**Figure 5.1**

**Examination Results for Year Group**

SAMP-G

RESULTS

| | M/F | | HI | MA | MA | MA | PH | CH | BI | AS | DR | LA | RO | HO | CD | GE | FR | CO | AR | GE | EN | EN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) ALLEN T | M | 118 | A | | | C | D | | B | E | | D | | | | | A | | | A | A | B |
| (2) BEATTY M | F | 91 | E | F | | | | E | | | | | D | D | | E | | | E | D | D | |
| (3) BENTLEY J | M | 0 | C | | C | | D | | D | | | | | | B | | C | | | B | C | C |
| (4) BRENNEN T | F | 97 | D | | C | | | E | D | | | F | | | | | D | | | C | C | D |
| (5) BORN F | F | 94 | D | | D | | | | C | | B | | | B | | | E | | | D | D | C |
| (6) BRADY G | M | 0 | G | F | | | | F | F | | F | | | | | | G | | | G | E | F |
| (7) BRAINS A | M | 119 | A | | | A | A | A | A | | | | | | | | B | | | A | A | A |
| (8) CAROL C | M | 94 | E | | E | | E | | D | C | | | | | | | F | | | D | D | D |
| (9) CARTER J | M | 109 | | | | C | C | | B | | | | | | C | | C | C | | C | C | C |
| (10) CARSON J | F | 97 | C | | | C | | D | C | | | | | B | | | D | | | C | C | C |

This abbreviated display ( *Figure 5.1* ) shows the individual pupils, as selected by calendar year or DFE year group, listed down the left hand side, their gender, their prior test score and their examination results all on the same line so that an individual's grades can be read off easily. The particular subjects taken are shown in abbreviated form above the columns of results. In this way one may scan down the results of a particular subject very easily for a quick visual impression of the subject's success or across from an individual to see that individual's overall performance.

It is important to show results alongside named individuals when discussing performance with teachers as they relate to people far more easily than figures. It is also important to have this raw data to hand when considering average figures later. For instance, when a pupil has achieved a high average grade but only sat one or two subjects, as discussed earlier (Chapter 4, p.72), the average score could be misleading. Similarly misleading would be an average grade that was reduced because of a rogue result in one subject for some reason, such as a personality clash between teacher and pupil for instance, when all the other grades are much higher. This situation would be apparent in the above display but not in a list of average grades.

This printout may appear very basic, and indeed it is, but the presentation of raw results data in this way is where most school examination officers and school managers start when looking at examination results; it is a familiar starting point to them. With the introduction of prior test scores alongside the raw data the first steps are taken towards linking summative attainment to prior attainment.

**Figure 5.2**

**Year group analysis**

sample

YEAR GROUP ANALYSIS

| NAME | EXAMS | A*-C | POINTS | MEAN | INDICATOR |
|---|---|---|---|---|---|
| ALLEN K | 4 | 0 | 5 | 1.25 | 70 |
| BEATTY W | 10 | 10 | 60 | 6.00 | 122 |
| BORROW ER | 10 | 6 | 50 | 5.00 | 95 |
| BREEN A | 6 | 2 | 20 | 3.33 | 78 |
| BORN F | 9 | 8 | 49 | 5.44 | 109 |
| BRADY B | 10 | 10 | 58 | 5.80 | 110 |
| BRAINS A | 10 | 9 | 54 | 5.40 | 105 |
| CARTER J | 10 | 6 | 45 | 4.50 | 0 |
| CARSON J | 10 | 8 | 51 | 5.10 | 106 |
| CHERRY B | 10 | 8 | 50 | 5.00 | 89 |
| THUMB T | 10 | 10 | 59 | 5.90 | 130 |
| TIMES ED | 10 | 7 | 50 | 5.00 | 121 |
| UNDER M-W | 9 | 4 | 39 | 4.33 | 89 |
| WARREN R | 10 | 9 | 54 | 5.40 | 102 |
| WHITELAW W | 10 | 10 | 65 | 6.50 | 130 |
| WHITTLE F | 10 | 2 | 41 | 4.10 | 103 |
| WILL I | 9 | 6 | 43 | 4.78 | 92 |
| WOOD U | 9 | 2 | 32 | 3.56 | 0 |
| YOO DID | 10 | 9 | 56 | 5.60 | 109 |

| | | | |
|---|---|---|---|
| Number of candidates | 40 | Number of entries | 380 |
| Female candidates | 15 | Female entries | 147 |
| Male candidates | 25 | Male entries | 233 |
| Subjects sat | 380 | Number of absences | 0 |
| Ave. entries per cand. | 9.50 | Female absences | 0 |
| Total A*-C | 272 | Male absences | 0 |
| Average no. A*-C | 6.80 | Year group mean | 5.01 |
| Standard dev. | 1.36 | Year group Ind. mean | 104.17 |
| I.Q.R. 5.92 to 4.11 | | Fem. Ind. 107.45 | Male Ind.102.26 |

| Grade | A* | A | B | C | D | E | F | G | U | A*-C |
|---|---|---|---|---|---|---|---|---|---|---|
| % | 2.11 | 7.37 | 28.68 | 33.42 | 16.05 | 6.84 | 3.68 | 1.58 | 0.26 | 71.58 |
| Nos. | 8 | 28 | 109 | 127 | 61 | 26 | 14 | 6 | 1 | 272 |
| Female | | | | | | | | | | |
| % | 2.72 | 11.56 | 33.33 | 36.05 | 12.24 | 3.40 | 0.68 | 0.00 | 0.00 | 83.67 |
| Nos. | 4 | 17 | 49 | 53 | 18 | 5 | 1 | 0 | 0 | 123 |
| Male | | | | | | | | | | |
| % | 1.72 | 4.72 | 25.75 | 31.76 | 18.45 | 9.01 | 5.58 | 2.58 | 0.43 | 63.95 |
| Nos. | 4 | 11 | 60 | 74 | 43 | 21 | 13 | 6 | 1 | 149 |

In this display ( *Figure 5.2* ), again abbreviated, next to the pupil names are shown the number of examinations taken by each individual, the number of grades A*- C (GCSE) or A-E (A level), the individual's points total, mean score (points divided by subjects entered) and indicator score. At the foot of the display

is given an analysis of the year group's performance, broken down by gender as well as for the whole group, including grade distribution. The percentages are calculated with the total subjects sat as the common denominator and therefore exclude any absences. Absences are shown on the sheet so that they may be noted but they are not included in any calculations, as mentioned earlier.

Absences are examinations for which pupils were entered but did not attend. Entries for examinations are usually made by the end of January of the examination year and so relatively late in a pupil's pre-GCSE education. No account is taken here of pupils who followed non-examination courses or courses leading to qualifications other than GCSE or A level.

