Leckie, G. B., & Goldstein, H. (2011). Understanding Uncertainty in School League Tables. Fiscal Studies, 32(2), 207 - 224. 10.1111/j.1475-5890.2011.00133.x

Understanding uncertainty in school league tables

George Leckie and Harvey Goldstein

Centre for Multilevel Modelling

University of Bristol

Abstract

In England, contextual value added (CVA) school performance tables are published annually by the government. These tables present statistical model based estimates of the educational effectiveness of schools together with 95% confidence intervals to communicate their statistical uncertainty. However, this information, particularly the notion of statistical uncertainty, is hard for users to understand. There is a real need to make school performance tables clearer. The media attempt to do this for the public by ranking schools in so-called 'school league tables', however they invariably discard the 95% confidence intervals and in doing so encourage the public to over interpret differences in schools' ranks. In this paper, we explore a simulation method to produce simple graphical summaries of schools' ranks which clearly communicate their associated uncertainty.

Keywords: contextual value added, institutional comparisons, multilevel model, performance indicators, school accountability, school choice, school effectiveness, school league tables, school performance tables, value added

Address for correspondence: George Leckie, Centre for Multilevel Modelling, Graduate School of Education, University of Bristol, 2 Priory Road, Bristol, BS8 1TX, UK.

E-mail: g.leckie@bristol.ac.uk

1. Introduction

1.1 *What is CVA?*

In England, the government publish annually secondary school performance tables to hold schools accountable and to assist parents choosing schools. The government's official measure of school effectiveness is its contextual value added (CVA) measure and it is this measure which we focus on in this paper. The CVA measure gives a separate score for the educational effectiveness of each school. These CVA scores are based on estimates derived from a statistical model that explicitly adjusts for a wide range of intake differences between schools in pupils' academic and background characteristics. (Appendix A describes this model). The CVA scores are presented with 95% confidence intervals to communicate their statistical uncertainty.

It is important to contrast CVA with the most prominent performance indicator published in the secondary school performance tables: the percentage of children, in Year 11 in each school, who gained five or more A* to C General Certificate of Secondary Education (GCSE) grades. This measure is frequently misinterpreted as a measure of the quality or educational effectiveness of schools. Such an interpretation is invalid as no recognition is made for intake differences between schools in pupils' academic abilities; the highest scoring schools are largely those that already had the highest achieving pupils when they entered secondary education.

1.2 *Who uses CVA and for what do they use if for?*

CVA is used for two main purposes. First, a range of users, including the public, local authorities, the national school inspection system and the government, all use CVA to hold schools accountable for how effectively they have educated their pupils. Second, parents use CVA to choose which school to send their child to. An important problem is that while the

published CVA scores provide relevant information for holding schools accountable, they provide misleading information for parents choosing schools.

The published CVA scores estimate how effective schools were for the cohort of children who have just completed their GCSE examinations (i.e. the relevant information for holding schools accountable). However, the relevant information for parents choosing schools is instead how effective schools are predicted to be for their child's cohort who are moving through the education system seven years after the cohort of children who have just completed their GCSE examinations. In Leckie and Goldstein (2009, 2011) we predicted schools' future CVA scores and demonstrated that these were so uncertain that almost no schools could be distinguished from the overall average, or from one another, with an acceptable degree of precision. The conclusion of that work was that school league tables are unreliable and misleading guides for school choice. (See Goldstein and Leckie, 2008, for a non-technical treatment of this issue).

Given that the published CVA scores are not appropriate for parents choosing schools, in this paper we largely discuss CVA, and present our simulation method and graphical approach to communicating the uncertainty in CVA, in terms of using CVA for school accountability purposes. However, in Section 4 of the paper we shall return to the issue of using CVA for school choice purposes and we shall explain how our simulation method and graphical approach can be modified so that they can also be used to assist parents choosing schools.

1.3 *Why is CVA so hard for users to interpret?*

CVA scores and their associated confidence intervals are hard for users to interpret and this difficulty is compounded by the lack of explanation of CVA given by the government. The first point of difficulty is that CVA scores measure the effectiveness of schools relative to a standardised national average of 1000 points. CVA scores therefore have no absolute

interpretation; schools with scores greater than 1000 are described as being more effective than the national average while schools with scores less than 1000 are described as being less effective than the national average.

