to very different GCSE outcome figures for these four schools.

Schools A and B have identical indicator means ( 95.91) but school B has 34.7% of its cohort in the critical 91-100 band, where pupils are capable of

gaining C grades, and only 33.45% of its cohort with ERT scores of below 91. School A in contrast has 39.63% of its cohort with ERT scores of below 91 and therefore far less likely to gain C grades in their GCSE examinations but more pupils with ERT scores in excess of 100 and likely to gain grades C or above. School E ( *Figure 7.12 c* ) has 41.22% of its cohort with ERT scores of above 100, a higher figure than the other three schools illustrated, but school H has the highest figure for pupils in the top two ability bands at 18.0%.

These high ability pupils should be able to gain high grades in a high number of GCSEs and so raise the mean GCSE score for the school as a whole. This, in part, accounts for School H having the highest GCSE mean of the four schools and the third highest GCSE mean of the eighteen schools in the 1996 sample.

**Figure 7.12 a**

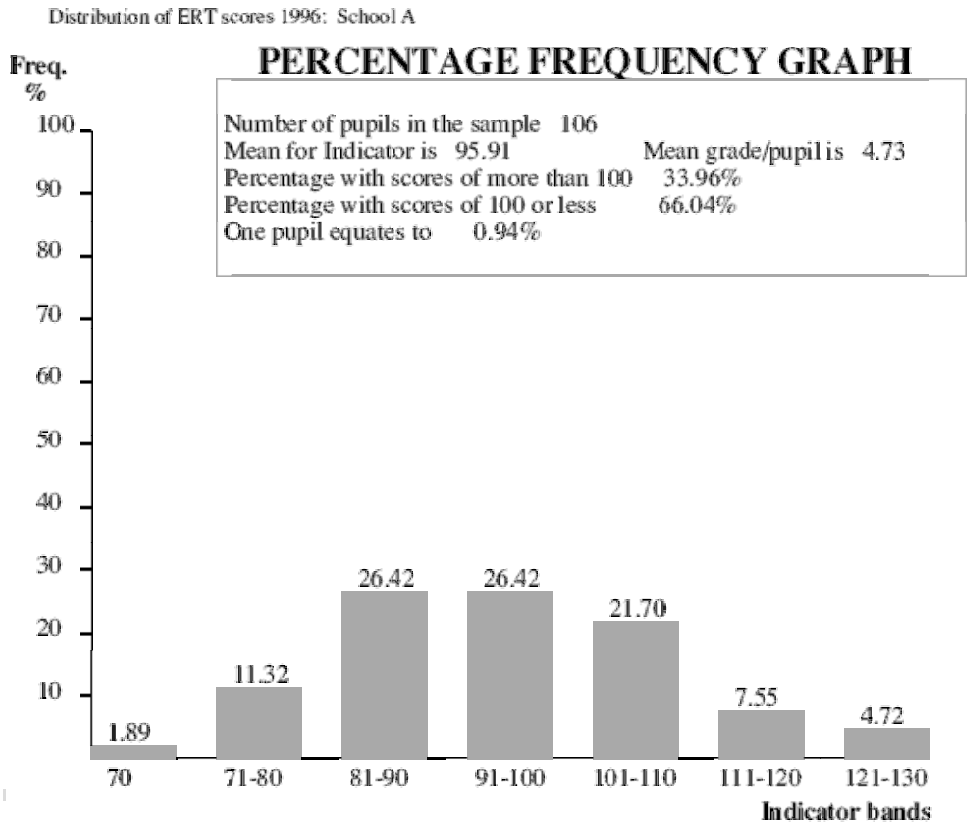



**Figure 7.12 b**

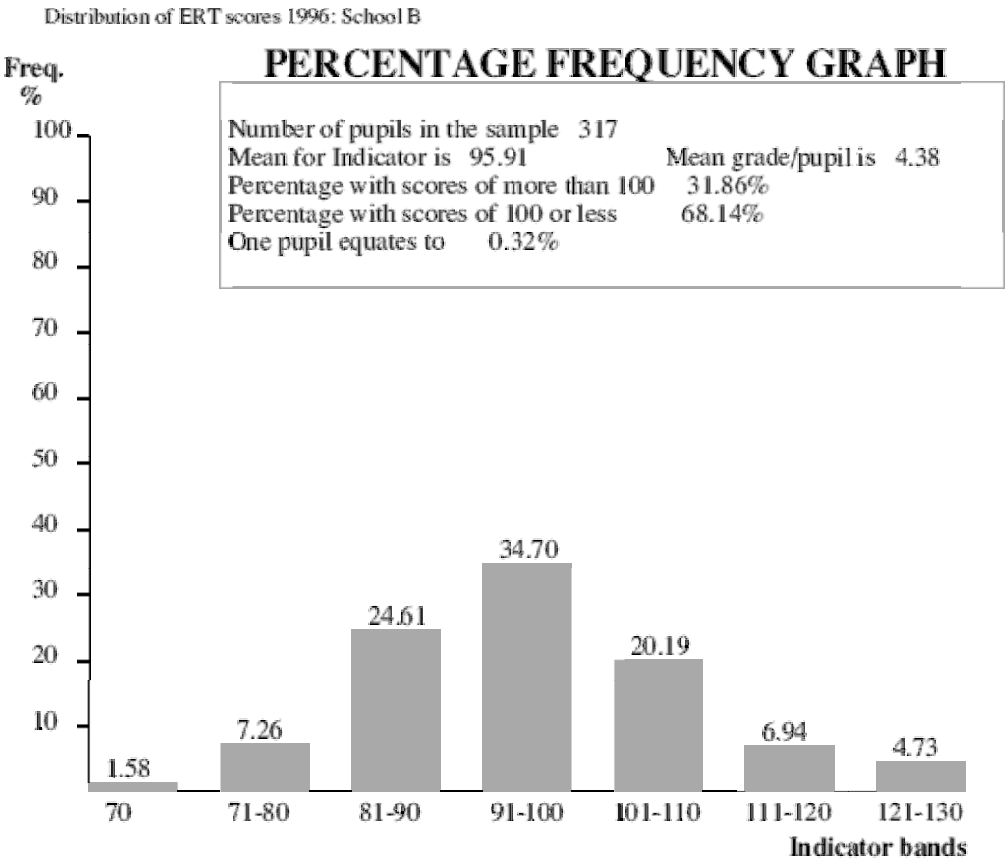


**Figure 7.12 c**

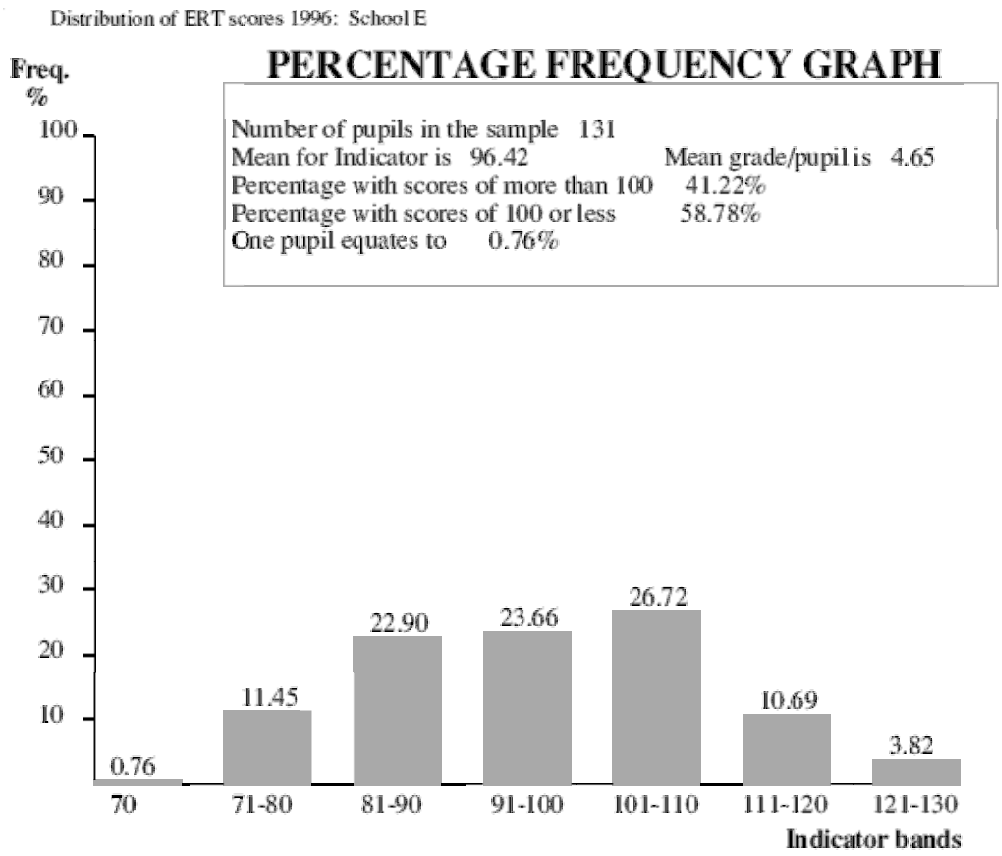



**Figure 7.12 d**

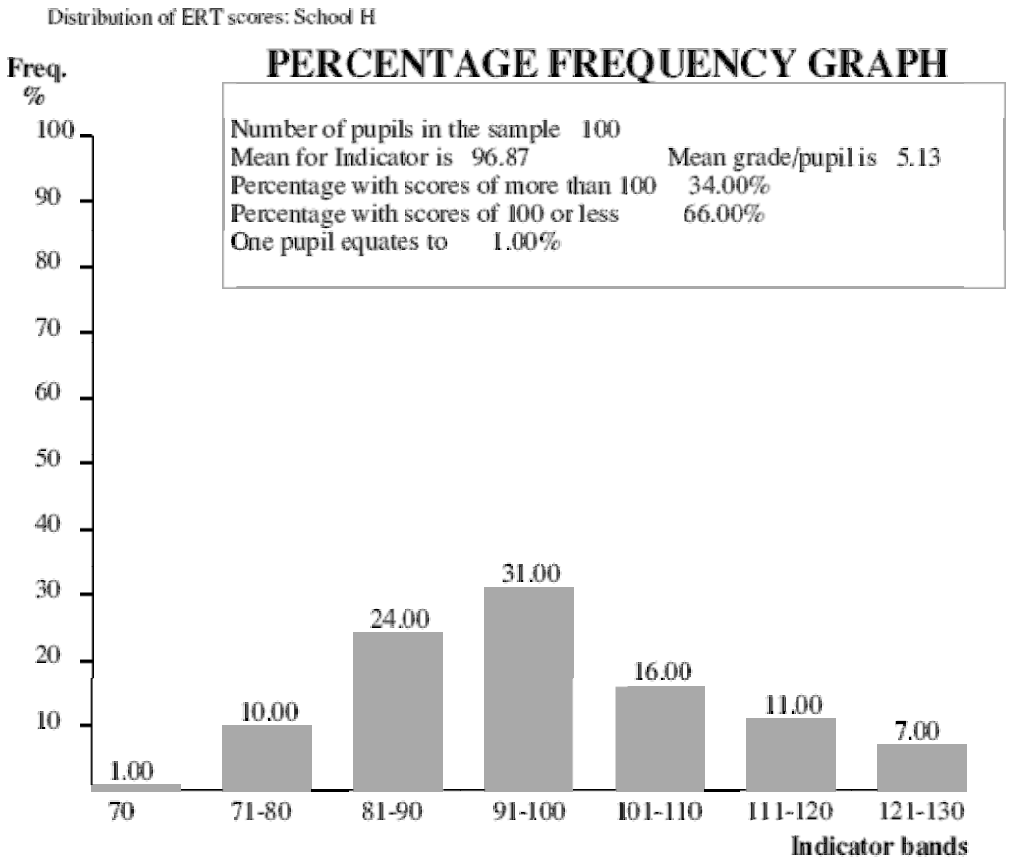


*Figure 7.13* shows this ability banding and performance quite clearly. Of the 18% of pupils in School H who had ERT scores in excess of 110, 4% gained average grades in excess of A grades, 10% averaged in excess of B grades but less than or equal to A, and 3% gained average grades above C but less than or equal to B. Tables such as that shown in *Figure 7.13* offer a simple but quite powerful illustration of schools' distributions of ability and the relative performance of those ability bandings, particularly when set against a similar table showing the performance of all the pupils in the combined sample of 18 schools, *Figure 7.14*.

## Figure 7.13

**Distribution of ERT scores 1996: School H**

GCSE grades by Indicator score banding

| Indicator band | Average grade (nos.) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A* | A | B | C | D | E | F | G | U | Pupils |
| 70- 80 | 0 | 0 | 1 | 1 | 4 | 5 | 0 | 0 | 0 | 11 |
| 81- 90 | 0 | 0 | 3 | 11 | 6 | 3 | 1 | 0 | 0 | 24 |
| 91-100 | 0 | 6 | 9 | 10 | 4 | 1 | 1 | 0 | 0 | 31 |
| 101-110 | 1 | 8 | 4 | 2 | 1 | 0 | 0 | 0 | 0 | 16 |
| 111-120 | 2 | 7 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 11 |
| 121-130 | 2 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 7 |
| Totals | 5 | 24 | 20 | 25 | 15 | 9 | 2 | 0 | 0 | 100 |

| Indicator band | Average grade ( % ) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A* | A | B | C | D | E | F | G | U | Pupils |
| 70- 80 | 0.0 | 0.0 | 1.0 | 1.0 | 4.0 | 5.0 | 0.0 | 0.0 | 0.0 | 11.0 |
| 81- 90 | 0.0 | 0.0 | 3.0 | 11.0 | 6.0 | 3.0 | 1.0 | 0.0 | 0.0 | 24.0 |
| 91-100 | 0.0 | 6.0 | 9.0 | 10.0 | 4.0 | 1.0 | 1.0 | 0.0 | 0.0 | 31.0 |
| 101-110 | 1.0 | 8.0 | 4.0 | 2.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 16.0 |
| 111-120 | 2.0 | 7.0 | 1.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 11.0 |
| 121-130 | 2.0 | 3.0 | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 7.0 |
| Totals | 5.0 | 24.0 | 20.0 | 25.0 | 15.0 | 9.0 | 2.0 | 0.0 | 0.0 | 100.0 |

Schools can consider the performance of particular ability bands in relation to similar ability bands in other schools, identify under performance, if any, in a specific ability band and then attempt to address this. For example, a school may be gaining very high percentages of pupils in the A*-C range but not a high percentage of A* /A grades. By looking at the percentage of pupils with ERT scores indicating that they ought to be capable of gaining A* or A grades

and then checking to see what those candidates actually achieved one can ascertain whether potential was being met and how such candidates stand against similar candidates in a larger sample.

This approach is far better than comparing the percentage of A* and A grades achieved in a school with the percentage of A* and A grades nationally, for there is no guarantee that the school's year group ability profile matches that of the national sample.

## Figure 7.14

**Distribution of ERT scores 1996: All pupils**

GCSE grades by Indicator score banding

| Indicator band | Average grade (nos.) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A* | A | B | C | D | E | F | G | U | Pupils |
| 70- 80 | 0 | 0 | 5 | 9 | 35 | 83 | 60 | 25 | 4 | 221 |
| 81- 90 | 0 | 4 | 27 | 108 | 196 | 154 | 56 | 7 | 3 | 555 |
| 91-100 | 0 | 33 | 193 | 296 | 235 | 86 | 27 | 5 | 2 | 877 |
| 101-110 | 6 | 123 | 263 | 226 | 66 | 11 | 4 | 0 | 0 | 699 |
| 111-120 | 15 | 117 | 112 | 62 | 7 | 2 | 1 | 0 | 0 | 316 |
| 121-130 | 43 | 68 | 43 | 10 | 2 | 0 | 0 | 0 | 0 | 166 |
| Totals | 64 | 345 | 643 | 711 | 541 | 336 | 148 | 37 | 9 | 2834 |

| Indicator band | Average grade ( % ) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A* | A | B | C | D | E | F | G | U | Pupils |
| 70- 80 | 0.0 | 0.0 | 0.2 | 0.3 | 1.2 | 2.9 | 2.1 | 0.9 | 0.1 | 7.8 |
| 81- 90 | 0.0 | 0.1 | 1.0 | 3.8 | 6.9 | 5.4 | 2.0 | 0.2 | 0.1 | 19.6 |
| 91-100 | 0.0 | 1.2 | 6.8 | 10.4 | 8.3 | 3.0 | 1.0 | 0.2 | 0.1 | 30.9 |
| 101-110 | 0.2 | 4.3 | 9.3 | 8.0 | 2.3 | 0.4 | 0.1 | 0.0 | 0.0 | 24.7 |
| 111-120 | 0.5 | 4.1 | 4.0 | 2.2 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 11.2 |
| 121-130 | 1.5 | 2.4 | 1.5 | 0.4 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 5.9 |
| Totals | 2.3 | 12.2 | 22.7 | 25.1 | 19.1 | 11.9 | 5.2 | 1.3 | 0.3 | 100.0 |

A further refinement is to produce the same tables, but broken down by gender, to check for unequal distribution of ability by genders, bearing in mind what has already been said about the achievement of girls compared to boys and the case study for School X (Chapter 6), which showed just such unequal distributions and the consequential disparity in performance.

Both gender and the distribution of pupil ability are highly relevant to my next point and that is the consistency of school performance over time.

In speaking with Headteachers and Senior Managers in schools who are leading the school performance issue in their schools, it is very soul destroying

for them and their staff to have made great efforts to raise the examination performance of the school only to see the headline performance measures, such as percentage of pupils achieving five or more GCSEs at grade C or above or the percentage of examinations sat graded at C or above, actually fall.
In schools which have only recently introduced value-added approaches to raising school performance there is often the response from staff that this year's cohort were "a poor lot" as though that excused the lower examination results. When challenged, what they cannot do is give any evidence to support their claims.

Their gut feeling may or may not be right but with the correct use of prior tests, such as the Edinburgh Reading Test, the school and its members of staff should know the ability of the pupil cohorts going through school and be able to set targets accordingly. If the current cohort are less able than previous year cohorts the school should know this and work to maximise the achievement of each pupil at whatever level; that is where the targeting of effort is likely to be most successful, rather than necessarily at the level of the whole school, but should in turn lead to better results at the whole school level.

