THE CENTRE FOR MARKET AND PUBLIC ORGANISATION

The Centre for Market and Public Organisation (CMPO) is a leading research centre, combining expertise in economics, geography and law. Our objective is to study the intersection between the public and private sectors of the economy, and in particular to understand the right way to organise and deliver public services. The Centre aims to develop research, contribute to the public debate and inform policy-making.

CMPO, now an ESRC Research Centre was established in 1998 with two large grants from The Leverhulme Trust. In 2004 we were awarded ESRC Research Centre status, and CMPO now combines core funding from both the ESRC and the Trust.

Centre for Market and Public Organisation
Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
http://www.bris.ac.uk/Depts/CMPO/

Tel: (0117) 33 10799
Fax: (0117) 33 10705
E-mail: cmpo-office@bristol.ac.uk

Accurate performance measure but meaningless ranking exercise? An analysis of the English school league tables.

Deborah Wilson and Anete Piebalga

January 2008
Working Paper No. 07/176

ISSN 1473-625X

CMPO Working Paper Series No. 07/176

Accurate performance measure but meaningless ranking exercise? An analysis of the English school league tables.

Deborah Wilson
and
Anete Piebalga

CMPO, University of Bristol

January 2008

Abstract

Parental choice among schools in England is informed by annually published school performance (league) tables. The 2006 league tables included a measure of contextual value added (CVA) for the first time. By explicitly accounting for the characteristics of a school's intake, CVA should provide a more accurate measure of the impact a school has on its pupils' progress, i.e. on school effectiveness. In this paper we use UK government administrative data to replicate CVA and other key performance measures in order to investigate the extent to which the current league tables provide the information necessary to support parental choice on the basis of school effectiveness. We find that while CVA does provide a more accurate measure of school performance or effectiveness, school rankings based on CVA are largely meaningless: almost half of English secondary schools are indistinguishable from the national average.

Keywords: education, performance measures, ranking

JEL Classification: I2, H4

Electronic version: **http://www.bris.ac.uk/Depts/CMPO/workingpapers/wp176.pdf**

Acknowledgements

Thanks to the Leverhulme Trust and ESRC for funding this research through the CMPO. Thanks also for comments on previous drafts to Simon Burgess, Harvey Goldstein, George Leckie, Carol Propper, Helen Simpson and participants at the 2007 IPMN Workshop, Oxford. All errors remain our own.

Address for Correspondence
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol
BS8 1TX
D.Wilson@bristol.ac.uk
www.bris.ac.uk/Depts/CMPO/

CMPO is jointly funded by the Leverhulme Trust and the ESRC

Introduction

The English education system has incorporated elements of parental choice for twenty years, and the 2006 Education and Inspections Bill aims to extend and strengthen the role of choice as a driver for improving educational outcomes. For choice to create the incentives for schools to improve their performance two conditions need to be satisfied (Rothstein 2006). First, school effectiveness needs to be at least a substantial determinant of parental choice; second, parents need to be able to distinguish effective from ineffective schools. In England parental choice is informed by in-depth inspection reports on individual schools and by the annually published school performance or league tables. This paper investigates the extent to which the current English secondary school league tables provide the information necessary to support parental choice on the basis of school effectiveness.

The 2006 league tables included two new performance measures (PMs) for the first time. Alongside the original target indicator of the proportion of a school's pupils gaining at least five passes at GCSE at grade C or above (5AC), there is now an indicator which distinguishes the sub-group whose five good passes include English and maths (5AC:EM)[1]. Also a new measure of contextual value added (CVA) has replaced the value added (VA) PMs which have been included in the league tables since 2002. These new indicators have been introduced partly in response to the shortcomings of previous PMs.

Since they were first published in 1992, 5AC has been the key PM in the English secondary school league tables. This is a summary measure of test score outcomes, combining information on a school's effectiveness with information on the composition of its intake. All other things being equal, a high ability intake will provide a high level output score, regardless of the effectiveness of the schooling provided. This creates potentially perverse incentives for schools who aim to achieve a high ranking in the league tables derived from this PM. These include various creaming strategies, both with regard to selection of pupils (Meyer 1997; West and Pennell 2000) and with regard to selective entry of different types of pupils into certain types of tests (Figlio and Getzler 2002; Cullen and Reback 2006; Wilson et al 2006). The specific 'target' nature of the 5AC measure may also create the incentive for schools to focus resources at those pupils on the borderline between grades D and C, possibly to the detriment of other pupils in the school (Deere and Strayer 2001; Burgess et al 2005; Wilson et al 2006). More generally, a ranking exercise based on raw test scores will favour schools with a higher ability composition. Given the known link between socio-economic status and attainment, this also has the potential to "demonise" schools in poor neighbourhoods while saying nothing about their effectiveness (Raey and Lucey 2003). The 5AC:EM performance measure addresses at least one potential gaming strategy as it prevents schools from entering borderline pupils from more vocational exams (GNVQs, worth four good GCSE passes) in order to boost the indicator (Wilson et al 2006).

Value added (VA) performance measures were introduced in 2002. They take account of pupils' prior attainment, and so provide a measure of a school's impact on the progress of a pupil cohort between two points in time. Value added PMs therefore provide a more accurate measure of school effectiveness (Propper and Wilson 2003; Wilson 2004); they should also reduce the incentive for creaming strategies, as the different abilities of a school's intake are explicitly taken into account. It is not clear, however, whether VA PMs impact on parents' choice of school (and hence on the incentives faced by schools). Certainly a sample of secondary school headteachers expected 5AC to still be the 'headline' figure and for them to therefore have to continue to focus on that indicator (Wilson et al 2006). This is relevant given that the two indicators lead to different rankings across schools (Wilson 2004; Goldstein 2001). The VA PM does not fully separate effectiveness from composition, however. While prior attainment is the most important predictor for future performance (Thomas 1998), other factors outside the school's control also impact on pupil results. These include gender, deprivation and high levels of pupil mobility (Yang et al 1999). Further criticisms of VA PMs include, first, the extent to which they hide evidence of differential effectiveness, i.e. differences in performance by any one school for its pupils of different abilities, which may limit its usefulness to any one parent trying to choose a school for their particular child (Thomas 2001; Wilson 2004). Second, the extent to which, in practice, meaningful comparison across schools according to their VA may be somewhat limited due to the extent to which their confidence intervals overlap (Goldstein 2001). Finally, there is evidence that VA is not as stable across years as PMs based on raw test scores (Kane and Staiger 2002; Gray et al 2001), which again may limit its usefulness as a means of evaluating and comparing school performance.

