Chapter 6 Case Studies

These case studies have been included in this thesis as illustrations of the utility of my research in the context of schools and practical work with teachers using examination results as a medium for considering school and department effectiveness. They arose from requests for help by the senior management of the schools in response to issues affecting their school. In the case of School X, and other similar requests, I was asked to help because I had most of the necessary examination data already, was experienced in looking at examination data and could make comparisons against larger data sets, plus I was an external agency and so free from the politics, personalities and general pressures acting upon a member of staff attempting such an analysis.

The case study involving School X illustrates some of the pressures operating upon the staff in schools, particularly the Senior Management, as Governors, themselves accountable for the actions of the school, seek to understand the reasons behind a particular set of examination results. This case study also highlights the complexity of apparently simple data. Even when considering examination results in terms of the ability of the pupils, which as yet the Government Performance Tables fail to do, one must also consider the distribution of that ability amongst the year cohort.

The second case study is useful because it highlights the concerns of ordinary teachers about statistical attempts to quantify the performance of their pupils, the "professional phobias" discussed in chapter 3 of this thesis. The particular teacher concerned is very experienced, well respected by his colleagues and generally perceived as gaining good results from his candidates. He had specific concerns about the nature of his subject which are mirrored to a greater or lesser degree by many teachers in their own subject areas when first coming to terms with this form of analysis.

The implications of using correlation statistics and their interpretation for

teachers who are not familiar with statistics were highlighted as general problems which in turn can hinder the implementation of any action shown as necessary by the analysis of the examination data.

In the second case study, based at my own school, whilst maintaining an element of detachment in my analysis, I have the benefit of "internal knowledge" of the member of staff concerned, having been his colleague for many years. This brings with it the understanding of a teacher's antipathy to statistics and a perception of his real concern for the success of his pupils and subject department.

Case study 1

School X and gender differences in performance

I was contacted by the Headteacher of this school shortly after the release of the 1994 GCSE examination results. Following a couple of years, 1992 & 1993, when the performance of the boys in relation to the girls appeared to be considerably worse. The governors of the school were concerned that yet again in 1994 the performance difference between boys and girls appeared great. (See *Figure 6.1*). In 1992 despite almost identical indicator scores 20% more girls than boys achieved five or more GCSEs at grade 'C' or above. In 1993 this difference in performance was repeated with 20% of boys and 49% of girls

Figure 6.1

School X Percentages 5+ A*-C and ERT scores

Year		All pupils	Boys	Girls
1995	5+A*-C	45%	44%	47%
	ERT	94.34	94.19	94.52
1994	5+A*-C	42%	33%	52%
	ERT	94.71	91.93	98.19
1993	5+A*-C	32%	20%	49%
	ERT	96.21	95.62	97.03
1992	5+A*-C	39%	29%	49%
	ERT	96.73	96.73	96.72
1991	5+A*-C	26%	23%	29%
	ERT	99.42	100.85	98.12

achieving five or more GCSEs at grade 'C' or above.

My task was to analyse the results of the school and see if I could find some reason in the data for there being a 19% difference in the percentages of boys and girls gaining five or more GCSE grades at 'C' or above, despite the school's best efforts to do something about the performance of boys.

My findings were as follows:-

a) In 1994 the average ability of of the girls, as judged by their Edinburgh Reading Test (ERT) scores was 98.19 which was some 6.26 points higher than the boys at 91.93. It is not unusual for schools to have the two genders with different ERT score averages because of the variation in the ability of the pupil intake in a comprehensive school in any given year. The ERT standardises both genders on the same basis and therefore one would logically expect the girls in this year group to do better than the boys given the strong correlations for ERT and GCSE success (See *Appendices A & B* for correlation figures).

b) Given that nationally girls of similar ability to boys are out-performing them at GCSE level (See Hedger and Raleigh, 1990 and SCAA, 1996, as discussed earlier in this thesis) then this gap in the performance of boys and girls at School X would be compounded.

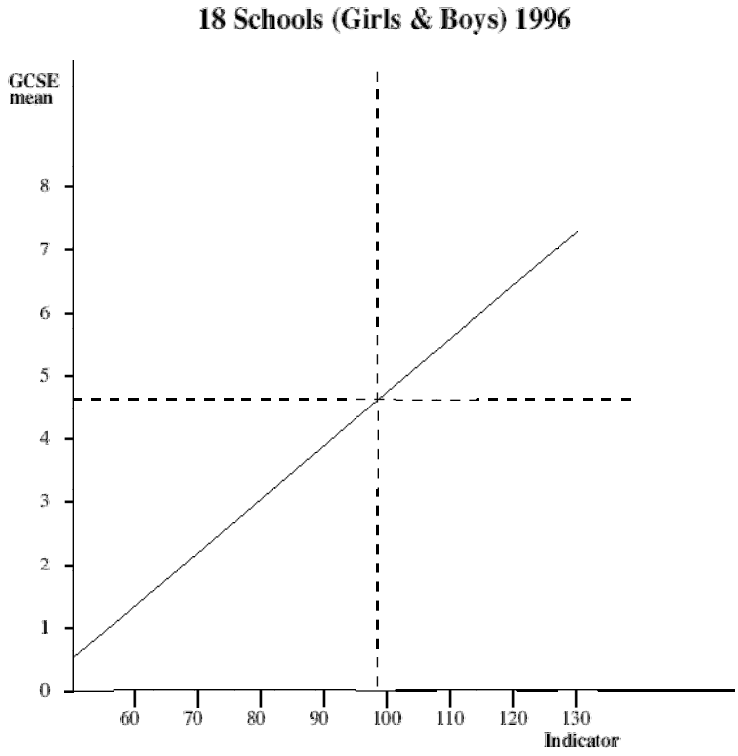
In 1994 nationally, according to DFEE annual examination results statistical bulletins, 47.8% of girls and 39.1% of boys in the Year 11 cohort achieved five or more GCSE grades at A* - C level. In 1993 the figure for girls was 45.8% and for boys 36.8. In 1992 the figures were 42.7% and 34.1% respectively. These figures confirm that nationally boys do less well than girls in GCSE examinations, at least in the higher grades A*- C.

c) Empirical evidence from my research into the correlation between ERT and GCSE (see *Figure 6.2* for the graph of the combined schools' sample, boys and

girls, for 1996 and Appendix G for examples of regression line graphs for each year of the research) would suggest that a minimum ERT score is necessary in the majority of cases for pupils to acquire an average GCSE score of above a 'C', this figure being lower for girls and higher for boys. At an average ERT score of 98.19 many of the girls were likely to be over this threshold figure and therefore capable of gaining 'C' grades whereas the majority of the boys were likely to be below the threshold figure and therefore unlikely to gain 'C' grades even if they performed well.