**Tracking GCSE performance against ERT over time**

Nevertheless, in the current atmosphere of competition, accountability and parental choice as discussed in the introduction to this thesis, schools will be judged on their overall figures.

I wanted to explore the variation in examination results from year to year at the level of the school unit and so plotted the performance ( mean GCSE grade ) of six schools against the ability ( mean ERT score ) of their year 11 cohorts over a period of five years (*Figures 7.15 a - 7.15 j, but Figures 7.15 k & 7.15 l cover four years only*). I also plotted lines for each gender so that gender performance could be considered in relation to the overall school performance.
Lastly I plotted exactly the same graphs but for the combined schools' samples

( 5 schools in 1991, 9 schools in 1992 &1993, 12 schools in 1994 & 1995 and 18 schools in 1996 ) over the same period of five years
(*Figures 7.15 m & 7.15 n*).

Bearing in mind what has just been said about distribution of pupil ability, these graphs considered in isolation are flawed in that they do not take account of the distribution of pupil ability other than in the pupil mean. One must also check the relative distribution of ability by gender to ensure that the mean is representative of each gender's potential.

Similarly the graphs do not take account of the relative numbers of the respective genders so I have included these beneath each pair of graphs. The possible effect that an imbalance in the numbers of girls and boys might have can be seen in some of the average ERT and GCSE score graphs in the relative position of the separate gender lines to the line for the whole school.

For example in 1993 and 1994 School B had more boys than girls, 25 more boys in 1993 from a year cohort of 149 and 16 more boys in 1994 from a year cohort of 133, with the result that the lines plotted for the average ERT score and GCSE score for the boys are closer than the girls' to the lines plotted for the year cohort because of the greater contribution made to the average by the larger number of boys.

In School D with a much smaller year group size this effect is also apparent in 1993 and 1994 where the line plotted for the boys' average ERT score is much closer to the line for the year cohort's average because of the larger number of boys. In 1993 School D's average GCSE score for boys is closer to that for the year cohort than the girls for the same reason but in 1994 the difference is far less marked, though still apparent. This is because of the excellent performance of the 15 girls at GCSE ( GCSE mean of 5.21 ) which meant that despite having a much lower average ERT score ( Girls' ERT mean 101.67 ) than the

boys ( Boys' ERT mean 107.72 ) they exceeded the average GCSE score of the boys ( Boys' GCSE mean 5.09 ) and obtained an average very close to that of the year cohort as a whole ( Year cohort GCSE mean 5.13 ).

If, taking account of prior ability, girls and boys performed equally well in GCSE examinations then, should there be a gap of, say, five or more points on the ERT scale between girls and boys in a year cohort within a school, one would expect the gender with the higher ERT score to have the higher GCSE score. However, as girls out-perform boys with similar ERT scores at GCSE level, if the girls had the higher ERT score then the gap between the GCSE performance of girls and boys would tend to be wider than if the boys had the higher ERT score.

**Figure 7.15 a**

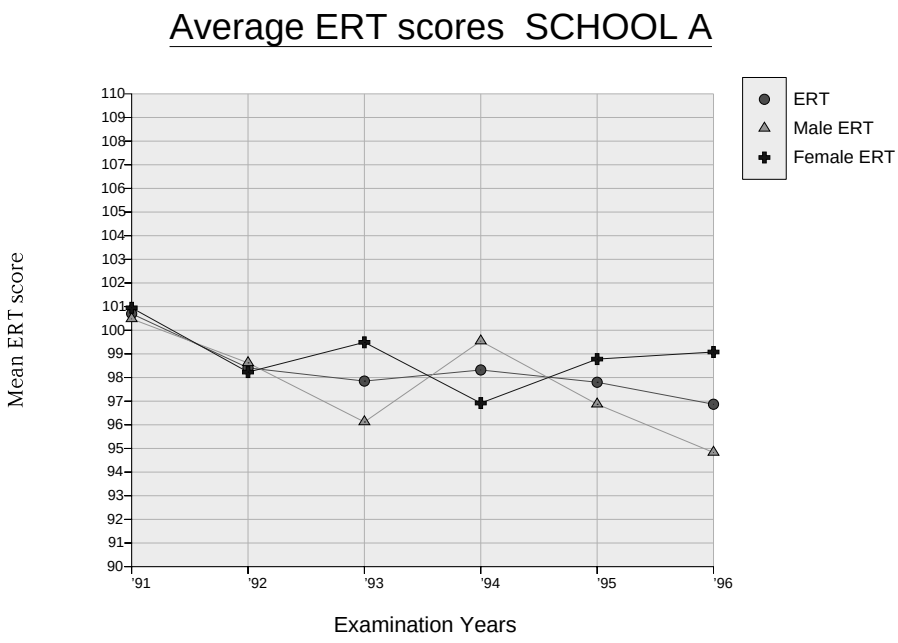
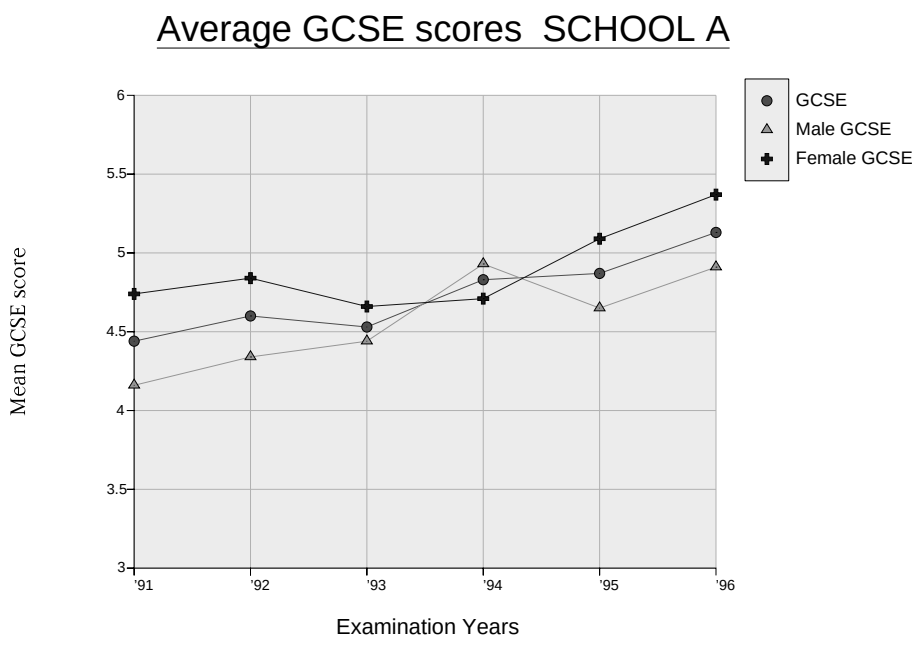



**Figure 7.15 b**




| **Pupil numbers** | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|
| Year cohort | 83 | 109 | 84 | 111 | 92 | 100 |
| Male | 44 | 52 | 41 | 59 | 47 | 52 |
| Female | 39 | 57 | 43 | 52 | 45 | 48 |

**Figure 7.15 c**

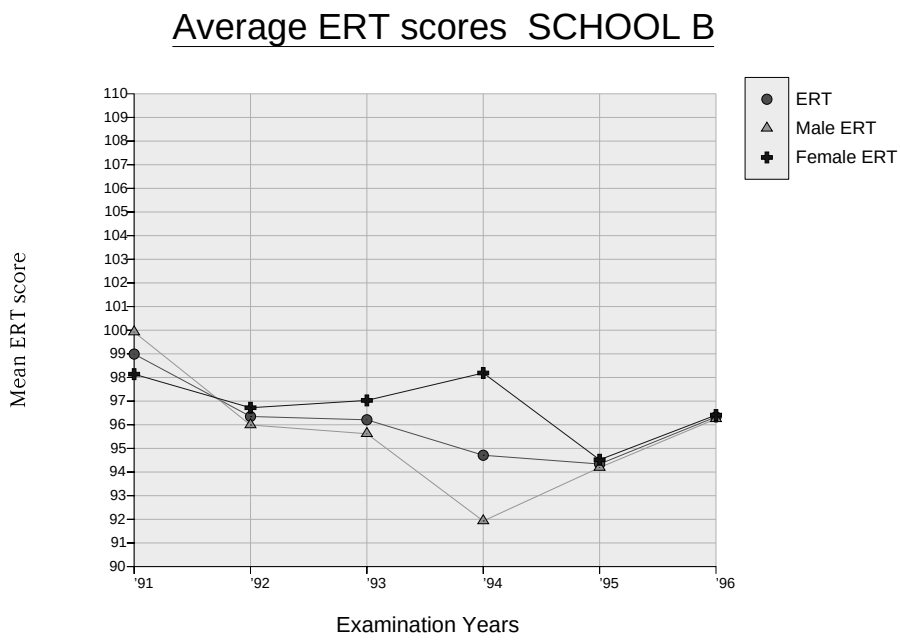
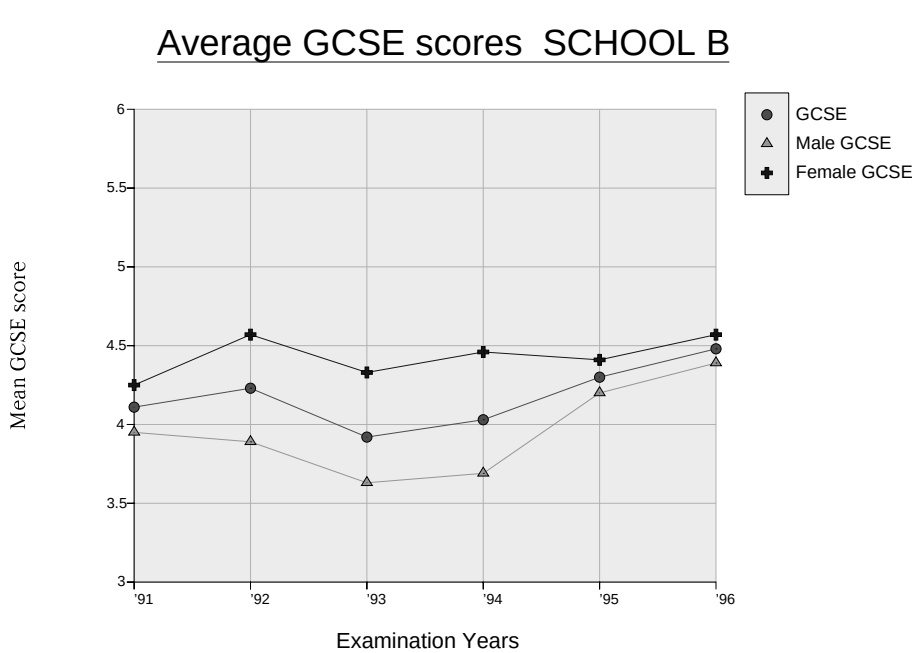



**Figure 7.15 d**




| **Pupil numbers** | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|
| Year cohort | 186 | 175 | 149 | 133 | 142 | 156 |
| Male | 89 | 89 | 87 | 74 | 79 | 75 |
| Female | 97 | 86 | 62 | 59 | 63 | 81 |

**Figure 7.15 e**

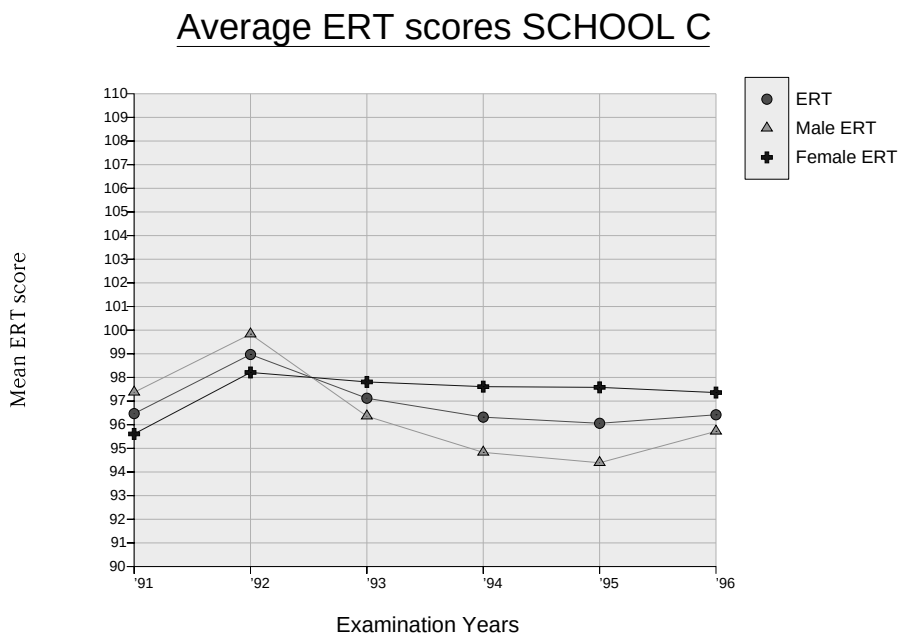
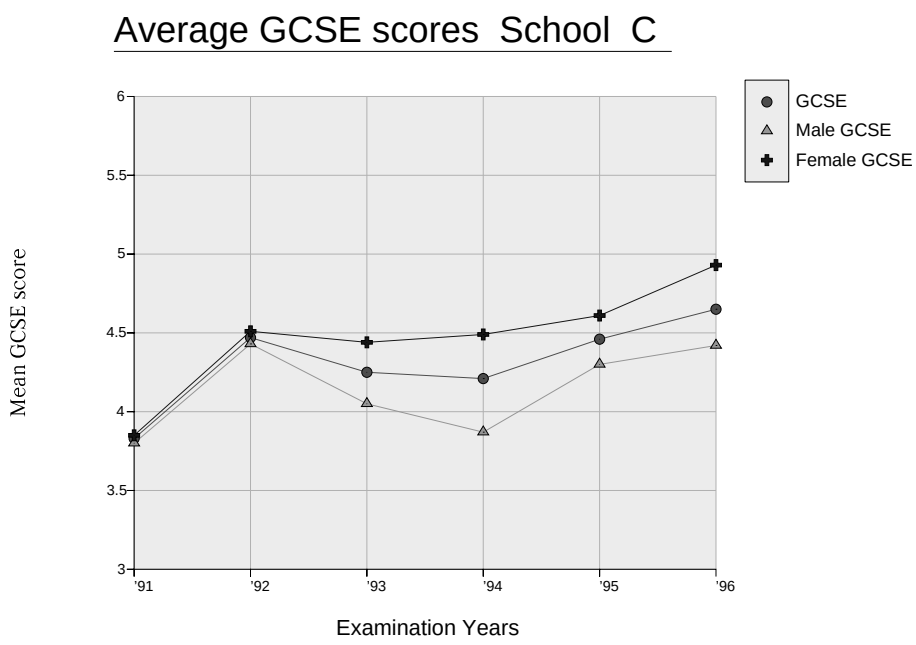



**Figure 7.15 f**




| **Pupil numbers** | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|
| Year cohort | 100 | 115 | 89 | 116 | 139 | 131 |
| Male | 49 | 54 | 42 | 54 | 66 | 75 |
| Female | 51 | 61 | 47 | 62 | 73 | 56 |

**Figure 7.15 g**

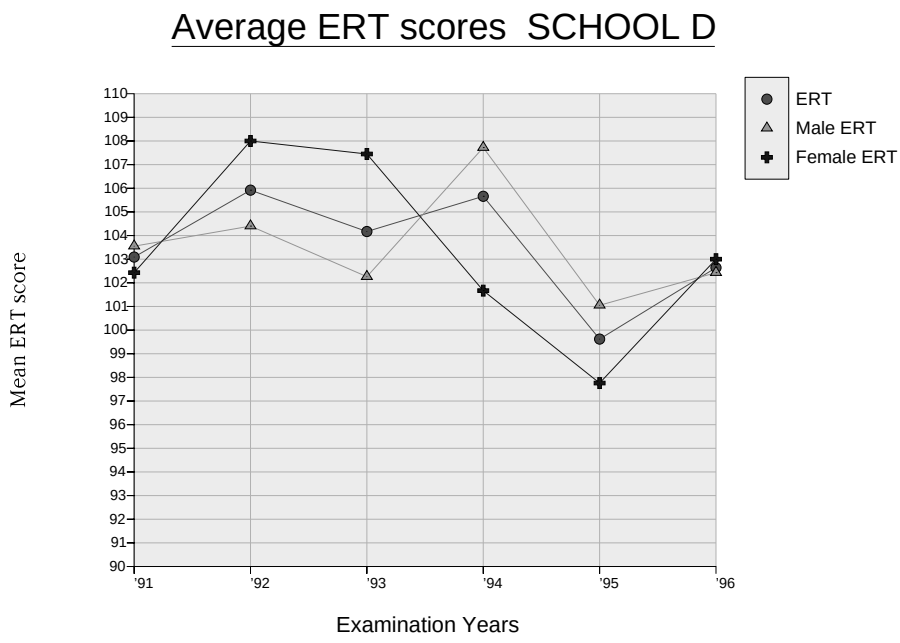
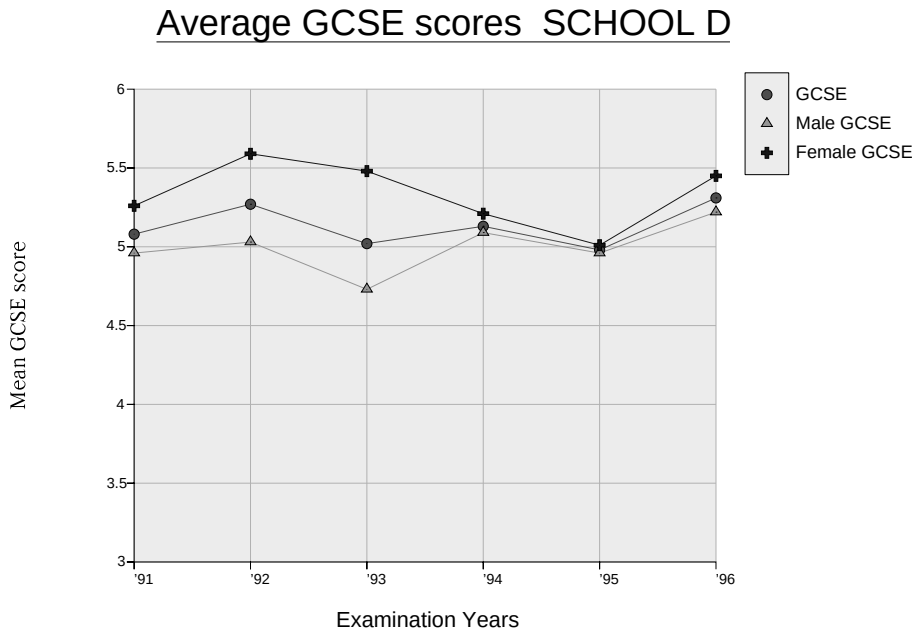