The second point of difficulty is that no attempt is made to communicate to users the units in which CVA scores are measured. For example, a school with a score of 1006 is six points more effective than the average school, but no indication is given as to whether a six point difference is a substantively large difference. The answer lies buried in a government technical document on how each pupil's GCSE performance is calculated (Appendix B gives the web link to this document): a six point difference in CVA is approximately equivalent to the average pupil in the more effective school scoring one grade higher on one of their GCSE examinations than the average pupil in the less effective school.

The third point of difficulty is that little attempt is made to communicate the concept of 95 per cent confidence intervals and how they should be used. In the CVA tables, each 95 per cent confidence interval quantifies how precisely each CVA score is estimated. Loosely speaking, each 95 per cent confidence interval gives the range of scores within which we are 95 per cent certain that the true CVA score for that school lies. Each published CVA score lies at the midpoint of its associated confidence interval and represents the best estimate for that school. The more precisely a school's CVA score is estimated, the narrower its confidence interval. For example, large schools have narrower confidence intervals than small schools and so the CVA estimates for large schools are more reliable than those for small schools. The main use of the 95 per cent confidence intervals is to judge whether each school is significantly more or less effective than the national average. Schools with scores greater than 1000 and whose 95 percent confidence intervals do not overlap with a score of 1000 are described as being significantly more effective than the national average while schools with scores less than 1000 and whose 95 percent confidence intervals do not overlap with a score of 1000 are described

as being significantly less effective than the national average. Those schools whose 95 percent confidence intervals include a score of 1000 are judged as not differing significantly from the national average.

The fourth point of difficulty is that the 95 per cent confidence intervals are only appropriate for comparing a single school to the national average as described above. However, these are not typically the types of inference that most users want to make. Instead, most users want to make bespoke comparisons between several specific schools of interest to them. For example, a head teacher might want to check how their school is performing relative to neighbouring schools. The provided 95% confidence intervals are not appropriate for either making multiple comparisons, or for comparing pairs of schools to one another.

There is clearly a great need for CVA to be better explained and communicated to end users.

1.4 *How do the media present CVA to the public?*

Each year the media republishes the government's school performance tables in the form of so-called 'school league tables'. (See Appendix B for web links to the BBC and The Guardian's CVA school league tables as examples.) These tables attract considerable public attention.

The media typically make two substantial changes when they present CVA. First, they discard the 95% confidence intervals. Second, they present schools in rank order of their CVA scores, placing the most effective schools at the top of the league table and the least effective schools at the bottom. To what extent these changes reflect the media's own lack of understanding of CVA, as opposed to their desire to simplify CVA for the public, is unclear. However, what is clear is that these distortions to the original information are highly misleading. By not presenting the confidence intervals, the media present CVA scores as if they were free from sampling error and therefore as if they were completely reliable estimates

of school effectiveness. This is far from true. Indeed, the sampling error of CVA scores is so great that, nationally, only around a half of schools are statistically distinguishable from the national average. By presenting schools in rank order of their CVA scores, the media encourage users to focus on schools' league table positions rather than on their estimated levels of educational effectiveness. Together, these two changes to the government's presentation of CVA encourage users to interpret even the smallest differences in schools' league table positions as genuine differences in their educational effectiveness.

In sum, the media's presentation of CVA is highly misleading. Their attempt to make CVA easier to understand by ranking schools is undermined by their failure to communicate the uncertainty in these rankings.

1.5 *What do we do in this paper?*

In this paper we explore a simulation method to produce simple graphical summaries of schools' ranks that simultaneously communicate their associated uncertainty. This method has been used previously by medical statisticians to communicate uncertainty in healthcare league tables (Goldstein and Spiegelhalter, 1996; Marshall and Spiegelhalter, 1998), but we are not aware of it having been applied to school league tables. We then demonstrate how this method facilitates bespoke graphical comparisons between schools of specific interest to the user. Such comparison tools are relevant for head teachers who want to compare how their school is performing relative to neighbouring schools and also to parents who wish to choose which school to send their child to.