Figure 6.2

Regression line for GCSE upon ERT



Number of pupils in the sample	2834		
Mean for X is	98.42	Mean for Y is	4.63
Standard dev. for X is	12.90	Standard dev. for Y is	1.51
Covariance is	14.09		
Coefficient of correlation is	0.73		
Coefficient of determination is	52.68%		
Standard error of estimation is	1.04		

It is not possible to give an exact ERT threshold score beyond which pupils with higher scores are guaranteed to gain an average GCSE grade in excess of 'C'. However, using the combined schools' sample of pupils (Boys & Girls) with ERT scores and GCSE results for 1996 and the regression line equation

for GCSE mean upon ERT score, an ERT score of 103 equated to a GCSE mean of 5.02 in a sample size of 2834 pupils with a standard error of prediction of 1.04 or just over a grade either side of the predicted 'C' grade.

Using the 1996 data for boys only an ERT score of 104 predicted a GCSE average grade of 4.98, just under a 'C' grade with a standard error of prediction of 1.07. Looking at the girls only sample an ERT score of 101, some three points lower than the boys, predicted a GCSE mean of 4.97, again just under a 'C' grade, with a standard error of prediction of 0.98.

In 1995 (sample size 1630) and 1994 (sample size 1487) an ERT score of 103 predicted GCSE means of 4.94 and 4.97 respectively with standard errors of prediction of 1.03 and 1.08 for boys and girls combined.

An ERT score of around 103 for a mixed gender sample would therefore seem to be a general guide to a threshold figure above which pupils could reasonably be expected to achieve a GCSE average grade of 'C' or better. A slightly lower figure for girls and higher for boys. (See *Figure 6.3*).

Figure 6.3

ERT scores and predicted GCSE mean grades

Year		ERT	GCSE mean	Sample size	Standard error
1996	All	103	5.02	2834	1.04
	Boys	104	4.98	1420	1.07
	Girls	101	4.97	1414	0.98
1995	All	104	5.03	1630	1.03
	Boys	105	4.97	818	1.03
	Girls	102	4.97	812	1.01
1994	All	103	4.97	1489	1.08
	Boys	105	4.99	761	1.05
	Girls	102	5.01	728	1.08

This scenario, whereby pupils of differing ability are not evenly distributed within a year cohort or by gender, is yet again evidence of the fallibility of using the percentage of pupils achieving five or more GCSE grades of 'C' or above as an indicator of the school's performance. The national performance

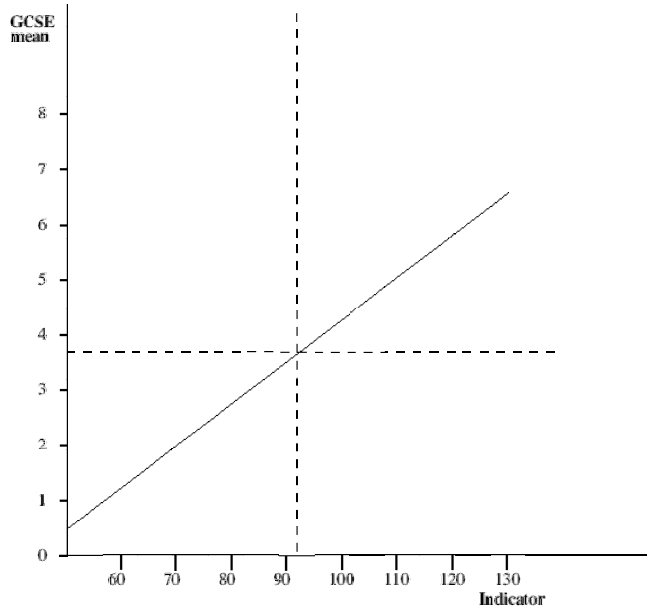
tables take no account of gender and the different performance of each gender nor pupil ability nor the distribution of pupil ability within schools.

d) The average GCSE grade for boys in School X was 3.69, a grade equivalence of E/D, compared to the girls' 4.46 or grade D/C. Whether these figures were good or bad, or whether the real gap in performance between girls and boys was larger than it should be, could be ascertained by comparing the average outcome scores in School X with what might be expected from their indicator scores using the regression line graphs for the larger sample of combined schools. In this way it would be possible to compare the expected performance of similar pupils in the large sample with what was achieved in School X. (See *Figures 6.4 - 6.7*).

When this was done it was found that the regression line for School X boys was fractionally below that for the twelve schools combined; it was close enough to be almost identical and certainly well within the standard error of estimation at just over a grade above or below the regression line. By the same token the girls' graph was again almost identical to that of the larger sample. From this I deduce that the performance of boys and girls at School X, as indicated by the regression line graphs, was not significantly different from what one would expect for pupils of their ability.