**Figure 7.15 h**




| **Pupil numbers** | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|
| Year cohort | 34 | 26 | 30 | 44 | 39 | 44 |
| Male | 20 | 15 | 19 | 29 | 22 | 28 |
| Female | 14 | 11 | 11 | 15 | 17 | 16 |

**Figure 7.15 i**

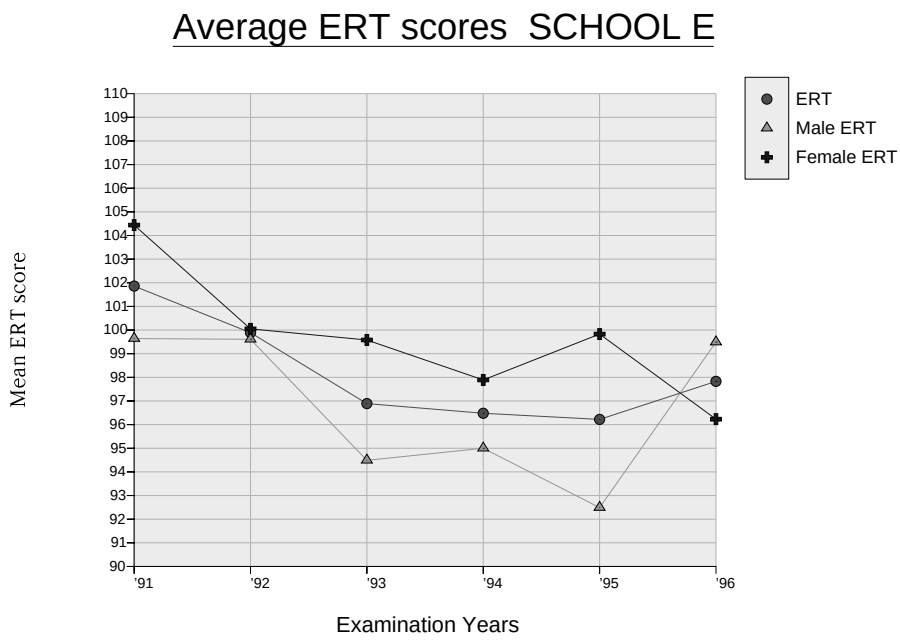
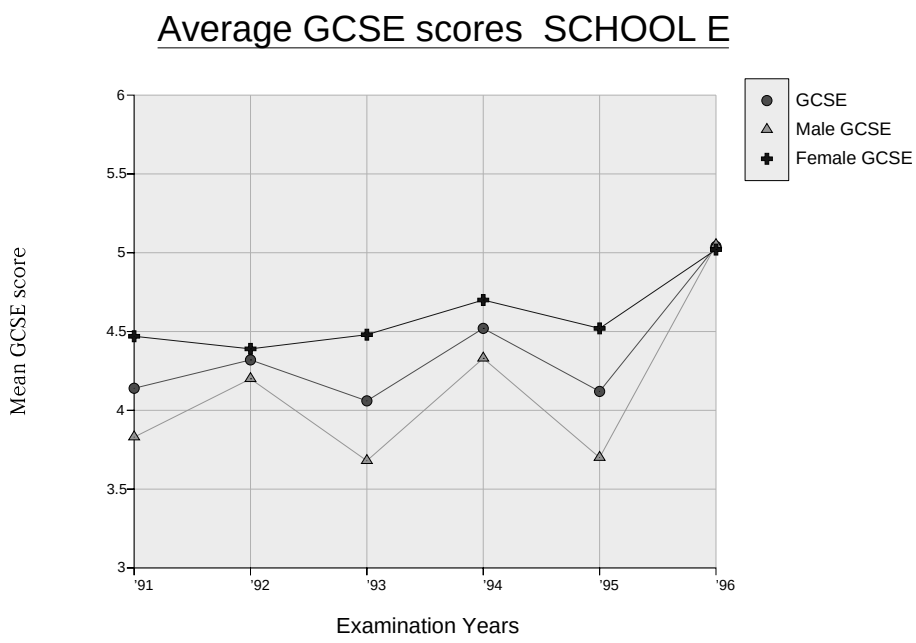



**Figure 7.15 j**




| **Pupil numbers** | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|
| Year cohort | 119 | 117 | 138 | 105 | 128 | 112 |
| Male | 64 | 44 | 73 | 52 | 63 | 55 |
| Female | 55 | 73 | 65 | 53 | 65 | 57 |

**Figure 7.15 k**

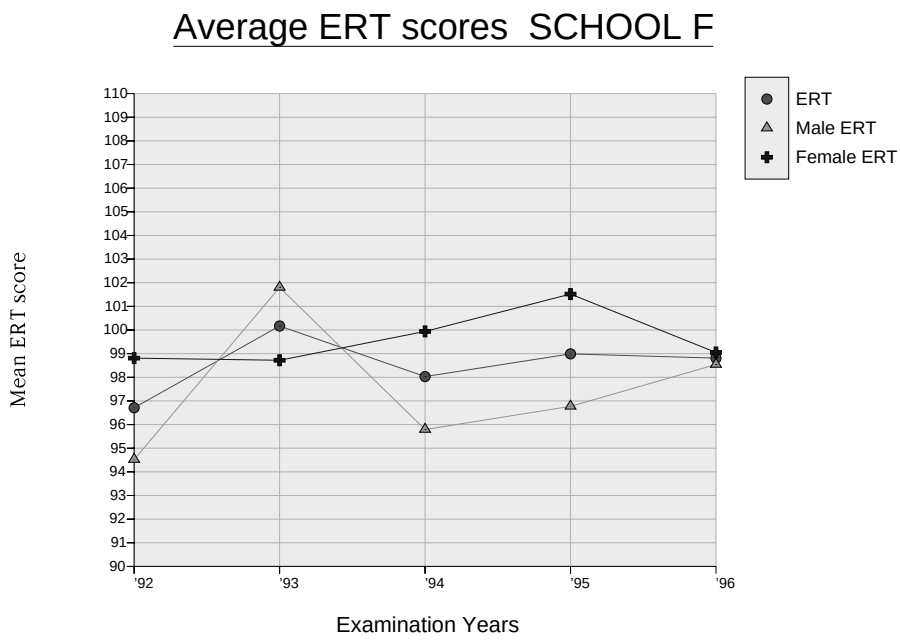
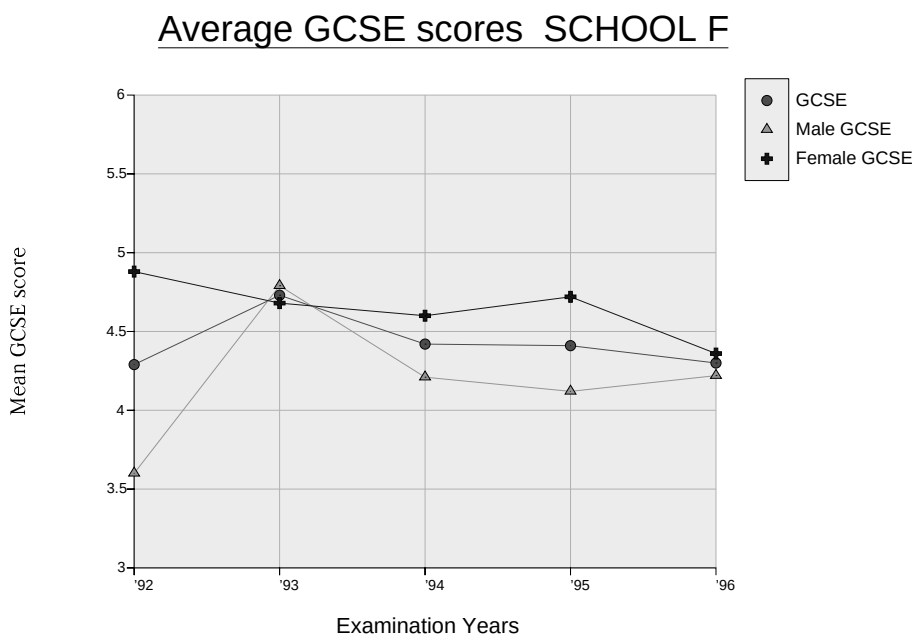



**Figure 7.15 l**




| **Pupil numbers** | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|
| Year cohort | | 145 | 140 | 156 | 177 | 200 |
| Male | | 71 | 66 | 72 | 94 | 95 |
| Female | | 74 | 74 | 84 | 83 | 105 |

**Figure 7.15 m**

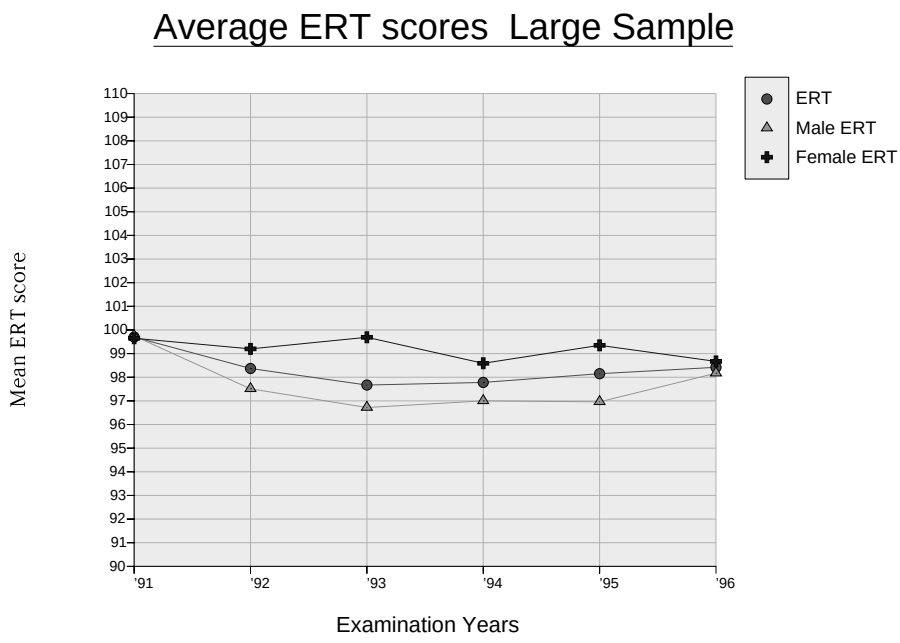
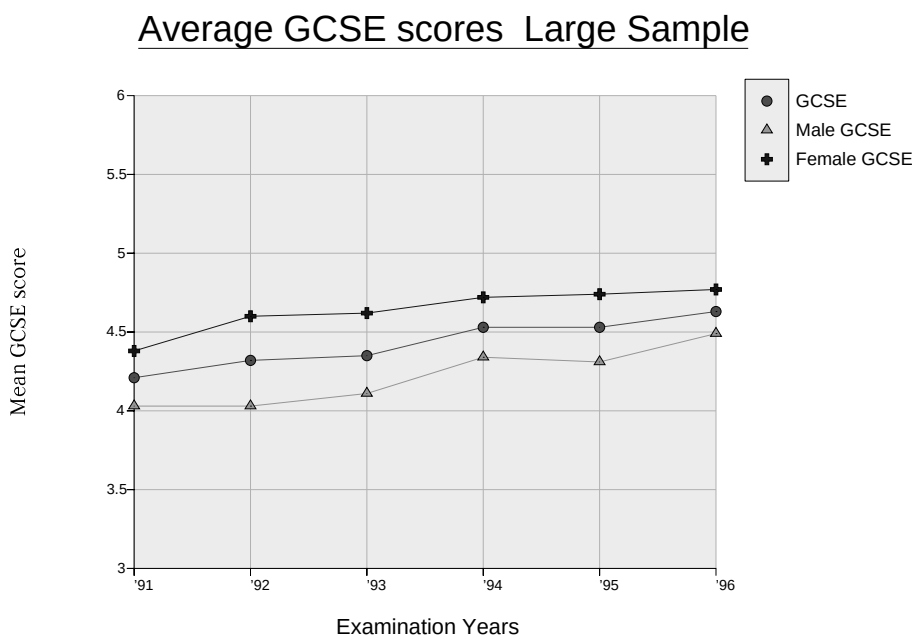



**Figure 7.15 n**




| **Pupil numbers** | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|
| Year cohort | 522 | 1092 | 990 | 1489 | 1630 | 2834 |
| Male | 266 | 534 | 513 | 761 | 818 | 1420 |
| Female | 256 | 558 | 477 | 728 | 812 | 1414 |
| **Number of schools** | 5 | 9 | 9 | 12 | 12 | 18 |

Looking at the graphs for the large sample of pupils from all schools with ERT and GCSE information (*Figures 7.15 m & 7.15 n*) it can be seen that in each year the girls already had higher ERT average scores than the boys in Year 8 in each year since 1991. The difference has narrowed in 1996. The spread of ability for the large samples over the range of years was reasonably normally distributed for each gender. This would suggest that the girls have established an educational advantage over the boys, as assessed by the ERT, in the later years of primary school and / or in the first year of secondary education. This advantage is continued at GCSE where the disparity in performance is increased.

In looking at the regression lines for GCSE mean upon ERT scores for the large combined schools' samples broken down by gender (See *Appendix G* ), the line for girls is consistently above that of the boys and widening as the ERT score increases. The difference is not large and less than the standard error of estimation but clearly visible and apparent each year. Whilst not statistically significant this trend is stable and persistent suggesting a real difference in the performance of girls and boys. The tables showing the banding of pupils by ability (ERT score) and GCSE mean, as per *Figure 5.9* in chapter 5, also show girls in the highest ability bandings out-performing boys in terms of percentages achieving the high GCSE mean grades in the various ability bandings.

Returning to the graphs of ERT and GCSE mean for individual schools over the years 1991 to 1996 it can be seen how the GCSE performance of each gender largely tracks the ERT score of that gender. School A (*Figures 7.15 a & 7.15 b*) has managed to raise its examination performance, in terms of GCSE mean grade, despite a falling average ERT score over the period shown but the positions of the genders relative to each other still reflect the trend shown in the ERT scores.

In 1994 the boys in school A obtained a higher GCSE mean than the girls. The boys had a higher average ERT score but as this was only just under 2.5 higher than the girls one would not have expected, based upon the general trend shown in the regression line graphs for combined schools' samples (*Appendix G*) and those graphs for individual schools (as in *Appendix F*), the difference in GCSE mean to be quite so marked (boys 4.93 girls 4.71). When looking at the distribution of ability for the two genders in 1994 it can be seen that although the mean ERT scores were not hugely different the distributions were quite different. 44.07% of the boys had ERT scores of greater than 100 compared to a figure of 30.77% of the girls with scores in excess of 100. Although the means were not that dissimilar the larger number of boys with high ERT scores meant that they were able to gain a higher GCSE mean than the girls despite the fact that girls of similar ability to boys would tend to out-perform them.

This case emphasises the need to consider the distribution of ability as well as the mean figures when evaluating performance, not just of the year cohort but also the respective genders in a mixed school and the contribution they make to the overall academic performance of the school. Even when comparing one's school with others having apparently similar intakes this may well not be the case. This must be borne in mind when seeking to "benchmark" as recommended by the Government in 1996 (DFEE, 1996) and likely to be adopted by the current Labour Government (DFEE, 1997).

In school D (*Figures 7.15 g & 7.15 h*) the population size is much smaller than the other schools and therefore likely to show more extreme variation. Whilst the school is relatively small, the numbers are also reduced in this particular school because a considerable number of the pupils arrived at the school after the Year 8 Edinburgh Reading Test and so did not have the necessary information to be included in the analysis. The school population is also unusual in that the proportion of girls is very small in relation to the boys. This

has the effect, both in the graph for ERT scores and GCSE means, of bringing the overall mean for the year groups closer to that of the boys than the girls because of the influence of the greater number of boys. This is clearly visible in years 1992, 1993 & 1994. Therefore in considering the performance of different schools one should consider the relative numbers of the different genders in case an imbalance, as here, exerts an influence upon the overall school mean. (See also the Case study of School X in Chapter 6.)

In two schools, each with normally distributed ability groups for boys and girls and similar mean levels of ability, if one had considerably more girls than boys and given the difference in performance at GCSE by girls and boys it is likely that the school with the greater percentage of girls would gain a higher GCSE mean. The reason for this disparity in performance between our two notional schools would not be apparent in any league or performance tables because the relative numbers of the respective genders are not shown beyond the fact that some schools are single sex and some are mixed.

There is considerable variation in the composition of the ability of the year groups in these schools from year to year whilst the average ERT score for the large sample of pupils from all the schools remains relatively constant.
Not only does the ERT average for the school change quite dramatically in some cases but the ERT average score for the genders within these schools can change even more markedly. Schools D, E and F demonstrate this with the change in the ERT score for boys in school E between 1995 and 1996 or school F between 1992 and 1993 being perhaps the most dramatic. With changes in the ability of pupils within a school so marked as these in the space of a year, it is unrealistic to expect to find consistency in the examination performance of schools except in terms of value-added performance which takes account of the prior ability indicator information and even then due notice must be taken of the gender mix within schools and the distribution of pupil ability by gender within the schools being compared.

**Subject differences / variation / stability**

In Sexey's School, and many others, there was the annual ritual staff meeting at the start of the autumn term to discuss the summer's examination results. At this meeting the head would select subject departments for especial praise because of *their excellent examination results*. No account was taken of the ability of the pupils or, in some staff's opinions, the level of difficulty of the subject. The exercise was invidious and one almost felt relief that your own department was not singled out for the odium such selection caused amongst the other staff.