For simplicity, we conduct all our analysis for schools in one local authority: Bristol. However, the data we use and that for other local authorities is publicly available and so the

interested reader can not only replicate our results, but they can repeat our analysis for schools in other local authorities. (See Appendix B for web links to the CVA data).

In Section 2 we reproduce the government's and the media's CVA tables for Bristol schools and describe what users can and cannot infer from each presentation. In Section 3 we describe the simulation method and introduce our simple graphical summaries of CVA ranks and their associated statistical uncertainty. Sections 2 and 3 of the paper are presented in the context of using CVA to hold schools accountable as we have argued that the published CVA scores are not appropriate for school choice purposes. However, our simulation method can be modified to give graphical summaries of CVA ranks that are relevant to parents choosing which school to send their children to. Section 4 therefore describes these modifications and compares these adjusted summaries to those relevant for school accountability purposes, which were presented in Section 3. Section 5 concludes.

2. The government's and the media's presentations of CVA

2.1 *The government's presentation of CVA*

Table 1 reproduces the government's 2010 CVA school performance table for the 19 state secondary schools in Bristol. (See Appendix B for web links to the CVA data). The first column presents the name of each school (which we have chosen to anonymise). The second column presents the CVA scores. The third and fourth columns present the lower and upper bounds for the 95% confidence intervals. The final column presents the number of pupils included in the CVA calculation in each school.

The table shows that 16 of the 19 schools in Bristol have CVA scores higher than 1000. This suggests that schools in Bristol tend to be more effective than the national average. However, careful examination of the confidence intervals reveals that just seven schools performed significantly differently from the national average; only seven schools' confidence

intervals do not overlap with a value of 1000. Four schools scored significantly higher than the national average (G, J, L and R) and three schools (C, E and K) scored significantly lower than the national average. Thus, only one-third of Bristol schools were statistically different from the national average. This figure is lower than that for the country as a whole where 56 percent of schools differ significantly from the national average. The substantial overlap in schools' confidence intervals suggests that it would also be hard to statistically distinguish many schools from one another.

The number of pupils included in the CVA calculation also varied across schools. The smallest school (school O) included 47 students compared to 210 students at the largest school (school C). The precision of the estimated CVA scores are an increasing function of school size and so large schools have more precise scores and therefore narrower confidence intervals than small schools. For example, while schools O and P both have a CVA score of 1009.6, the score for school P is based on more pupils and so is estimated more precisely leading it to have a narrower confidence interval.

2.2 *The media's presentation of CVA*

Table 2 reproduces the media's typical presentation of the government's data. The 95% confidence intervals and school size are omitted from the presentation and the schools are ranked from the most effective to the least effective according to their CVA scores.

By omitting the confidence intervals, the table erroneously suggests that schools' CVA scores are completely reliable. This is not the case. Moreover, ranking the schools from the most effective to the least effective, combined with the aforementioned difficulty in interpreting the magnitudes of the CVA scores, explicitly encourages users to focus on schools' league table positions or rankings rather than their estimated levels of effectiveness. However, these ranks, just as the CVA scores they are based upon, have sampling error which must be

communicated to the user. In the next section, we present a simulation method to derive confidence intervals for rankings.

3. A simulation method to allow a graphical presentation of CVA

3.1 *Simulating and presenting confidence intervals for CVA ranks*

The simulation method explored in this section closely follows that presented by Goldstein and Spiegelhalter (1996) and Marshall and Spiegelhalter (1998) in their analysis of healthcare league tables. We start by considering the sampling distributions for the CVA scores. The CVA scores are predicted values from the CVA statistical model (Appendix A describes this model). As with parameter estimates in multiple regression, these predicted values are assumed to have normal sampling distributions. These sampling distributions are also assumed to be independent, which in view of the large number of schools included in the CVA model is approximately true. Figure 1 presents normal sampling distributions for three schools' CVA scores, G, N and O, which were chosen as they are the closest three schools to the authors' university work address. The mean of each sampling distribution is simply given by that school's CVA score, while its variance is derived from that school's 95 per cent confidence interval. Specifically, each school's variance is calculated as the square of its standard deviation where the latter is calculated as that school's 95 per cent confidence interval divided by twice the standard normal deviate (i.e. 2×1.96). This calculation reflects the fact that 95% of the probability mass of a normal distribution lies within approximately 1.96 standard deviations of the mean.