The average entries per candidate is the sort of indicator figure for a school to look at over a number of years to see what the trend is. For instance, if Gray *et al*. (1996) are correct and increasing the number of examination entries per pupil is a factor in schools becoming more successful then one would see the figure for average entries per candidate rise. If, on the other hand, schools were maximising the attainment of the more able and minimising the failure of the less able by a selective entry policy this figure might remain largely static. The entry policy could be checked by scanning down the column showing individual pupils' number of examinations taken and then across to the column showing the pupil's indicator score to see whether it was the able pupils who were being entered for high numbers of exams and the less able restricted to a minimum, usually more than five now that the national performance tables' key indicator for GCSE is number of pupils achieving five or more GCSEs.

The standard deviation gives some idea of the spread of the pupil mean grades either side of the mean grade for the year cohort. This figure could be monitored along with the distribution of pupil ability, shown in a later printout, to see if it reflects the range of ability within the year cohort. The greater the spread of ability the greater one would expect the spread of individual pupil mean grades to

be.

The IQR, or interquartile range, shows the range of grades obtained by the middle 50% of candidates, ignoring the extremes of attainment at both ends of the scale and assuming a normal distribution.

The year group mean is the average grade achieved by the group as a whole and could be used to make a comparison with previous years, but the year group indicator mean should also be borne in mind as an indication of the general ability of the group. In tracking performance over time it is useful to plot these two figures and see trends. One would expect the average grades to move in relation to the average indicator scores, the higher the indicator average the higher the average grade. If the average grade achieved seems to rise more quickly than the average indicator this may be evidence of improved results, but it could also be the examinations getting easier as was discussed earlier in this thesis with reference to the review of examination standards over time (SCAA, 1996b).

It should not be forgotten, although not mentioned in the literature, that the average grade for the year will also be affected by the distribution of pupil abilities and the entry pattern associated with that spread. For example, if the majority of pupils were of average ability or below but a few pupils were very able and sat a large number of examinations then they would raise the average grade achieved beyond what might have been expected from the average indicator score. In very large samples this would not occur because the spread of abilities would generally be normally distributed but in individual schools, even as large as 200 pupils, this is not always the case.
This sort of consideration must be borne in mind when discussing results with Heads or Senior Managers in trying to help them understand the dynamics of examination results within their own schools.

The average (mean) indicator scores for males and females are shown separately because it is quite common for these two figures to be quite different in a school year cohort. This would help to explain apparent differences in the distribution of grades, also shown by gender, for girls and boys. This will be an important factor in the case study to be discussed later in the thesis looking at gender performance in a particular school.

More detailed comparisons can be seen by looking at the performance of the genders separately and considering the correlation graphs for the separate genders and the distribution of pupils according to ability by gender.

The grade distributions for the group, girls and boys are shown by the percentage figures and may be used to consider whether girls are doing as well or better than boys. This could be useful information for an equal opportunities discussion.

The numbers of girls and boys are shown. It is a national trend that girls tend to do rather better at GCSE than boys with the possible exception of certain subjects such as the sciences ( SCAA, 1996 ), but in addition to considerations as to why this should be so it is important to look at the male and female indicator scores to check that there is not a large difference between the average abilities of the genders which could make any gap seem larger than it is in reality. (See case study on School X, Chapter 6 ).

Look also at the distribution of the abilities by gender using the Percentage Frequency Graphs as shown in *Figure 5.8* but by separate genders. Even though one gender may have the higher average, if the other gender has the majority of its number clustered above the critical level needed to gain GCSE / A level success then it may score more highly on certain indicators such as percentage of grades A*- C at GCSE for instance.

The average number of grades A*-C (GCSE), or A-E (A level), and the average

entries per candidate are both indicators of examination performance but can reflect the ability of the group as well as the performance of the school.

The number of absences could also be important. If this figure starts to become significant then one must look at such things as teachers allowing pupils to drop subjects after entry and pupils failing to deliver required coursework, besides the obvious failure to attend the examination.

Another analysis sheet given to schools is that giving a breakdown of results as per DfEE regulations as in *Figure 5.3* below.

**Figure 5.3**

**Analysis of Results for DFE reports**

SAMP-G

**GCSE POINTS TALLY A*=8 A=7 B=6 C=5 D=4 E=3 F=2 G=1**

| Points | All | Girls | Boys |
|---|---|---|---|
| 0 - 9 | 0 | 0 | 0 |
| 10 - 19 | 1 | 0 | 1 |
| 20 - 29 | 1 | 0 | 1 |
| 30 - 39 | 9 | 3 | 6 |
| 40 - 49 | 21 | 10 | 11 |
| 50 - 59 | 13 | 3 | 10 |
| 60 - 69 | 3 | 1 | 2 |
| 70 - 79 | 0 | 0 | 0 |
| TOTALS | 48 | 17 | 31 |
| Ave. points per cand. | 45.15 | 45.00 | 45.23 |

| Year group | 5+ A*-C | 5+ A*-G | 1+ A*-C | 1+ A*-G | No grade | Absent |
|---|---|---|---|---|---|---|
| ALL | 31 (65%) | 48 (100%) | 45 (94%) | 48 (100%) | 0 (0%) | 0 (0%) |
| Male | 21 (68%) | 31 (100%) | 29 (94%) | 31 (100%) | 0 (0%) | 0 (0%) |
| Fem. | 10 (59%) | 17 (100%) | 16 (94%) | 17 (100%) | 0 (0%) | 0 (0%) |

Various ranges of points are shown and the numbers of pupils who achieved points totals within those ranges. Again a breakdown is shown according to gender, including the average points per candidate.

As per DFEE requirements, the number of pupils achieving 5 or more GCSE grades A*-C, 5 or more grades A*-G, 1 or more grades A*-C and 1 or more

grades A*-G are shown including the percentage of the year cohort that each figure represents. In addition, figures are given for those who did not obtain any grade and those who were absent from all their examinations. This is not required at A level.

It is possible to use these figures for comparison between different years, but again one must not forget to consider the indicator of ability for each year compared. The higher the ability of the group the greater the percentage of pupils gaining five or more GCSEs at level C or above one would expect to see. Once 100% of pupils have gained or exceeded a C grade in five GCSEs this ceases to be a useful measure of school examination performance for it cannot discriminate between schools where, for example, 100% of pupils gained A grades in ten GCSEs and a school where 100% of pupils gained C grades in five GCSEs

With regard to A/AS levels, schools that enter candidates for a higher than average number of examinations are almost by default going to do better, if judged by total A/AS points alone, as in the national performance tables, for their pupils will tend to acquire more points. Gray *et al.* (1995) found this to be the case and considered the use of an alternative measure,

> "The outcome measure we have concentrated on so far in this report has been the candidate's 'total GCE A/AS point score'. Not surprisingly, candidates who were entered for more examinations at GCE A/AS level were likely to do better than those who were entered for a smaller diet of, say, just two or three GCE A levels or AS equivalents. By contrast, the 'average GCE A/AS grade' outcome measure is designed to focus on the 'quality' of passes a candidate obtains; it is not directly influenced by the decisions institutions made about how many examinations to prepare their students for."
>
> "..... use of the alternative measure of, 'average GCE A/AS grade', to assess outcomes would also affect interpretations of the apparent

'effectiveness' of a large number of institutions." (Gray, Allnutt, Gardner, Blackham and Frost, 1995).