Most importantly there was no recognition of a job well done with pupils who had performed outstandingly but were never going to gain the top grades. There was no acknowledgement of the ability of the pupils and their examination performance in relation to that ability.

Approaching the problem of comparing subject department performance from the level of the individual pupils, my first step was to quantify what very many teachers do each year and that is to look at what the pupils who did my subject achieved in their other subjects. With this approach the individual pupil is the common baseline against which subject departments could compare themselves. For each pupil a GCSE mean was calculated against which each subject could compare the grade achieved in their area.

Figure 5.5 in chapter 5 shows a German subject department printout listing individual pupils who took German, their German GCSE grade, their average (mean) GCSE grade and the numerical difference between German grade and mean grade produced by subtracting the mean from the points equivalent of the subject grade. Alongside this is the pupil' indicator score, such as ERT score.

Averages of the subject grades, the pupil GCSE means and the pupil indicator

scores in each subject area were calculated and shown on the subject printout as discussed in chapter 5. These were then compiled to produce a list of subject departments showing their averages and differences as in Figure 5.4.
Whilst this summary document gave some notion of comparison between different subjects within one school, and performance was shown, both in relation to other subject areas teaching the same pupils and to the average ability of the group, there remained the question of whether one should expect particular subjects to gain higher differentials over other subjects because they were "easier". By easier I mean that in many different schools over a number of years pupils who sat the "easy" subject would consistently average higher grades in that subject than in the other subjects they sat.

Looking at consistency over time would allow Heads of subject and school managers to see if the differential of one subject over others were maintained or fluctuated year on year. However, in a single school too many other variables come into play such that this comparison over time is very difficult to quantify. A better approach is to look at a larger sample for trends, such as a number of schools and how particular subjects compared to other subjects in these schools.

The average ability of groups taking subjects such as Maths, English or Science GCSE tend to reflect the ability of the year group as in most cases all pupils will sit these subjects under the requirements of the National Curriculum. Subjects with smaller numbers, not core National Curriculum subjects, will vary even within the annual cohort variation because of option choice restrictions operating on the pupils wishing to take different subjects, any entry standards that may be imposed by certain subject areas, restricted group sizes and the general popularity of a subject.

Gender variation will also play a part in some subjects with gender imbalance and consequent gender effects in the performance of the groups.

The smaller the group size the greater the influence of individual performance within the group leading to heightened variation swings in both prior ability and outcome performance. Many of these points I have already stated in this thesis in the context of school year cohorts, but these points are worth making again in the context of individual subject areas for they are equally valid if not more so when numbers of subject candidates are generally smaller than those of year cohorts.

Comparison of like with like, for example Drama results with Drama results of departments having similar ability groupings in other schools, seemed much fairer and potentially more illuminating when comparing subject department effectiveness. This desire to be compared, like with like, was echoed by the vast majority of teachers with whom I have discussed examination results and department performance since 1991.

Fitz-Gibbon has done a great deal of research into A level subject comparisons with the A Level Information Service (ALIS) project based at Newcastle and now Durham University. She has argued strongly (Fitz-Gibbon, 1992) that there is considerable variation in the performance of A level subject departments from year to year, even accounting for the ability of the pupil intake, and there are significant differences in the level of difficulty between A level subjects, expostulating whether, "a 'D' in Mathematics worth a 'B' in English?" in which case it would be extremely unfair on a Mathematics department to expect their candidates to gain the same grades as English candidates of similar ability as judged by their average GCSE score. The problems of establishing comparability of examination grades between subjects, that a C grade in one subject is equally difficult to get as a C in another, make the comparison of subject department performance in schools even more difficult.

Fitz-Gibbon (1996) goes further in discussing the problems of estimating

grades for English A level and judgements being made on subject teachers when actual grades do not match with the estimated grades,

> "This is an important illustration of two principles: that the broad features of the system must be known before individuals are judged and that the competencies of teachers can only be considered in comparison with similar teachers *teaching the same subject* to similar pupils."

To this end I compiled tables and supplied them to schools from 1992 onwards as shown in Appendix A and the examples given in *Figures 7.16 a, 7.16 b & 7.16 c*. *Figure 7.16*a shows the 1996 GCSE results of 21 Drama departments. Three departments highlighted in bold italic type used a prior test other than ERT. The subject departments are ranked according to their GCSE mean grade but this does not imply that the department at the top of the list was more effective, simply that the department at the top of the list had pupils who averaged a higher GCSE mean than those departments below. For schools wishing to consider the effectiveness of their subject departments there are a number of ways of using these tables.

**Figure 7.16 a**

| | Subject grade | Group grade | Diff. | Ind. mean | Group size | Syllabus | Board |
|---|---|---|---|---|---|---|---|
| Drama | ***6.47*** | ***4.93*** | ***1.54*** | ***94.97*** | ***30*** | ***1248*** | ***SEG*** |
| | 6.33 | 5.18 | 1.15 | 98.01 | 75 | 1255 | SEG |
| | 6.15 | 5.03 | 1.12 | 98.87 | 41 | 1248 | SEG |
| | 6.05 | 4.93 | 1.12 | 101.32 | 22 | 1248 | SEG |
| | 6.00 | 4.98 | 1.02 | 97.25 | 23 | 1248 | SEG |
| | 5.91 | 4.57 | 1.34 | 96.19 | 80 | 1248 | SEG |
| | ***5.90*** | ***5.71*** | ***0.19*** | ***115.00*** | ***20*** | ***1281*** | ***SEG*** |
| | 5.89 | 5.01 | 0.88 | 104.54 | 62 | 1248 | SEG |
| | ***5.86*** | ***4.25*** | ***1.61*** | ***99.62*** | ***21*** | ***1698*** | ***ULEAC*** |
| | 5.76 | 4.49 | 1.27 | 99.57 | 46 | 1248 | SEG |
| | 5.71 | 4.68 | 1.03 | 99.80 | 59 | 1698 | ULEAC |
| | 5.67 | 4.46 | 1.20 | 92.58 | 33 | 1248 | SEG |
| | 5.57 | 4.96 | 0.61 | 98.89 | 47 | 1698 | ULEAC |
| | 5.42 | 5.18 | 0.24 | 102.00 | 19 | 1698 | ULEAC |
| | 5.40 | 4.87 | 0.53 | 100.17 | 50 | 1248 | SEG |
| | 5.35 | 4.69 | 0.66 | 100.97 | 34 | 1248 | SEG |
| | 5.33 | 4.14 | 1.19 | 97.68 | 24 | | |
| | 5.03 | 4.71 | 0.32 | 102.47 | 65 | 1248 | SEG |
| | 4.76 | 3.62 | 1.15 | 88.00 | 21 | 1248 | SEG |
| | 4.72 | 4.24 | 0.48 | 98.79 | 92 | 2325 | MEG |
| | 4.65 | 4.44 | 0.22 | 96.50 | 43 | 1698 | ULEAC |

In figure 7.16a the highest differential over the achievements of the pupils in all the subjects they sat was achieved by the department listed ninth with a differential of 1.61 for the 21 pupils who sat drama in that school. This department could claim that it was most effective in that it had the largest positive differential over the other subjects its pupils sat of all the Drama departments in 1996. This differential is also dependent upon the quality of the other subject departments in that particular school. If they, or some of them, were less effective than generally the case in other schools it could be said that it was easier for the Drama department to look good.

Discounting the department ranked first in the list because its pupils did not have ERT scores using another test instead, the department ranked sixth has a better differential than those subject departments above it despite having a lower average ERT score and more pupils (80) than any other school bar one. By this measure the sixth ranked school could claim to be the most effective department, taking account of the ability of its pupils as judged by the ERT score. It also has a higher average subject grade than many of the other schools' drama departments with more able pupils.

Before coming to any conclusion about the effectiveness or otherwise of a subject department it is necessary to look at a number of years' results to see if the differences over other departments are maintained or simply one offs. Small departments will tend to fluctuate more because of their small numbers and the increased influence of individual pupils' results. Schools might wish to restrict their comparisons to school departments of a similar size.

The syllabus followed by the majority of departments was Southern Examining Group (SEG) 1248, four schools followed a University of London syllabus and one school followed a Midland Examining Group syllabus. There is insufficient evidence here to comment upon the difficulty of the different

syllabuses but the very popularity of the SEG syllabus, particularly if repeated year on year, would suggest that those Heads of department not using it might wish to consider it and revisit their reasons for using the syllabus they do. Consideration as to whether the most effective departments, comparable to their own in size and / or pupil ability, are using the same or different syllabuses would seem sensible. It might be thought that Heads of Subject would be considering the suitability of examination syllabuses on a regular basis but, from the conversations I have had with Senior Managers in other schools, this is not necessarily so. Without such information, as presented in these tables, Heads of Subject have no objective way of knowing how suitable particular examination syllabuses are for pupils of different abilities other than to try them.

As can be seen by looking down the table at the column of subject differentials (*See Appendix C page ii* ), in Drama every school's Drama department achieved a positive difference over the average for all the subjects their candidates sat. In 12 out of the 21 departments this difference was well over a grade. This pattern was repeated every year from 1992 to 1996 which strongly suggests that the pupils who sit Drama GCSE examinations find it "easier" than the other subjects they take. As was discussed earlier in this chapter (pages 152-153) the correlation between ERT and GCSE results in Drama is not as high as in many other subjects, suggesting that the skills assessed and assessment methods used in Drama have less in common with those of the ERT than many other subjects.

Using the same method, looking at the subject differential over the average grade achieved in all the subjects sat by pupils doing the particular subject, one can look at the results for English Language GCSE in 1996 ( *Figure 7.16b* ) and see that 11 departments out of 24 departments had negative differences and of these all but one department had negative differences of less than half a grade. Of the departments with positive differences, only one had a positive

difference in excess of half a grade. Again this pattern in 1996 reflects that of previous years indicating that pupils found English language of average difficulty (*See Appendix C page iii*).

**Figure 7.16b**

| Subject grade | Group grade | Diff. | Ind. mean | Group size | Syllabus | Board |
|---|---|---|---|---|---|---|
| English Language | | | | | | |
| ***5.84*** | ***5.83*** | ***0.00*** | ***117.17*** | ***122*** | ***1611*** | ***NEAB*** |
| 5.38 | 5.31 | 0.07 | 102.64 | 45 | 1611 | NEAB |
| 5.33 | 5.27 | 0.06 | 105.13 | 165 | 1611 | NEAB |
| 5.20 | 5.02 | 0.18 | 98.72 | 122 | 2400 | SEG |
| 5.19 | 5.04 | 0.14 | 98.16 | 118 | 1611 | NEAB |
| 5.02 | 4.79 | 0.23 | 96.60 | 206 | 2400 | SEG |
| 4.95 | 4.70 | 0.25 | 101.88 | 226 | 2400 | SEG |
| 4.89 | 4.28 | 0.62 | 99.23 | 208 | 1510 | MEG |
| 4.83 | 4.71 | 0.12 | 95.91 | 118 | 1611 | NEAB |
| 4.82 | 5.07 | -0.25 | 96.87 | 104 | 1611 | NEAB |
| 4.62 | 4.65 | -0.03 | 102.49 | 224 | 1510 | UCLES |
| 4.59 | 4.35 | 0.25 | 95.26 | 101 | 1202 | ULEAC |
| 4.56 | 4.58 | -0.02 | 99.74 | 236 | 1611 | NEAB |
| 4.56 | 4.83 | -0.27 | 97.65 | 192 | 1202 | ULEAC |
| 4.53 | 4.40 | 0.13 | 96.42 | 167 | 2400 | SEG |
| 4.40 | 4.67 | -0.27 | 97.28 | 130 | 2400 | SEG |
| 4.38 | 4.47 | -0.09 | 98.13 | 147 | 1611 | NEAB |
| ***4.20*** | ***4.62*** | ***-0.42*** | ***99.16*** | ***90*** | ***2400*** | ***SEG*** |
| 4.13 | 4.31 | -0.17 | 96.67 | 208 | 1611 | NEAB |
| ***4.09*** | ***4.25*** | ***-0.16*** | ***97.36*** | ***154*** | ***1611*** | ***NEAB*** |
| 4.03 | 3.96 | 0.06 | 96.38 | 114 | | |
| ***3.78*** | ***4.39*** | ***-0.61*** | ***91.85*** | ***129*** | ***2400*** | ***SEG*** |
| ***3.56*** | ***3.06*** | ***0.50*** | ***88.89*** | ***61*** | | |
| ***2.93*** | ***3.36*** | ***-0.43*** | ***87.02*** | ***59*** | | |

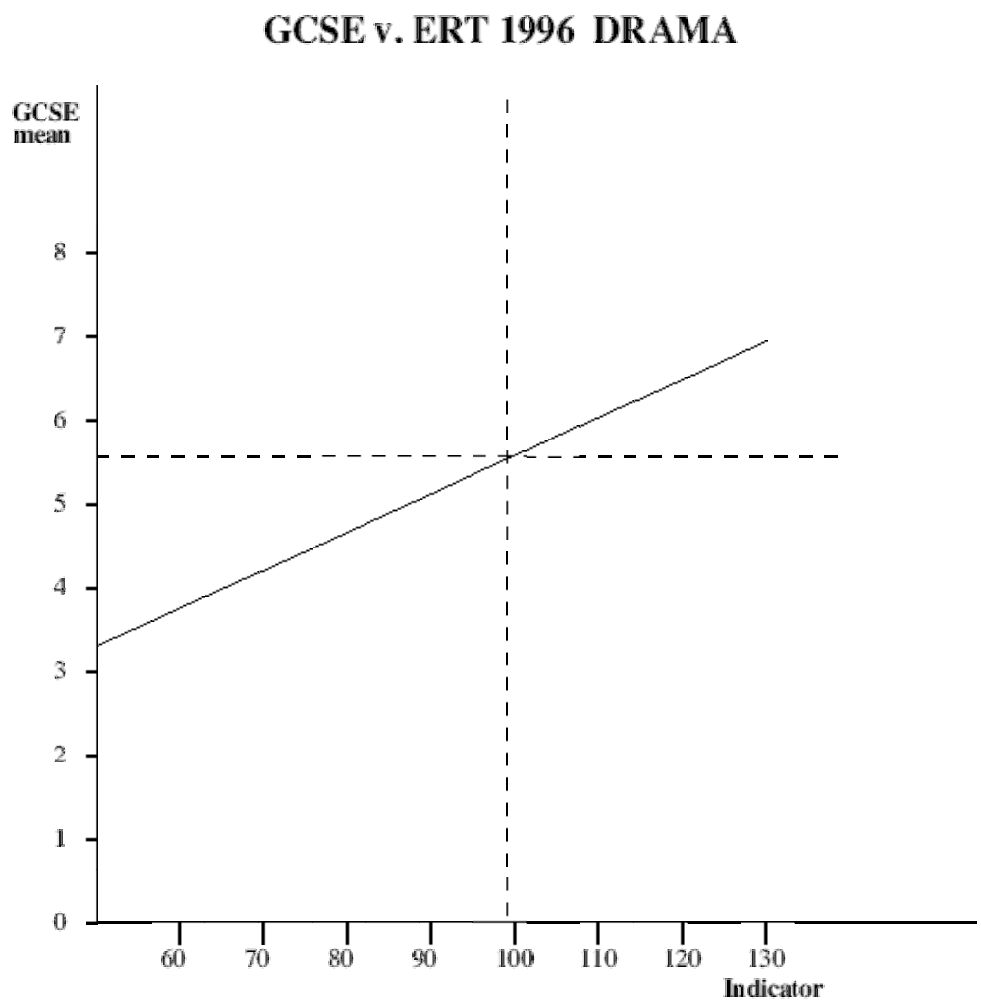
In Maths on the other hand ( *Figure 7.16c* ) the subject differences are largely negative in relation to the other subjects sat by the pupils taking Maths. Twenty-four of the twenty-eight department groups listed show negative differences, thirteen of them by over half a grade. This pattern is repeated each year from 1992 to 1996 and indicates that pupils sitting Maths GCSE generally found it more "difficult" than the other subjects they sat.

**Table 7.16 c**

| | Subject grade | Group grade | Diff. | Ind. mean | Group size | Syllabus | Board |
|---|---|---|---|---|---|---|---|
| Maths | | | | | | | |
| | ***5.97*** | ***5.85*** | ***0.12*** | ***117.17*** | ***120*** | ***1384*** | ***ULEAC*** |
| | 5.29 | 5.31 | -0.02 | 102.64 | 45 | 1384 | ULEAC |
| | 5.27 | 5.43 | -0.16 | 106.55 | 143 | 1632 | NEAB |
| | ***4.99*** | ***4.82*** | ***0.17*** | ***95.37*** | ***89*** | ***2410*** | ***SEG*** |
| | 4.56 | 4.82 | -0.26 | 97.29 | 195 | 1631 | NEAB |
| | 4.55 | 4.68 | -0.14 | 101.74 | 227 | 1666&2410 | UCLES/SEG |
| | 4.52 | 4.58 | -0.06 | 99.72 | 236 | 2410 | SEG |
| | 4.46 | 5.02 | -0.56 | 98.72 | 121 | 1384 | ULEAC |
| | 4.42 | 5.24 | -0.82 | 99.71 | 106 | 1663 | UCLES |
| | 4.40 | 5.08 | -0.68 | 96.68 | 102 | 1663/1666 | ULEAC/UCLES |
| | ***4.28*** | ***4.66*** | ***-0.37*** | ***99.79*** | ***88*** | ***1384*** | ***ULEAC*** |
| | 4.11 | 4.26 | -0.14 | 98.95 | 208 | 1663 | SEG |
| | 4.07 | 4.62 | -0.55 | 96.77 | 133 | 2410 | SEG |
| | 4.05 | 4.74 | -0.69 | 96.86 | 95 | 2410 | SEG |
| | 4.04 | 3.96 | 0.08 | 96.38 | 116 | | |
| | 4.02 | 4.83 | -0.82 | 97.18 | 198 | 2410 | SEG |
| | 3.94 | 4.31 | -0.37 | 94.55 | 105 | 1666 | UCLES |
| | 3.87 | 4.65 | -0.78 | 102.51 | 224 | 1660 | UCLES |
| | 3.86 | 4.35 | -0.48 | 96.23 | 340 | 1663&1660 | UCLES |
| | ***3.83*** | ***4.25*** | ***-0.42*** | ***97.23*** | ***149*** | ***2410/1666*** | ***SEG/MEG*** |
| | 3.72 | 4.48 | -0.76 | 98.35 | 137 | 2410/1666 | SEG/MEG |
| | 3.66 | 4.40 | -0.74 | 96.41 | 166 | 1666&2410 | UCLES/SEG |
| | 3.62 | 4.28 | -0.66 | 96.32 | 211 | 2410 | SEG |
| | ***3.52*** | ***3.32*** | ***0.21*** | ***86.18*** | ***61*** | | |
| | ***2.80*** | ***3.31*** | ***-0.51*** | ***83.34*** | ***45*** | ***1666*** | ***UCLES*** |
| | ***2.59*** | ***3.06*** | ***-0.46*** | ***88.89*** | ***59*** | | |
| | 1.65 | 2.64 | -0.99 | 88.67 | 17 | 1666 | UCLES |
| | 1.60 | 2.34 | -0.74 | 79.00 | 10 | 1666 SMP | UCLES |

Another way of looking at the comparative difficulty of these subjects is to study the regression line graphs for all the pupils who sat these GCSEs in 1996. *Figures 7.17a, 7.17b and 7.17c* show the regression line graphs for GCSE Drama, English Language and Maths results against Edinburgh Reading Test scores. The number of pupils who sat Drama was some 2000 pupils less than the other two subjects and the correlation coefficient was lower but the average ability of the three subject groups, expressed in terms of mean ERT score, and the distributions of pupil ability are very similar ( See *Figure 7.18* ).