In Figure 1, the sampling distribution for school G is plotted furthest to the right and has the tightest variance as, out of the three schools, it has the highest CVA score and the most pupils. However, while the CVA scores in Table 1 suggest that school G is the best school, school O the second best school and school N the third best school, the figure shows substantial

overlap in the three sampling distributions; it is unclear to what extent the schools' levels of educational effectiveness can be statistically distinguished from one another.

To apply the simulation method, we appeal to a Bayesian interpretation of the sampling distributions. Under a Bayesian interpretation, each distribution provides the probability distribution of each possible CVA score for that school. We run the simulation method for a large number M (we have used $M = 10{,}000$) of iterations where at each iteration the method samples at random a value from each of the 19 schools' distributions. We then rank the schools at each iteration based on their simulated values. Given the clear overlap in the distributions, schools' ranks will often change from iteration to iteration.

One way to communicate the statistical uncertainty in schools' ranks is to plot these ranks together with their associated 95% confidence intervals. The lower and upper bounds for these 95% confidence intervals are given by the 2.5^{th} and 97.5^{th} percentiles from their respective distributions of M simulated ranks. Figure 2 plots the simulated 95 per cent confidence intervals for schools, ranked from the most effective (positioned at the top of the vertical axis) to the least effective (at the bottom of that axis) according to their CVA scores, against their simulated 95 per cent confidence intervals (on the horizontal axis). Thus, school L, the most effective school, is positioned at the top of y-axis while school C, the least effective school is positioned at the bottom of the y-axis.

From Figure 2, the statistical uncertainty involved in ranking schools is immediately apparent. Nearly all schools' ranks have wide and substantially overlapping confidence intervals. For schools placed in the middle of the league table, the rankings are particularly unreliable. For example, school M is ranked 10^{th} best, but its "true" rank has a 95% confidence interval ranging from the 4^{th} best school to the 16^{th} best school. Figure 2 demonstrates that even for school accountability purposes the CVA scores only contain limited information. In Section

4 we shall see that there is considerably more uncertainty involved in using CVA scores for school choice decisions.

3.2 *Simulating probabilities for making bespoke comparisons*

In the introduction, we explained that the 95% confidence intervals associated with the published CVA scores are only appropriate for comparing a single school to the national average; they are not appropriate for making multiple comparisons, or for comparing pairs of schools to one another. Here we show how the simulation method can be adapted to make bespoke comparisons between several specific schools. We again consider schools G, N and O. Using the M simulated ranks for each of these three schools, we calculate the proportion of iterations for which each school was ranked 1st, 2nd or 3rd. These proportions can then be interpreted as the probability that each school is the best, second best or third best school of the three respectively. Figure 3 reports these probabilities.

Figure 3, clearly shows that while school G appears to be the best school, this is far from certain. The probability that school G is the best school is just 0.69, compared to a probability of 0.26 that school O is the best school and a probability of 0.05 that school N is the best school.

An advantage of Figure 3 over Figure 2 is that it communicates uncertainty in terms of probabilities rather than through confidence intervals; probabilities are perhaps an easier concept for the public to grasp than confidence intervals. However, probabilities are not the only way in which the results of these bespoke comparisons can be expressed. For example, an equivalent way to summarise these results is to say that school G has a 7 in 10 chance of being the best school, while the chance that school O is the best is 1 in 4 and the chance that school N is the best is 1 in 20. Another alternative is to report the results in terms of the odds that each school is the best school. For example, the odds that school G is the best school are 2.23 to 1 $(= 0.69/(1 - 0.69))$ or approximately 9 to 4.