The introduction of contextual value added (CVA) can be seen as an attempt to further separate effectiveness from composition and so provide a more accurate measure of the actual impact of a school on pupil progress. CVA explicitly takes account of various factors that are independent of schools but which are known to impact on educational outcomes, both at the pupil and the school level, including gender, deprivation and peer group. We discuss how it is constructed below and assess how the published school average CVA performs with regard to the criticisms levelled at its VA predecessor.

The evolution of PMs in the English school performance tables can therefore be seen as a move to better reflect school effectiveness, which is supportive of the government aim of improving outcomes or standards. What is less clear is the extent to which the new measures are able to provide meaningful comparisons of individual schools' effectiveness, and thus provide the more accurate information necessary to support parental choice as an effective driver for such improvements.

In this paper we use UK government administrative data to replicate the key PMs and resultant school rankings for English state secondary schools to provide evidence on two issues. First, we consider the extent to which CVA does isolate school effectiveness from composition. To do this we explore differences between the 5AC, 5AC:EM and CVA rankings, and investigate how CVA depends on different school characteristics. Second, we address the extent to which CVA itself informs parents about the relative performance (effectiveness) of secondary schools. We first investigate whether the published school average CVA hides differences in a school's effectiveness across pupils of different abilities, and then explore the extent to which a ranking exercise based on CVA is meaningful. Here we consider both the degree to which schools can be meaningfully ranked in any one year, and how stable these rankings are over a three-year period.

Method

The general aim of a value added PM is to isolate the impact made by a school on the progress of a cohort of pupils between two points in time. In practice, in England, this means progress between successive Key Stages of the National Curriculum[2]. There are various ways in which value added can be calculated (Wilson 2004). The Department for Children, Schools and Families (DCSF; until recently the Department for Education and Skills (DfES)) uses a national 'median line' approach (DfES 2005): the value added for each individual pupil is the difference between her own output point score and the median (middle) output point score achieved by others with similar input point scores (prior attainment). So, for example, a measure of the value added by secondary schools takes Key Stage (KS) 2 scores as the input (taken at age 11 just prior to entry into secondary school) and KS4 (or GCSE, taken at the end of compulsory schooling at age 16) as the output.

The principal remains the same when calculating contextual value added, although now each pupil is compared to peers who not only have similar input point scores but who are also similar across a range of other contextual factors which are known to impact on educational attainment. A prediction is made for each pupil based on nationally observed patterns, and her CVA score is then the difference – positive or negative – from this prediction. The pupil-level CVA model is able to isolate the impact of each of these contextual factors, and shows how well each pupil is performing relative to similar pupils in her national cohort. The pupil level CVA scores are aggregated to give a school level CVA PM. In the secondary school league tables there is a CVA measure for each school measuring pupil progress between KS2 and KS4.

So what are the contextual factors included in the DCSF calculation? In addition to prior attainment (here, KS2), the model includes the following pupil characteristics:

- gender,
- special educational needs status,
- eligibility for free school meals (a proxy for low income),
- ethnicity,
- English as an additional language,
- mobility (movement into schools at non-standard transfer times),
- within-year age,
- whether the pupil has been 'in care' while at school,
- IDACI measure of deprivation of the pupil's home neighbourhood[3].

The model also includes two school level attainment variables: the average KS2 score of the pupil's cohort and a measure of its spread (standard deviation), which try to take account of the impact of a pupil's peer group on her progress through school. This reflects the body of evidence that peer group matters for individual attainment (see Tough and Brooks (2007) for a recent review). The DCSF model does not capture all the factors which have an impact on pupil attainment, however. There is evidence, for example, that better predictions are obtained if account is taken of pupil performance prior to KS2 tests (Goldstein and Sammons 1997). Because the contextual factors that are included in the model are measured at two levels, pupil and school, the appropriate statistical technique employed to calculate CVA is multi-level modelling (MLM) (Aitkin and Longford 1986; Snijders and Bosker 1999; Raudenbush and Bryk 2002; Goldstein 2003).

We replicate the DCSF methodology using government administrative data. Specifically we use data from the PLASC/NPD dataset. The Pupil Level Annual Schools Census (PLASC), part of the National Pupil Database (NPD) is an administrative dataset collected by the DCSF. As it is a Census it contains information on all pupils in state schools in England. PLASC comprises information on pupil characteristics such as those listed above, which can be linked to each pupil's test score history, all of which are contained in the NPD (Barker 2006). We consider three cohorts of pupils. We primarily report results from the cohort who took their GCSE (KS4) exams in 2004, their KS3 tests in 2002 and KS2 tests in 1999. We repeat the analysis for the 2005 and 2006 cohorts (results available on request where not reported in the paper). We use pupils attending mainstream schools for which we have a full set of individual characteristics and home postcode (required to match in the IDACI score of their local neighbourhood). Our 2004 cohort analysis sample contains 444,559 pupils[4].

Table 1 shows the coefficients we derive from a two-level (secondary school and pupil) multilevel model for our 2004 cohort. We also report some descriptive statistics of the estimation sample. The model enables us to predict the total GSCE score (capped at the best eight GCSE grades as in the DCSF calculation) for an individual pupil with any set of characteristics. For example, as expected, the higher the average KS2 score of a pupil, the higher her predicted GCSE score[5]. A pupil eligible for free school meals has a predicted GCSE score that is lower by 20 GCSE points (the difference between a grade A and a grade C, for example) than an otherwise equivalent non-FSM pupil. If she lives in a deprived neighbourhood as measured by the IDACI coefficient she is further disadvantaged. Girls perform better than boys by 16.4 GCSE points, and minority ethnic groups generally perform better than their white peers[6]. Pupils with unusual school mobility, i.e. those who transfer between schools at non-standard times, appear to be at a large disadvantage (we return to this result later). Finally, there appears to be an advantage in attending schools with a higher ability cohort (higher average KS2 score), and with a peer group of similar ability, as shown by a negative coefficient on the KS2 standard deviation variable.

Using this model we can derive a predicted GCSE score for each pupil with a specific set of characteristics. This is then compared to the actual GCSE score of each pupil with those characteristics, and the CVA for that pupil is the difference – positive or negative – between the two. The CVA PM for each school is then calculated by averaging across the individual CVA scores of all pupils in the relevant school cohort. This school-level average is then multiplied by a shrinkage factor, which depends on the number of pupils in a school's cohort and which brings the school residuals (CVA scores) of smaller schools closer to the national average[7]. Finally, 95% confidence intervals are calculated for each school's CVA measure. This is to take account of the uncertainty involved in using a set of test results, achieved by one set of pupils on one day, as a measure of the underlying effectiveness of the school. The school's CVA represents its departure – positive or negative – from the national average (zero) that is the prediction given by the fixed parameters derived from the model and described in Table 1. Two points to note here. First, schools for which the 95% confidence interval does not include zero have a performance that is significantly different at the 5% level to the national average[8]. Second, the standard 95% confidence intervals calculated from the model – and those published in the government's school performance tables – enable statistically meaningful comparisons to be made between individual schools' performance and the overall national average, but not directly between the performance of pairs (or more) of schools. We return to both these points below.

