Monitoring school performance: A statistical critique of ‘expected progress’ with comparison to multilevel ‘value-added’ models

George Leckie
Graduate School of Education
University of Bristol

Work funded by a UK ESRC early career grant to statistically critique England’s school performance tables
The UK has a long history of publishing school performance tables summarising students’ test results.

Over time, increasingly sophisticated measures have been introduced culminating in 2006 with contextual value-added (CVA), a multilevel model based value-added measure.

However, in 2011 the Government withdrew CVA replacing it with expected progress (EP), a simpler value-table approach.

In this talk we:
1. question Government’s reasons for withdrawing CVA
2. raise three statistical concerns with current EP measure
3. describe Progress 8, a new VA measure to start in 2017
Background
KS2 tests and GCSE examinations

- Primary phase of education (ages 4–11, years R–6)
  - KS2 tests in English and maths
  - Measured as a fine point score, but discretised into national curriculum levels for reporting: W, 1, 2, 3, 4, 5

- Secondary phase of education (ages 11–16, years 7–11)
  - GCSE examinations in English, maths and other subjects
# School performance tables

**Topline**

<table>
<thead>
<tr>
<th>School name</th>
<th>School type</th>
<th>% of pupils making expected progress</th>
<th>% achieving 5+ A*-C GCSEs (or equivalent) including English and maths GCSEs</th>
<th>% achieving the English Baccalaureate</th>
<th>% achieving grades A*-C in English and maths GCSEs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>England - state funded schools only</strong></td>
<td></td>
<td>70.4% 70.8%</td>
<td>55.1% 58.2%</td>
<td>60.6%</td>
<td>61.3%</td>
</tr>
<tr>
<td><strong>Local Authority</strong></td>
<td></td>
<td>64.3% 62.8%</td>
<td>46.2% 58.1%</td>
<td>51.6%</td>
<td>52.3%</td>
</tr>
<tr>
<td>Ashton Park School</td>
<td>Foundation School</td>
<td>57% 59%</td>
<td>49% 43%</td>
<td>48%</td>
<td>10% 48%</td>
</tr>
<tr>
<td>Bedminster Down School</td>
<td>Academy - Converter Mainstream</td>
<td>38% 59%</td>
<td>40% 46%</td>
<td>40%</td>
<td>36% 8% 36%</td>
</tr>
<tr>
<td>Bridge Learning Campus - Secondary</td>
<td>Foundation School</td>
<td>54% 49%</td>
<td>34% 40%</td>
<td>37%</td>
<td>0% 37%</td>
</tr>
</tbody>
</table>
Gov. justifications for publishing school performance tables

1. To support parental **school choice** based on schools’ ability to teach the national curriculum, and to therefore create competition and a free market in education (Education Reform act 1988)

2. To hold schools publically **accountable** for their results

3. To promote **school improvement** via school self-reflection and the identification of effective practices being employed in successful schools

- In contrast to the US, the UK Government has not yet attempted to measure and report **teacher effects**
Contextual value-added
2006-2010
Contextual value-added
2006-2010

- The withdrawn CVA measure was derived from a two-level random-intercept model, a simplified version of which can be written as

\[ y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_{ij} + \beta_3 \bar{x}.j + u_j + e_{ij} \]

- \( y_{ij} \) denotes the GCSE score of student \( i \) in school \( j \)
- \( x_{ij} \) denotes their KS2 score
- \( z_{ij} \) denotes their free school mean status
- \( \bar{x}.j \) denotes their school’s mean KS2 score; included to adjust for peer group effects and give ‘Type B’ school effects for accountability (Raudenbush & Willms, 1995)
- \( u_j \) denotes their school’s effect (i.e., CVA score)
The 2010 formula for predicting a pupil’s KS4 outcome (capped “best 8” score with an additional bonus for attainment in each of)