Bespoke comparisons between other triplets of schools will lead to different results from those presented in Figure 3. For some triplets of schools it will be easier to accurately rank the schools while for other triplets it will be harder. The precision with which schools can be ranked depends on the schools' CVA scores and the number of pupils that they are based upon. The further apart schools' CVA scores, the more accurately the schools can be ranked. Similarly, the larger the schools, the more precisely estimated their CVA scores will be and so again the more confident we will be in making rankings.

In Figure 3, we have compared three schools. However, similar figures can be produced for comparing two schools to one another or for comparing four or more schools to one another. For example, when we simply compare school G and school O, we find that the probability that school G is better than school O is 0.73. More complicated probabilities, should they be of interest to the user, can also be calculated. Returning to our three school comparison, we can calculate the joint probability that school G is better than school O which in turn is better than school N and this is 0.63.

4. Using school league tables to inform school choice

One of the main justifications given by the government for publishing CVA scores is that they provide information to parents choosing which school to send their child to. Leckie and Goldstein (2009, 2011) highlighted a fundamental limitation of using the published CVA scores for this purpose. The relevant information for parents choosing is how effective schools are predicted to be for their child's cohort. However, the published CVA scores instead describe how effective schools are estimated to be for a cohort of children who entered secondary schooling seven years prior to the time when the children of parents choosing will enter their schools. For example, consider the 2010 CVA school league table. This table is

based on the cohort of children who started secondary schooling in Autumn 2005 and who took their GCSE examinations in Summer 2010. The cohort of children who will use the 2010 tables to choose secondary schools, on the other hand, will only enter their secondary schools in autumn 2012 and will take their GCSE examinations in summer 2017. Thus, the relevant information for those choosing is not how effective schools were for the 2010 cohort, but rather how effective schools are predicted to be for the 2017 cohort. (See Goldstein and Leckie, 2008, for a non-technical treatment of this issue).

Our earlier work (Leckie and Goldstein, 2009) presented a formula for predicting schools' future CVA scores based on their current CVA scores. There was so much uncertainty in these predictions that almost no schools could be distinguished from the overall average, or from one another, with an acceptable degree of precision. The conclusion of this work was that the CVA school league tables are unreliable and misleading guides for school choice. In subsequent work (Leckie and Goldstein, 2011) we confirmed that these conclusions still held even when predictions are based on multiple years of CVA school league tables.

Given the importance of the conclusions in our earlier work, in this section we repeat the analysis presented in Section 3 using schools' predicted 2017 CVA scores rather than their estimated 2010 CVA scores. To do this, we first calculate the predicted 2017 CVA scores and their associated confidence intervals using the prediction formula we presented in Leckie and Goldstein (2009) (see Appendix A for details). Then we apply the simulation method as before to obtain sampling distributions for the rank of each school's predicted 2017 CVA score. Figure 4 plots schools against the simulated 95 per cent confidence intervals in the same way as in Figure 2.

The schools in Figure 4 are ranked in the same order as in Figure 2. This is expected as both figures are based on the same underlying data, the 2010 CVA scores. The key difference

is that the 95% confidence intervals in Figure 4 are considerably wider than their corresponding ones in Figure 2. This result reflects the fact that inferences about schools' future levels of effectiveness are far less certain than inferences about their current levels of effectiveness. The 2010 school league table simply does not contain enough information to be able to make accurate predictions about schools' ranks in seven years time. In other words, there is considerably more uncertainty in ranking schools for school choice purposes than there is in ranking schools for school accountability purposes.

In Section 3, we also described how the simulation method could be adapted to make bespoke comparisons between several specific schools. These are very much the type of comparisons that parents make. Indeed, in Bristol, and in many other local authorities, parents are required to express their 1st, 2nd and 3rd preference for the secondary school that they wish to send their child to. (See Appendix B for a web link to the Bristol local authority application form.) In Figure 5 we therefore present the results from a bespoke comparison of schools G, N and O in terms of their predicted ranks for the 2017 cohort.