We construct school-level CVA measures for state maintained secondary schools in England. Following the DCSF methodology we estimate the model on mainstream (non-special) schools only. We also delete middle schools (see endnote 2) and schools that have particularly high pupil mobility, which led to them having highly biased estimates (we return to this issue below; see also Goldstein et al (2006)). We omit schools with less than 20 pupils in the cohort, for whom we have full data as described above, due to the statistical uncertainty in CVA and other performance measures for such small cohorts. Our 2004 analysis sample contains 2,940 schools.

For each school in our sample we also construct the two key published target indicators: the percentage of pupils gaining at least 5 GCSE passes or equivalent at grade C or above (5AC) and the percentage of pupils gaining the same number of good passes including English and Maths (5AC:EM). We then calculate school rankings based on each of these three PMs, for 2004, 2005 and 2006.

Results

1 The extent to which CVA isolates school effectiveness from school composition

(a) Comparing PMs and the resultant rankings

First we consider both the relationships between the alternative performance measures published for each secondary school and between the resultant rankings. Tables 2a and 2b show the correlations between the PMs and the rankings respectively[9]. As expected, there is a high correlation between the two PMs based on raw GCSE scores. It is interesting to consider the relationship between 5AC and 5AC:EM in more detail. Figure 1 plots these two performance measures for each school. While the overall correlation coefficient is high at 0.93, virtually all schools lie below the 45 degree line, showing that it is easier to hit the target which has no restrictions regarding the subjects that can be included. For some schools the discrepancy between the two measures is quite pronounced: it is possible for a school's score to drop from over 90% of pupils hitting the target to around 30% doing so, once the English and maths requirement is included. This suggests that some schools may indeed have been using the GNVQ strategy discussed above to boost their 5AC performance measure.

Referring again to Tables 2a and 2b, the correlation between either of the target indicators and CVA is much lower, which suggests that ranking schools on the basis of these different PMs would yield different, and conflicting, results. This is further illustrated in Figure 2. Here we plot school rankings based on CVA against their ranking on the 5AC measure. Each triangle represents a secondary school. Schools are lower in the rankings the further they are from the origin. The figure shows that there is very little relationship between the two: it is possible to be ranked high on CVA and low on 5AC and vice versa. The 5AC measure incorporates information on both school composition and effectiveness. By attempting to isolate the latter, the CVA PM produces very different rankings.

(b) Relationship between rankings and school characteristics

Here we explore further the extent to which CVA produces different rankings from 5AC for different school types[10,11]. In Figure 3 we categorise schools according to the percentage of their pupils who are from low income households, defined by their eligibility for free school meals (FSM). In general terms, schools in high income areas have outperformed those in low income areas, according to the 5AC measure. Here, we see if this is still the case with CVA, which controls for the (low) income of a school's intake. We distinguish three sections of the FSM distribution across schools: low, medium and high (which correspond to schools in the lower quartile, inter-quartile range and upper quartile of the distribution respectively). Again we plot CVA versus 5AC rankings for each school, with the lower ranked schools further from the origin. The majority of low-FSM schools are above the 45 degree line drawn on the figure, which shows that low-FSM schools will be ranked higher on 5AC than on the CVA PM, with the reverse being true for high-FSM schools. By taking account of the link between low income and low attainment, the CVA measure enables schools with a high proportion of FSM pupils to achieve a higher ranking, other things being equal. Similar pictures are obtained when we categorise schools according to the average deprivation of their pupils' home neighbourhoods (see endnote 3), and according to the proportion of their pupils with special educational needs.

Figure 4 again plots CVA versus 5AC rankings for each school, this time distinguishing schools according to their admissions policy: comprehensive, academically selective (i.e. grammar schools), secondary modern. One enduring feature of the 5AC league tables is how well grammar schools perform and how highly they are ranked. Here we see if this is still the case when the PM controls for school composition. In Figure 4, schools are again lower in the rankings the further they are from the origin. The striking feature is the grouping of grammar schools – schools that explicitly select pupils by ability – at the left hand side of the figure: grammar schools all rank very highly on the 5AC measure, but once key features of their intake is taken into account, there is much more variation in their performance. This highlights one reason for the introduction of the CVA PM: performance measures based on raw output 'flatter' schools with a high ability intake which, by definition, include grammar schools. We get a very similar picture when we look at single-sex versus co-educational schools due to the fact that the majority of single-sex schools are selective.

There are, however, two indicators included in the CVA model that might create a misleading picture of schools' relative performance. These are the mobility indicators: 'Pupil joined school after September in Year 10' and 'Pupil did not join school in July/August/September in Years 7, 8 and 9'[12]. The inclusion of indicators for pupils who transfer between schools at non-standard times reflects evidence that this increased mobility has a negative impact on these pupils' attainment. This should therefore be factored out of any measure attempting to isolate school effectiveness.

The particularly low attainment of the children of travellers provides one example of this (Ofsted 2003; Equalities Review 2007). Because the CVA measure takes this into account, schools with high proportions of 'mobile' children will, all other things equal, be ranked higher on this measure than on the traditional VA or raw output measures. This may create an anomaly in the case of newly-opened schools if they admit large numbers of pupils at non-standard times. If these pupils do not have the lower attainment associated with 'traditionally mobile' pupils, the school's CVA measure will be upward biased and their performance will be artificially high. This is particularly relevant in the current English system, where some so-called 'failing' schools are being replaced by/re-opened as new schools[13].

The key question is whether late-joining pupils in 'high-mobility' schools (such as those that are newly opened) are in fact similar in terms of attainment to late-joining pupils in other schools. The lower attainment of the latter group is accounted for within the CVA model. If, however, the attainment of the former is in fact not lower on average, the estimates of CVA for these 'high-mobility' schools will be upward biased.

To test whether late joining pupils in 'high-mobility' schools experience the same disadvantage as other late joining pupils, we rerun the CVA model and allow for the interaction of the 'high-mobility' school indicator with the pupil level mobility indicators. We find that late joiners in average schools are more disadvantaged than late joiners in 'high-mobility' schools. For example, the disadvantage of joining school after September in Year 10 for an average school is 68.61 GCSE points (two passes at grade D), while it is only 4.09 and not significant for a high-mobility school[14]. The two types of 'mobile' pupil are not the same, and hence the CVA league tables may therefore overestimate the performance of 'high-mobility' schools simply due to this anomaly. This may be a particular issue when comparing the performance of new schools with more established schools.

