

## Does your value-added depend on which set you teach?

### Class values - The issues

In every school that takes part in the Yellis project there is probably someone - the head, one of the deputies, the assessment manager or a head of department - who averages residuals by teaching groups to obtain 'class values'. These are then treated as measures of the 'value-added' for the teachers concerned.

- They may be used
- by teachers in their Threshold applications,
  - by team leaders in setting objectives under Performance Management,
  - by heads to show inspection teams that they are on top of self-evaluation,
  - and perhaps soon to judge who shall progress up the Upper Pay Scale.

Most schools will treat class values with the necessary statistical caution. The 95% confidence limit around the average residual for a teaching group of 25 might typically be +/- 0.4 either side of the value. Hence any class average less than +/- 0.4 may be taken as indicating average progress. But class values are often larger than this and, even if they are not, successive years of positive or negative values will begin to paint a convincing picture of teaching effectiveness.

### The evidence

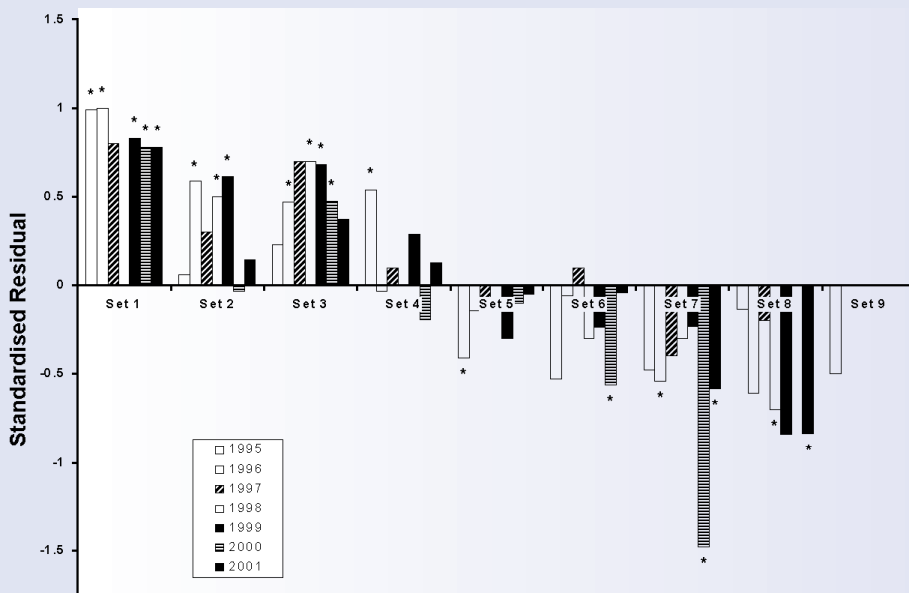
Steve Rogers and John Critchlow have made a particular study of class values. They have found strong evidence that, when groups are set by ability, there is a marked bias towards positive values for top groups, through zero for middle groups, to negative values for the lowest ability groups. Fig. 1, a chart for Mathematics in one school over six years, illustrates the point.

Similar results are found for all subjects where groups are set. Class values for top sets are almost invariably positive, while even the most effective teachers will struggle to reach zero when it is their turn to take the bottom set.

### When are class values derived from value-added data valid?

- An explanation of why this happens is given below. This shows that class values are only valid
- if the groups are of mixed ability, or
  - if students have been set strictly according to the test scores that serve as the baseline for the value added calculations (e.g. the Yellis test if you are using Yellis).

Fig.1 Average residuals for Mathematics sets over six years.

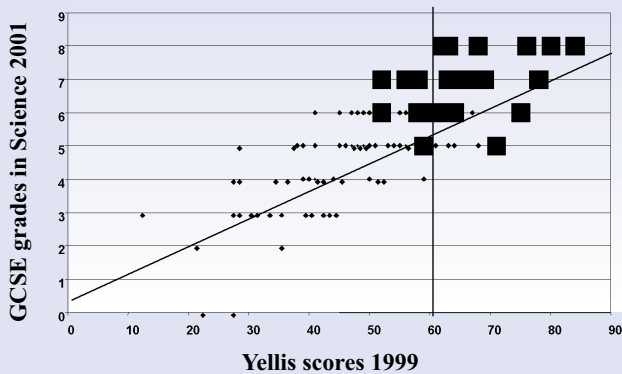


(asterisks indicate where values are statistically significant)

Furthermore, if these conditions are not met, the same setting bias will occur in class values derived by any other value-added method, e.g. the Autumn Package, Chances Graphs or the values produced by SIMS based on comparing the GCSE grades of the same students in all their other subjects.