330.761 + 0.3923*(KS2 APS squared) - 6.072 *KS2 APS + 0.925*(KS2 English points - KS2 APS) + 1.443*(KS2 maths points - KS2 APS)
4 - 26.336 (if in care) - 49.266*IDACI score - 26.618 (if School Action) - 65.226 (if Action Plus or Statemented)
5 - 72.169 (if joined after Sept Y10) - 24.054 (if joined not in July/Aug/Sept yr 7-9)
6 + 16.294 (if female) - 11.501*(age within year where 1 Sept= 1.00, 31 Aug = 0.00)
7 + for EAL pupils only (-35.507 - 0.190*(KS2 APS squared) + 7.190 *KS2 APS )
8 + ethnicity coefficient from table below
9 + for FSM pupils only (-22.410 + FSM/ethnicity interaction from table below)
10 -0.385 * cohort KS2 average point score - 6.124* cohort KS2 APS standard deviation

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Ethnicity coefficient</th>
<th>FSM/ethnicity interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>White British</td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Irish</td>
<td>2.748</td>
<td>-5.2177</td>
</tr>
<tr>
<td>Traveller of Irish heritage</td>
<td>-48.992</td>
<td>0.4921</td>
</tr>
<tr>
<td>Gypsy/ Roma</td>
<td>-67.775</td>
<td>30.2742</td>
</tr>
<tr>
<td>Any other white background</td>
<td>13.115</td>
<td>15.2550</td>
</tr>
<tr>
<td>White and Black Caribbean</td>
<td>-2.002</td>
<td>14.1580</td>
</tr>
<tr>
<td>White and Black African</td>
<td>10.583</td>
<td>8.7812</td>
</tr>
<tr>
<td>White and Asian</td>
<td>10.776</td>
<td>3.8260</td>
</tr>
<tr>
<td>Any other mixed background</td>
<td>8.339</td>
<td>17.7861</td>
</tr>
<tr>
<td>Indian</td>
<td>25.627</td>
<td>17.8612</td>
</tr>
<tr>
<td>Pakistani</td>
<td>18.877</td>
<td>17.3433</td>
</tr>
<tr>
<td>Bangladeshi</td>
<td>24.033</td>
<td>17.7587</td>
</tr>
<tr>
<td>Any other Asian background</td>
<td>29.186</td>
<td>23.0746</td>
</tr>
<tr>
<td>Caribbean</td>
<td>14.167</td>
<td>20.8229</td>
</tr>
<tr>
<td>Black African</td>
<td>32.129</td>
<td>21.2376</td>
</tr>
</tbody>
</table>
### Local Authority: Bristol, City of

<table>
<thead>
<tr>
<th>Institution Name</th>
<th>Measure centred around 1000</th>
<th>Limit of Key Stage 2 to 4 CVA Confidence Intervals</th>
<th>Coverage</th>
<th>Number of qualifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Authority Average</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>England (state funded schools only)</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>England (all schools)</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>Andalusia Academy Bristol - Bristol</td>
<td>1003.2</td>
<td>1012.8</td>
<td>98%</td>
<td>11.2</td>
</tr>
<tr>
<td>Ashton Park School - Bristol</td>
<td>988.5</td>
<td>997.8</td>
<td>98%</td>
<td>9.2</td>
</tr>
<tr>
<td>Bedminster Down School - Bristol</td>
<td>1003.4</td>
<td>1013.9</td>
<td>98%</td>
<td>10.4</td>
</tr>
</tbody>
</table>

\[ \hat{u}_j^{EB} \pm 1.96 \times \overline{SE}(\hat{u}_j^{EB}) \]
Gov. reasons for ending CVA
Schools white paper 2010

1. ‘[CVA] is difficult for the public to understand’

- Agreed, but maybe more could have been done to explain and communicate CVA?

- For example, one had to delve deep into the technical documentation to find out what the CVA unit of measurement was

- Also, the methodology underlying CVA is the same as that used in Hong Kong (SVAIS) and simpler than that underlying other school performance measures published around the world (e.g., Tennessee’s TVAAS or Australia’s similar schools methodology)
Gov. reasons for ending CVA (cont’d)
Schools white paper 2010

2. ‘recent research shows [CVA] to be a less strong predictor of success than raw attainment measures’

- Not entirely clear what the Government are trying to say here (they don’t cite the research they refer to)

- Sounds like they are saying that a student’s GCSE score \(y_{ij}\) is more strongly predicted by their KS2 score \(x_{ij}\) than by their school’s CVA score \(u_{j}\)