We investigate this by comparing the CVA performance of newly opened schools in the year in which they open to that in the subsequent years. The question we are asking is whether their CVA is upward biased due to the anomalous high mobility in their year of opening. We find evidence of such an upward bias: all but one of the newly-opened schools in our sample experience a significant drop in their CVA in their second year of existence, with new schools from 2004 maintaining this lower level of performance in the third year[15]. While this evidence is suggestive rather than conclusive, it implies that caution may be required in evaluating the performance of new schools in terms of their CVA.

2 The extent to which CVA informs parents about the relative effectiveness of schools

(a) Does school-level CVA hide evidence of differential effectiveness?

We investigate the extent to which schools are differentially effective with respect to the CVA measure across pupils of different ability, to see how informative the published school average measure is for individual parents. As a proxy for ability we use a pupil's average Key Stage 2 (KS2) score (average across English, maths and science, taken in the final year of primary school at age 11). We split the national KS2 distribution into three parts: the top 25% ('high ability'), the bottom 25% ('low ability') and the middle 50% ('medium ability') and allow schools' CVA performance to vary among their pupils according to their positions in this KS2 distribution[16]. In Table 3 we report the correlations between schools' performance for different ability pupils. We find some evidence of differential effectiveness across pupils of different abilities. Correlations between schools' performance for different ability groups are as follows: 0.85 between low and medium ability, 0.86 between high and medium ability and 0.61 between low and high ability. These results concur with previous evidence (Thomas et al 1997; Goldstein and Thomas 1996) and show that the published school average CVA does hide evidence of schools' differential effectiveness for pupils of different abilities.

(b) How stable are CVA performance measures over time?

In Table 4 we investigate the proportions of schools that exhibit large swings in their CVA performance from year to year, according to the following categories: significantly negative, average and significantly positive. Significantly negative (positive) means the CVA and confidence interval lie wholly below (above) the zero line. Two points to note from this table are that approximately 60% of schools stay in the same category across years (figures along the main diagonals), and that there are only small numbers of schools switching from significantly positive to significantly negative or vice versa (highlighted in bold italics; see also Thomas et al (2007))[17]. This suggests a high degree of stability.

Consider now, however, Tables 5a and 5b, in which we report the correlations across the three years for which we have data, both for CVA and the resultant rankings respectively. At around 0.7 for successive years, and 0.6 across the two-year gap, these summary figures suggest a lower, and decreasing, degree of stability over time (compared to around 0.9 and 0.87 for 5AC). A key point here is the inevitable time-lag involved in using such performance measures for parental choice (Gray et al 2001). The most recent published information available to parents choosing between secondary schools is based on the school's performance for the cohort of pupils who entered the school six years previously. Our results suggest that this lack of stability over time may limit the extent to which current CVA PMs can be used to predict future performance, and hence limit the usefulness of CVA to parents choosing between schools.

(c) To what extent can schools be ranked according to CVA?

In Figure 5 we plot the 2004 CVA PM and its confidence interval for each secondary school. We rank schools from high to low CVA across our sample[18]. The scale on the y axis is in GCSE points[19]. A CVA value of zero means that, on average, the school performs as expected for its pupils, given their characteristics. The zero line on the graph therefore represents the national average level of CVA. The figure shows that the confidence intervals of a large proportion of the schools cross the zero line. It is only possible to state that a school's CVA is significantly better (worse) than the national average if they have a score above (below) zero and a confidence interval that does not include zero. Only 51.8% of secondary schools in England can be judged to be significantly different from average according to the CVA measure

(25.7% better and 26.1% worse). In other words, 48.2% of secondary schools cannot be judged to be significantly different from the national average. We obtain extremely similar results for our 2005 and 2006 cohorts.

In general, however, the focus is not on comparing a school's performance with the national average. School performance tables are published with a view to help parents compare the performance of pairs or groups of schools, across a Local Authority or other geographical area, for example, in order to inform parental choice. Crucially, the confidence intervals illustrated in Figure 5 – which replicate those published by the government – only enable comparison with the national average; they do not enable meaningful pairwise comparison of schools. The correct confidence intervals (referred to as overlap intervals) which enable such pairwise comparisons can be constructed, however, and, under certain assumptions, can be adjusted to account for comparison between more than two schools (Goldstein & Healey 1995; Goldstein 2005; Goldstein and Thomas 1996).

We calculate such pairwise 95% confidence intervals and plot these with school CVA in Figure 6. The performance of any two schools is significantly different from one another if their confidence intervals do not overlap. From this figure we see that the majority of pairwise comparisons of schools will not be meaningful. Any ranking exercise across groups of schools that is based on their CVA scores alone may therefore be extremely misleading.

Conclusion

A contextual value added performance measure was published for the first time in the 2006 English secondary school league tables. By accounting for factors which impact on pupils' attainment, but which are outside the control of the school, the aim of CVA is to provide a more accurate measure of school effectiveness, i.e. the impact of a school on the progress of its pupils. What is less clear is the extent to which schools can be meaningfully ranked according to CVA, and hence the extent to which it facilitates parental choice along this dimension. This paper provides evidence on both these questions.

Overall, school-level CVA is a more accurate performance measure than its predecessors. It better isolates school effectiveness from composition, and so provides a better measure of the actual impact of a school on the progress of its pupils (although we have identified a problem regarding the mobility indicator in the case of newly-opened schools). The underlying pupil-level methodology from which it is constructed additionally provides information which is being used within schools as an internal performance management tool[20]. There is evidence of differential effectiveness, and a question mark over its stability over time, however, both of which reduce its usefulness as a tool for parental choice.

But our analysis highlights further major problems with the current use of CVA to rank and compare individual schools. Given that almost half the schools cannot be statistically distinguished from the national average, any ranking exercise based on these numbers will be largely spurious. This tendency will increase when the number of pupils within schools is small, and the resulting confidence intervals large. This casts doubt on the usefulness of the UK government's introduction of CVA for primary schools from 2007, given that their average cohort size is approximately one quarter that of secondary schools.

Moreover, the confidence intervals published by the government (but, interestingly, not generally reproduced by the print media) in the 2006 school performance tables do not enable pairwise or group comparison between schools. Even if the correct confidence intervals were published, our analysis shows that the performance of most schools cannot be meaningfully distinguished from that of other schools, again showing the largely spurious nature of any subsequent ranking exercise.