Whilst I would agree that an "average grade" solves many of the problems when comparing candidates' GCSE results, the problem is not so easily resolved when looking at A level results where the number of examinations taken by the candidates are generally much fewer. The usual number of A level examinations taken by candidates in non-selective state schools is three with some taking two and some taking four. This is more usually a reflection of the candidate's strength and subject choices than a constraint of the school entry policy with the exception of the entries for General Studies as discussed in chapter 4 of this thesis. In independent schools the most able candidates may be entered for as many as five A level examinations. Maths is a particular subject where in a modular scheme the opportunity to gain credit for two or more A levels with appropriate extra module choices is a case in point that reflects the possible effect of subject choice rather than general school entry policy.

At the level of the school unit the decision to use the average GCE grade as the indicator of "*apparent* effectiveness" would affect, in particular, those institutions with the most able candidates because the distinction between schools where candidates were averaging A and B grades in two and three A levels and those schools where candidates were averaging A and B grades in four and five A levels would be lost.

In the A level analysis data I return to schools, I include both average grade and average points per candidate but I also include a figure giving the average number of examination entries per candidate to add to the contextual information of the school cohort.

The table below (*Figure 5.4*) shows each subject, its average (mean) grade achieved by the pupils doing the subject, the average grade achieved by those

**Figure 5.4**

**Comparison of GCSE subject mean grade vs. group mean**

SAMP-G

| Subject | Subject Grade | Group Grade | Diff. | Ind. Mean | Group Size |
|---|---|---|---|---|---|
| HISTORY | 4.69 | 4.87 | -0.19 | 106.60 | 35 |
| MATHS F | 2.37 | 3.51 | -1.14 | 102.86 | 8 |
| MATHS I | 4.60 | 4.58 | 0.02 | 105.58 | 15 |
| MATHS H | 5.92 | 5.68 | 0.24 | 109.10 | 25 |
| PHYSICS | 4.65 | 5.04 | -0.38 | 105.71 | 26 |
| CHEMISTRY | 4.91 | 5.22 | -0.31 | 108.83 | 23 |
| BIOLOGY | 5.37 | 4.98 | 0.39 | 106.95 | 48 |
| ASTRONOMY | 3.50 | 5.65 | -2.15 | 121.67 | 4 |
| DRAMA | 5.00 | 4.09 | 0.91 | 98.57 | 9 |
| LATIN | 4.20 | 5.67 | -1.47 | 113.78 | 10 |
| ROMAN CIVILIZATION | 5.60 | 4.98 | 0.62 | 114.25 | 5 |
| HOME ECONOMICS | 6.00 | 4.73 | 1.27 | 99.25 | 9 |
| CDT | 4.47 | 4.73 | -0.27 | 105.31 | 15 |
| GERMAN | 6.50 | 5.53 | 0.97 | 117.50 | 4 |
| FRENCH | 4.52 | 4.98 | -0.46 | 106.95 | 48 |
| COMPUTER STUDIES | 4.73 | 5.17 | -0.44 | 106.64 | 11 |
| GEOGRAPHY | 5.12 | 4.96 | 0.16 | 106.41 | 41 |
| ENGLISH LANGUAGE | 5.35 | 4.98 | 0.37 | 106.95 | 48 |
| ENGLISH LITERATURE | 5.15 | 4.98 | 0.16 | 106.95 | 48 |
| H & A OF MUSIC | 5.33 | 5.04 | 0.30 | 125.00 | 3 |

same pupils but in all the subjects they entered, the difference between the two last figures, and the indicator average score for the pupils doing that subject. This is a very useful table for one can see at a glance the average grade achieved in all the subjects sat. The group grade gives an indication of the ability of the students sitting each subject judged by their actual performance in these particular examinations.

The differential shows how each subject fared with the pupils who studied it compared to the other subjects those particular pupils sat. The indicator average (mean) score gives a measure of the ability of the pupils who sat a particular subject judged by whatever test was administered previously. The group size is the number of pupils who actually sat the examination and is included so that any conclusions about the subject departments based on average grade can take into account how many pupils were taught. Obviously it would be unfair to compare the averages of a department having a hundred candidates with a department which only had five or less for example.

In a very large department other factors to take into account would be the

teaching group sizes and setting or streaming arrangements that operated in the department and school, particularly if one wanted to look at the effectiveness of teachers within the department or particular reasons for good or bad results.

We should not expect the column of differences in the table to add up to zero: it will not, for each group of students sitting a particular subject may sit a different range of subjects according to the options scheme in place in the particular school.

We should also beware of making an assumption about a particular subject's performance until we have checked on the numbers sitting it. Very small numbers in a subject can lead to distorted performance figures. A single aberrant result, a pupil who under performs badly or a pupil who excels all expectations can each make an exaggerated contribution to the overall figures when numbers are very small. In such cases it is more important to look at the performance of the individuals who took the subject as shown in the subject department sheet shown later.

Some subjects are simply harder than others. By this I mean that pupils with a given indicator score will, on average, gain higher grades in some subjects than others. This view is supported by Fitz-Gibbon (1992) with regard to A level subjects and is discussed in greater detail with regard to GCSE subjects and subject differentials in chapter 7 of this thesis. The comparison of examination boards, syllabuses and subjects was raised as a matter of concern in chapter 3 (page 42ff.) of this thesis when reviewing the relevant literature with reference to DFE (1992a), Goldstein (1982), Mortimore and Byford (1981), Torrance (1986), SCAA (1996 and 1996a). Therefore, it is unfair to criticize the performance of school subject departments in comparison to a subject that is less difficult. It is better to look at trends over a number of years to see if a particular year's results are aberrant and then search for reasons, for example a change in syllabus, different teachers, smaller numbers in the group etcetera.

By becoming involved in a consortium approach to the analysis of examination results, it is possible to compare your own school's figures, as per this table, with those of other schools. It then very quickly becomes apparent which subjects are harder than others, consistently maintaining a lower differential when comparing its average pupil grade to the performance of the same pupils in other subjects, and the performance of a department can be judged in relation to the same subject area in other schools with similar ability groupings.

