A response to Allen and Burgess:

"Evaluating the provision of school performance information for school choice"

This paper appeared on the CMPO website in June 2010
(http://www.bristol.ac.uk/cmpo/publications/papers/2010/wp241.pdf).

It claims to show, using an analysis of the National Pupil database, that for parents who use school league tables to choose 'highly performing' secondary schools their children will on average have better GCSE results than choosing a school at random.

This paper has not appeared in a peer reviewed journal and as such its findings need to be viewed with caution. In particular it suffers from the following problems.

The first point to note is that there are no formal statistical comparisons in terms of significance tests in the paper so that we cannot evaluate whether the results really do reflect reality.

The procedure adopted can be summarised as follows:

For each student in the data set a local neighbourhood is selected within which a set of schools is available among which the student has a choice of attendance. All comparisons are then made within these neighbourhoods – choice sets. This is an innovative aspect of the design.

For each school in the data set a separate prediction model (multiple regression) of GCSE score is set up based upon student and area level characteristics and including Key stage 2 test scores taken 6 years earlier (2003) before starting secondary school. This is based on 2009 (+ 2008) GCSE data.

A series of parental (decision rules) are presented – for simplicity we consider the proportion of pupils gaining 5 or more A*-C grades at GCSE – results are similar for other outcomes. These are data available in 2003.

For each pupil the paper computes the predicted probability of the outcome, for every school in the choice set, using the prediction formula for that school. For each pupil, the average predicted outcome over this set is then computed. The paper effectively defines a 'good' school as the average predicted probability for the schools in the top half of the ranking (based on the predictions) of the schools in the choice set.

The analysis then looks at choosing the 2003 school with the 'best' decision rule (highest proportion of A*-C grades) and working out its 2009 predicted score and seeing in what proportion of cases this results in choosing a 'good' school. This proportion is then computed for all the other schools in the choice set. For each pupil in each school the 'counterfactual' is to suppose that they could have attended any other school – the

constraint being that the schools taking part should have a 'similar' composition. That is, they are loosely matched in terms of the predictor variables, but not in terms of the GCSE score. The authors show that the former proportion is higher than the latter and conclude that choice matters. They also repeat the analysis using 2003 value added scores and show a much weaker or non-existent difference.

There are several problems with this analysis.

First, the lack of any statistical tests is important since the predictions are based solely on the numbers in each school (over 2 years) and so the predictions have a great deal of uncertainty attached to them

Secondly, and somewhat more technically, the actual prediction formulae are based upon the (joint) distribution of all the variables (including GCSE) in each school. Applying the formula estimated for school A to the pupils in school B who are part of the counterfactual reallocation, therefore makes little sense, since they will have a different (joint) distribution and hence the prediction does not apply to them.

Thirdly and importantly, it Is well known that there is a relatively high (0.7 - 0.8) correlation between average GCSE scores across a 6 year period. Thus, a 2003 school with a high proportion of A*-C grades will also tend to have high GCSE scores 6 years later. Let us consider, therefore, taking a set of pupils from an average 2003 school and predicting how they would have performed in the best school (in terms of GCSE grades) in 2003, using an optimum 2003 prediction equation (similarly to what the authors do for 2009). Clearly, if the prediction is almost perfect then we are accurately predicting the actual performance in 2003 of the *best* school. It is hardly surprising that this will be higher than if we chose to assign the pupils to an average school! Since the 2003 performance is highly correlated with that in 2009 we would also find that the same school in 2009 was doing better than average. Clearly, the actual predictions are less than perfect and we are dealing with 2009 and not 2003, but we still end up with a result that favours the 'best' school, as the authors indeed find. Yet this is just what we expect – an artefact of the way the study has been designed and a reflection of high correlations for overall performance across time. By contrast, the typical 'school effects' analysis would take as outcome (in 2009) the 'value added' performance, that is the *difference* between the actual performance (GCSE grades) and the predicted value given prior achievements etc.

Finally, therefore, we are led to conclude that this study has such flaws that it should not be considered as contributing in any serious way to the debate on league tables and choice. In particular, the work of Leckie and Goldstein (2009, 2011) shows that future predictions, using value added league table models, have such large measures of uncertainty associated with them that parental choice based on these effectively does not exist.

Harvey Goldstein

University of Bristol