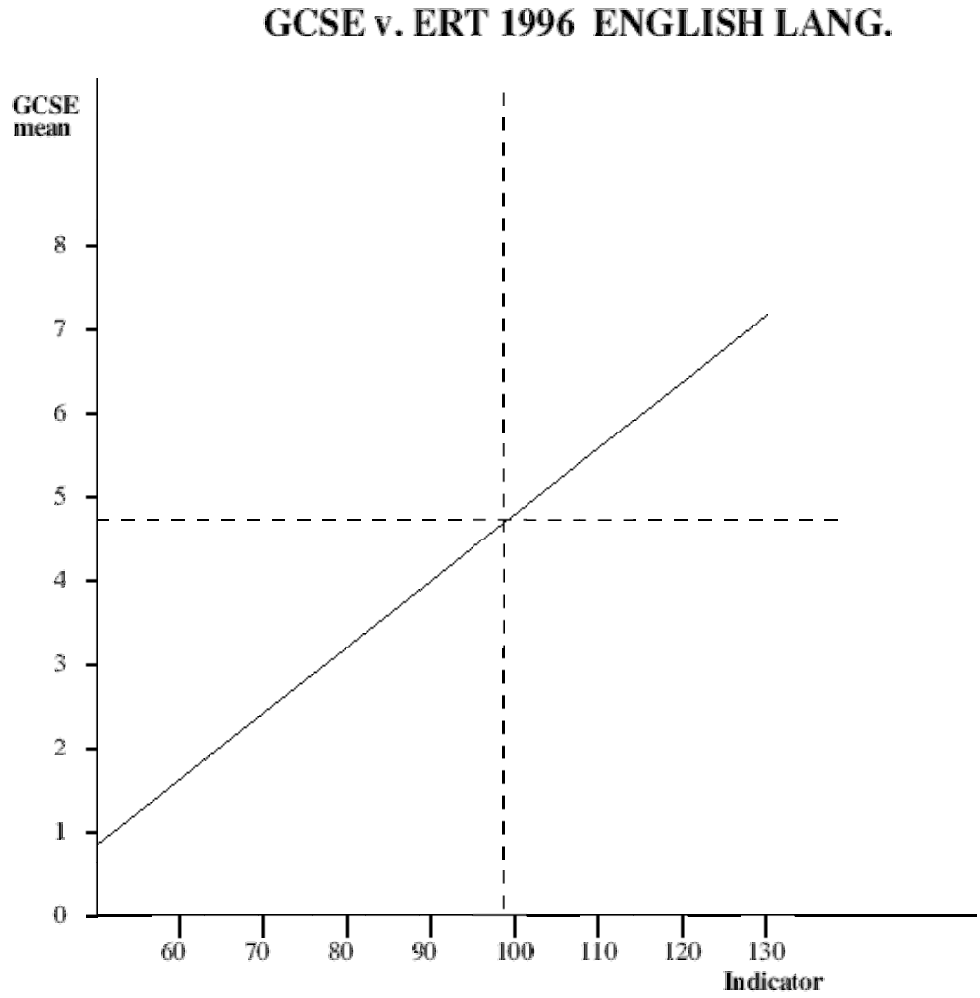
**Figure 7.17 a**




Number of pupils in the sample 709

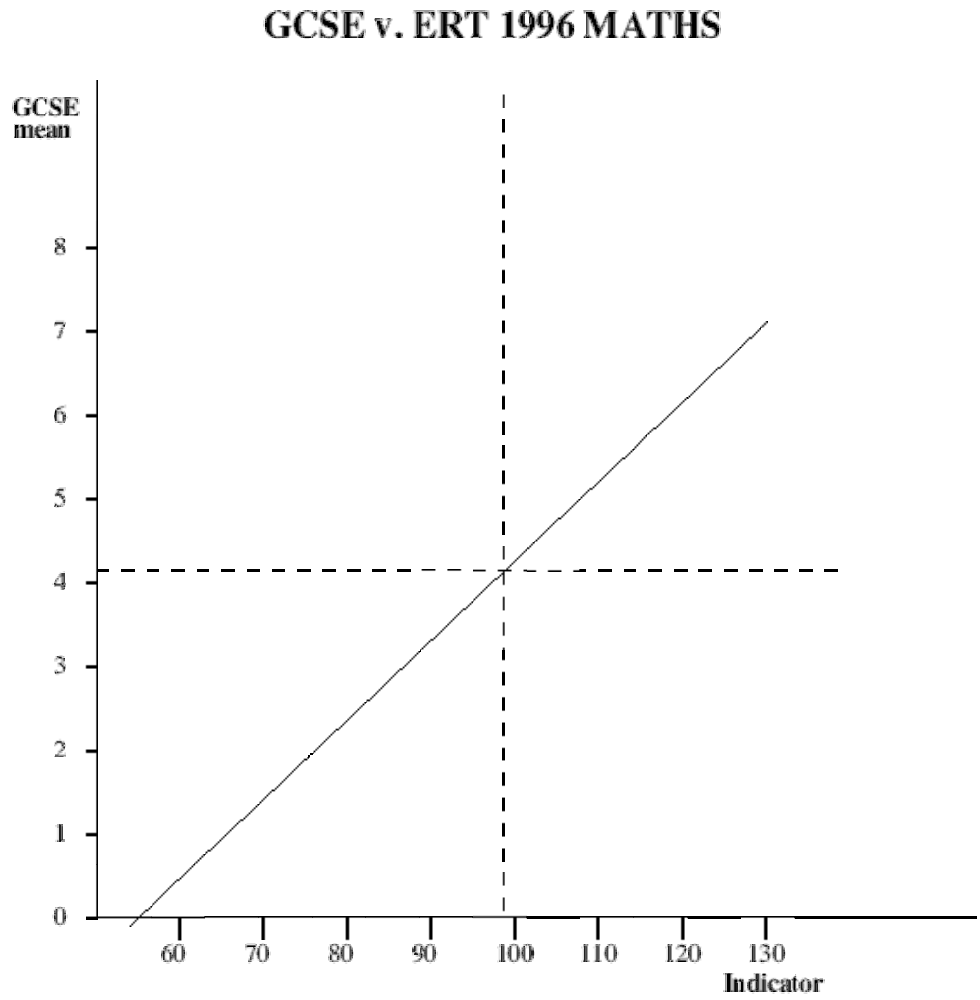
| | | | |
|---|---|---|---|
| Mean for X is | 99.07 | Mean for Y is | 5.56 |
| Standard deviation for X is | 12.45 | Standard dev. for Y is | 1.25 |
| Covariance is | 7.06 | | |
| Coefficient of correlation is | 0.45 | | |
| Coefficient of determination is | | 20.45% | |
| Standard error of estimation for Y upon X is | | 1.12 | |

**Figure 7.17 b**




| | | | |
|---|---|---|---|
| Number of pupils in the sample | 2767 | | |
| Mean for X is | 98.73 | Mean for Y is | 4.72 |
| Standard deviation for X is | 12.77 | Standard dev. for Y is | 1.44 |
| Covariance is | 12.87 | | |
| Coefficient of correlation is | 0.70 | | |
| Coefficient of determination is | | 48.82% | |
| Standard error of estimation for Y upon X is | | 1.03 | |

**Figure 7.17 c**




| | | | |
|---|---|---|---|
| Number of pupils in the sample | 2746 | | |
| Mean for X is | 98.65 | Mean for Y is | 4.16 |
| Standard deviation for X is | 12.67 | Standard dev. for Y is | 1.75 |
| Covariance is | 15.22 | | |
| Coefficient of correlation is | 0.69 | | |
| Coefficient of determination is | | 47.18% | |
| Standard error of estimation for Y upon X is | | 1.27 | |

**Figure 7.18**

**Combined Schools' Subject data 1996**

| | *Drama* | *English Lang.* | *Maths* |
|---|---|---|---|
| **Sample size (n)** | 709 | 2767 | 2746 |
| **Correlation (r)** | 0.45 | 0.70 | 0.69 |
| **Mean ERT** | 99.07 | 98.73 | 98.65 |
| **Std. dev. ERT** | 12.45 | 12.77 | 12.67 |

Despite the similarity in the ability of the groups the average GCSE grade for Drama is much higher than the other two subjects at 5.56, midway between B and C grades. The average GCSE grade for English Language was 4.72, just below a C grade, and for Maths the mean grade was 4.16, just above a D grade. A pupil with an ERT score of 99 who sat both GCSE Maths and Drama on the basis of this data could expect to gain a C in Drama but only a D in Maths.

Further analysis of the graphs and their slopes reveals that, plus or minus the standard error of estimation, in Drama a candidate with an ERT score of 70 might expect a GCSE grade of a D whereas the Maths candidate with a similar ERT score would be unlikely to gain an E grade.

The regression line graphs for Drama and Maths are different, Drama's line of best fit slope being somewhat flatter but with a higher point of intercept on the *y* axis. This means that Drama pupils with ERT scores of 130 are likely to gain A grades at GCSE, plus or minus the standard error of estimation, whereas Maths pupils with similar ERT scores are likely to do slightly better, the line indicating an average just slightly above an A grade.

At the top end of the ability range Maths candidates do just as well as Drama candidates, possibly slightly better, but at any point of the ability range below 120 on the x axis (ERT) Drama candidates do considerably better than similar candidates taking Maths.

The regression graphs and pupil ability distribution graphs for even larger samples, combining the results of candidates for the years 1994 to 1996, produce almost identical results, emphasising the consistency of these findings. Maths 1994-1996 gives a sample size of 5763 with a mean ERT of 98.54, a mean GCSE grade of 4.15 and a standard deviation of 12.78 for the ERT scores. English Language 1994 - 1996 gives a sample size of 5817 with a mean ERT of 98.48, a mean GCSE grade of 4.66 and a standard deviation of 12.89 for the ERT scores. Drama 1994 - 1996 gives a sample size of 1442 with a mean ERT of 99.14, a mean GCSE grade of 5.55 and a standard deviation of 12.84 for the ERT scores.

This information, taking account of pupil ability and the distribution of ability within the sample populations, shows that candidates of similar ability in terms of ERT scores do not achieve as high GCSE grades in Maths as in Drama or English, except for the most able candidates. Even in English Language, a main core element of the National Curriculum with very similar sample size and distribution of pupil abilities, candidates achieve higher GCSE grades than in Maths across the range of pupil abilities.

Taking the regression line for *y* ( GCSE mean ) upon *x* ( ERT score ) graphs of all subjects and their combined school samples in 1996, I produced a table of ERT scores and the range of GCSE grades from A* to G they were likely to produce, predicting the GCSE grades from the ERT scores ( *Figure 7.19* ). This table was then used at my own school to produce target grades for each pupil in the various subjects with the proviso that we expected the pupils to better these. This process is only at an early stage and it is too soon to come to any conclusions but staff have found the process of reviewing pupil progress in relation to the the targets interesting and stimulating. Mindful of error margins and the lower reliability at the extremes of the ability range the table is useful in highlighting the potentially different outcomes for the same pupils attempting different subjects.

Initial comments from teaching staff are that the target grades broadly agree with their own estimation of the pupils' likely future attainment based on teacher assessment of current work, behaviour and attitudes. However, it is in the discussion of the exceptions, the pupils whose current work does not match the statistically derived target, that has been stimulating in drawing on teachers' experience and knowledge of the individuals concerned.

The initial data and the feedback from subject teachers were discussed at a meeting of the Form Tutors, Boarding House representatives, Special Educational Needs Co-ordinator and Key Stage Co-ordinator, all of whom were able to contribute to the discussion of the individual pupils' progress. From this meeting recommendations on suggested courses of action went back to the whole teaching staff. Where a pupil was struggling in one particular subject or in all his / her subjects teachers were made more aware. Communication was improved. In some cases staff were told to expect more from certain individuals, in other cases specific strategies were prescribed, such as help with reading or comprehending questions, and in a very few cases staff were told to ignore the prior test information because the overwhelming evidence from other sources indicated different expectations. The important thing was that there had been some initial input which then aided discussion of academic progress in relation to likely outcomes and the continued monitoring of progress.

I believe it was important to involve the pastoral side of the school, for if a pupil is not happy they are not likely to make the best progress academically that they could. It may possibly be the lack of academic progress that is making a pupil unhappy and the staff responsible for the pupil's welfare should know how the academic side of school life is going.

## Figure 7.19

**GCSE target grades predicted from ERT score range based on 1996 results**

| Subject (n) | * | A | B | C | D | E | F | G | U | SE | r |
|---|---|---|---|---|---|---|---|---|---|---|---|
| All subjects (2834) | | | | | | | | | | 1.04 | 0.73 |
| Art (1156) | | | 114-130 | 89-113 | 70-88 | | | | | 1.37 | 0.35 |
| Business Studies (234) | | | 124-130 | 107-123 | 91-106 | 74-90 | 70-73 | | | 1.45 | 0.41 |
| Child Studies (165) | | 121-130 | 108-120 | 94-107 | 81-93 | 70-80 | | | | 1.19 | 0.57 |
| Drama (709) | | 120-130 | 98-119 | 76-97 | 70-75 | | | | | 1.12 | 0.45 |
| Design Technology (827) | | 127-130 | 113-126 | 100-112 | 86-99 | 73-85 | 70-72 | | | 1.7 | 0.48 |
| Electronics (79) | | | 130 | 115-129 | 100-114 | 85-99 | 70-84 | | | 1.48 | 0.53 |
| English Language (2767) | | 122-130 | 109-121 | 96-108 | 84-95 | 71-83 | 70 | | | 1.03 | 0.7 |
| English Literature (2178) | | 122-130 | 108-121 | 94-107 | 80-93 | 70-79 | | | | 1.16 | 0.61 |
| Food (530) | | 129-130 | 113-128 | 97-112 | 81-96 | 70-80 | | | | 1.22 | 0.52 |
| French (1555) | | 124-130 | 113-123 | 102-112 | 91-101 | 80-90 | 70-79 | | | 1.38 | 0.63 |
| Geography (1189) | 130 | 120-129 | 110-119 | 99-109 | 89-98 | 79-88 | 70-77 | | | 1.33 | 0.66 |
| German (745) | | 121-130 | 108-120 | 95-107 | 83-94 | 70-82 | | | | 1.36 | 0.58 |
| History (874) | | 124-130 | 113-123 | 102-112 | 91-101 | 79-90 | 70-78 | | | 1.42 | 0.63 |
| Humanities (536) | | 121-130 | 112-120 | 102-111 | 92-101 | 83-91 | 73-82 | 70-72 | | 1.41 | 0.66 |
| Information Technology (299) | | | 120-130 | 90-119 | 70-89 | | | | | 1.29 | 0.29 |
| Italian (30) | | 125-130 | 115-124 | 104-114 | 93-103 | 82-92 | 71-81 | 70 | | 0.99 | 0.73 |
| Maths (2746) | | 124-130 | 113-123 | 103-112 | 92-102 | 82-91 | 71-81 | 70 | | 1.27 | 0.69 |
| Media Studies (160) | | | 118-130 | 105-117 | 92-104 | 78-91 | 70-77 | | | 1.17 | 0.58 |
| Music (276) | | | 112-130 | 92-111 | 73-91 | 70-72 | | | | 1.53 | 0.41 |
| Physical Education (346) | 130 | 116-129 | 102-115 | 89-101 | 75-88 | 70-74 | | | | 1.2 | 0.59 |
| Religion (264) | | 123-130 | 114-122 | 104-113 | 94-103 | 84-93 | 74-83 | 70-73 | | 1.44 | 0.67 |
| Science Double (2236) | | 121-130 | 109-120 | 98-108 | 86-97 | 74-85 | 70-73 | | | 1.17 | 0.68 |
| Science Single (465) | | 129-130 | 117-128 | 105-116 | 93-104 | 81-92 | 70-80 | | | 1.16 | 0.65 |
| Sociology (129) | 130 | 117-129 | 105-116 | 92-104 | 80-91 | 70-79 | | | | 1.18 | 0.6 |
| Spanish (164) | | 130 | 119-129 | 108-118 | 97-107 | 87-96 | 76-86 | 70-75 | | 1.46 | 0.62 |
| Statistics (53) | | 128-130 | 110-127 | 91-109 | 73-90 | 70-72 | | | | 1.08 | 0.48 |

*Figure 7.19* shows the GCSE subjects and the sample size for each in brackets alongside. Under each column headed by a GCSE grade are the various ranges of ERT scores likely to result in pupils gaining that grade. Because of the statistical regression towards the mean, the extremes of the GCSE grades are likely to be under predicted. Note that only three subject areas would seem to be indicating potential A* grades, and then only for the highest possible ERT score, namely Geography, Physical Education and Sociology. This is a characteristic of the statistical method of regression whereby predictions tend towards the mean. A* grades are indicative of exceptional performance and so will tend to be under predicted in terms of numbers of pupils who will gain them.