- But predicting GCSE success was never the aim of CVA. The aim was to measure the effects schools actually had on their students
Gov. reasons for ending CVA (cont’d)
Schools white paper 2010

3. ‘[CVA] also has the effect of expecting different levels of progress from different groups of pupils on the basis of their ethnic background, or family circumstances, which we think is wrong in principle’

- CVA explicitly recognized the reality, which is that some student groups do make less progress than others and that this must be adjusted for if we are to make fair comparisons between schools
Gov. reasons for ending CVA (cont’d)  
Schools white paper 2010

4. ‘It is morally wrong to have an attainment measure which entrenches low aspirations for children because of their background’

- Yes, CVA lowered the bar that disadvantaged students had to cross for schools to be rewarded. But, surely this incentivised schools to focus on their disadvantaged students?
Gov. reasons for ending CVA (cont’d)
Schools white paper 2010

5. ‘We should expect every child to succeed and measure schools on how much value they add for all pupils, not rank them on the ethnic make-up of their intake…’

- But, by explicitly adjusting for the ethnic make-up of schools intakes, CVA actually removed the influence of ethnicity from schools’ rankings
Expected progress
2011-2016
### Expected progress 2011-2016

<table>
<thead>
<tr>
<th>KS2 outcome</th>
<th>GCSE Grade</th>
<th>U</th>
<th>G</th>
<th>F</th>
<th>E</th>
<th>D</th>
<th>C</th>
<th>B</th>
<th>A</th>
<th>A*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working towards level 1</td>
<td>Expected progress not made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
</tr>
<tr>
<td>Level 1</td>
<td>Expected progress not made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
</tr>
<tr>
<td>Level 2</td>
<td>Expected progress not made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
</tr>
<tr>
<td>Level 3</td>
<td>Expected progress not made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
</tr>
<tr>
<td>Level 4</td>
<td>Expected progress not made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
</tr>
<tr>
<td>Level 5</td>
<td>Expected progress not made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
<td>Expected progress made</td>
</tr>
</tbody>
</table>

- EP is defined as the percentage of students achieving their target GCSE grades
- An example of a value table model (transition matrix or categorical model) of growth popular in some US states
Statistical concern 1
Borderline effects

<table>
<thead>
<tr>
<th>KS2 level</th>
<th>U</th>
<th>G</th>
<th>F</th>
<th>E</th>
<th>D</th>
<th>C</th>
<th>B</th>
<th>A</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working towards level 1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Level 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Level 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Level 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Level 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Level 5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Value table** values are binary (EP is a *threshold* measure)
- No partial rewards for just missing target grades; no additional rewards for surpassing target grades
- Schools are perversely incentivised to concentrate on borderline students (within each KS2 level) (c.f., 5+ A*-C)
Statistical concern 2
Dependency on prior attainment

- Schools with higher attaining intakes will do better on EP; EP under-adjusts for school differences in prior attainment
EP also has an illogical sawtooth (zig zag) dependency on prior attainment with sharp discontinuities in probability of making EP.
Statistical concern 3

Statistical uncertainty

- EP makes no attempt to quantify and communicate the statistical uncertainty in measuring school effects.

- There is no obvious way for users to establish whether measured differences between schools, or differences from national averages and floor standards, are meaningful, or whether they more likely reflect the variations of chance.

- Consider a school with 180 students where 70% make EP. The associated 95% Wald binomial confidence interval ranges from 63% to 77% and so the school has a ±7 percentage point margin of error.
Expected progress vs. value-added
Partial vs. full adjustment for prior attainment

- The strong dependency of EP on prior attainment suggests EP only partially adjusts for school differences in KS2 scores and is therefore biased in favour of schools with high prior attaining intakes.

- We explore this by comparing schools’ EP ranks to those based on a simple VAM which aims to fully adjust for school differences in KS2 score.