# Issues & Evidence

**Fig. 2:**  
Vaule-added graph for Science - 2001 - Set 1 highlighted



### Why does this happen?

The setting bias arises because averaging value-added residuals for an ability set involves using the data in a way that is not justified by the value-added methodology. This can be seen by looking at the plot above of GCSE grades for Science in one school against Yellis scores, where the national regression line is also shown and points referring to students in set 1 have been highlighted, (Fig. 2).

All value-added measurements depend by definition on a like-for-like comparison between one group of students and a larger, more representative sample with identical scores on some baseline measure. Using Yellis, the school's overall value-added score for Science (0.4 in this case) is obtained by comparing each point on the graph with the average performance of all the students in the national sample, as represented by the regression line. It would be equally valid to carry out this comparison for all the students within a particular range of Yellis scores, for example those represented by the 30 points to the right of the vertical line.

These denote students who would have been in the top set for Science if sets had been constructed according to Yellis test scores. We could call this the top Yellis set. Their average residual is 0.3, i.e. virtually identical with the department's performance overall.

Students were, however, put into Science sets by the department on the basis of their performance in Science rather than their Yellis scores. Setting is universally preferred to streaming precisely because students have different strengths and weaknesses in different subjects. As the graph shows, many students to the right of the vertical line are good enough at Science to have earned a place in the top set, but a few were not. As might be expected, the latter gained relatively weaker GCSE grades in Science, leaving their more scientifically able peers with a higher average residual. There are also several students to the left of the line who were placed in set 1 but would not have been on the basis of their Yellis scores. They too would be expected to gain better Science grades than other students with similar

Yellis scores, (i.e. in Yellis set 2) and it is clear from the graph that their average residuals will be higher. It is not therefore surprising that the class value for all students in the top Science set is 0.8, giving an artificially high measure of their performance.

Thus, the process of setting according to subject-specific criteria inevitably filters out into the top set students who should achieve high residuals for that subject. Exactly the reverse arguments apply to the bottom set, giving a negative average residual, with a gradation through zero on moving down the sets. The bias will always be in the same direction but it is impossible to know its magnitude in any particular case. It is only possible to draw tentative, qualitative conclusions from deviations from the expected trend. A middle set with a marked positive class value or a bottom set close to zero will have done rather well; a top set close to zero will, on the other hand, have probably performed rather poorly.

Dr Steve Rogers is deputy head of Woldgate School in the East Riding of Yorkshire and can be contacted at [s.g.rogers@btinternet.com](mailto:s.g.rogers@btinternet.com).

Dr John Critchlow is headteacher of Bedale High School in North Yorkshire, whose Beacon School activities include the use of data to improve performance, and can be contacted at [jecritchlow@aol.com](mailto:jecritchlow@aol.com).

## Information Systems from the CEM Centre

YEAR 13 **Alis** Post 16 Baseline Testing

YEAR 12 A-level Information System

YEAR 11 **Yellis** YELLIS Baseline Test

YEAR 10 Year 11 Information System

YEAR 9 **MidYIS** Year 9 Secondary Baseline

YEAR 8 Middle Years Information System

YEAR 7 Secondary Baseline

YEAR 6 **PIPS** Year 6 PIPS

YEAR 5 PIPS

YEAR 4 **PIPS** Year 4 PIPS

YEAR 3 PIPS

YEAR 2 **PIPS** Year 2 PIPS

YEAR 1 Performance Indicators in Primary Schools

Reception PIPS End of Reception PIPS Baseline Assessment

Curriculum, Evaluation and Management Centre

## Measuring what matters.

www.cem.dur.ac.uk