A question for policy that arises from this analysis, therefore, is (how) can the performance information contained within the CVA measure be used as part of a broader performance management regime for schools? Our analysis suggests some form of categorisation of schools, relative to the national average CVA. Consider again Figure 5. Schools could be categorised as having significantly positive performance (CVA and confidence interval wholly above the zero line); significantly negative performance (below the zero line) and average performance (confidence interval crossing the zero line). Such a categorisation is indeed possible, but can only be used in absolute and not relative or comparative terms. This is because it is not possible to compare either positively or negatively performing schools with those in the average category due to the larger confidence intervals – and hence higher likelihood of being categorised as 'average' – associated with smaller schools. Such an absolute categorisation would enable schools at the extremes of the distribution to be identified, however (Goldstein 1997; 2001), which could provide a starting point for a dialogue in which schools have to account for their performance. Contrary to being part of a parental choice agenda, this accords more with the notion of a 'performance intelligence' approach to performance management, whereby ranking positions provide the basis for dialogue between provider and policy maker, rather than being explicitly linked to specific targets or penalties[21].

Providing accurate information on school performance (effectiveness) is a necessary, but not sufficient, condition for facilitating parental choice along that dimension, and hence for creating an effective driver for real improvement. A broader question, outside the scope of this paper, is the extent to which parents actually choose schools on the basis of effectiveness, rather than with consideration of other aspects of the school such as location, ethos, or the composition of the student body. On the basis of the evidence presented here, however, it seems that much work is needed to produce meaningful rankings and thus ensure that even the necessary condition is met.

References

Aitkin, M and Longford, N (1986) Statistical modelling issues in school effectiveness studies, *Journal of the Royal Statistical Society*, 149(1): 1-43.

Barker, A (2006) Presentation to the PLASC User Group Workshop, http://www.bris.ac.uk/Depts/CMPO/PLUG/events/wkshop_8_3_06/anna.ppt, accessed 24.4.7.

Burgess, S, Propper, C, Slater, H and Wilson, D (2005) *Who wins and who loses from school accountability? The distribution of educational gains in English secondary schools*, CMPO Working Paper 05/128, CMPO, University of Bristol.

Cullen, JB and Reback, R (2006) *Tinkering towards accolades: school gaming under a performance accountability system*, NBER Working Paper 12286, Cambridge MA, NBER.

Deere, D and Strayer, W (2001) *Putting schools to the test: school accountability, incentives and behaviour*, Department of Economics, Texas A&M University (March).

DfES (2005) *About CVA*, http://www.dfes.gov.uk/performancetables/pilotks4_05/aboutcva.shtml, accessed 26.1.6.

Equalities Review (2007) *Fairness and Freedom: The Final Report of the Equalities Review*, http://www.theequalitiesreview.org.uk/publications.aspx , accessed 17.4.7.

Figlio, DN and Getzler, LS (2002) *Accountability, ability and disability: gaming the system*, NBER Working Paper 9307, Cambridge MA, NBER.

Goldstein, H (1997) Value added tables: the less-than-holy grail, *Managing Schools Today*, 6: 18-19.

Goldstein, H (2001) Using pupil performance data for judging schools and teachers: scope and limitations, *British Educational Research Journal*, 27(4): 433-442.

Goldstein, H (2003) *Multilevel Statistical Models*, London: Arnold Publications.

Goldstein, H, Burgess, S and McConnell, B (2007) Modelling the effect of pupil mobility on school differences in educational achievement, *Journal of the Royal Statistical Society, Series A*, 170(4): 941-954.

Goldstein, H and Healey, J, R (1995) The Graphical Presentation of a Collection of Means, *Journal of the Royal Statistical Society*, 158(1): 175-177.

Goldstein, H and Thomas, S (1996) Using Examination Results as Indicators of School and College Performance, *Journal of the Royal Statistical Society*, 159(1): 149-163.

Goldstein, H and Sammons, P (1997) The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multilevel analysis, *School Effectiveness and School Improvement*, 8: 219-230.

Gray, J, Goldstein, H and Thomas, S (2001) Predicting the future: the role of past performance in determining trends in institutional effectiveness at A level, *British Educational Research Journal*, 27: 391-406.

Kane, TJ and Staiger, DO (2002) The Promise and Pitfalls of Using Imprecise School Accountability Measures, *Journal of Economic Perspectives*, 16(4): 91-114.

Meyer, RH (1997) Value-added indicators of school performance: a primer, *Economics of Education Review*, 16(3): 283-301.

Ofsted (2003) *Provision and support for traveller pupils*, London, HMI 455.

Propper, C and Wilson, D (2003) The use and usefulness of performance measures in the public sector, *Oxford Review of Economic Policy*, 19(2): 250-267.

Raudenbush, SW and Bryk, AS (2002) *Hierarchical linear models: appliations and data analysis methods*, 2nd edition. Newbury Park, CA: Sage.

Reay, D and Lucey, H (2003) The limits of 'choice': children and the inner city, *Sociology*, 37(1): 121-142.

Rothstein, J (2006) *Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions*, American Economic Review, 96(4): 1333-1350.

Snijders, TAB and Bosker, RJ (1999) *Multilevel analysis: an introduction to basic and advanced multilevel modelling*, London, Sage Publications.

Thomas, S (1998) Value added measures of school effectiveness in the United Kingdom, *Prospects*, 28(1): 91-108.

Thomas, S (2001) Dimensions of secondary school effectiveness: comparative analyses across regions, *School Effectiveness and School Improvement*, 12: 285-322.

Thomas, S, Sammons, P, Mortimore, P and Smees, R (1997) Differential secondary school effectiveness: comparing the performance of different pupil groups, *British Educational Research Journal*, 23(4): 451-469.

Times Educational Supplement (2002) *League table bonus attracts schools to vocational option*, 23 August.

Tough, S and Brooks, R (2007) *School admissions: fair choice for parents and pupils*, London, IPPR.

West, A and Pennell, H (2000) Publishing school examination results in England: incentives and consequences, *Educational Studies*, 26(4): 423-436.

Wiggins, A and Tymms, P (2002) Dysfunctional effects of league tables: aA comparison between English and Scottish primary schools, *Public Money and Management*, 22(1): 43-48.

Wilson, D (2004) Which ranking? The impact of a 'value added' measure of secondary school performance, *Public Money and Management*, 24(1): 37-45.

Wilson, D, Burgess, S and Briggs, A (2005) *The dynamics of school attainment of England's ethnic minorities*, CMPO Working Paper 05/130, CMPO, University of Bristol.

Wilson, D, Croxson, B and Atkinson, A (2006) 'What gets measured gets done': headteachers' responses to the English secondary school performance management system, *Policy Studies*, 27(2): 153-171.

Yang, M, Goldstein, H, Rath, T and Hill, N (1999) The use of assessment data for school improvement purposes, *Oxford Review of Education*, 25: 469-483.

---

[1] GCSE exams are taken at the end of compulsory schooling in England, at age 16.