**Analysis by Subject**

In *Figure 5.5* The names of the pupils who sat a particular subject are listed along with the grade they achieved, their examination mean (average) result, the

**Figure 5.5**

sample
SUBJECT RESULTS GERMAN

| | Grade | GCSE Mean | Diff. | Ind. score |
|---|---|---|---|---|
| BORN F | C | 5.44 | -0.44 | 109 |
| BRADY B | B | 5.80 | 0.20 | 110 |
| BRAINS A | B | 5.40 | 0.60 | 105 |
| CARTER J | C | 4.50 | 0.50 | 0 |
| CHERRY B | D | 5.00 | -1.00 | 89 |
| CLIMBER S | D | 2.70 | 1.30 | 92 |
| COURT T | * | 6.90 | 1.10 | 114 |
| DOIT ALL | D | 4.30 | -0.30 | 106 |
| FIDDLE DD | B | 4.70 | 1.30 | 0 |
| WARREN R | B | 5.40 | 0.60 | 102 |
| WHITELAW W | C | 6.50 | -1.50 | 130 |
| WHITTLE F | E | 4.10 | -1.10 | 103 |
| WILL I | C | 4.78 | 0.22 | 92 |
| YOO DID | B | 5.60 | 0.40 | 109 |

| | | | |
|---|---|---|---|
| Number of candidates | 32 | Number of entries | 32 |
| Female candidates | 13 | Female entries | 13 |
| Male candidates | 19 | Male entries | 19 |
| Number of absences | 0 | Subject ave. grade | 5.09 |
| Female absences | 0 | Group mean grade | 5.10 |
| Male absences | 0 | Year mean grade | 5.01 |
| | | Group mean Ind. score | 106.25 |

| Grade | A* | A | B | C | D | E | F | G | U | A*-C |
|---|---|---|---|---|---|---|---|---|---|---|
| % | 3.12 | 6.25 | 28.12 | 34.37 | 18.75 | 6.25 | 3.12 | 0.00 | 0.00 | 71.87 |
| Nos. | 1 | 2 | 9 | 11 | 6 | 2 | 1 | 0 | 0 | 23 |
| Female | | | | | | | | | | |
| % | 7.69 | 7.69 | 23.08 | 53.85 | 7.69 | 0.00 | 0.00 | 0.00 | 0.00 | 92.31 |
| Nos. | 1 | 1 | 3 | 7 | 1 | 0 | 0 | 0 | 0 | 12 |
| Male | | | | | | | | | | |
| % | 0.00 | 5.26 | 31.58 | 21.05 | 26.32 | 10.53 | 5.26 | 0.00 | 0.00 | 57.89 |
| Nos. | 0 | 1 | 6 | 4 | 5 | 2 | 1 | 0 | 0 | 11 |

difference between the last two to indicate whether the pupil did as well in this

subject as they did on average in all their subjects, and their indicator score to give some indication of their ability as judged by a prior attainment score, such as the Edinburgh Reading Test.

In this way the performance of the individuals who sat a subject is not lost within the general statistics. Individuals who shine in particular subjects can be spotted and this may even be useful in considering further options such as taking the subject further (A level or degree) or avoiding it like the plague.

Beneath the pupil listing are the standard grade distribution figures, again broken down by gender, the subject average grade for comparison with the year average grade and the average grade for these same pupils in all their subjects (group mean grade).
(N.B. Where a subject is sat by all the year group the last two grades mentioned will actually be the same).

Looking at the grade distribution and average ability for the subject group one can check to see if they are as expected. Further information on the distribution of pupil ability is given in the Frequency Distribution Graph (see later). The percentage of grades at the A*-C level will very much depend upon the spread of pupil ability in the group, particularly in small groups where the ability and performance of each individual play a much more significant part than in large groups.

Figures are given separately for girls and boys so that the relative performance of the genders may be considered. Given that girls in many subjects now outperform boys at GCSE level, even when considering pupils of like abilities (SCAA, 1996), and are beginning to outperform boys at A level (SCAA, 1996a), then it is worth looking at the proportion of girls and boys actually sitting the subject and its effect on the overall performance of the subject area.

The number of absences are shown and if large it is worth seeking out reasons why. It may be that there are a significant proportion failing to hand in the required coursework or that students being encouraged not to attend the examination in order to massage the results statistics.

In subjects that are not a compulsory part of the curriculum and for which pupils opt it is worth looking at the numbers taking the subject. For example, we can check the trend in numbers to see if it is attracting its percentage of the year group, or whether numbers of pupils opting for the subject are dwindling. Pupils are perceptive and a sudden drop off in numbers taking a once popular subject is likely to indicate that something is wrong in the delivery of the subject but, importantly, could also reflect curricular pressure from other areas. By curricular pressure I mean situations such as the changing of school option blocks so that optional subjects are set against each other preventing pupils from doing both and forcing pupils to make choices. Such a scenario may come about because of the staffing situation at a school, external pressure from government making elements of the curriculum compulsory or specifying minimum amount of time that a subject should be taught, or the school's decision to introduce a new subject option. Both scenarios, poor delivery of the subject and curricular pressure, should be investigated to prevent a drop in standards or loss of a subject area.

In larger schools this analysis sheet can usefully be broken down even further to look at particular teaching groups. A number of schools have used these sheets in such a way as "one offs" or year on year checks. If setting operates within the school, then over time it is possible to investigate the effectiveness of some members of staff with particular ability groups, or to check the setting arrangements in terms of pupil prior abilities and eventual outcomes.

Another use is to look at the examination paper tiering that pupils were entered for (GCSE) and whether some pupils could have been more helpfully placed in

different groupings. This retrospective analysis can then be used to inform future decisions on entry policy. In schools where certain subjects are experimenting with single sex teaching groups this sort of analysis is useful in helping gauge the difference, if any, that such groupings make, being careful to take into account the abilities of the respective groups as well as their gender.

**Spearman Rank Correlation**

This table ( *Figure 5.6* ) shows the pupils listed in rank positions according to their performance in the examinations. Against each pupil is shown their indicator score and rank position according to that score, then their examination mean performance and their rank position according to that.

At the end of the list is the co-efficient of correlation between the indicator score and the examination mean result. The higher this co-efficient value (closer to 1.0) the greater the predictive efficiency of the indicator test score. However, once it is accepted that the indicator test score does have predictive validity it is the exceptions to the expected that are more interesting.

For instance, in this example candidate number 5, John E, was ranked 21st on his test score but on GCSE results was ranked 4th= . This was a great improvement over the expected position and the reasons for this improvement should be discussed. It could be a faulty prior test score or lucky examination results but I have found in looking at such individuals from Sexey's School and discussing results with the Heads or Deputies of other schools that such individuals are often well known either because they have worked incredibly hard, were thought to be more able than the prior test had indicated, or with help had overcome some learning difficulty, such as reading problems, that had held back their early education. Similarly candidates who had done much worse than their prior ability test score had indicated were also well known because of particular problems with their learning such as social / behavioural, or they were thought to be lazy/ lacking motivation. Whilst the first of these problem areas is almost

always investigated to some degree by schools the second area can be too easily accepted without consideration of reasons for lack of motivation, such as the setting of work at an inappropriate level of difficulty.