Figure 6.4

School X 1994 GCSE Boys only

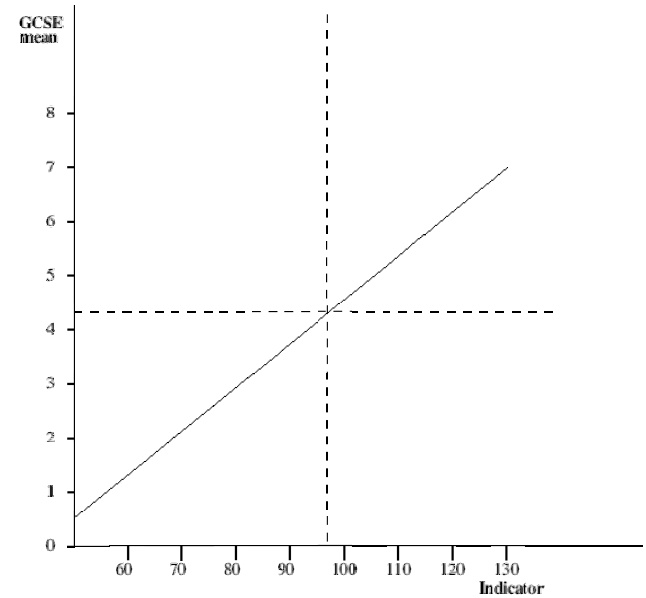


Number of pupils in the sample	74	
Mean for X is	91.93	Mean for Y is 3.69
Standard deviation for X is	13.80	Standard deviation for Y is 1.45
Covariance is	14.53	
Coefficient of correlation is	0.72	Coefficient of determination is 52.38%

Standard error of estimation for Y upon X is 1.00

Figure 6.5

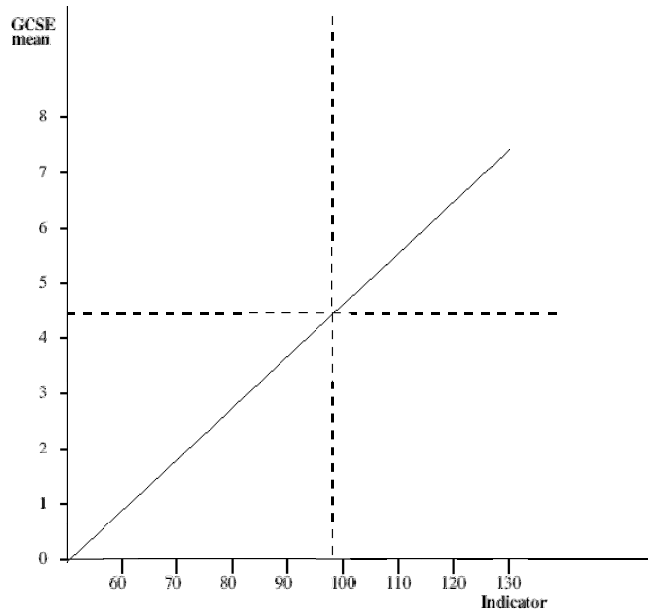
12 Schools GCSE Boys only



Number of pupils in the sample	761	
Mean for X is	97.00	Mean for Y is 4.34
Standard deviation for X is	13.89	Standard deviation for Y is 1.54
Covariance is	15.64	
Coefficient of correlation is	0.73	Coefficient of determination is 53.30%

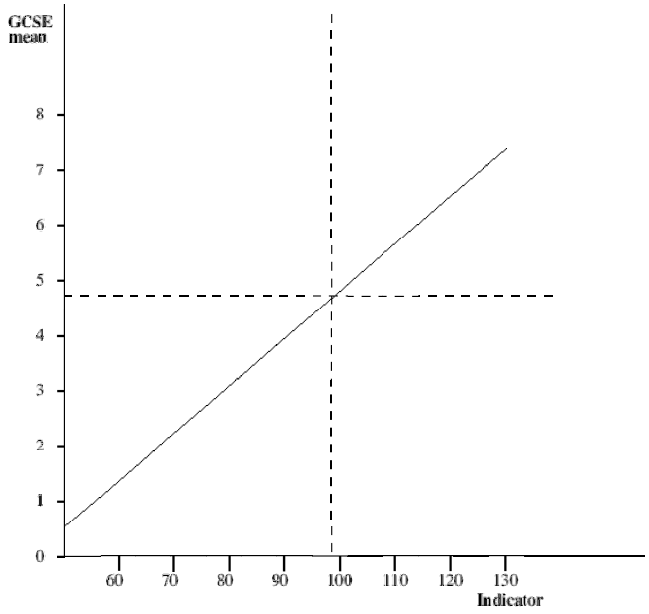
Standard error of estimation for Y upon X is 1.05

Figure 6.6
School X 1994 GCSE Girls only



Number of pupils in the sample	59	
Mean for X is	98.19	Mean for Y is 4.46
Standard deviation for X is	11.68	Standard deviation for Y is 1.50
Covariance is	12.73	
Coefficient of correlation is	0.73	Coefficient of determination is 53.05%
Standard error of estimation for Y upon X is 1.03		

Figure 6.7
12 Schools GCSE Girls only



Number of pupils in the sample	728	
Mean for X is	98.59	Mean for Y is 4.72
Standard deviation for X is	12.14	Standard deviation for Y is 1.50
Covariance is	12.64	
Coefficient of correlation is	0.69	Coefficient of determination is 48.12%
Standard error of estimation for Y upon X is 1.08		

e) The distribution of ability by gender is also important. By plotting the frequency distribution of various ability bandings one can gain some idea of the actual ability spread for each gender rather than simply relying upon a mean figure which could in itself be misleading.
(See *Appendix G, "Combined Schools' samples 1996 - 1992",* for correlation and distribution graphs for combined schools with ERT information by year and by gender).

By looking at the two graphs following (*Figures 6.8a & 6.8b*), one can see quite easily the difference in the spread of ability, as indicated by the pupils' ERT scores and the strong correlation between these and success at GCSE level, between the girls and boys at School X in 1994. Almost 30% of the boys had ERT scores of over 100 compared to the girls where almost 36% had scores of above 100.

An ERT score of 100 is a good score to use as a benchmark, for the majority of pupils with this score or above will have a good chance of achieving an average C grade or better in their GCSE examinations whereas those pupils with scores of less than 100 are unlikely to average C grades in their GCSE examinations.