[2] All pupils in English state schools follow the National Curriculum, which comprises four Key Stages. They take Key Stage (KS) tests in primary school at ages 7 and 11 (KS1 and KS2). At age 11 they move to secondary school and take KS3 tests at age 14 and KS4 (GCSE) at age 16. A small number of schools ('middle schools') don't follow this pattern of transfer between primary and secondary school at the age of 11. We omit middle schools from our analysis sample.

[3] Income Deprivation Affecting Children Index (IDACI) measures the proportion of children under the age of 16 in an area living in low income households. IDACI is measured at Lower Layer Super Output Area (LLSOA). A lookup table of LLSOA to IDACI can be obtained from http://www.communities.gov.uk/odpm/SOA/IDACandIDAOP2004.xls .

[4] There are a few unavoidable differences between our data and those used in the DCSF CVA calculations. Due to confidentiality issues we were not able to obtain the 'in care' variable, nor each pupil's full date of birth (we have month within year). Also our ethnicity categorisation is less detailed: we do not have indicators for Irish, traveller of Irish heritage, Gypsy Roma or any other white background ethnicities.

[5] Return to KS2 is [-2.92*KS2] + [0.33*$(KS2)^2$], which is positive overall.

[6] Wilson et al (2005) provide a detailed analysis of the relative progress of pupils of different ethnicities through both primary and secondary schooling.

[7] Schools with smaller numbers of pupils have a much higher chance of having an outlying residual by chance alone; the shrinkage factor takes account of this.

[8] In practice the published CVA PMs are centred around 1000, presumably to avoid publishing negative CVAs. In this paper we centre the PMs around zero.

[9] As CVA allows for uncertainty around the mean school residual and the true school residual can lie anywhere along this uncertainty interval, we use the mean residual for ranking purposes.

[10] We have also carried out this exercise with the 5AC measure including English and maths. As this is so highly correlated with 5AC, we obtain very similar results.

[11] To obtain a complete picture of how schools with different characteristics perform according to the different PMs, we need to consider what happens when we allow all their characteristics to vary simultaneously in a multivariate setting. The results from doing so are not reported here but are available on request; they support the results illustrated in the Figures.

[12] Secondary school years are numbered 7 through 11; pupils take KS3 in Year 9 and GCSEs in Year 11. So, for example, the indicator ‘Pupil joined school after September in Year 10’ means that the pupil joined the school less than two years before taking her GCSEs.
[13] These new schools are known as Academies, and have a non-government sponsor. See http://www.standards.dfes.gov.uk/academies/ for more details of the Academies programme.
[14] Full results available from the authors on request.
[15] Full results available from the authors on request.
[16] Details of the modelling framework used are available from the authors on request.
[17] Note that the other off-diagonal numbers may be driven by schools having small numbers of pupils; we return to this point later.
[18] Throughout this paper we rank schools across our sample, i.e. across the whole of England. An alternative would be to rank schools within their local authority (LA), weighted by pupil numbers (see Wilson (2004)). In fact the two methods of ranking produce highly correlated results.
[19] These are as calculated in the 2004 points system (see http://www.dfes.gov.uk/performancetables/nscoringsys.shtml).
[20] See, for example, the information available at http://www.standards.dfes.gov.uk/performance/1316367/CVAinPAT2005/?version=1
[21] This point arises from comments made by colleagues at the IPMN Workshop, Oxford, August 2007.

Table 1: Pupil level CVA model

	Effect	Mean	Min	Max
Constant	**92.15 (11.82)****	1	1	1
KS2 fine grade average points score	**-2.92 (0.26)****	26.74	15	36
KS2english – KS2 average point score fine grades	**2.58 (0.07)****	-0.57	-11.94	7.19
KS2maths – KS2 average point score fine grades	**0.54 (0.07)****	-0.12	-10.03	10.39
Quadratic of KS2 Average Point Score	**0.33 (0.01)****	730	225	1296
Free school meal flag	**-20.02 (0.32)****	0.12	0	1
Income Deprivation Affecting Children Index score	**-61.35 (0.73)****	0.21	0	0.99
SEN action	**-37.42 (0.38)****	0.08	0	1
SEN action + statement	**-56.14 (0.48)****	0.05	0	1
Pupil joined school after September Year10	**-68.54 (0.88)****	0.01	0	1
Pupil joined school not in July/August/September Yr7,8,9	**-20.24 (0.49)****	0.04	0	1
Female	**16.42 (0.21)****	0.50	0	1
Age in months	**-1.13 (0.03)****	5.44	0	11
First language: Other or believed to be other than English	**23.84 (0.66)****	0.08	0	1
Black Caribbean	**9.66 (0.85)****	0.01	0	1
Black African	**29.38 (1.09)****	0.01	0	1
Black other	**4.19 (1.53)****	0.00	0	1
Indian	**21.91 (0.86)****	0.02	0	1
Pakistani	**22.98 (0.93)****	0.02	0	1
Bangladeshi	**27.78 (1.34)****	0.01	0	1
Asian	**20.06 (1.50)****	0.00	0	1
Chinese	**28.76 (1.74)****	0.00	0	1
Mixed White & Black Caribbean	**-2.01 (1.13)***	0.01	0	1
Mixed White & Black African	**3.52 (2.52)**	0.00	0	1
Mixed White & Asian	**11.78 (1.53)****	0.00	0	1
Mixed Other	**4.55 (1.15)****	0.01	0	1
Other	**23.39 (1.34)****	0.01	0	1
Unclassified ethnic group	**-13.14 (0.59)****	0.04	0	1
KS2 average point score of cohort	**2.85 (0.33)****	26.74	22.26	32.95
KS2 standard deviation in cohort	**-5.30 (0.89)****	3.53	0.91	5.07

Varience between schools	**305.78 (8.86)**
Varience within schools	**3833.85 (8.21)**

*- significant at 10%
**- significant at 5%

Table 2a: Correlations between alternative PMs

	5AC	5AC:EM	CVA
5AC	1		
5AC:EM	0.93	1	
CVA	0.43	0.27	1

Table 2b: Correlations between rankings based on alternative PMs

	5AC	5AC:EM	CVA
5AC	1		
5AC:EM	0.92	1	
CVA	0.46	0.33	1

Table 3: Correlations of school CVA performance across different sections of the ability distribution

	low ability	medium ability	high ability
low ability	1		
medium ability	0.85	1	
high ability	0.61	0.86	1

Table 4: Proportions of schools exhibiting large swings in CVA across years

		2005 CVA		
2004 CVA		negative	average	positive
	negative	66.84%	30.77%	***2.39%***
	average	18.96%	62.30%	18.74%
	positive	***3.60%***	28.27%	68.13%

Note: The figures are based on the percentage of schools with a particular 2004 performance. For example: 66.84% of schools with significantly negative performance in 2004 have significantly negative performance in 2005.