**Figure 5.6**

**SPEARMAN RANK CORRELATION**

| | | Score1 | Rank1 | Score2 | Rank2 |
|---|---|---|---|---|---|
| 1 | OSMOND D | 125.00 | 3.00 | 69.00 | 1.50 |
| 2 | COURT T | 114.00 | 7.50 | 69.00 | 1.50 |
| 3 | WHITELAW W | 130.00 | 1.50 | 65.00 | 3.00 |
| 4 | BEATTY W | 122.00 | 4.00 | 60.00 | 4.50 |
| 5 | JOHN E | 99.00 | 21.00 | 60.00 | 4.50 |
| 6 | THUMB T | 130.00 | 1.50 | 59.00 | 6.00 |
| 7 | BRADY B | 110.00 | 9.00 | 58.00 | 7.00 |
| 8 | MASTER M | 114.00 | 7.50 | 57.00 | 8.00 |
| 9 | YOO DID | 109.00 | 10.50 | 56.00 | 9.00 |
| 10 | BORN F | 109.00 | 10.50 | 49.00 | 10.00 |
| 11 | WARREN R | 102.00 | 18.50 | 54.00 | 11.50 |
| 12 | BRAINS A | 105.00 | 15.00 | 54.00 | 11.50 |
| 13 | SMITHS C | 101.00 | 20.00 | 53.00 | 13.00 |
| 14 | HOPPER G | 108.00 | 12.00 | 52.00 | 14.00 |
| 15 | CARSON J | 106.00 | 13.50 | 51.00 | 15.00 |
| 16 | LINTON C | 102.00 | 18.50 | 45.00 | 18.00 |
| 17 | HOWLE R | 121.00 | 5.50 | 50.00 | 18.00 |
| 18 | TIMES ED | 121.00 | 5.50 | 50.00 | 18.00 |
| 19 | BORROW ER | 95.00 | 22.50 | 50.00 | 18.00 |
| 20 | CHERRY B | 89.00 | 26.50 | 50.00 | 18.00 |
| 21 | GREEN P | 95.00 | 22.50 | 48.00 | 21.00 |
| 22 | WILL I | 92.00 | 24.50 | 43.00 | 22.00 |
| 23 | UNDER M-W | 89.00 | 26.50 | 39.00 | 23.00 |
| 24 | DOIT ALL | 106.00 | 13.50 | 43.00 | 24.00 |
| 25 | GOODYEAR T | 104.00 | 16.00 | 37.00 | 25.00 |
| 26 | WHITTLE F | 103.00 | 17.00 | 41.00 | 26.00 |
| 27 | BREEN A | 78.00 | 29.00 | 20.00 | 27.00 |
| 28 | CLIMBER S | 92.00 | 24.50 | 27.00 | 28.00 |
| 29 | NICK RS | 84.00 | 28.00 | 22.00 | 29.00 |
| 30 | ALLEN K | 70.00 | 30.00 | 5.00 | 30.00 |

| | | | |
|---|---|---|---|
| The Correlation factor is | 0.74 | | |
| The Indicator Median is | 104.50 | Mean is | 104.17 |
| Indicator Mode is | 130.00 | Frequency is | 2.00 |
| The Outcome Median is | 5.05 | Mean grade/pupil is | 5.06 |
| Outcome mode is | 5.00 | Frequency is | 5.00 |
| Mean points / pupil is | 47.87 | The number of pupils is | 30 |

Even with a correlation co-efficient as high as 0.74 as in this example, only just under 55% of the variance in outcome scores could be said to be directly attributable to the variance in prior indicator scores so other factors play a large part in the performance of candidates in examinations.

It is at this level of analysis that the involvement of teachers, people who knew the individuals concerned in their learning environment, is extremely helpful for their personal knowledge can enlighten and explain outcomes that would not have been predicted by prior test information alone. This involvement of the teachers has an effect upon their professional development as they monitor their own effectiveness from year to year.

Figures are also given for the indicator score mean, mode and median values. These can give some idea of the distribution of the indicator scores. We have to beware of putting too much weight on the modal score as it is possible to have a number of modes within any sample. When using the different types of "average" I often found it necessary to remind teachers that the modal score is the most commonly occurring score, the median represents the mid-range score and the mean is the average score calculated by summing the individual scores and dividing by the number of individuals.

The same figures are also shown for the outcome score, be this GCSE mean grade or A level points.

Providing the correlation is good, this table is very useful for one can run down the list and see at a glance whether pupils have performed as you would have expected them to, better or worse, by looking at their respective rankings for the indicator and the examination performance.

For instance, the pupil who was ranked 21 on the indicator score but was ranked equal 4th on the examination results has shown considerable improvement whereas a pupil who was ranked equal 5th on the indicator score but has slipped to equal 16th has done less well. An alternative hypothesis would be that for the individual pupil the indicator result was not typical and therefore this would highlight the need for the use of other tests to act as a check.

An additional point worth noting here is that it may be the examination mean that is not indicative of the pupil's ability. The mean can be distorted by a rogue result caused by illness, or a personality clash with a particular subject teacher, or the fact that the pupil only took a very few examinations, so we need to check what each pupil actually achieved as shown on the sheet listing the actual grades obtained.

In reviewing this particular sheet it is also an ideal opportunity to consider other pastoral factors which may have had an effect on a pupil's outcome score. Consultation with the pupil's Tutor / Year Head can be useful to check whether personal circumstances, home problems, glandular fever, a particularly torrid romance or whatever may have played their part in reducing the candidate's overall performance. An effective tutorial programme, involving regular discussion with pupils, may well serve to highlight problems impinging on academic progress and therefore give the school the chance to offer help, where it can, in ameliorating the problem and lessen its effect upon learning. These discussions could also offer a method of checking otherwise unchallenged views of teachers on the motivation of pupils.

Particularly studious pupils, those who show great personal determination or who benefit from an educationally resource rich home, may well do better than one would expect from their indicator scores. Factors such as these all play their part and should be considered when trying to establish reasons for the performance or under performance of pupils in examinations.

In the light of hindsight, one can check the correctness of a pupil's subject option choices, the setting arrangements if any which applied, or the suitability of the paper levels for the pupil's ability. These issues are discussed in more detail in the case studies later in this thesis. Areas such as these should be explored by schools and the findings addressed if future year cohorts are to benefit from the experiences of a particular year cohort. In this way, knowledge gained from

summative assessment may be used in a formative manner and qualitative input, rather than purely quantitative, can play its part.