In the larger sample for twelve schools with ERT information in 1994 the percentage of pupils with ERT scores of more than 100 was almost 40% for boys and 42% for girls. By this benchmark the averages for the pupils of both genders at School X were less able than the averages for the twelve schools, some 10% less of the boys being above the critical ERT score of 100 and 6% of the girls.

Furthermore, at School X almost 23% of the boys were in the range of ability below an ERT score of 81, a point below which Somerset LEA would consider pupils merited Special Educational Need support, whereas only just over 3% of

the girls were in this banding. For the larger sample of twelve schools, which included School X, the percentage of boys with ERT scores below 81 was 12.88% and the figure for girls was 6.04%. In this particular low ability range there were considerably more boys as a percentage of the school population at School X, almost 10% more, than in the larger sample whereas the percentage of girls in this banding was less than the larger sample, just over 3% less. This breakdown again emphasises the disparity in the ability of the two genders in School X.

To emphasise just how important it is to consider the distribution of ability within a school, in 1995 the percentage of boys with an ERT score of greater than 100 at School X was even lower at almost 27% (See *Figure 6.8c*) and yet the school achieved much better results than the previous year, averaging a third of a grade per pupil better across all GCSEs taken. The average GCSE grade for the school in 1994 was 4.03, just above a D grade; in 1995 the average GCSE grade was 4.30, almost a third of a grade higher but still below a C grade. In 1995, however, only 12½% of the boys had ERT scores of below 81 compared to 23% in 1994 and yet the mean ERT for the whole school, girls and boys combined, at 94 was virtually identical for the two years - 94.71 in 1994 and 94.34 in 1995. Merely considering average ability for the year group would have hidden the large differences in the makeup of the year group. In 1994 there were differences in the average abilities of the two genders and different distributions of ability within the two genders. In 1995, although the average ability for the year group remained much as in 1994, the ability of the boys improved both in the average ERT score (91.93 in 1994 to 94.19 in 1995) and in the reduced proportion of less able candidates whereas the average ability of the girls fell from 98.19 in 1994 to 94.52 in 1995 (See *Figures 6.8b & 6.8d*).

Important factors such as the distribution of pupil ability within schools do play a key part in the overall performance of the school year cohort. This is

illustrated by the example of School X but as these important changes in year cohort composition from one year to the next still did not raise the average ability of the year cohort to a level where the pupils were likely to achieve C grades, then the indicator figures as used by the Government to compile performance figures, the number of pupils achieving five or more grades at C or above, are not going to reflect these changes in the nature of the ability of the school cohort or the performance of those cohorts.

Research into school effectiveness should consider the distribution of pupil ability within schools rather than just the "mean on mean" approach, comparing mean indicator score with mean outcome score, but such analysis is not apparent in much of the research literature.

On the basis of my findings, I was able to reassure the Headteacher and Governors of the school that the apparent gross disparity in the performance of the boys and girls at School X could be explained in relation to the respective abilities of the genders and there was no need for any drastic change in policy, other than to take a more analytical approach in comparing the relative abilities of the two genders. This advice was given further support by the 1995 GCSE results for the school which saw the gap between the boys' average grade per pupil and the girls' average grade per pupil narrow from 0.77 in 1994 to 0.21 in 1995 as the abilities of the two groups became much closer.

