Educational league tables and spurious comparisons

For two years (January 2012 and January 2013) the Department for Education has released detailed performance tables for Secondary schools in a new format:

(http://www.education.gov.uk/schools/performance/index.html)

This note addresses one particular innovation in these tables which could lead to misleading comparisons because of a couple of statistical problems that have not been recognised.

Following the release in 2012 the following points were conveyed to the Department, via the UK Statistics Authority with the suggestion that suitable care was taken to avoid these problems in the future. It appears that no notice has been taken and the 2013 tables exhibit the same deficiencies.

A new table is produced for each school, and an extract from this for two London Schools is as follows taken from the 2012 tables.

Table 1. School HW. KS4 outcome by KS2 groups: Those below level 4 (Low attainers); those at level 4 (middle attainers); those above level 4 (high attainers). Number of low attainers =32(14%).

	All pupils	Low attainers KS2	Middle attainers KS2	High attainers KS2
Percentage achieving 5+ A*-C GCSEs (or equivalent) including English and maths GCSEs	68%	9%	67%	96%
Percentage achieving A*-C in English and maths GCSEs	69%	9%	68%	97%

Table 2. School BD. KS4 outcome by KS2 groups: Those below level 4 (Low attainers); those at level 4 (middle attainers); those above level 4 (high attainers). Number of low attainers =24 (27%).

	All pupils	Low attainers KS2	Middle attainers KS2	High attainers KS2
Percentage achieving 5+ A*-C GCSEs (or equivalent) including English and maths GCSEs	45%	17%	48%	94%
Percentage achieving A*-C in English and maths GCSEs	45%	21%	48%	94%

If we look at the first row and second column of each table this reports the proportion with 'good' GCSE results for the low attaining pupils at KS2 (those below level 4). The intention is to show how schools 'progress' low attainers (as well as the middle and high attainers shown in the other columns). In effect this is an attempt to introduce what is known as 'differential effectiveness' or 'differential value added' into the reporting. Such more detailed

presentation of data is welcome in principle, and has long been advocated in the school effectiveness literature which has shown how important it can be when a proper analysis of the data is carried out (see e.g. Goldstein, H. (2001). "Using pupil performance data for judging schools and teachers: scope and limitations." British educational Research Journal 27: 433-442.).

In the above tables it appears that the second school (BD) has twice the percentage of low attainers gaining 'good' GCSEs than school 1 (HW) with the implication that it really may be doing a better job of educating such pupils than HW. However, these percentages are, respectively, equivalent to 4 & 3 pupils (17% of 24 and 9% of 32 respectively)! If just one extra pupil from HW were to gain a 'good' GCSE and one from BD to move out of that category then we would see both schools with effectively the same percentage (12%). In other words there is no real difference and a formal statistical comparison verifies this. This issue is relevant to many comparisons for individual schools.

In 2013, for the same two schools the respective numbers of low achievers gaining a 'good' GCSE are 7 and 3 equivalent to 28% and 25% - again no real differences. Because of the small numbers involved, the apparent improvement for HW is in fact not statistically significant so that the data supply little evidence of any real improvement.

There is also a further, more subtle, problem with these and similar comparisons.

We see that BD has twice the percentage of low attainers (27%) as HW (14%). Hence the mean KS2 score for the BD pupils as a whole can be expected to be less than the mean for those at HW, and this is in fact the case. If, for the sake of argument, we assumed a standard 'normal' underlying distribution of scores, then the mean score *for the low attainers* would turn out to be -1.59 and -1.22 respectively, and this will be true more generally. Thus, for this reason alone we would expect HW to do better at GCSE since the 'low attainer' group in BD in fact contains more extreme lower attaining pupils than in the HW low attainer group. Similar issues arise for comparisons based upon the separate 'expected progress' tables where groups are defined in terms of KS2 levels and the associated percentages attaining specified GCSE grades are used to define the extent of progress.

Clearly, knowledge about differential value added effects is important and an attempt to introduce such estimates is welcome. Nevertheless given the way these results have been published for all schools, without regard for small numbers involved and different actual attainment within the defined attainment groups, makes it perfectly possible for people to use these to draw unwarranted conclusions about observed differences. These differences might be between school within a local authority, or for example, between selective and non-selective schools or between academies and non-academies, and such comparisons are already happening. They may appear to have face validity but are quite likely to be spurious. The Department for Education really needs to act to avoid the drawing of misleading and unfair inferences.

Harvey Goldstein

28/02/2013