**Scatter graphs and Correlation**

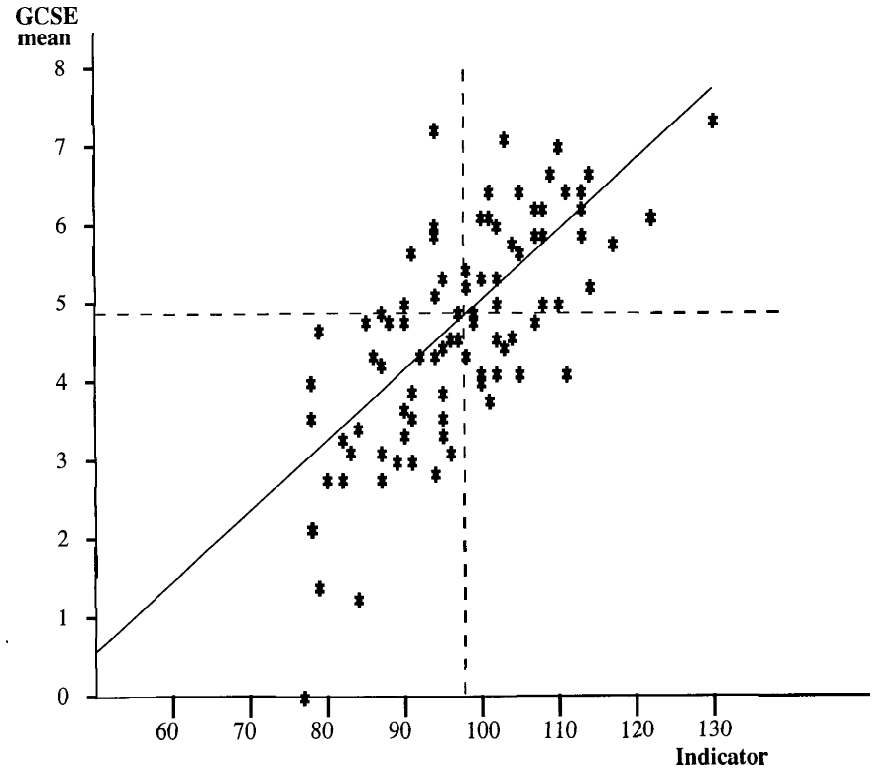
The graph in *Figure 5.7* plots the indicator score along the x axis and the examination score along the y axis. Individual pupil performances are shown as asterisks and a line of best fit (regression line for y upon x), for predicting individual pupil GCSE mean grades from ERT scores, plotted between them giving a visual representation of the correlation between the two variables.
It is the distribution of the points plotted as asterisks that shows the correlation, not the regression line itself. If the "scatter" is long and thin following the regression line sloping from bottom left to top right then the correlation co-efficient is likely to be high, nearer to a value of plus 1. If the scatter is similar but sloping from top left to bottom right then the correlation may still be high but negative, nearer to a value of minus 1. If the scatter resembles a round cloud centred upon the point representing the means for both x and y axes then there is likely to be little or no correlation, the co-efficient having a value near to zero.

A high correlation co-efficient is no measure of the quality of the examination results achieved by the pupils in a particular sample. It is quite possible to have a high correlation co-efficient and yet the slope of the regression line be relatively flat and / or reach a low value on the y axis. For instance, if the extreme right of the regression line reached a point equivalent to 5 on the y axis then this would mean, assuming a high correlation co-efficient, that even the most able pupils were on average only likely to achieve GCSE means of 'C' grade.
The closer the extreme top right of the regression line gets to a point equivalent to a value of 8 on the y axis (GCSE mean) then the more likely it is that pupils with maximum indicator scores will achieve maximum GCSE mean scores.

That the slope of the regression line is less steep can also be a good sign.
For example, in a school where the less able pupils, those with indicator scores of

between 70 and 90 for argument's sake, do particularly well for their ability the regression line will intersect the y axis at a higher point than in a school where the less able pupils achieve very little. Assuming that the more able pupils in the school also perform well, the brightest achieving GCSE means of 8, equivalent to A* grades in all their GCSEs, the slope of the line will be flatter reflecting the excellent performance of the lower ability pupils and the possible ceiling effect of

**Figure 5.7**




| | | | |
|---|---|---|---|
| Number of pupils in the sample | 92 | | |
| Mean for X is | 97.80 | Mean for Y is | 4.87 |
| Standard dev. for X is | 11.13 | Standard dev. for Y is | 1.40 |
| Covariance is | 11.08 | | |
| Coefficient of correlation is | 0.71 | | |
| Coefficient of determination is | 50.28% | | |
| Standard error of estimation is | 0.99 | | |

the GCSE mean scale stopping at a value of 8.

Individual points plotted farther away from the regression line than the majority of the other plotted points will reduce the correlation because they either exceed or fall below the general level of performance shown by the majority of pupils.

The extent to which the majority of plotted points follow the trend indicated by the regression line is shown by the standard error of estimation. The standard error of estimation given for *Figure 5.7* is shown as 0.99 meaning that approximately 68% of the sample will be within plus or minus just under a GCSE grade either side of the regression line.

Pearson's Product Moment correlation method is used to calculate this correlation co-efficient and may produce a slightly different value to that arrived at using Spearman's formula, the latter being distorted if too many scores are tied. Guilford (1973) comments, "For the same basic data, the (Spearman) rho coefficients are systematically a bit lower than the corresponding Pearson r's, but the maximum difference, which occurs when both coefficients are near 0.50, is less than 0.02."

Pearson's method is better suited to the correlation of two variables which both operate on linear scales, as in this case, but we must be mindful of the restrictions which apply to the use of Pearson's product moment correlation,

> "The derivation of the formula for the Pearson r assumes that (1) the scores have been obtained in independent pairs, each pair being unconnected with other pairs; (2) the two variables correlated are continuous; and (3) the relationship between the two variables is rectilinear." ( Guilford, 1973).

It is important that the regression line and correlation be seen against the scatter graph to check that the relationship between indicator and outcome variables is indeed linear.

The vertical dotted line in *Figure 5.7* shows the average indicator score (in this case the Edinburgh Reading Test but it could be any valid test) and the horizontal dotted line shows the average examination performance. These are useful in that it is immediately obvious that all those plots in the top right quadrant were better than average in both their indicator score and their examination results. Those

plots in the bottom left quadrant were worse than average in both respects. Those plots in the bottom right quadrant had higher than average indicator scores but worse than average examination results, whereas those plots in the top left quadrant had worse than average indicator scores but better than average examination results.

The co-efficient of correlation is also shown as are standard deviation figures for the values on the x and y axes. These latter figures give some idea of the spread of their respective values and the covariance gives a measure of the shared variation in the two variables; the larger this figure is within a particular data set the more linear the appearance of the scattergraph.

The coefficient of determination shows statistically how much of the variation in outcome could be attributed to the variation in input. By squaring the co-efficient of correlation and multiplying by 100 it is possible to produce a figure showing the percentage of variance in one variable that can be attributed to its linear relationship with the other variable. This is the co-efficient of determination. For example, a correlation co-efficient of 0.71 would produce the following (0.71 x 0.71) x 100 = 50.41% so we could say that just over 50% of the variation in the GCSE average grades of the pupils can be explained statistically by the variation in the Edinburgh Reading Test results. The standard error of estimation gives the error margin to be taken account of in any estimate of likely outcome based upon a given indicator score.