The column headed SE indicates the standard error of estimation for predicting GCSE grades from ERT scores, using the regression line *y* upon *x* with ERT scores on the *x* axis. When looking at the predicted grades for the respective subjects, these must be considered in the range of plus or minus the standard error. The column headed r shows the correlation coefficient for the various subjects, indicating the strength of the relationship between ERT scores and the GCSE grades.

Using this table (*Figure 7.19*) one can see that a pupil with an ERT score of 75, for the sake of an example, would be predicted a D grade in Art, Drama, Information Technology, Music, Physical Education and Statistics.
In Electronics, French, Geography, History, Humanities, Italian, Maths, Media Studies, Religion and Single Science this same candidate would be predicted an F grade or even a G in Spanish.

There are clearly differences in the predicted outcomes of the different subjects for pupils with a particular ERT score suggesting that some subjects are easier and others harder. This information is helpful in advising pupils on their choices of subject for GCSE examination when particular grades are required to further their career options. This advice must be given in the light of the pupils' particular interests and talents. A pupil who requires five grade C GCSEs and has an ERT score of 100 would be better advised to take Geography, where he or she would be within in the range of ERT scores likely to gain a C grade, rather than Spanish where the likely outcome would be a D grade. This advise would be changed, however, if the pupil already had some positive experience of Spanish or was a native speaker and therefore was not a typical candidate. This issue was discussed in Chapter 6 with the case study of the French Department.

It is important to note here that the information in this table is a general

indication and not an exact prediction. Due notice must be taken of the small sample sizes in some subjects, such as Italian, and the error margins which tend to be plus or minus over a grade either side of the grade indicated by the regression line graph. It was because this indicated grade for a specific ERT score could be part way between GCSE grades that I gave ranges of ERT scores likely to lead to specific GCSE grades. It is possible that the accuracy of the prediction process would be improved by producing separate regression graphs for the respective genders and then using these to produce gender specific tables and predictions. I have not yet tested this theory.

The consequences of these findings on subject differences are potentially far reaching, for the evaluation of school effectiveness and the consistency of schools' performance in examinations depend not only on ascertaining the ability of the year cohort, the distribution of that ability, the composition of the year cohort by gender and the distribution of ability by gender, but also on the range of subjects the pupils attempt, the numbers of pupils taking various optional subjects and the ability of the pupils attempting those optional subjects.

Drama as an optional subject can cater for less able candidates far better than subjects similar to Maths in the skills and abilities they require of the pupils. In 1996 of the 6.1% of Drama candidates with ERT scores of between 70 and 80 only 3.4% in that ability band failed to gain a C grade in GCSE Drama. 24% of Drama candidates gained A or A* grades and 2.7% of these were in the bottom two ability bands.

Of the 2746 Maths candidates in 1996 7.1% were in the ERT band 70-80 and 0.3% of candidates were in this lowest ability band and gained C or above at GCSE. Only 7.8% of Maths candidates gained A or A* grades, none of whom were in the lower two ability bands.

In 1996 not one of the schools which supplied me with examination data had a Drama department that failed to achieve a positive differential over the other subjects its candidates sat, in terms of average GCSE grade per pupil (See *Appendix C*). This tendency is repeated year on year.

Carefully steering less able candidates into subjects such as Drama rather than letting them attempt harder subjects would have a marked effect upon the headline outcome measures, such as percentage of pupils attaining five or more GCSE grades C or above, used to produce national performance tables. More pupils would stand a chance of gaining 5 C grades but these would not necessarily be in the core subjects of Maths, Science or English Literature.

**Subject variation within and between schools**

Looking at the performance of subject departments, there is often more variation between schools comparing individual subject area performance across the ability range, French department with French department for example, than there is between schools comparing their GCSE performance across pupil abilities using regression analysis.

At the analysis level of the school unit in 1996 GCSE results all regression lines were very similar in seventeen of the eighteen schools with ERT information, both in the angle of slope and the fact that they were all within one standard deviation of the mean for the combined sample of eighteen schools, with the one exception of Sexey's School ( See *Appendix F* ). Sexey's, as has previously been explained, has a very small sample size and in 1996 there were predicted discrepancies between prior test information and actual results for at least two pupils which markedly reduced the school correlation figure.

In GCSE subjects which are part of the National Curriculum compulsory core, such as English and Maths, most schools will normally enter virtually all their year cohort for examination. Science, although part of the core, is slightly different in that it is common practice for schools to enter different ability groups for different syllabuses, such as Single Science, Double Award Science or the separate sciences of Biology, Chemistry and Physics.
As pupil numbers for English and Maths are virtually the same as for the year cohort these subjects tend to be more stable, in terms of the variation between schools and from year to year within the school once pupil ability is taken account of, than those subjects with smaller numbers.

Variation does still occur. In 1996 one school's English language department results were very good and its pupils exceeded the grades that would have been

expected of them had they performed in accordance with the average for the eighteen schools' English Language results in 1996. This is most clearly shown in *Figure 7.20a* where the higher of the two regression lines in the top right hand quadrant represents the particular school and the lower line is that of the combined sample for eighteen schools. It can be seen that the individual school's line is not above throughout its length but dips below the combined school line in the lower left hand quadrant.

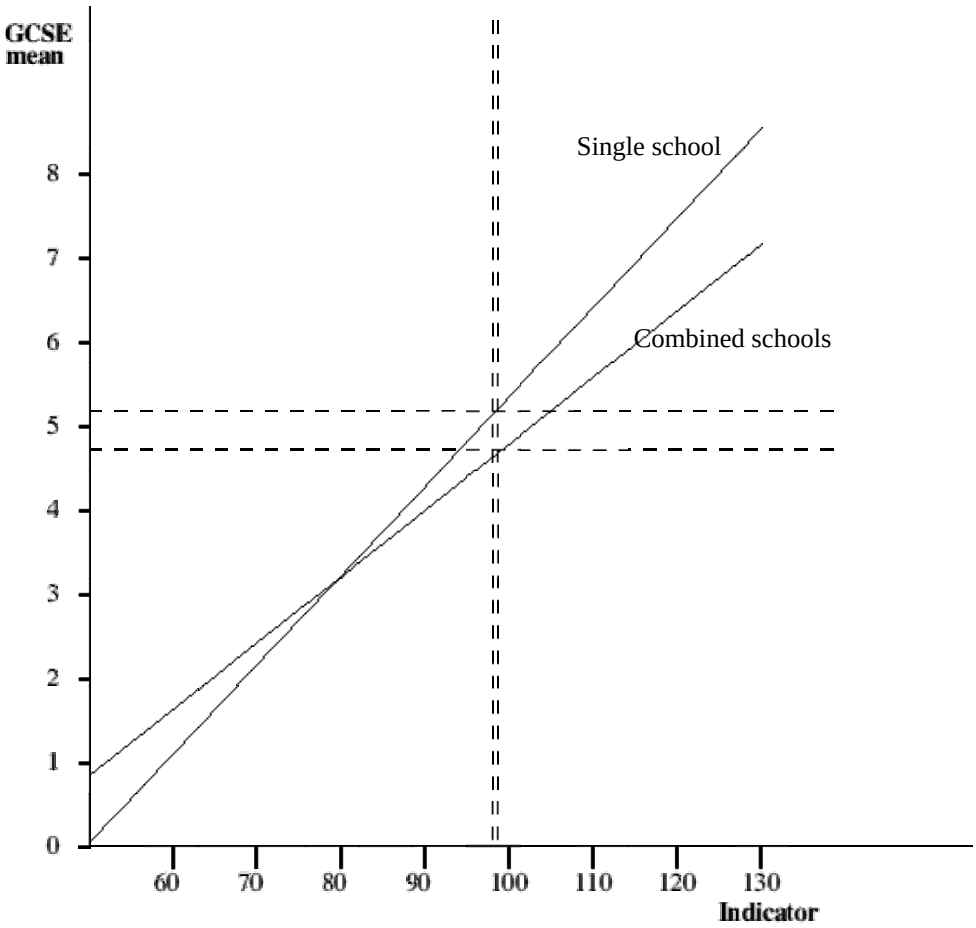
The divergence of the lines at any point below 70 on the x axis should be ignored for there were no pupils with such scores. That the lines come close together in the lower left hand quadrant shows that with lower ability pupils the school's English Language department was performing at around average expectations but as the ability of the pupils increased so the performance of the English Language candidates increased beyond the level expected of average performance. This department was differentially effective, gaining increasingly better results with the more able pupils.

This was not the case in 1995 where, as *Figure 7.20b* shows, this particular school's regression line for English was below the average for the combined twelve schools' English Language results, only reaching average performance with its very brightest pupils. The less able the individual school's English Language candidates the farther they fell below the average performance of the candidates in all the schools. The individual school's average GCSE grade for its English Language pupils was over a grade higher in 1996 than it was in 1995.

The average ERT score for the English group in 1995 was 96.24 with a standard deviation of 14.72 compared to 98.73 and a standard deviation of 12.05 in 1996. The correlation between ERT and the English Department results is equally strong in both years at 0.77.

**Figure 7.20 a**

**English Language Single School against Combined School sample 1996**




***Combined Schools' data***

| | | | |
|---|---|---|---|
| Number of pupils in the sample | 2767 | | |
| Mean for X is | 98.73 | Mean for Y is | 4.72 |
| Standard deviation for X is | 12.77 | Standard deviations for Y is | 1.44 |
| Covariance is | 12.87 | | |
| Coefficient of correlation is | 0.70 | | |
| Coefficient of determination is | 48.82% | | |
| Standard error of estimation for Y upon X is | 1.03 | | |

***Individual School's data***

| | | | |
|---|---|---|---|
| Number of pupils in the sample | 109 | | |
| Mean for X is | 98.16 | Mean for Y is | 5.19 |
| Standard deviation for X is | 12.05 | Standard deviations for Y is | 1.67 |
| Covariance is | 15.43 | | |
| Coefficient of correlation is | 0.77 | | |
| Coefficient of determination is | 58.62% | | |
| Standard error of estimation for Y upon X is | 1.08 | | |

Mindful of the need to consider the effect of gender upon examination results, I examined the results for this school in 1996 and 1995 broken down by gender. In 1996 the mean ERT of the 53 boys was 99.89 with a standard deviation of 12.56 and they obtained a mean GCSE score of 5.11. The coefficient of correlation was 0.83. The 56 girls had a lower ERT mean of 96.52 with a standard deviation of 11.30 and they obtained a GCSE mean of 5.27. The correlation coefficient of 0.72 was lower than the boys' but still strong.
In 1996, despite having a lower ERT than the boys, the girls obtained a higher average GCSE score than the boys. The spread of pupil abilities amongst the two genders was not dissimilar and the higher average grade of the girls as expected bearing in mind the performance of girls generally noted elsewhere in this thesis.

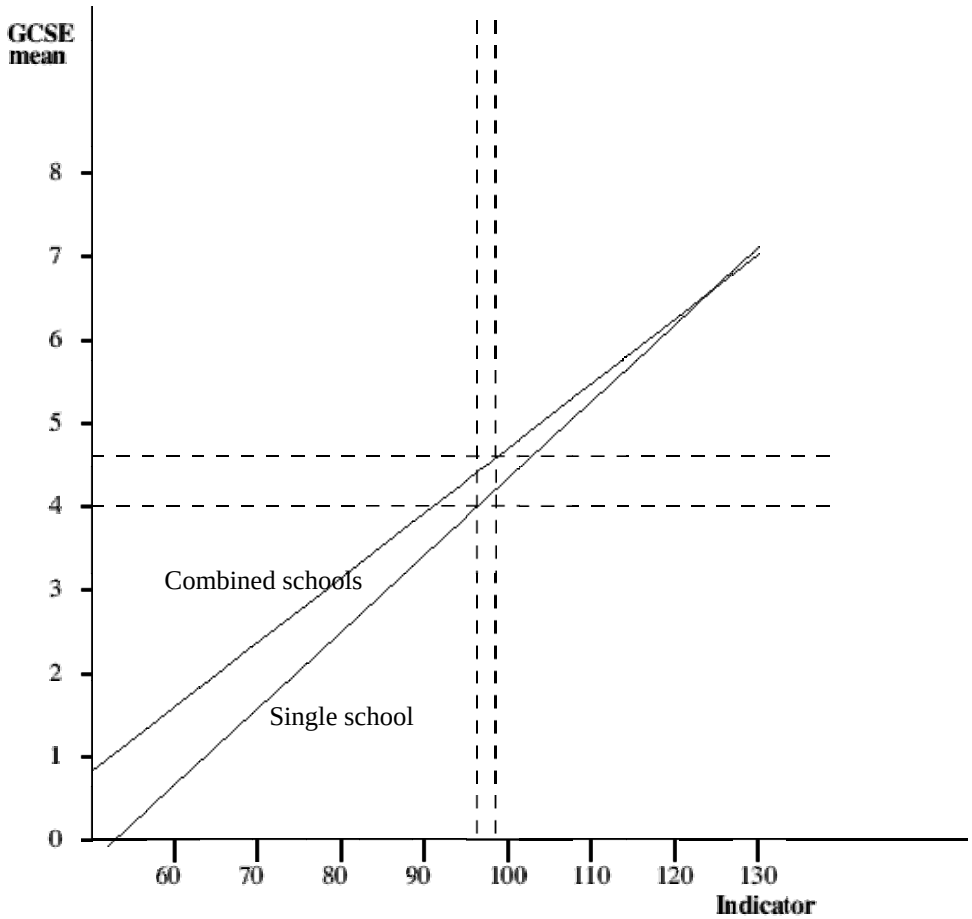
In 1995, however, the 63 boys had a mean ERT of 92.49 with a standard deviation of 14.22 and a GCSE mean of 3.38. The correlation between ERT and GCSE remained high at 0.73. In contrast, the 64 girls had an ERT mean of 99.92 with a standard deviation of 13.23. As a group they were much more able than the boys and obtained a GCSE mean of 4.64, a grade and a third higher than that of the boys.

Comparing the regression line graphs for the boys and the girls, that of the girls meets that of the boys when the ERT score is 70 but thereafter the gap between the two lines widens increasingly until when the ERT score is 130 the girls' line is well over a grade higher than that of the boys.

The distribution of pupil ability was such that in the boys' group just over 50% of the boys had ERT scores of 90 or less compared to just over 21% of the girls. There were considerably more girls, as a proportion of the group, capable of gaining higher GCSE grades compared to the boys.

**Figure 7.20 b**

**English Language Single School against Combined School sample 1995**




***Combined Schools' data***

| | | | |
|---|---|---|---|
| Number of pupils in the sample | 1593 | | |
| Mean for X is | 98.44 | Mean for Y is | 4.60 |
| Standard deviation for X is | 13.00 | Standard deviations for Y is | 1.46 |
| Covariance is | 13.11 | | |
| Coefficient of correlation is | 0.69 | | |
| Coefficient of determination is | 47.60% | | |
| Standard error of estimation for Y upon X is | 1.06 | | |

***Individual School's data***

| | | | |
|---|---|---|---|
| Number of pupils in the sample | 127 | | |
| Mean for X is | 96.24 | Mean for Y is | 4.02 |
| Standard deviation for X is | 14.22 | Standard deviations for Y is | 1.71 |
| Covariance is | 18.63 | | |
| Coefficient of correlation is | 0.77 | | |
| Coefficient of determination is | 58.93% | | |
| Standard error of estimation for Y upon X is | 1.09 | | |

The very much lower performance of the boys reduced the overall performance of the school's English Department as a whole both in terms of average GCSE grade and in terms of performance across the ability range as indicated by the regression line. The more able boys under-performed in relation to what would normally be expected of their ability expressed in terms of ERT score.
There is no obvious reason why the performance of the boys in 1995 in this school's English department, and indeed year cohort for the under-performance was reflected there also, was so much lower except perhaps that the preponderance of lower ability boys depressed the expectations of the more able boys or their teachers or both. This is an area which would merit more research into the effects of ability distribution within a teaching group upon their expectations and those of their teachers and the consequent effect upon actual results.

By way of illustrating the greater variance amongst smaller, non-core curriculum, GCSE subjects I refer to a number of schools' History results.
At the level of the combined schools' sample the data set is still reasonably large and the regression lines very similar from year to year ( see *Figure 7.21* ).

**Figure 7.21**

**Data from Combined Schools' History results**

| **Year** | **1996** | **1995** | **1994** | **1993** |
|---|---|---|---|---|
| *No. of pupils* | 874 | 613 | 664 | 444 |
| *Mean ERT* | 102.77 | 100.75 | 99.93 | 99.93 |
| *SD ERT* | 12.79 | 13.60 | 12.76 | 13.67 |
| *Mean GCSE* | 4.64 | 4.46 | 4.52 | 4.18 |
| *SD GCSE* | 1.83 | 1.89 | 1.80 | 1.77 |
| *Covariance* | 14.72 | 16.32 | 14.83 | 16.47 |
| *Coefficient of correlation* | 0.63 | 0.64 | 0.65 | 0.68 |
| *Coefficient of determination* | 39.63% | 40.48% | 41.67% | 46.18% |
| *Standard error of estimation* | 1.42 | 1.45 | 1.38 | 1.30 |

The number of pupils increased each year as more schools submitted their data for analysis. The mean ERT is around 100 in three of the four years shown, increasing to 102.77 in 1996 and the spread of ability is broadly the same as shown by the Standard Deviation for the ERT scores which ranges from 12.76

to 13.67. The mean GCSE ranges from 4.18 to 4.64, approximately a D grade at GCSE with the highest mean in 1996 as would be expected from the highest mean ERT score which is also in that year.

The coefficients of correlation are consistent in the range of 0.63 to 0.68 indicating that some 40% of the variance in GCSE mean score can be attributed to the ERT scores; slightly more in 1993 at 46.18%.
The standard error in estimating GCSE grades from ERT scores is in the range of 1.30 to 1.45.