Figure 6.8a

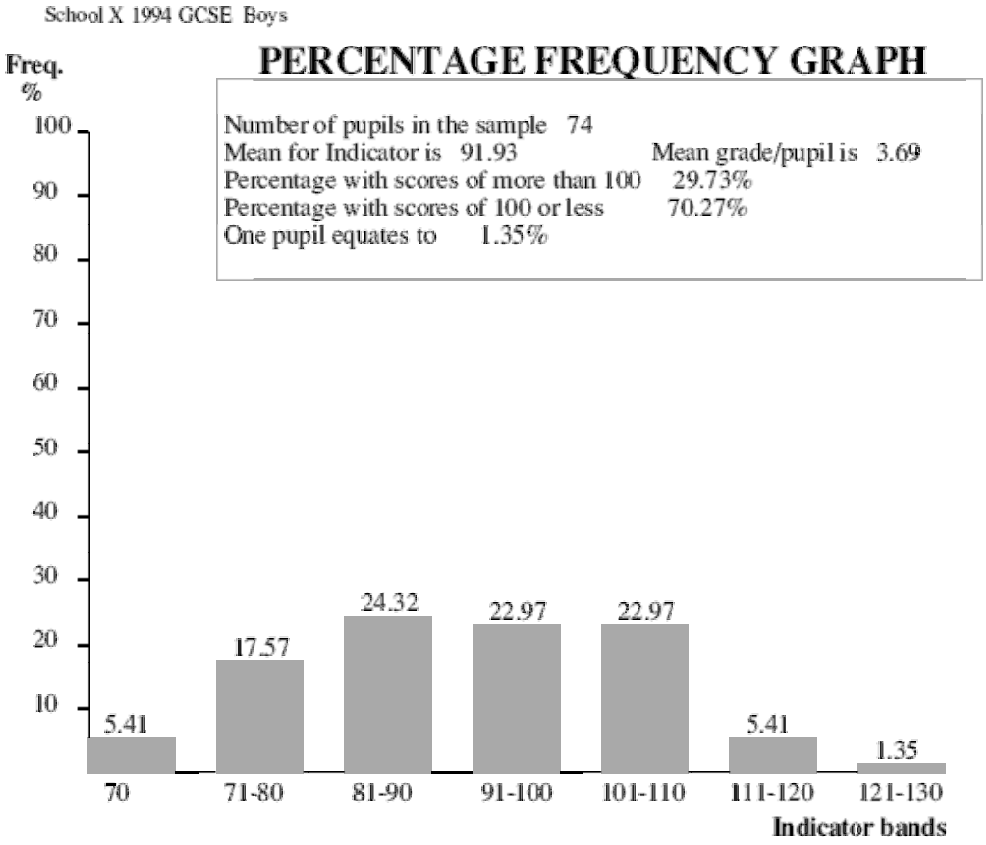


Figure 6.8b

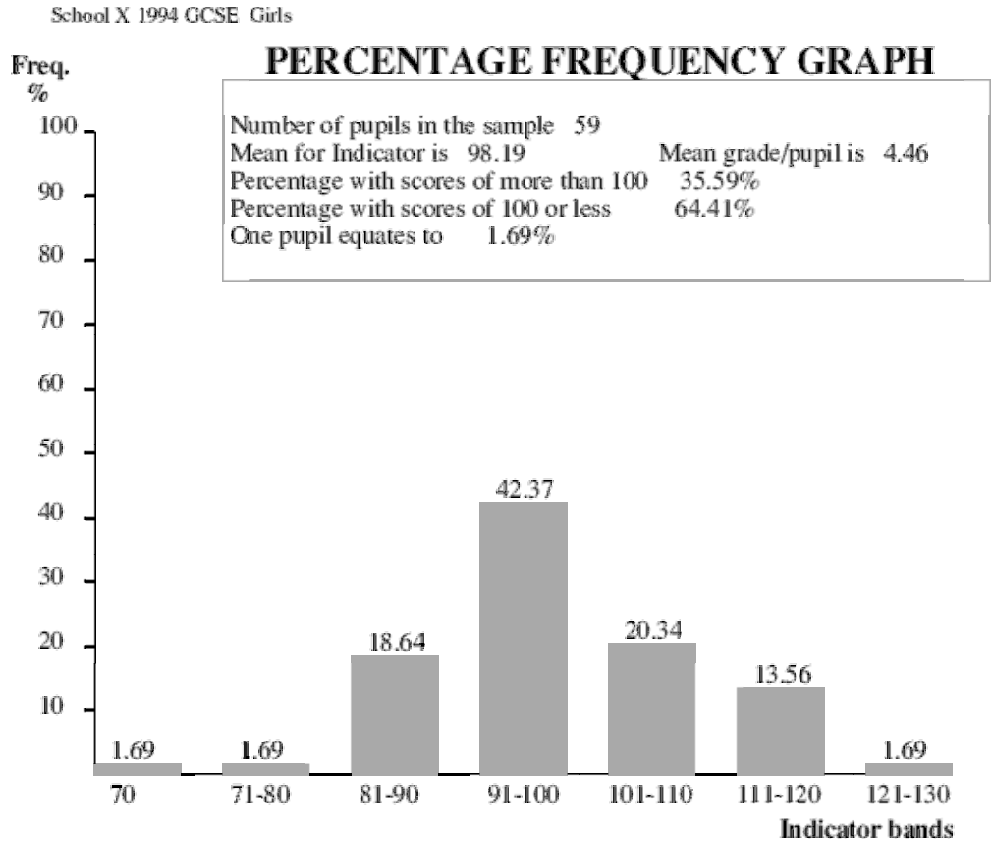


Figure 6.8c

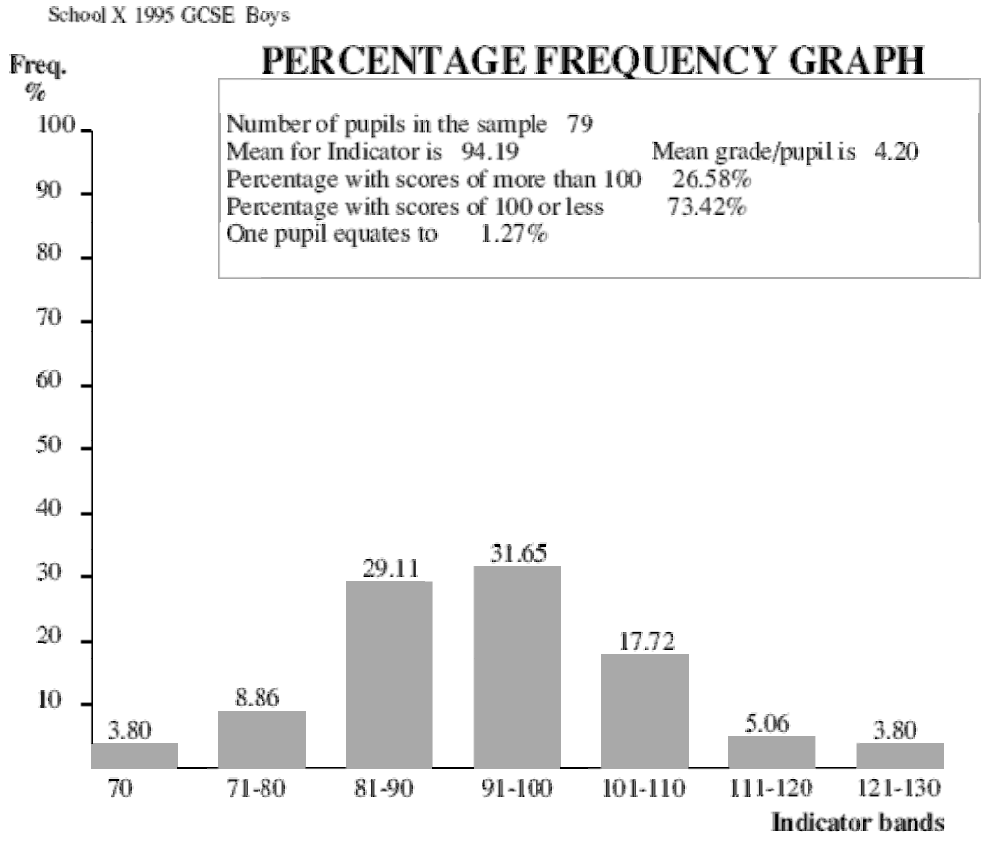
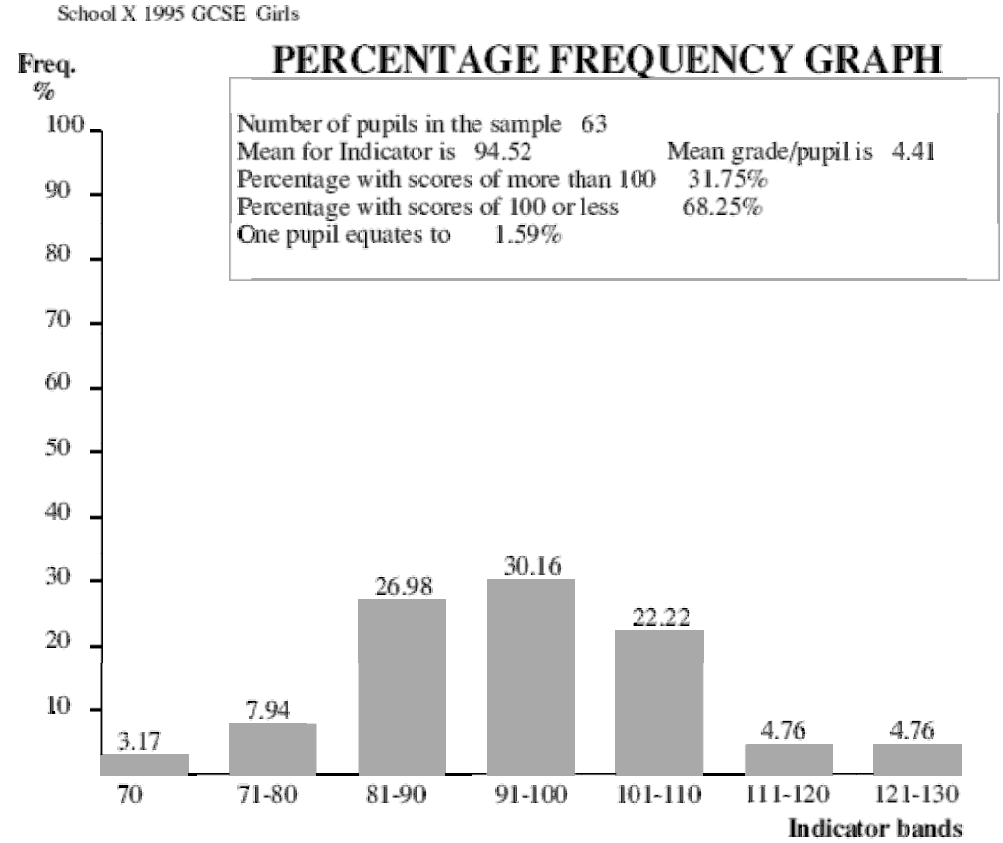