By selecting a value for the indicator score it is possible to read off the diagonal line what the equivalent examination result would be. For instance, by using past performance as an indicator for the future and given that you would know a year group's indicator score before their results it would be possible to feed this in and see what you might expect them to achieve. Similarly one might feed in an individual's indicator score and see if what s/he actually achieved was better or worse than the average for someone of that ability as indicated by the regression

line.

One must be aware that the line of best fit may seem somewhat skewed because several pupils may be plotted on exactly the same point therefore making it seem that there were far fewer plots on one side of the line than the other.

An interesting exercise, discussed later in this thesis, is to produce a graph from the combined results of a number of years, finding an average indicator score and the examination mean. One can then enter in the actual average indicator score for a particular year and see if that year's average examination result was higher or lower than the average shown by the regression line.
If the line for the particular year being considered is higher throughout its length than the line for the combined years then it can be considered a successful year although how successful would depend how far it was above the average line. One would expect there to be variation each year and for one year to be judged significantly better than the average the line would have to be above that of the average by more than the standard error of the combined sample.

In using a consortium approach to this analysis one is able to compare one's own school graph with that produced by pupils using the same indicator but in all participating schools. In this way one can gain a measure of the school's examination performance, value added or not as the case may be. Of course there are other factors to take into account but at least one would have an indicator that takes account of the potential differences in ability of students in different schools. Another use is to produce a similar graph for a specific subject area, again combining the results from a number of schools, and then using this as a baseline against which to compare the graph for the same subject area but in one school.

If an individual school has a regression line that runs above that of the combined sample for its entire length then it has achieved better results with similar ability

pupils across the ability range. If the individual school's graph perhaps runs above the combined schools' line for pupils with test scores of, say, 70 - 95 but then dips below the combined schools' line for pupils with scores above 95, this latter scenario would possibly indicate that more able pupils did less well in this particular school's subject area than the average performance for such pupils in the combined schools' sample for that subject.

By considering the combined schools' subject graphs and using these to predict grades for pupils of given indicator scores, allowing for the error of prediction (See *Appendix B*) , then once one knows a pupil's test score it is possible to set target grades in each subject area they are studying and consider their current performance in relation to the predicted grade. Obviously there are many factors that play their part in a pupil achieving their eventual subject grades, and there is the danger of creating a self-fulfilling prophecy, but used sensibly and with discretion there is the potential here for identifying early on in their secondary education gross under or over-achievers, relative to the other pupils in the sample, and doing something about it.

**Frequency Graph**

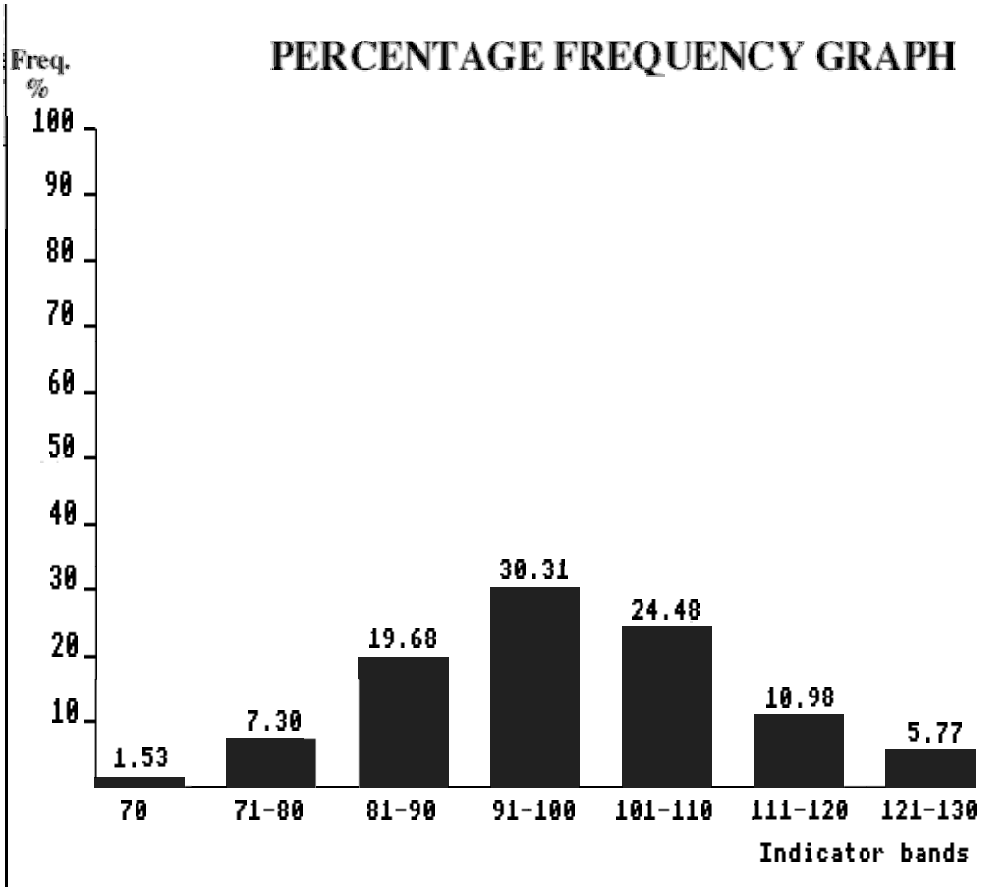
This display ( *Figure 5.8* ) shows the relative frequencies of pupils with different indicator scores in a representative sample. The indicator score along the x axis is the Edinburgh Reading Test (ERT) which operates on a scale of 70 to 130 but is here broken down into ability bandings. The scale on the y axis shows the percentage frequency of pupils with scores in any of the seven ability bandings. The exact percentage figure for each ability banding is shown above the respective columns. The bar chart therefore shows the spread of pupils in the sample with scores in each ability banding expressed as percentages of the sample population. For example, 30.31% of this sample had ERT scores in the range of 91-100.

The larger the sample population (number of pupils) the more likely the graph is

to show an approximately normal distribution curve if one were to join the mid-points of the tops of each column with an imaginary curved line interpolating.

Although the Edinburgh Reading Test is on a continuous scale from 70 to 130 I chose to plot the frequency of pupils in particular ability ranges rather than, say, the number of pupils with an indicator score of 97. My reasoning behind this

**Figure 5.8**




is that in a normally distributed school sample of 200 pupils the frequency of even the modal value is often no more than 10 which equates to a percentage frequency of 5%. Many frequency values would be fractions of one percent and this would not give much visual impact. If the y axis scale were stretched to emphasise the difference between such small figures, when a school sample that was not normally distributed was shown, the scaling of the y axis would have to be adjusted to accommodate the larger percentages and this would give misleading visual comparisons. Also, in a school environment teachers and managers tend to think in terms of the number of pupils in particular ability

bandings and this bar chart graph reflects that fact.
Separate graphs can be produced according to gender, particularly useful when looking at gender differences in examination outcomes.