Figure 5, like Figure 4, maintains the ordering of the schools and so school G continues to appear to be the best school, followed by school O then school N. However, the probabilities that each of these schools will be the best school are far more similar than before. Now, the probability that school G, N or O is the best school is 0.44, 0.23 and 0.33 respectively. This compares to probabilities of 0.69, 0.05 and 0.26 for when we compare schools' ranks for the 2010 cohort. Thus, we are far less certain that school G will be the most effective of the three schools for the 2017 cohort than we were that school G was the most effective school for the 2010 cohort. Once again, the 2010 school league table simply does not contain enough information to be able to make accurate predictions about schools' ranks in seven years time.

And again we see that there is considerably more uncertainty in ranking schools for school choice purposes than there is for school accountability purposes.

5. Conclusion

The issue of statistical uncertainty is a fundamental aspect of reporting CVA, but is typically ignored by the media. There is therefore a real need for the government to do more to make its CVA school performance tables easier to understand. In this paper we explored one possible approach. We presented a simulation method that enables schools to be ranked, but in a way which communicates the uncertainty in making such rankings through simple statements about the chance that each school has the highest score rather than through the current use of confidence intervals. Our approach appears particularly suited to making the type of bespoke comparisons made by parents when choosing which local school to send their child to. However, for school choice purposes, it is important to stress a point that we made in earlier work which is that the government's CVA school performance tables substantially understate the uncertainty in predicting schools' future levels of effectiveness and that ignoring this additional uncertainty will also potentially mislead the public.

We hope that this paper encourages a debate about the government's and the media's presentation of CVA scores and, in particular, about the communication and public understanding of uncertainty in school league tables. It is a straightforward task to implement our proposed approach within a simple piece of software that could be run from a web site, and we are having discussions about taking this proposal forward.

Acknowledgements

We are grateful for the helpful comments provided by the Guest Editors and also for those provided by the referees. This research was funded under the UK Economic and Social Research Council's National Centre for Research Methods programme.

Appendix A

The CVA scores and their associated 95% confidence intervals presented in the government's CVA school performance tables are derived from a multilevel statistical model (Goldstein, 2010). The model is fitted to the GCSE performances (age 16) of all pupils in state maintained mainstream schools in England for a given cohort. (See Appendix A for a web link to details as to how each pupil's GCSE performance is calculated.) The model is a two-level (pupils nested within schools) random intercepts multilevel model where adjustments are made for pupils' achievements at intake (age 11) and for a range of pupil background characteristics. (See Appendix A for a web link to the full model specification). This model can be fitted in many statistical software packages, including those that specialise in multilevel models such as MLwiN (Rasbash et al., 2009). Specifically, the CVA scores are posterior estimates of the school random effects from this model while the 95% confidence intervals are given by the CVA scores ± 1.96 times their associated 'comparative' standard errors. In multilevel models, random effects are centred on zero and so approximately half the posterior estimates will be positive and half will be negative. As a final step, the government centre their CVA scores on 1000, presumably to avoid potential public confusion over assigning half the schools in the country negative scores.

In Section 4, we predict schools' 2017 CVA scores using the prediction formula presented in Leckie and Goldstein (2009). (See equation 4 of that paper.) In calculating these predictions, as in our earlier work, we assume a correlation of 0.64 for the correlation in CVA scores seven-cohorts-apart and we assume that the between-school and within-school variance parameters

are constant over time and therefore equal to the 2010 CVA model parameter estimates of 415.302 and 4426.127 respectively. (See Appendix A for a web link to these published parameter estimates). The substantive points made in this paper are not sensitive to these assumptions.

Appendix B

This appendix gives the web references to all data sources used in our analysis. All data is publicly available.

The government's 2010 CVA school performance tables for Bristol local authority are accessible at:

- http://www.education.gov.uk/cgi-bin/performancetables/group_10.pl?Mode=Z&Type=LA&Begin=b1&No=801&Base=e&Phase=1&F=1&L=50&Year=10&Key=4&Order=asc

The data can be accessed as a Microsoft Excel spreadsheet or as a comma separated variable file at:

- http://www.education.gov.uk/performancetables/schools_10/xls_10/801.xls
- http://www.education.gov.uk/performancetables/schools_10/csv_10/801.csv

Note that the number of pupils included in the CVA calculation is not provided in the above files, but can be calculated by multiplying the percentage of pupils in each school included in the CVA calculation (CVACov) by the number of pupils at the end of Key Stage 4 (KS4TotPup) and then dividing by 100.