Figure 6.8d



Case study 2

Problems with Correlation as a concept --- The French Department

A very experienced and well respected colleague, the Head of the Modern Languages Department, has maintained a strong but discrete antipathy to suggested correlation between GCSE mean grades and A level results in French. As the staff as a whole have come round to making more use of GCSE mean grades as indicators of likely attainment in their subject areas and of A level success in general, the Head of Modern Languages has come under pressure to make more use of these statistical indicators.

His objections were broadly as follows:

1. GCSE French results were not indicative of French A level success so why should a general basket of subject results be any more accurate in predicting A level success in French?

2. The quality of French GCSE result was very dependent upon the examination syllabus followed. He placed little faith in Modular French courses because of their lack of linguistic content. He was also able to refer to the examination results of individuals and departments in other schools where the Modular examination results were high but the ability of the pupils, as judged by their other examination results and Edinburgh Reading Test results, was weak. This, he claimed, strongly suggested that the Modular courses were "soft options" and provided little linguistic grounding even for able candidates.

3. The influence of native French speakers in the families of some candidates, or the fact that families have taken regular holidays in France, means that some candidates are almost bilingual and at a distinct advantage over other candidates of similar ability in the examinations. The average GCSE grade used as an indicator of likely success takes no account of such factors.

4. A natural inclination to believe that people, their characters, work ethic and innate linguistic ability have more to do with success in studying French language than how well candidates did in their GCSE English, Maths, Science, Technology and so on.

5. There appeared to be no consistency in the correlations for French A level at Sexey's with correlation co-efficients as in *Figure 6.9.*

Figure 6.9

Figures for correlation between pupil GCSE means and A level French grades at Sexey's School

Year	***Pearson's r***	***Sample size***
1996	0.74	8
1995	0.18	4
1994	0.79	13
1993	0.77	12
1992	0.46	18
1991	0.76	6
1991-1996	0.65	61

My problem was to convince my colleague of the utility of GCSE mean grades as indicators of likely A level potential, even in French, and that 'correlation' has limitations which mean that in some circumstances it is less useful, or even a hindrance, in considering examination performance.

This last point in itself is a problem for I now appeared to be saying if the "statistics" suited my purpose I would use them and if they didn't I wouldn't. To my colleague without a working knowledge of statistical significance, error margins, correlation and other statistical terms, this could seem rather like a courtroom prosecution lawyer choosing to cite the evidence which suited the prosecution case and being dismissive of the rest.

The presentation of school effectiveness information to non-statisticians is of great importance for if they do not believe what they are being told nor are they convinced that it is of relevance to them and their teaching they will not act upon that information.

In many ways I agree with what the Head of French was saying about other factors being involved in the examination results which pupils achieved and my own view is, both by natural inclination and by my findings from the statistics, that the students and their characters have a great deal to do with their eventual results. Even with a correlation co-efficient of 0.79 only some 62% of the variance in examination grades is explained by GCSE mean grades. However, that is a very useful percentage, when coupled with a knowledge of the pupil's academic background, attendance, behaviour, and helpful in our teaching of pupils. It means that the GCSE mean of a pupil can give teachers helpful guidance in considering current A level performance in the classroom with likely expectations in the examination proper and appropriate targets can be set. Such targets can then be reviewed in the light of progress made during the course.