At the individual school level, variation in examination results for History between schools is more apparent (See *Appendix H* for the individual school History GCSE regression line and scattergraphs in 1996).
School A's History department in 1996 had a mean GCSE score of 6.40 (B/A) in contrast to the combined schools' GCSE mean of 4.64 (C/D), a difference of almost two grades. School A's History department pupils had a mean ERT score of 108.87 with a standard deviation of 10.21 compared to the combined schools' mean ERT score of 102.77 with a standard deviation of 12.79. School A therefore had a considerably more able group of pupils and one would have expected them to achieve a higher GCSE mean.

When one compares the performance of School A's History department with that of the combined schools' sample in terms of like ability with like using the regression line graphs it can be seen that the line of School A runs parallel to that of the combined schools but just over a grade higher indicating that pupils of similar ability would be likely to gain higher grades in School A than on average in the larger sample. It should be noted, however, that the range of ability in School A is limited, only one pupil had an ERT score of below 90, and therefore this was an able group. The standard error of estimation in School A was 0.93, just under a grade, whereas the standard error of estimation for the combined schools was 1.42, almost a grade and a half, which would

indicate that School A's pupil performance in History was within the range expected, plus or minus one standard deviation, from the combined schools' data.

School L had a very similar average ERT score ( 108.13 with standard deviation of 13.07 ) to School A but a mean GCSE score of 4.36, just over a D grade. In comparison to the combined schools' sample the regression line for school L ran just under a GCSE grade below that of the larger sample with the gap narrowing slightly at the top of the ability range (130 on the x axis). Comparing the regression lines for schools A and L, the line for L runs almost two GCSE grades below that of School A. Whilst the performances of both schools are within a range plus or minus one standard deviation from the combined schools' sample they are both markedly different from each other with quite similar pupil intakes.

Using regression graphs in this way to show the relative effectiveness of one school with another, or one department with the average for all departments in that subject area, is very useful because the visual information can be taken in quickly and is more easily understood by the non-statistical teacher. An idea of performance across the range of ability is easily shown in contrast to trying to show such differences purely in numerical form.

Studying such graphs over a number of years one may be able to see patterns and trends. If a particular school maintains a small performance advantage over the average, one that is within one standard error of the mean for the larger sample, consistently year on year, then common experience would suggest that the school is better than the average even if the margin of its superiority is not statistically significant in individual years.

Fitz-Gibbon (1996), discussing A level performance, puts the case for practical interpretation of regression line data rather than relying purely upon statistical

statements,

> "*Are departments differentially effective to a substantive degree*? Note that the question posed above is not 'Are departments differentially effective to a *statistically significant degree*?' The sizes of differences which are important to schools cannot be determined *a priori* by reference to levels of statistical significance habitually employed in other studies, but will come to be understood over the years as people work with the data."

I believe that a common sense attitude is needed in interpreting examination data, one that takes account of the evidence presented but is not limited in vision by a blinkered adherence to statistical process. Those individuals working within schools and with performance data year on year will gain a feel for the figures and their interpretation within their own school's context. This intimate knowledge cannot be the same for those conducting large scale projects over short timescales.

**Improvements over time**

There have been very few studies conducted into school effectiveness over a prolonged period of time. Gray *et al.* (1996) refer to Teddlie and Stringfield (1993) as being the notable exception. This latter being a ten year study of elementary schools in Louisiana.

Stressing the importance of a longer term view on school improvement Gray *et al.* refer to Fullan (1991) concluding that: "significant change in the form of implementing specific innovations can be expected to take a minimum of two or three years; bringing about institutional reforms can take five or more years."

Gray and colleagues go on to report on their study of five examination cohorts passing through over thirty schools in one LEA. They find that approximately 25% of schools were improving significantly, some at a faster and some at a slower rate than other schools, over a period of five years but make the point that even in a large sample this is relatively few schools in these categories to consider. The statistical significance of any claims based upon such a small sample is therefore limited.

Gray's estimates for school effectiveness, as he points out, "are for the average pupil in each school; the picture would be slightly more complex if the performances of pupils who were at either the upper or lower ends of the achievement scale in each school were to be presented."

This approach to the analysis, using averages, precludes more detailed analysis of possible differential effectiveness for different ability pupils. Ten years earlier Gray *et al.* (1986) highlighted the problems with statistical analysis of school performance, particularly regression analysis and relying on mean values. This becomes directly relevant when they go on to discuss "The effects of changes in entry policy" of some schools and the not surprising finding that schools which increased the number of examination entries per pupil improved

their overall GCSE performance. (See Chapter 3 page 47 of this thesis.) Gray's reliance on "means" makes it impossible to ascertain whether the increase in entries per pupil was an increase across all pupils, regardless of ability, or an increase of entries for those more able pupils in order to maximise their contribution to the overall school statistics.

Blakey and Heath, in Reynolds and Cuttance (1992) also found that schools with more generous entry policies achieved better results than schools with comparable pupils but more restrictive entry policies, but again no evidence is given to show how the increased entry policy operated. (See Chapter 3 pages 56-57 of this thesis). Much of the improvement referred to is acknowledged to have occurred in a single year, 1991, when schools were responding to the pressures of nationally imposed "League Tables".

Gray acknowledges that school entry policies are not simple and reflect many issues in schools but it would have been helpful to know if the more able pupils were being entered for more examinations and the less able being entered for less. One might generously consider that schools were thus allowing the least able to concentrate on maximising their attainment in key areas whilst allowing the most able to stretch themselves. On the other hand a more cynical interpretation, in the light of the pressures which schools are under, also comes to mind, that potential failures were being minimised and the talents of the most able being exploited to raise the average number of grades A*- C per pupil.

Establishing whether schools have improved over time is far from simple even when considering a single school. No single indicator is sufficient because of the variables operating within and without the school vary from year to year. The average GCSE grade per pupil tends to track the average prior ability indicator score, as was seen in the graphs, Figures 7.15i to 7.15n. The composition of the school in terms of gender mix and the distribution of the

pupil ability within a given year cohort can have quite major effects on the average performance for the school as can be seen from graphs, Figures 7.15i to 7.15n, and in particular the case study for School X.

Analysis of school examination performance in relation to other schools, judgements on school effectiveness, must take into account factors such as the distribution of pupil ability within a school, within the respective genders in the school unit, and the relative numbers of each gender in the school. Researchers should consider using statistical techniques such as multi-level modelling to take account of these factors. Schools themselves, and the practitioners within them, can gain useful insights from the techniques illustrated in this thesis to compare like with like as much as is possible and highlight differences in the nature of the pupil samples where they occur.

The Schools Curriculum and Assessment Authority (SCAA) have attempted to look at 'O' level / GCSE grades over a period of years from 1988 to 1995 (SCAA, 1996a). This entails the analysis of around five million subject entries each year. Since 1988 there has been a marked increase in the number of pupils achieving five or more GCSEs. With such a large sample size one might expect the sample to be normally distributed in terms of ability and relatively stable year on year. However, a report on Standards in Education (SCAA 1996b) besides echoing the previously mentioned report's comments on the marked improvement of girls over the years since the introduction of GCSE notes changes in demographic trends which may impinge upon the nature of the examination cohorts,

> "Between 1972 and 1979, the proportion of legitimate live births in England and Wales to those in social classes I and II rose by a fifth, from 22.8% to 27.6%. Given an examination in which grades were not restricted by predetermined proportions (see below), these changes in the patterns of births might be expected to produce increasing proportions of higher grades at GCSE 16 years later. Between 1988 and 1995, the

> proportion of 16-year-olds in England achieving five or more GCSE grades C or better increased by over two-fifths, from 29.9% to 43.5%. The change in the pattern of births could provide part of the explanation for changes in GCSE and, possibly GCE A level results over recent years." (SCAA 1996b).

The above reference makes some rather large assumptions about the impact of social class on examination performance, the continued membership of social class categories by the parents after the birth of their children, the continued financial well-being of the parents in period of increasing unemployment, and the chances of those marriages / partnerships surviving in a period of ever increasing divorce / separation with concomitant effects on children.
Social class is in any case a poor indicator of academic attainment in relation to prior attainment (Thomas and Mortimore, 1996, amongst others), useful where information on pupil prior attainment is lacking. That said, if a national population can be subject to such changes over time then it is not surprising that the much smaller unit of the school shows the changes in the nature of its intake which I have observed with consequent changes in examination outcomes. The variation in schools' ERT scores, broken down by gender, their pupil numbers and mean GCSE scores over the years 1991 - 1996 were discussed in this on pages 172 - 185 and illustrated in Figures 7.15 a-n.

Comparing a school's examination performance with national figures over time is, therefore, likely to be far less useful than it would at first appear, because of the lack of a common baseline measure against which to measure the performance of all pupils. The annual publication of national results can provide benchmark examination figures against which to compare the individual school's figures, such as percentages of GCSE grades obtained, but, as the ability of the school cohort varies from year to year, true performance, examination results in relation to ability of the year cohort, cannot be

established for a national sample lacking baseline information. An individual school therefore has no means of comparing its true examination performance except by joining a consortium of schools with similar baseline indicator information.

Over a number of years I have built up tables of data pertaining to Sexey's School as shown in *Figure 7.22*. The thinking behind this was to show the interaction of a number of variables operating within the school.

**Figure 7.22**

**GCSE RESULTS at Sexey's School**

| Year | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|---|---|---|---|
| No. of candidates | 46 | 51 | 56 | 48 | 43 | 43 | 40 | 45 | 43 | 45 |
| No. of subjects sat | 402 | 454 | 503 | 435 | 420 | 423 | 380 | 434 | 360 | 385 |
| Average entries per cand. | 8.83 | 8.96 | 9.04 | 9.08 | 9.88 | 9.86 | 9.50 | 9.67 | 8.42 | 8.56 |
| Average passes per cand. | 5.50 | 5.63 | 5.52 | 6.12 | 6.47 | 7.70 | 6.78 | 6.93 | 5.93 | 6.73 |
| Year mean score | 4.42 D/C | 4.81 C/D | 4.60 C/D | 4.98 C | 4.99 C | 5.27 C | 4.98 C | 5.11 C | 5.05 C | 5.31 C |
| Highest mean score / pupil | 6.56 | 6.67 | 6.78 | 7.00 | 7.00 | 6.70 | 6.70 | 7.00 | 6.67 | 7.22 |
| Lowest mean score / pupil | 0.43 | 2.12 | 1.00 | 1.78 | 1.50 | 2.11 | 1.25 | 1.71 | 2.40 | 3.47 |
| % 5+ A*-C | 65 | 69 | 64 | 65 | 65 | 86 | 78 | 80 | 63 | 84 |
| % passes ( A-C ) | 62.94 | 63.22 | 61.43 | 67.59 | 66.19 | 78.25 | 71.32 | 71.89 | 70.83 | 78.70 |
| E.R.T. mean | 108.12 | 106.57 | 103.07 | 106.95 | 103.29 | 105.92 | 104.17 | 105.66 | 99.62 | 102.64 |
| Standard dev. (GCSE mean) | 2.05 | 1.44 | 1.55 | 1.43 | 1.52 | 1.25 | 1.32 | 1.41 | 1.17 | 1.17 |
| Spearman rank correl. | 0.62 | 0.75 | 0.65 | 0.54 | 0.59 | 0.74 | 0.73 | 0.67 | 0.67 | 0.49 |
| Pearson moment correl. | 0.64 | 0.75 | 0.71 | 0.57 | 0.61 | 0.70 | 0.76 | 0.69 | 0.67 | 0.45 |

This table is not all inclusive by any means and could be added to by giving more data on the respective genders. It does, however, show how the average ERT score fell from a peak of 108.12 in 1987 to a low of 99.62 in 1995 before rising again to 102.64 in 1996. Despite the low ERT score in 1995 the mean GCSE grade was still higher than in 1987 ('O' level), the percentage of pupils gaining five or more grades at C or above was lower in 1995 than 1987 but only by 2% and the percentage of examinations awarded C or above was higher in 1995 than in all the years from 1987 to 1991.

What is not shown on this table is the action that was taken by the school in response to the information on the ERT scores of the pupils who were going to take their GCSE examinations in 1995. Prior knowledge allowed the school to tailor the curriculum and option choices available to suit the nature of the candidates coming through the system. Support was given to less able candidates and they were not compelled to follow the curriculum which was suited to more able pupils. The number of GCSE examinations they had to sit was reduced allowing them to have extra time on the core subjects so avoiding needless failures, extra stress and allowing them to maximise their chances in the subjects they did study. Despite the lower headline figures, as a school we were very pleased with the performance of our pupils and the overall results. We as a staff knew the pupils' strengths and weaknesses and were able to celebrate their success at the appropriate level for them.

Setting targets for a school based upon the examination performance of the previous few years, some sort of rolling average, is not appropriate when the nature of the examination cohorts is changing within the school. There is little point in seeking to improve upon the average examination performance of the last three years, (GCSE mean, percentage of examinations graded A*-C, percentage of pupils gaining five or more GCSEs at grade C or above, or whatever), if the ability of the examination cohorts is falling.

A more productive way of looking at schools and whether they are improving is by considering their regression line graphs with scattergraphs to show the individual pupils. Using this method one may see at once on the same piece of paper the distribution of the pupils by ability and outcome, with the proviso that some pupils may be superimposed on top of each other, the mean indicator (vertical dotted line) and outcome (horizontal dotted line) scores for the cohort, the average performance of the cohort across the ability range (diagonal solid line) expressed as the regression line, and the performance of individual pupils who were particularly good or poor for their prior ability scores are obvious as

outliers on the scattergraph and can be identified by those working in the school.

As the regression line reflects performance across the ability range it is more representative of the school's performance than the mean GCSE score. This allows schools to compare their regression line with that of a larger consortium sample or against their own results for previous years.
This same sort of comparison can be done for individual subjects.

On the following pages I include some regression graphs for six individual schools showing the 1991 data and then their 1996 data. It can be seen, in the way that the regression lines for 1996 are generally steeper and reach a higher point on the GCSE mean axis, that for most of the six schools the results are a considerable improvement on those of 1991, particularly in the improved results of the higher ability candidates, those with ERT scores in excess of 110.

In my initial discussions with the Heads or Deputies of schools which became involved in this analysis it was very common, when studying the initial analysis, for them to remark that they had not realised certain pupils in their schools with high ERT scores were so able and obviously under performing. Looking at Figures 7.23a, 7.24a, 7.25a, 7.26a, 7.27a, 7.28a it can be seen that each school had a number of candidates, represented by the asterisks, with high ERT scores but lower than average GCSE mean outcomes. These candidates are shown in the bottom right hand quadrant of the graphs as being above average in ability and below average in attainment for the school. School B in particular had a large number of such candidates. In 1996 all schools had reduced the number of candidates in this sector and brought the performance of candidates still in this sector nearer to the average outcome line, represented by the horizontal dotted line which is also higher in most schools shown for 1996 than it was in 1991.

As the use of the analysis information was implemented in schools, in reviews of examination performance Heads of Subject were also called to account for the performance of pupils in relation to prior ability. This issue has clearly been addressed and able pupils encouraged to aspire to high GCSE results, in a number of schools the difference between their GCSE mean when starting to use the analysis information and their 1996 results is in the order of just over a grade per pupil across all their GCSE subjects, which constitutes a considerable improvement.

Importantly, in many schools, not just those illustrated here, the improvement has been not only at the top of the ability range but across the full range of pupil ability. This has resulted in higher average GCSE scores for the schools, in some cases despite falling average ERT scores.

Whether these schools which have been using the analysis information I provide them with have improved at a greater rate than the improvement seen nationally at GCSE level is difficult to prove because of lack of comparable baseline data against which to judge improvement. Nationally the percentage of examination entries graded A*-C has risen from 48.5% in 1991 when there were 4,947,593 subject entries to 54.0% in 1996 when there were 5,475,872 subject entries. Schools are under considerable pressure from public expectations to match these figures or at least the improvement they show.

As I have shown, because of the variability in pupil ability in individual school pupil cohorts, even a drop in a school's overall percentage of grades A*- C may represent a real improvement in terms of performance relative to pupil ability. A measure such as percentage of subject entries graded A* - C ignores the achievement of pupils gaining grades D - G which may be more relevant to some schools and their pupil intake.

Certainly in relation to the performance of those schools which have joined in

most recently, those schools which have been involved longest are more effective. Some evidence for this has already been shown in *Figures 7.8 & 7.10.*

It remains to be seen if the findings of the Value Added National Project, looking at Key Stage 3 average results as an indicator of potential GCSE success, will be able to provide the necessary baseline data for a value-added comparison between schools on a national level (Fitz-Gibbon, 1995, Trower and Vincent, 1995, Fitz-Gibbon, 1997).