The BBC and Guardian 2010 CVA school league tables are accessible at:

- http://www.bbc.co.uk/news/special/education/school_tables/secondary/10/html/cva_801.stm?compare=
- http://www.guardian.co.uk/education/table/2011/jan/11/bristol-gcse-alevel-tables

Details of how each pupil's GCSE performance is calculated can be found at:

- http://www.education.gov.uk/performancetables/schools_10/points.pdf

The full multilevel model specification and parameter estimates for the CVA model (see Coefficients tab of the Excel spreadsheet) are accessible at:

- http://www.education.gov.uk/performancetables/schools_10/rr24.xls

The Bristol local authority application form for transfer to secondary education in September 2011 is accessible at:

- http://www.bristol-cyps.org.uk/schools/pdf/secondary-application-11-12.pdf

References

Goldstein, H. (2010) *Multilevel statistical models*. Wiley.

Goldstein, H. & Leckie, G. (2008). School league tables: what can they really tell us? *Significance*, 5, 67-69.

Goldstein, H. & Spiegelhalter, D. J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 159*, 385-443.

Leckie, G. & Goldstein, H. (2009) The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*, 835-851.

Leckie, G. & Goldstein, H. (2011) A note on "The limitations of using school league tables to inform school choice". *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. *Forthcoming*.

Marshall, E. C. and Spiegelhalter, D. J. (1998). Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. *BMJ*, 316, 1701-1704.

Rasbash, J., Charlton, C., Browne, W.J., Healy, M. and Cameron, B. (2009) MLwiN Version 2.1. Centre for Multilevel Modelling, University of Bristol.

Table 1 The government's presentation of CVA

School	CVA score	Lower limit	Upper limit	Number of pupils
A	1003.2	993.7	1012.8	176
B	1001.0	990.7	1011.4	150
C	970.3	961.5	979.1	210
D	1004.7	994.7	1014.7	160
E	988.5	979.3	997.8	187
F	1003.4	992.8	1013.9	142
G	1015.7	1005.8	1025.5	164
H	1004.2	994.4	1013.9	169
I	1006.0	996.2	1015.8	167
J	1013.4	1004.4	1022.3	203
K	979.5	969.1	989.9	147
L	1036.2	1026.6	1045.8	174
M	1005.0	994.2	1015.8	134
N	1002.0	987.0	1017.0	65
O	1009.6	992.4	1026.7	47
P	1009.6	998.1	1021.2	117
Q	1007.4	995.4	1019.3	109
R	1028.2	1015.0	1041.4	87
S	1010.8	998.5	1023.1	102

Table 2 The media's presentation of CVA

Rank	School	CVA score
1	L	1036.2
2	R	1028.2
3	G	1015.7
4	J	1013.4
5	S	1010.8
6	O	1009.6
7	P	1009.6
8	Q	1007.4
9	I	1006.0
10	M	1005.0
11	D	1004.7
12	H	1004.2
13	F	1003.4
14	A	1003.2
15	N	1002.0
16	B	1001.0
17	E	988.5
18	K	979.5
19	C	970.3