To deal with my colleague's comments in turn: in most subjects, French included, it is unlikely that the candidates' results in a single subject at GCSE will correlate very highly with the A level results in the same subject.
If one were to consider only the GCSE grade in French as an indicator of potential success in French A level, the selective nature of the intake into French A level courses would mean that in the majority of cases only pupils with GCSE grades A*, A or possibly B would be allowed to take the A level course. This only allows a maximum of three indicator levels with which to consider A level performance. The restriction in the indicator range alone is likely to produce a very poor correlation.

There is insufficient detail in the GCSE result to discriminate between

candidates' true abilities in the particular subject and the very broad banding of attainment represented by a particular GCSE grade. The majority of French A level candidates will have passed GCSE French with a grade between A (more recently A*) and C with most having achieved A*, A or B grades. The ability banding within these grade ranges is wide such that there may be considerable differences in ability between a candidate at the bottom of the B grade range and one at the top but this distinction is not apparent once the grades have been awarded. At A level the grade range expands to seven levels (A - E, U & N). An average of the GCSE grades obtained, with distinctions made between the scores of pupils to two decimal places (ie. a GCSE mean of 6.54), increases the differentiation on the indicator scale against which A level results can be compared from 9 levels (A*-G & U) to potentially 800 levels (8.00 to 0), although in the majority of cases A level candidates will have GCSE means ranging from 4.50 to 8.00.

The fact that A level candidates coming from other schools will have covered different syllabuses with different linguistic emphases and different teachers also means that the linguistic competence of pupils with apparently the same GCSE grade may be quite different.

From consideration of the GCSE examination results of other schools, the ability of the subject groups as judged by Edinburgh Reading Test, the pupils' performance in the other subjects they sat in relation to French and the syllabuses followed (See compiled GCSE results for subject departments from all schools involved in this research in *Appendix C*) it is apparent that modular French courses do appear to award higher grades than other syllabuses to pupils of comparable general ability. This is likely to be because of the greater emphasis on oral elements of the language such as conversational French rather than grammar and syntax. It is also worth recalling Satterly's comments, referred to in the literature review of chapter 3, regarding coursework and terminal assessment via examination and the not inconsiderable problems of

ensuring reliability of assessment (Satterly, 1989).

Candidates for A level French, with its greater emphasis on written work, linguistics and literature, who have followed a modular French GCSE course are likely to find the change difficult, perhaps more difficult than those candidates who had followed more traditional GCSE courses and therefore the correlation between pupil attainment in a range of GCSE French syllabuses and pupil attainment at A level will be low because the the different GCSE syllabuses are assessing different skills.

For these reasons I encouraged my colleague to consider the average GCSE grade achieved by pupils in all their GCSEs as an indicator of general ability, the logic behind this being that a pupil with a high average GCSE mean as well as an A grade in French GCSE, be it a modular course or not, is more likely to adapt to the demands of A levels than a pupil with an A grade in French but lower average GCSE grade. The average GCSE grade represents a measure of general academic skills which are applicable to A level study rather than a measure of specific skills which are not.

Undoubtedly the advantages enjoyed by a French A level candidate from a bi-lingual background or one who has spent a considerable time in France are going to be great. This advantage is unlikely to be apparent in the average GCSE score of potential A level candidates, except perhaps in their high GCSE French grade but even here not as clear as it could be because of the limitations of the GCSE grading system. The extent to which national figures for A level examination results are inflated by native or near native speakers is not known so we cannot know how disadvantaged non-native speakers are in the examinations. In the school environment such native speakers are not common, being the exception rather than the rule.

Over the years 1993 - 1996 the correlation figures for pupils' mean GCSE score

per entry and A level French grades in the combined schools' sample are illustrated in *Figure 6.10.* Even in the combined schools' sample the numbers taking A level French are not particularly large. The correlations are reasonably strong but even the highest figure in the four year period when squared to give the coefficient of determination indicating that almost 45% of the variation in A level grades could be accounted for by variation in pupils' GCSE mean grades.

Figure 6.10

Correlation between pupil GCSE mean grades and A level French grades combined schools' sample

Year	*Pearson's r*	*Sample size*	*Standard error of prediction*
1996	0.63	127	2.42
1995	0.59	103	2.41
1994	0.67	88	2.60
1993	0.62	57	2.36

The standard errors of prediction indicate that approximately 68% of candidates at A level would achieve grades in the range of plus or minus one and a quarter grades from that predicted by line of regression for A level grades upon GCSE means.

This is useful information for the subject teacher to know even if there are other factors, specific to the study of French, to take into account.
Teachers of A level French will very quickly identify those candidates with the advantage of bi-lingual backgrounds and should normally expect them to do better in the subject, given that they work equally hard, than their less advantaged peers of similar general ability. Their performance should exceed what would, on average, be expected from pupils with similar GCSE mean grades and therefore in any correlation study they are likely to reduce the correlation co-efficient. The fact that the correlation co-efficient is reduced because of the exceptional performance of a candidate or candidates with extra

advantages does not negate the usefulness of the GCSE mean as an indicator.

The same is also true for pupils with particular learning difficulties, such as dyslexia, visual or auditory problems. Teachers would be aware of the candidates' problems and should not be too surprised if these candidates' results are not as high as would be expected from candidates with similar GCSE mean grades.

Pupils' characters and general work ethic do have a major part to play in the quality of results they ultimately achieve as does attendance for example.
If pupils do not attend a significant number of lessons then they are likely to perform less well than pupils of similar general ability who do attend.

Much of this is common sense and such information is taken on board either consciously or subconsciously by most good teachers but when shown statistics, particularly correlation statistics, those same teachers often react negatively because of their experience of pupils who were exceptions.
Some teachers seem to think that because one states that there is a correlation between mean GCSE grades and A level grade attained this implies direct causation without variance as though the predicted outcomes are set in stone. Furthermore some believe that the use of indicative measures for target setting will limit the aspirations of those pupils who might do better and that the use of correlation techniques denies the existence of exceptions to the general trend.