- GCSE grade, $y_{ij}$: $U = 2, G = 3, \ldots, A^* = 10$

- KS2 score, $x_{ij}$: $[0,36)$

- The VAM approach avoids EP’s borderline effects problem and facilitates communication of statistical uncertainty.
Model specification

- We allow for the non-linear relationships between the GCSE and KS2 scores seen in the data

\[ y_{ij} = \beta_0 + \beta_1 I(x_{ij} = 3) + \beta_2 I(x_{ij} = 9) + \beta_3 I(x_{ij} = 15) + \beta_4 x_{ij} + \beta_5 x_{ij}^2 + \beta_6 x_{ij}^3 + \beta_7 \bar{x}_j + \beta_8 \bar{x}_j^2 + u_j + e_{ij} \]

- A cubic relationship between \( y_{ij} \) and \( x_{ij} \) with discrete departures from this relationship at low levels of prior attainments

- A quadratic relationship between \( \bar{y}_j \) and \( \bar{x}_j \) to account for positive peer group effects (but more on this later)
Within and between relationships between GCSE score & KS2 score

- Between relationship is stronger than within relationship
Student VA dependency on prior attainment?

- In contrast to EP, the average value of $\hat{r}_{ij} = y_{ij} - x'_{ij} \hat{\beta}$ varies little with $x_{ij}$ (i.e., no overall bias, limited discontinuities)
Usual story: school effects are inherently imprecise; middle 1/3 cannot be distinguished from the national average.
And if we ignored prior attainment...

- Much greater variability in unadjusted school means shows importance of accounting for school prior attainment differences
Spearman rank correlations

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th>Mathematics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Means</td>
<td>EP</td>
<td>VA</td>
<td>Means</td>
</tr>
<tr>
<td>English</td>
<td>1</td>
<td>0.83</td>
<td>0.58</td>
<td>0.79</td>
</tr>
<tr>
<td>EP</td>
<td>0.83</td>
<td>1</td>
<td>0.79</td>
<td>1</td>
</tr>
<tr>
<td>VA</td>
<td>0.58</td>
<td>0.79</td>
<td>1</td>
<td>0.74</td>
</tr>
</tbody>
</table>

- EP is more highly correlated with the unadjusted mean GCSE grade/score than with the adjusted mean GCSE grade/score

- EP therefore appears closer to being a pure attainment (i.e., status) measure of school performance than the pure progress (i.e., growth or improvement) measure it purports to be
Difference in ranks: VA vs. EP

- As expected, EP is strongly biased in favour of schools with high prior attaining intakes (especially Grammar schools)
Limitations of VAM
Selection bias

- A simplified version of our exploratory VAM can be written as

\[ y_{ij} = \beta_0 + \beta_1^w x_{ij} + \beta_1^c \tilde{x}_j + u_j + e_{ij} \]

- Castellano et al. (2014) argue that in such models \( \tilde{x}_j \) and \( u_j \) will tend to be positively correlated leading \( \beta_1^c \) to be biased upwards and for \( \hat{u}_j \) to therefore be over-adjusted

- The highest prior attaining students are likely to select into schools with the highest quality teachers & vice versa (e.g., via residential moves, a big deal in England)

- See also Balou et al. (2004) and McCafrrrey et al. (2003)
Limitations of exploratory VAM
Measurement error bias

- Harker and Tymms (2004) argue that measurement error in $x_{ij}$ will also bias up any true positive peer group effect $\beta_1^C$ and therefore again lead $\hat{u}_j$ to be over-adjusted.

- This becomes clear when we rewrite this VAM as

$$y_{ij} = \beta_0 + \beta_1^W (x_{ij} - \bar{x}_j) + \beta_1^B \bar{x}_j + u_j + e_{ij}$$

- $\beta_1^B$ will suffer from less attenuation bias than $\beta_1^W$, as averaging $x_{ij}$ to the school-level would be expected to average much of the measurement error away.

- One solution would be to specify a measurement model for $x_{ij}$ and therefore $\bar{x}_j$ (e.g., Lüdtke et al, 2010).
Progress 8
2017-20??
Progress 8
2017-20??