		2006 CVA		
2005 CVA		negative	average	positive
	negative	63.07%	34.14%	***2.79%***
	average	17.72%	64.55%	17.72%
	positive	***2.28%***	35.44%	62.28%

Note: Figures based on the percentage of schools with a particular 2005 performance.

		2006 CVA		
2004 CVA		negative	average	positive
	negative	60.08%	33.69%	***6.22%***
	average	19.82%	61.31%	18.87%
	positive	***2.85%***	38.99%	58.15%

Note: Figures based on the percentage of schools with a particular 2004 performance as in the top panel

Table 5a: Correlations between CVA across years*

	CVA 2004	CVA 2005	CVA 2006
CVA 2004	1		
CVA 2005	0.74	1	
CVA 2006	0.64	0.75	1

Table 5b: Correlations between CVA rankings across years*

	CVA ranking 2004	CVA ranking 2005	CVA ranking 2006
CVA ranking 2004	1		
CVA ranking 2005	0.72	1	
CVA ranking 2006	0.62	0.73	1

*Note: these correlations are calculated for the 2,826 schools that are in our analysis sample for each of the three years.

Figure 1: Comparing performance measures: 5AC versus 5AC:EM

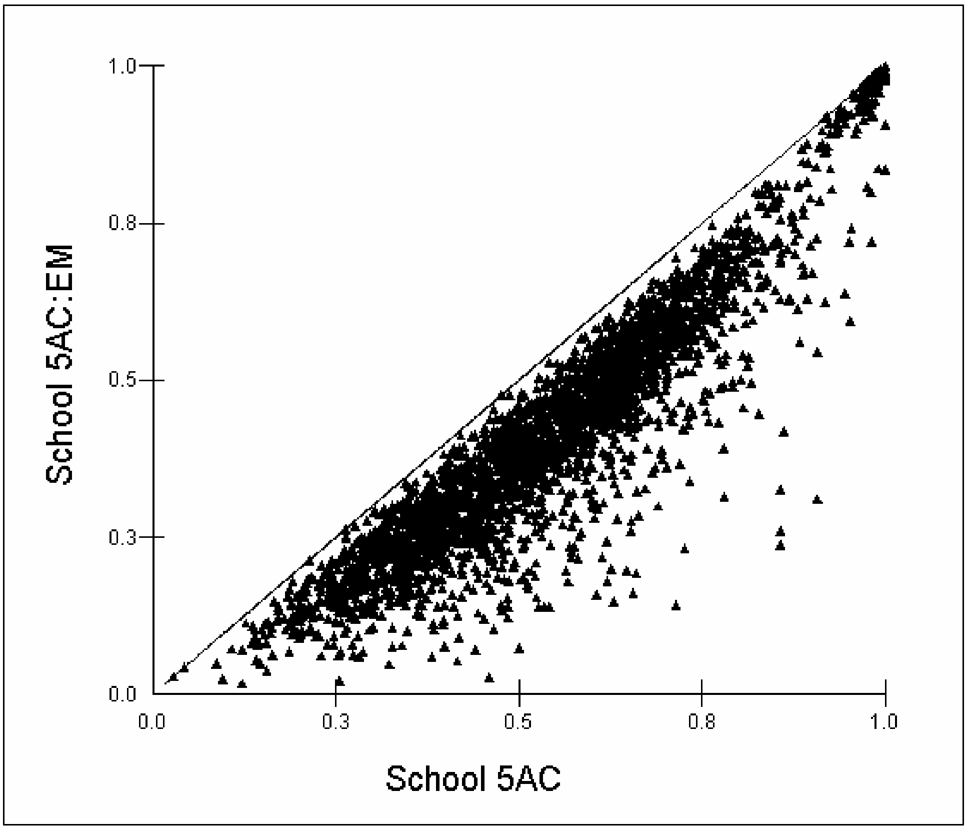

Figure 2: Correlations between school rankings: 5AC with CVA

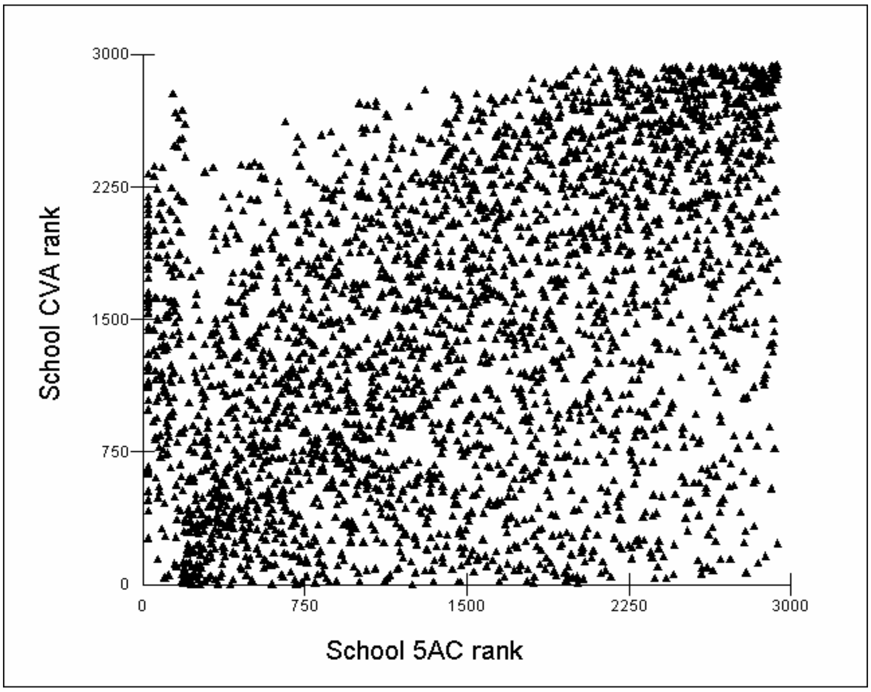

Figure 3: CVA versus 5AC rankings by proportion of FSM-eligible pupils

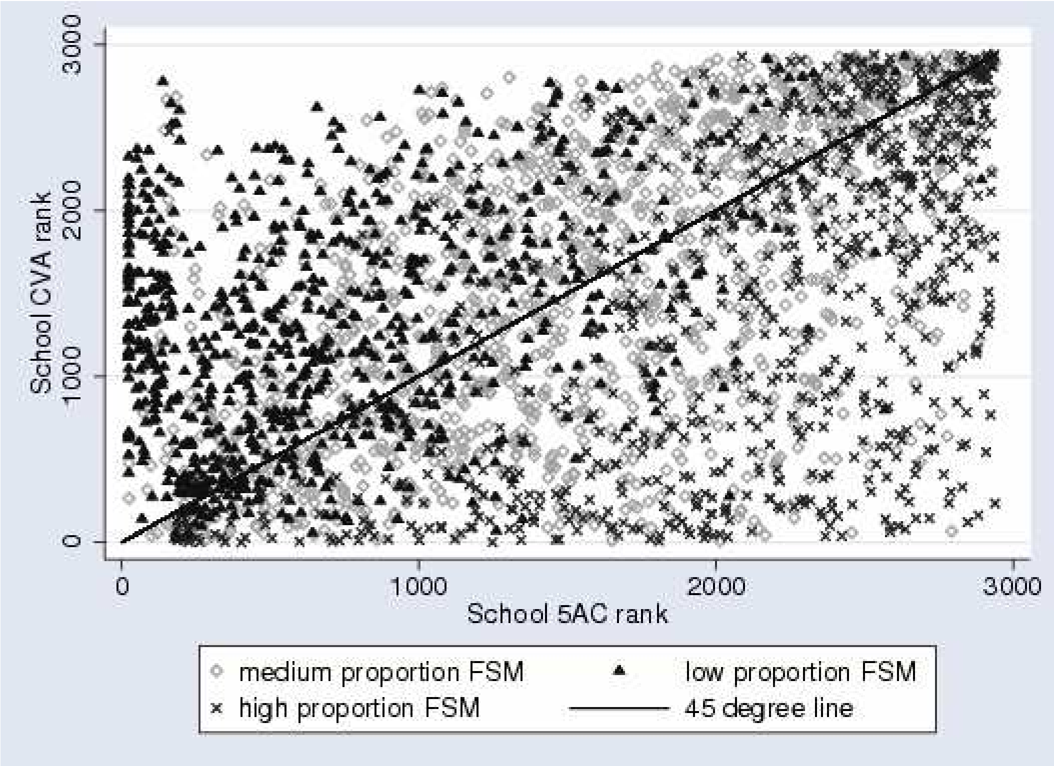

Figure 4: CVA versus 5AC rankings by admission policy

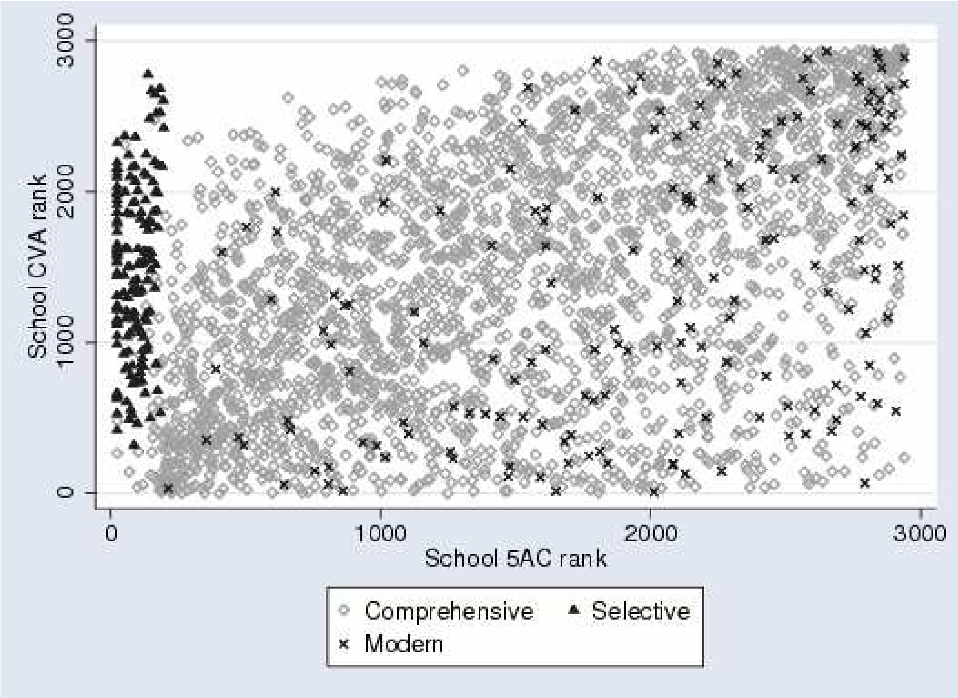

Figure 5: School level CVA with 95% confidence interval

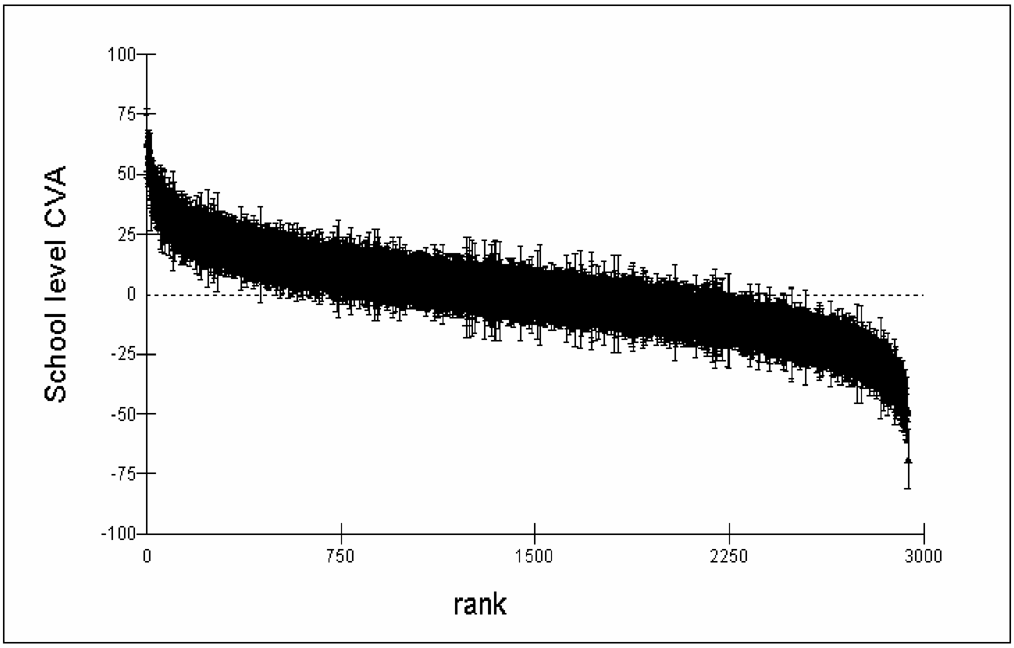

Figure 6: Pairwise 95% confidence intervals