Figure 1 Sampling distributions for schools G, N and O's CVA scores

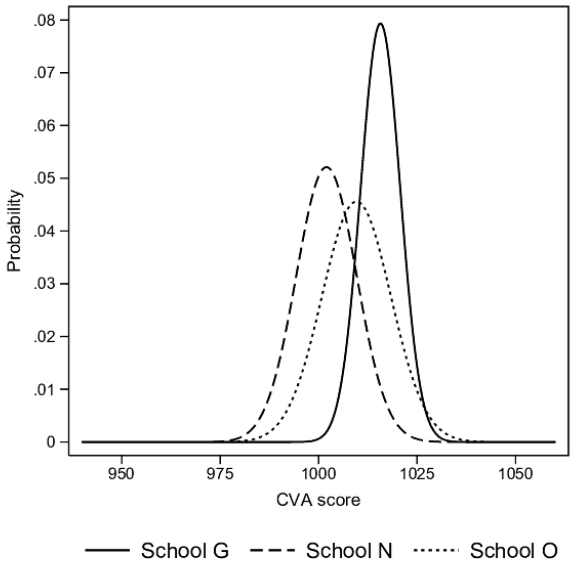

Figure 2 Figure 2 CVA ranks with simulated 95 per cent confidence intervals

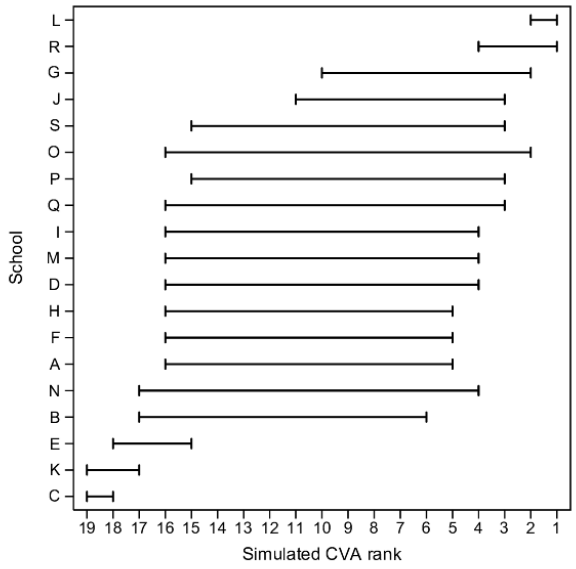

Figure 3 Probabilities that school G, N and O are ranked 1st, 2nd or 3rd

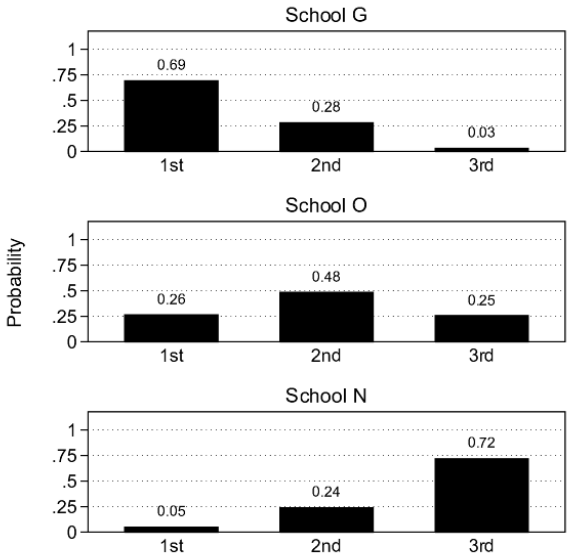

Figure 4 CVA ranks with simulated 95 per cent confidence intervals for the 2017 cohort

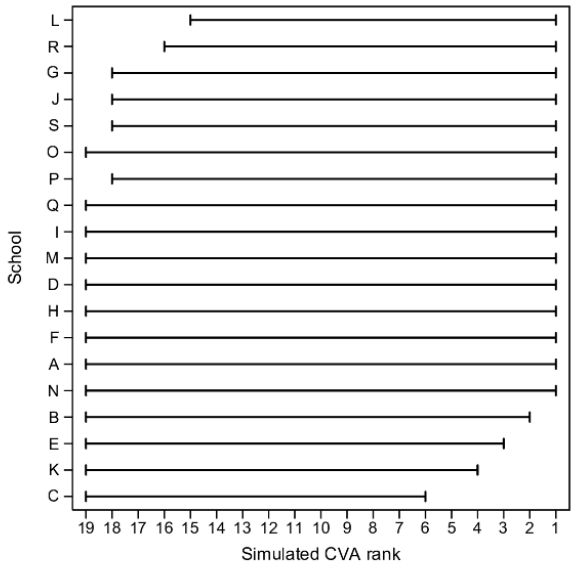

Figure 5 Probability that school G, N and O are ranked 1st, 2nd or 3rd for the 2017 cohort