The utility of these particular graphs is clearly shown later in this thesis in the case study involving School X where the spread of ability within the school is very different when comparing boys with girls. The graphs demonstrate this well (*Figures 6.8a and 6.8b, page 130*). With small populations the graph can be particularly useful in showing the spread of ability within a given group. There may be quite marked differences in the distribution of ability between two groups even though the mean ability indicator scores are nearly the same.

**GCSE grades by pupil ability banding**

This table ( *Figure 5.9* ) shows a breakdown of the GCSE results for pupils, in this case those who had Edinburgh Reading Test (ERT) results as their prior test, according to various prior ability bandings and also bandings according to their average GCSE grade. Thus it is possible to see how many pupils in ability band 70-80 obtained average the average GCSE grade of approximately "D" and their percentage of the whole combined school population. In this case 35 and 1.2% respectively.

This analysis sheet can be compared with and complement the bar graph of pupil ability distribution entitled "Frequency graphs" shown just previously. This sheet is not as graphic but does include the additional information about the average outcomes of each ability banding.

This table is available for the individual school population, as well as larger combined school sample populations as shown here, and can also be produced for respective genders and individual subject areas.

**Figure 5.9**

Schools with Edinburgh Reading Test information

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 |

NOTES

*The average grade is calculated for each pupil.*

*A pupil in the A* column will have an average greater than 7pts.*
*Pupils in the A column will have averages greater than 6 but less than or equal to 7 etcetera.*

*The points scale from which the averages are calculated is A*=8 A=7 B=6 C=5 D=4 E=3 F=2 G=1 U=0*

*These figures are based only on those pupils who had prior scores.*

The GCSE grades are based upon average grades achieved by individual pupils and are not the number of specific grades achieved. For example, in the above table, 5 candidates in the ability band 70-80 achieved average grades in excess of a C grade (5.00) but no greater than a B (6.00). These pupils must, therefore, have average grade scores in the range of 5.1 - 6.00 to be grouped in the B grade column. To achieve this average these candidates must have scored at least some Bs and possibly some As.

The pattern of the numerical spread as shown in tables such as *Figure 5.9* reflects that shown by the Scatter graphs (such as *Figure 5.7*) as would be expected. In many ways both sheets show the same information but in different ways. With this information a school can look at how it is performing with specific ability bands year on year. For instance if a school had chosen to focus

particularly on those pupils in the 91 - 100 range in order to secure more C grades then the results of that policy should be apparent in the data above.

What is very apparent from the above data is that there are pupils in the low ability bands who are capable of gaining a good set of GCSE results but they are relatively rare. Teachers must use their experience and skill and be prepared to say that in these particular cases the test score is misleading, the pupil is more able than his / her 'intake' score suggests and teacher expectations of that pupil's potential should be raised appropriately.

**Comparisons by subject area between schools**

The following abbreviated display ( *Figure 5.10* ) of English department results from different is schools compiled in such a way that each line represents a single school. The performance of each English department in relation to other departments in their own school is shown in the third column of figures by the difference between the subject average grade and the group grade achieved by all the pupils who sat English in all their examinations. A positive difference would show that the pupils who sat English averaged higher grades in English than they did, on average, in their other GCSE subjects, a negative difference would show that they did worse.

**Figure 5.10**

| Subject | Subject grade | Group grade | Diff. | Ind. mean | Group size | Syllabus | Exam board | School |
|---|---|---|---|---|---|---|---|---|
| English | 5.84 | 5.83 | 0.00 | 117.17 | 122 | 1611 | NEAB | |
| language | 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 | |
| | 3.78 | 4.39 | -0.61 | 91.85 | 129 | 2400 | SEG | |

There will be a margin of error in using the mean value to compare departments' scores such that minor differences of a fraction of a grade should be disregarded, particularly if the numbers taking the subject were small. However, as the numbers taking the subject increase, some subject departments such as English being the entire year cohort which could be as many as 200 pupils, so the significance of any differences increases. Another factor to consider would be the consistency of any differences year on year in comparison to other schools and to other subjects within the same school.

The Indicator mean column is the average prior test score for the English group. This is the measure of prior ability for the group. Different schools' English departments then have a common base against which to judge the performance of their pupils. One would expect a department with a higher average indicator test score to get a higher average GCSE grade.

The group size for each school is shown because it is important, particularly in subjects other than English where numbers may not be so large, to consider whether the department performance reflects the efforts of perhaps one teacher with a small group or a large department with several members of staff.
If schools wish, further analysis can be done to show the performance of teaching sets by breaking the subject group into teaching groups and running the analysis again to show averages and differentials for the separate teaching groups but schools must be warned that with the reduced sample sizes the margin of error which must be taken into account when making any judgement increases. The exercise is still useful to see how teaching sets are broken down by ability or truly mixed, gender groupings and so forth.

The syllabus number and examination board are included so that schools can be aware of what courses are being followed elsewhere and also to spot if a particular syllabus seems to be getting good results.

In 1996 for the first time this printout was produced with the schools ranked according to the average subject grade of their departments in each subject area. Previously I had simply mixed up the order to avoid identification of schools and departments but felt that this jumbled order was no longer necessary and the ranking would aid comparisons between school departments. This ranking does not imply any superiority of departmental performance. The margin of error in using the mean grade achieved by the pupils taking the subject in a given school could be such that adjacently ranked schools could reverse their order. One would expect a department whose pupils were, on average, more able as judged by the indicator score to achieve a higher average grade.

The more interesting comparisons come from looking at the differentials achieved by the pupils taking the particular subject over the other subjects they sat (3rd column), or looking at the average abilities of the department groups in relation to the average subject grade achieved by the department's candidates. For example in the English Language results shown above, the fourth ranked school in terms of average subject grade was ranked ninth out of twenty-one in terms of average indicator score. The school ranked eleventh in terms of average subject grade was ranked fourth in terms of average indicator score and one would have expected this department to have obtained a higher average subject grade because of the strong positive correlation between ERT and academic achievement generally but also ERT and performance in GCSE English ( See *Appendices A & B* respectively).

When the full analysis for all subjects is returned to a school, only that particular school is identified, the identity of all the other schools participating remains anonymous. Similar analysis sheets to this were also made available for A level results.

The above analysis sheets and printouts are examples of the sorts of analyses that

were produced by my software and made available to participating schools. They form the basis of my research, the analysis tools with which I intend to explore those aims outlined in Chapter 2 of this thesis.

My interpretation of the data made available to me by schools and my findings will be discussed in the next two chapters.