**Figure 7.23a**

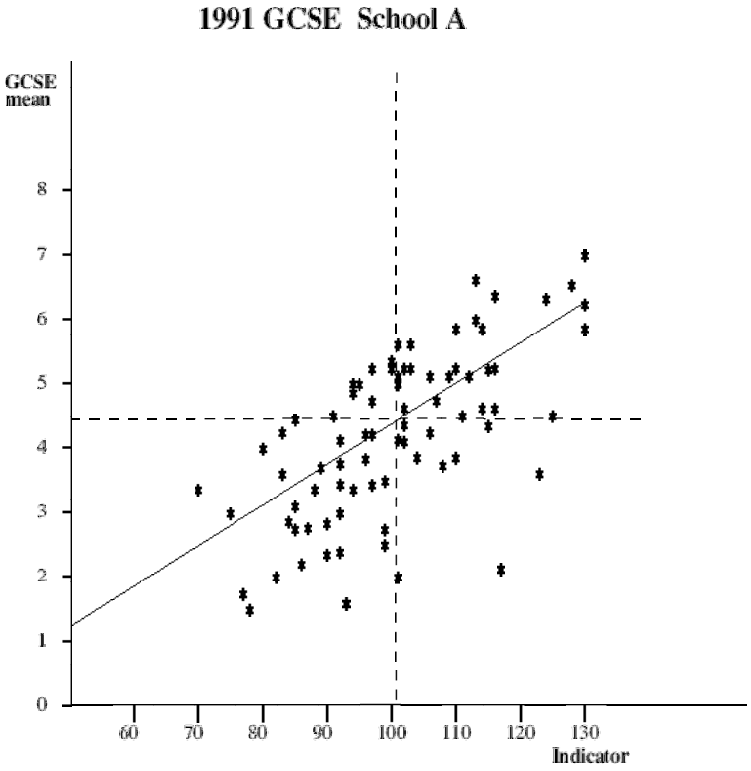



**Figure 7.23b**

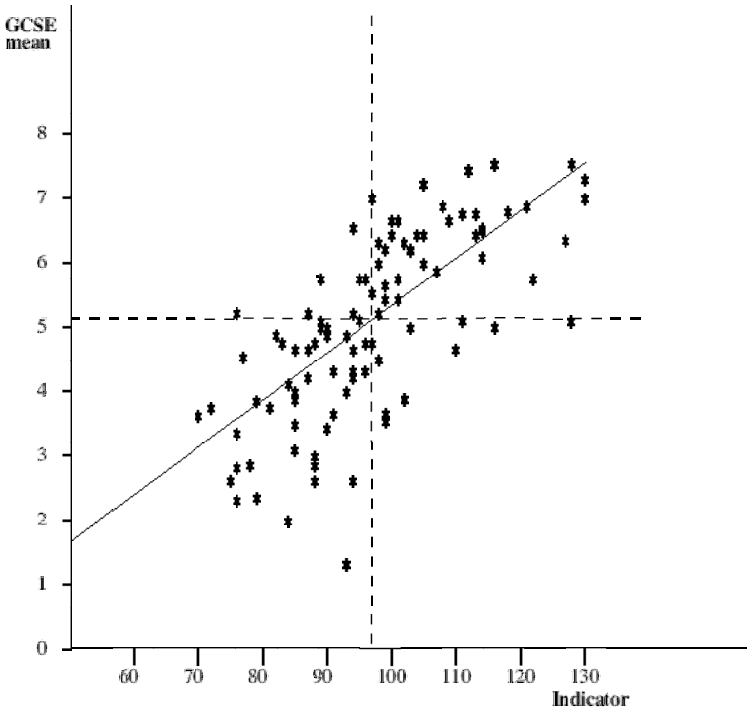


**Figure 7.24a**

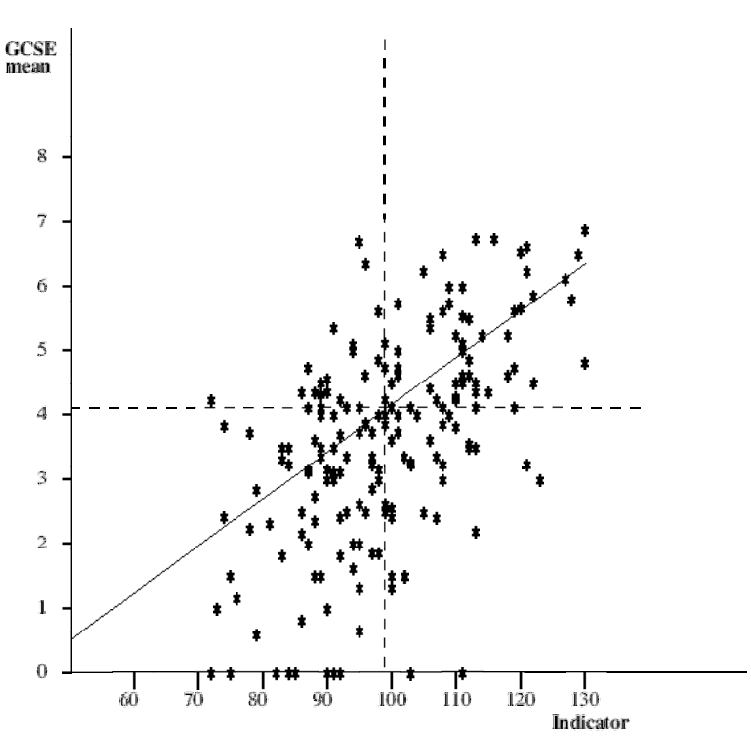



**Figure 7.24b**

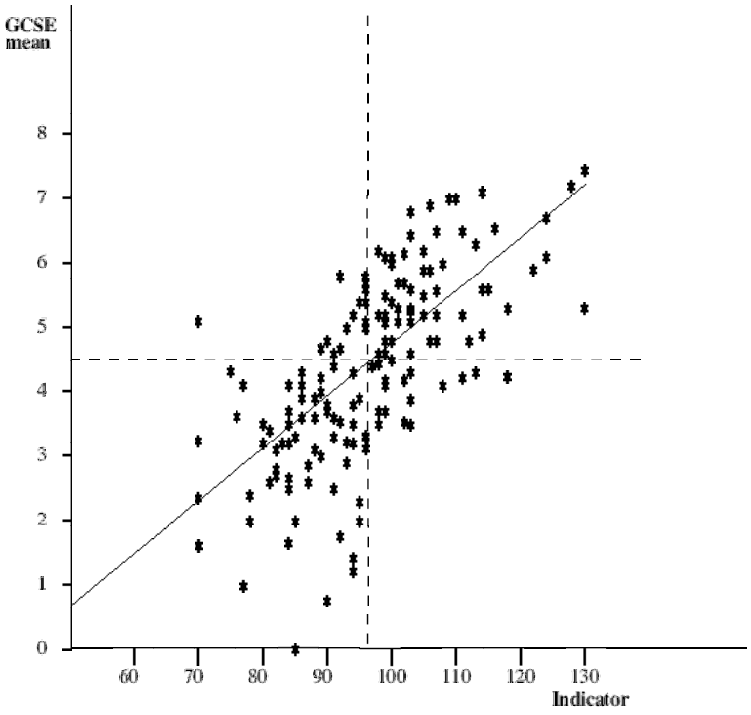


**Figure 7.25a**

1991 GCSE School C

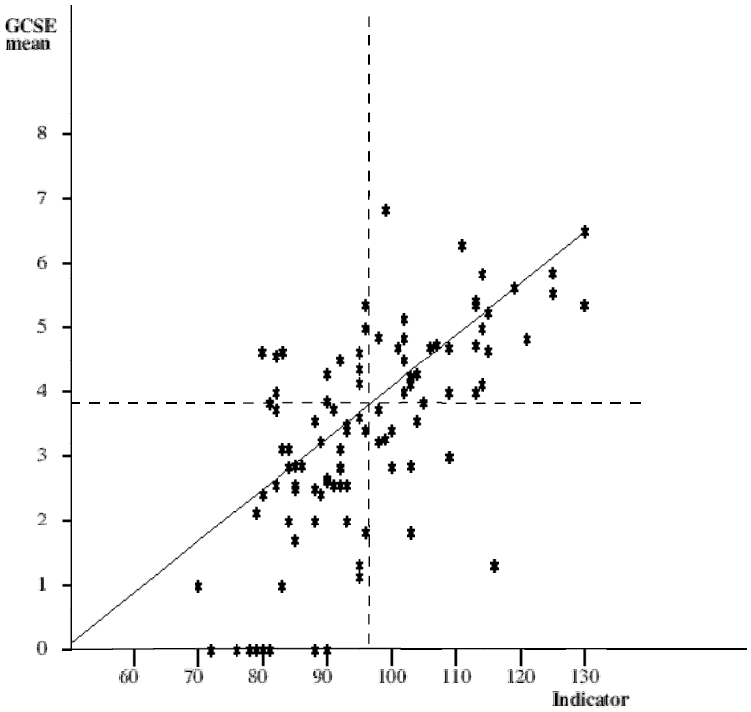



**Figure 7.25b**

1996 GCSE School C

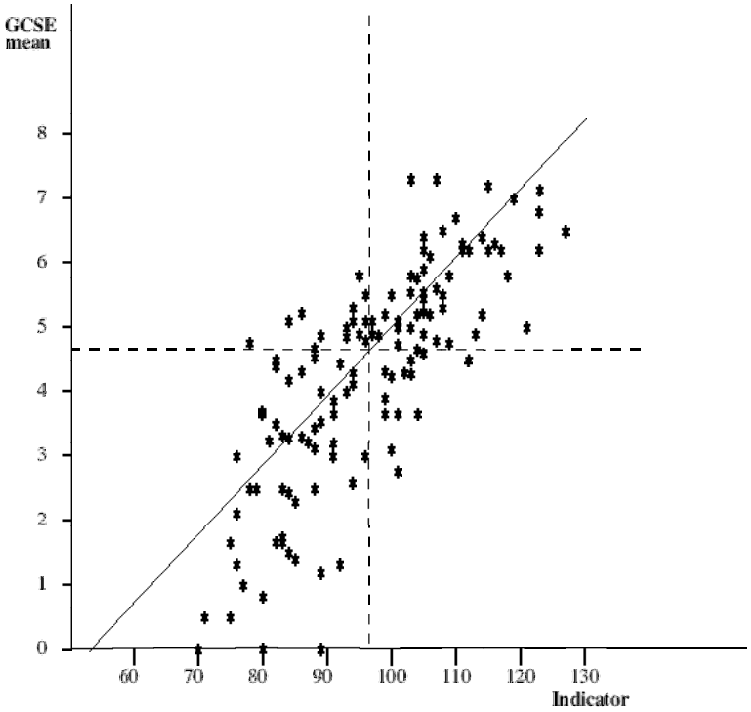


**Figure 7.26a**

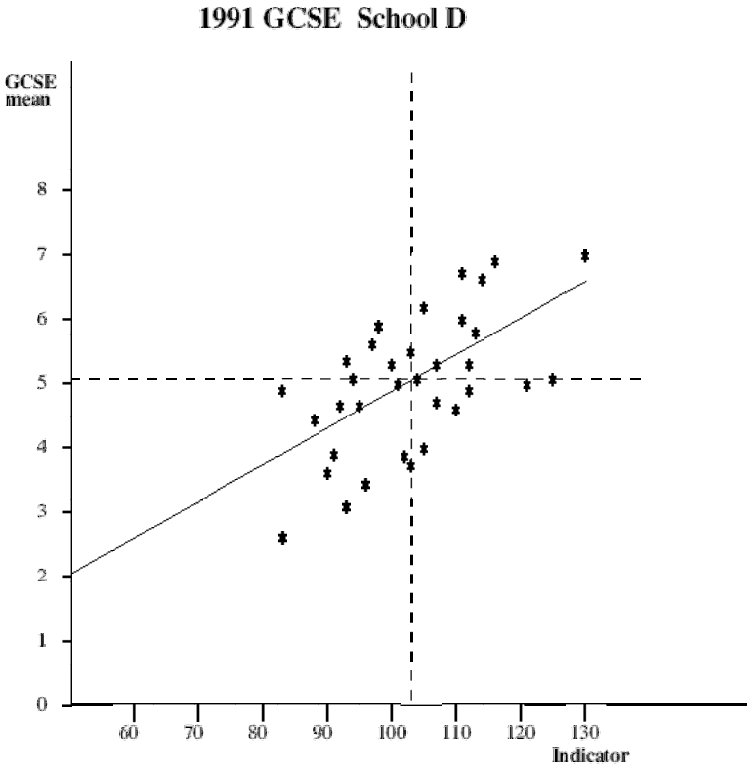



**Figure 7.26b**

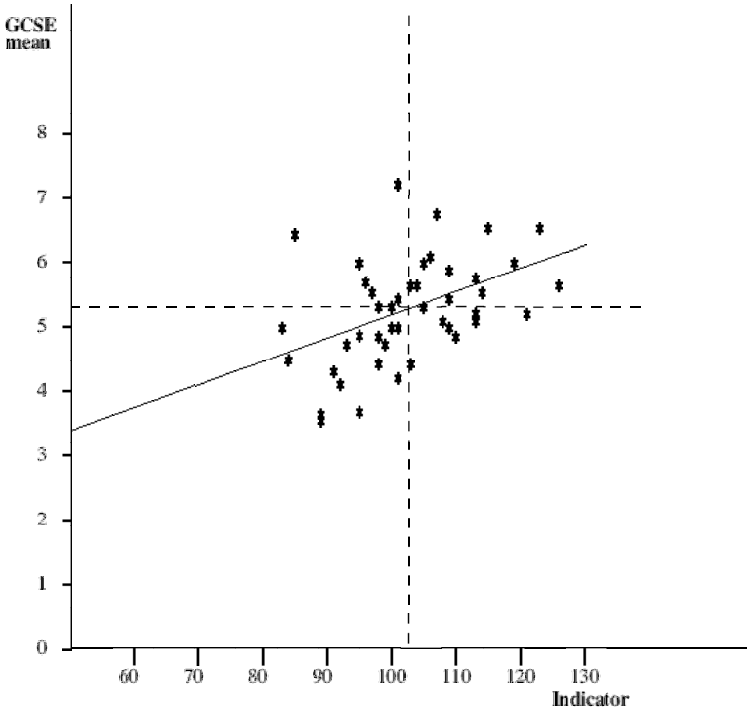


**Figure 7.27a**

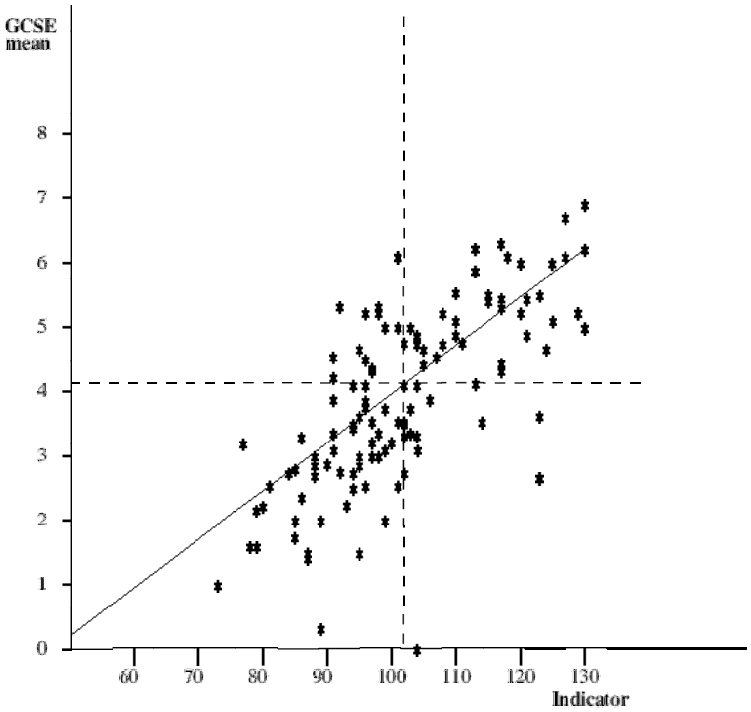



**Figure 7.27b**

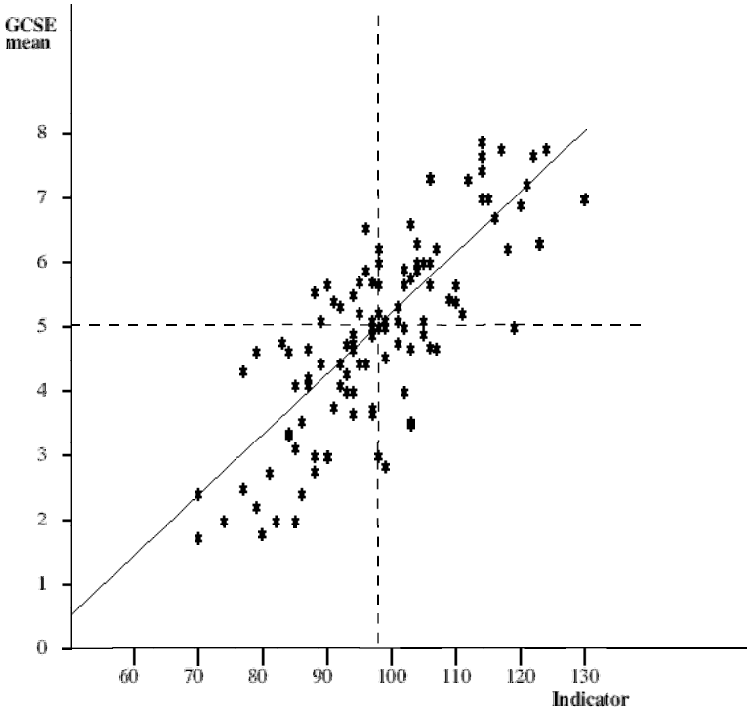


**Figure 7.28a**

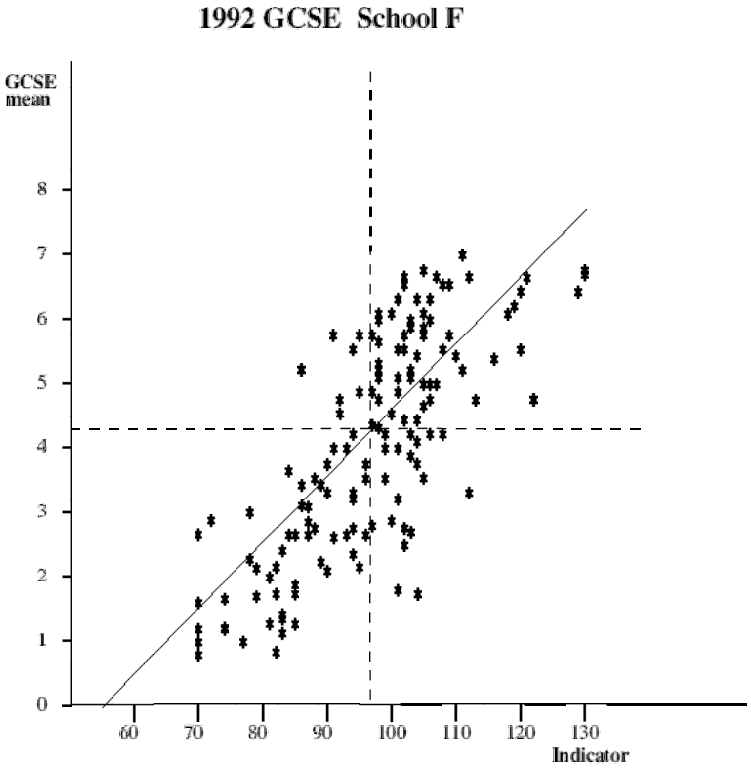



**Figure 7.28b**