Correlation does not imply causation. A correlation coefficient merely shows the degree of relationship between one variable and another. The fact that a pupil has a high ERT score does not *guarantee* that they will receive a high average GCSE grade, although the probability is high provided that they also attend lessons, study hard, are not struck down with a debilitating condition, their home background remains relatively stable and so on.

In interpreting graphs showing the correlation between one variable and another it is very important to acknowledge the exceptional results. On the subject department scatter graphs I produce, the exceptions are very clear as they are the points plotted farthest away from the regression line. In looking at the results of a single department for one year the sample size is likely to be small and so the influence of individuals is enhanced. When the result of an individual is different from what might have been expected the correlation co-efficient is reduced. The smaller the number of pupils involved the greater the correlation co-efficient must be for there to be any statistical confidence in the relationship between the two variables.

The correlation co-efficients for the French department A level results at Sexey's School over the period 1991 to 1996 have been variable but so have the numbers taking the subject. That the correlation for a particular year was low does not imply that the results were poor or that the indicator variable is of little use. Similarly, if the correlation coefficient were high this does not necessarily indicate that the results were good.

Looking at Sexey's results in French over a number of years, or at the larger sample for a number of schools combined, the correlation between GCSE mean grade and A level grade is strong. Where there is variation from the performance expected of individuals it is usually for reasons that the teacher is well aware of, such as bi-lingual background, work ethic (good or bad), attendance (good or bad) and so on. In the case of Sexey's examination results at A level, if one removes the exceptions, for that is what they are, then the general trend remains and appears stronger.

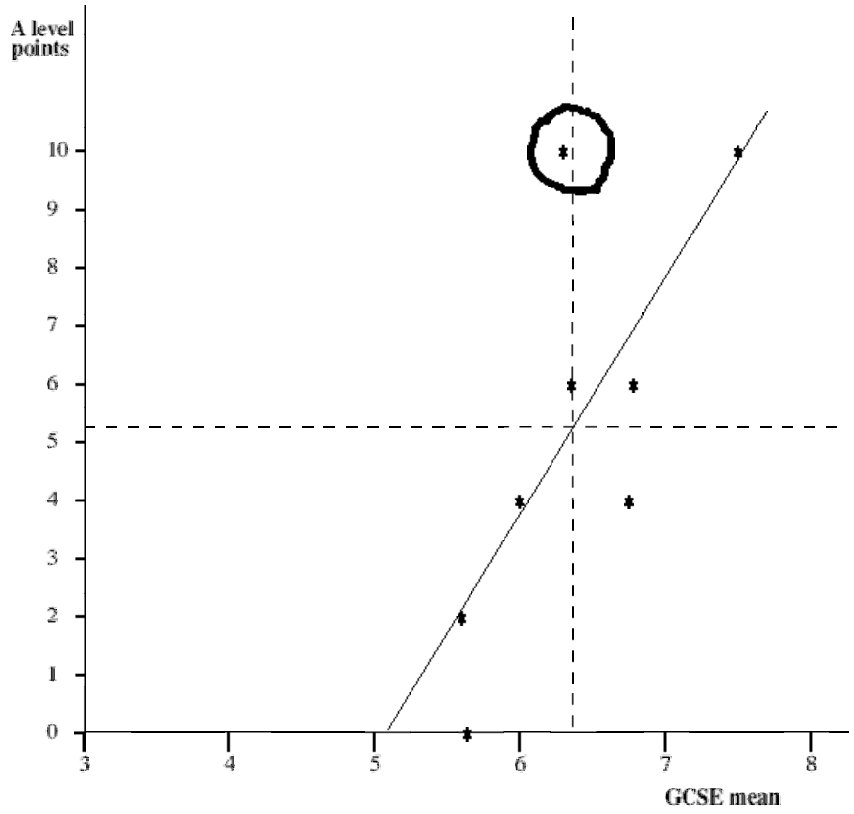
In the example below, *Figure 6.11*, which shows the Sexey's School French A level results for 1996, the circled result was exceptionally good and better than what would have been expected for a pupil of that GCSE mean score.
The correlation was high, even including the the exceptional result, at 0.74.

By removing this exceptional result and re-doing the correlation calculation the co-efficient increased to 0.91 and the co-efficient of determination (the amount of variation in A level points that can be attributed to variation in GCSE mean scores expressed as a percentage) rose from 54.36% to a very high 82.05%.

The higher correlation co-efficient for the department once the exceptional result was removed does not mean that the department's results were better. In fact because the exceptional result was better than the average for the group the average A level grade for the department fell from 5.25 to 4.57.

Figure 6.11

Sexey's School A level French results 1996



Of course if the exceptional result had been worse than the average for the group then excluding it would have raised the average performance of the group.

In answer to teachers' worry that correlation ignores the individual and his / her particular talents or weaknesses, the correct use of correlation along with regression line and scattergraph can serve to highlight individuals and their exceptional results, good or bad. Discussion of the reasons for the exceptional result can lead on to the development of teaching techniques to encourage other pupils to emulate the performance of the exceptional pupil who did well and avoid the problems of the exceptional pupil who did badly.

Correlation techniques used sensibly and properly are not about denying the uniqueness of the individual pupils and their teachers. Rather, they show the general trend and highlight the exceptional, both good and bad.

These two case studies were included because they are useful in illustrating key areas in taking school effectiveness data and using it for school improvement. The first case study, School X, shows how an apparently simple request regarding gender performance is far from simple, particularly when aggregated to the level of the school unit and when trying to make comparisons with national benchmark figures. National figures do not take account of the year to year variation in the makeup of their year cohorts apparent in many schools.

The second case study highlights many of the concerns held by teaching staff regarding the translation of human performance into statistical form. Issues such as correlation, small sample sizes, factors other than prior academic attainment which impinge upon examination success, and above all else the human element are all important and must not be ignored. Successful implementation of school improvement within schools depends upon being able to convince experienced staff of the utility of school effectiveness data and that human issues will not be ignored.