- The Government recently announced that they will withdraw EP replacing it in 2017 with Progress 8, a new value-added measure derived from a multiple linear regression model, a simplified version of which can be written as

\[ y_{ij} = \beta_0 + \beta_1 x_{ij} + r_{ij}, \quad \hat{r}_j = \sum_{i=1}^{n_j} \hat{r}_{ij} \]

- Progress 8 will adjust for student prior attainment, but ignore the effects of all other student background characteristics

- Progress 8 scores will be school averages of the OLS residuals (i.e., no shrinkage) and will be presented with 95% CIs
Importantly, no adjustment will be made for $\bar{x}_{.j}$ implying Government interest has shifted towards measuring ‘Type A’ school effects for school choice purposes (Raudenbush & Willms, 1995), although this has not be stated in any documentation.

However, when we omit $\bar{x}_{.j}$ from the model, $\beta_1$ becomes a weighted average of $\beta_1^B$ and $\beta_1^W$ and so Progress 8 will also over-adjust for $x_{ij}$ to the extent to which peer group effects (and selection into schools and measurement error) exist.

In this instance, a fixed-effects model would appear preferable (Castellano et al., 2014).
Conclusion
Conclusion

- Government reasoning for withdrawing CVA appear odd
- CVA’s successor, EP, appears fundamentally flawed
  1. Likely to focus schools’ efforts on borderline students
  2. Severely dependent on prior attainment
  3. Fails to communicate statistical uncertainty
- A conventional VAM could address these flaws, but would still suffer from selection bias as well as other statistical (and more general) concerns which arise when making quantitative school comparisons
- Progress 8 appears to aim to measure Type A school effects in which case a fixed effects model appears preferable
Future work
### Pupil Progress

<table>
<thead>
<tr>
<th>School name</th>
<th>% making expected progress in English</th>
<th>% making expected progress in maths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All pupils</td>
<td>Low attainers</td>
</tr>
<tr>
<td>England - all schools</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>England - state funded schools only</td>
<td>70.4%</td>
<td>45.2%</td>
</tr>
<tr>
<td>Local Authority</td>
<td>64.3%</td>
<td>41.9%</td>
</tr>
</tbody>
</table>

**Note:**

- **Click on headings to sort figures in ascending/descending order.**
- **Schools without data will be displayed below those with data.**

**Schools**

- Andalusia Academy Bristol
- Ashton Park School
- Badminton School
- Bedminster Down School
- Bridge Learning Campus
- Bridge Learning Campus - Secondary
- Brislington Enterprise College
- Bristol Brunel Academy
- Bristol Cathedral Choir School
- Bristol Free School
- Bristol Grammar School
- Bristol Metropolitan Academy
- Bristol Steiner School

**KS4 2013 Results data last updated on 20 Mar 2014**

**Click on the box next to a school/college to select it for comparison - once you have selected all required schools/colleges click here: Compare**
### KS4 2013 Results / Closing the Gap: time series - Sorted by School name, in ascending order.

#### KS4 2013 Results data last updated on 20 Mar 2014

Click on headings to sort figures in ascending/descending order. Schools without data will be displayed below those with data.

Displaying 1 - 50 of 50 schools.

<table>
<thead>
<tr>
<th>School name</th>
<th>% achieving A*-C in English and maths GCSEs</th>
<th>% achieving 5+ A*-C GCSEs (or equivalent) including English and maths GCSEs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All pupils</td>
<td>Disadvantaged pupils</td>
</tr>
<tr>
<td>England - all schools</td>
<td>60.0%</td>
<td>NA</td>
</tr>
<tr>
<td>England - state funded schools only</td>
<td>61.3%</td>
<td>41.8%</td>
</tr>
<tr>
<td>Local Authority</td>
<td>52.7%</td>
<td>33.3%</td>
</tr>
</tbody>
</table>

**Schools**

- Andalusia Academy Bristol
- Ashton Park School
- Badminton School
- Bedminster Down School
- Bridge Learning Campus
- Bridge Learning Campus - Secondary
- Brislington Enterprise College
- Bristol Brunel Academy
- Bristol Cathedral Choir School
- Bristol Free School
- Bristol Free School

Visit the [website](http://www.education.gov.uk/cgi-bin/schools/performance/2013/group.pl?qtype=LA&superview=sec&view=aat&set=5&sort=&ord=&tab=72&no=801&pg=1) for more information.
End of talk – Thank you

g.leckie@bristol.ac.uk
bristol.ac.uk/cmm/team/leckie.html
References
References


My previous work on school league tables


