

The limitations of using school league tables to inform school choice

George Leckie and Harvey Goldstein
Centre for Multilevel Modelling
University of Bristol

Introduction

- School league tables (i.e. school report cards) rank schools by:
 - Schools' average test scores
 - Estimates of **school quality** based on **statistical models**
- They are published:
 - To hold schools accountable
 - To inform parental **school choice**
- They are now published in many countries
 - Australia, Canada, **England**, US,...

The English education system

- Two phases
 - Primary schooling from ages 5 to 11
 - **Secondary schooling** from ages 11 to 16
- Two main tests/exams
 - At age 11 children take English and maths tests
 - At age 16 children take **GCSE exams**

A brief history of England's school league tables

- **1994 onwards:** Schools' averages GCSE exam results
 - Unfair since schools differ in the quality of their intakes
 - Not model based, so no statement of statistical uncertainty
- **2006 onwards:** Contextual value-added (CVA) scores
 - Adjusts for the intake achievement of students
 - Based on a multilevel model, so scores are published with 95% confidence intervals



Achievement and attainment tables 2009

Choose a set of tables

Secondary School (GCSE and equivalent) ▼

Go

The tables

User guide
and resources

Glossary and
abbreviations

Secondary School (GCSE and equivalent) ▶ South East Local Authorities ▶

Local Authority : Oxfordshire

[Print this page ▶](#)

[Download the PDF booklet for this LA ▶](#)

Cohort

KS4 Results

Key Stage 2 to 4 CVA

Expected progress from KS2-KS4

Year on year comparisons

Absence

Background

School	Level 2 (5+ A*-C) including English and maths GCSEs	% of pupils achieving								Average total point score per pupil
		English and maths Skills at Level 2	English and maths Skills at Level 1	Level 2 (5+ A*-C)	Level 1 (5+ A*-G)	2 grades A*-C in science	A*-C in a modern foreign language	at least an A*-G short course in a modern foreign language	at least one qualification	
Banbury School - Banbury	40%	47%	94%	68%	91%	82%	14%	24%	98%	384.5
Bartholomew School - Witney	66%	70%	99%	71%	99%	64%	45%	82%	100%	458.0
Bicester Community College - Bicester	43%	48%	98%	63%	98%	39%	14%	43%	99%	422.2
Blessed George Napier Catholic	62%	62%	98%	78%	97%	59%	38%	40%	100%	420.4



Achievement and attainment tables 2009

Choose a set of tables

Secondary School (GCSE and equivalent) ▾

Go

The tables

User guide
and resources

Glossary and
abbreviations

Secondary School (GCSE and equivalent) ▾ South East Local Authorities ▾

Local Authority : Oxfordshire

 [Print this page ▾](#)

 [Download the PDF booklet for this LA ▾](#)

Cohort

KS4 Results

Key Stage 2 to 4 CVA

Expected progress from KS2-KS4

Year on year comparisons

Absence

Background

School	Measure centred around 1000	Limit of Key Stage 2 to 4 CVA Confidence Intervals		Coverage % of pupils at the end of Key Stage 4 included in CVA calculation	Number of qualifications Average number of qualifications taken by pupils in Key Stage 2 to 4 CVA calculation
	CVA measure based on progress between Key Stage 2 and Key Stage 4	Upper	Lower		
Banbury School - Banbury	993.0	1001.8	984.2	95%	10.1
Bartholomew School - Witney	987.4	998.0	976.8	96%	11.0
Bicester Community College - Bicester	994.3	1004.0	984.7	97%	11.4
Blessed George Napier Catholic School and Sports College - Banbury	1001.0	1012.4	989.5	94%	10.0

[News Front Page](#)[Election 2010](#)[World](#)[UK](#)[England](#)[Northern Ireland](#)[Scotland](#)[Wales](#)[Business](#)[Health](#)[Education](#)[League Tables](#)[Science & Environment](#)[Technology](#)[Entertainment](#)[Also in the news](#)[Video and Audio](#)[Have Your Say](#)[Magazine](#)[In Pictures](#)[Country Profiles](#)[Special Reports](#)[Related BBC sites](#)

Page last updated at 09:30 GMT, Wednesday, 13 January 2010

[✉ E-mail this to a friend](#)

Secondary schools in Oxfordshire

This is an alphabetical list. The columns show the proportion of pupils attaining five or more good GCSEs and equivalents at the end of Key Stage 4 including English and maths, the Level 2 contextual value added (CVA) score, the average A/AS-level and equivalent points and the Level 3 CVA.

Click the name of an institution for a page detailing its performance.

[Click here for a fuller explanation of the figures.](#)

School	Sort by GCSE %	Level 2 CVA	Sort by A/AS average points	Level 3 CVA
Banbury School	40	993	647.3	1018.3
Bartholomew School	66	987.4	810.6	998.7
Bicester Community College	43	994.2	678.3	994.9
Blessed George Napier Catholic School and Sports College	62	1001	760.6	1006

Limitation 1

Past performance is no guarantee of future performance ... in many cases the value of the investment can fall as well as rise

Seven years out of date!

- The 2009 school league table report schools' performances for the **2009 GCSE cohort**
- However, parents want to know schools' performances for the **2016 GCSE cohort**
- Inferences about the future performances of schools will be **far less precise** than inferences about their current performances

The CVA model is a two-level multilevel model

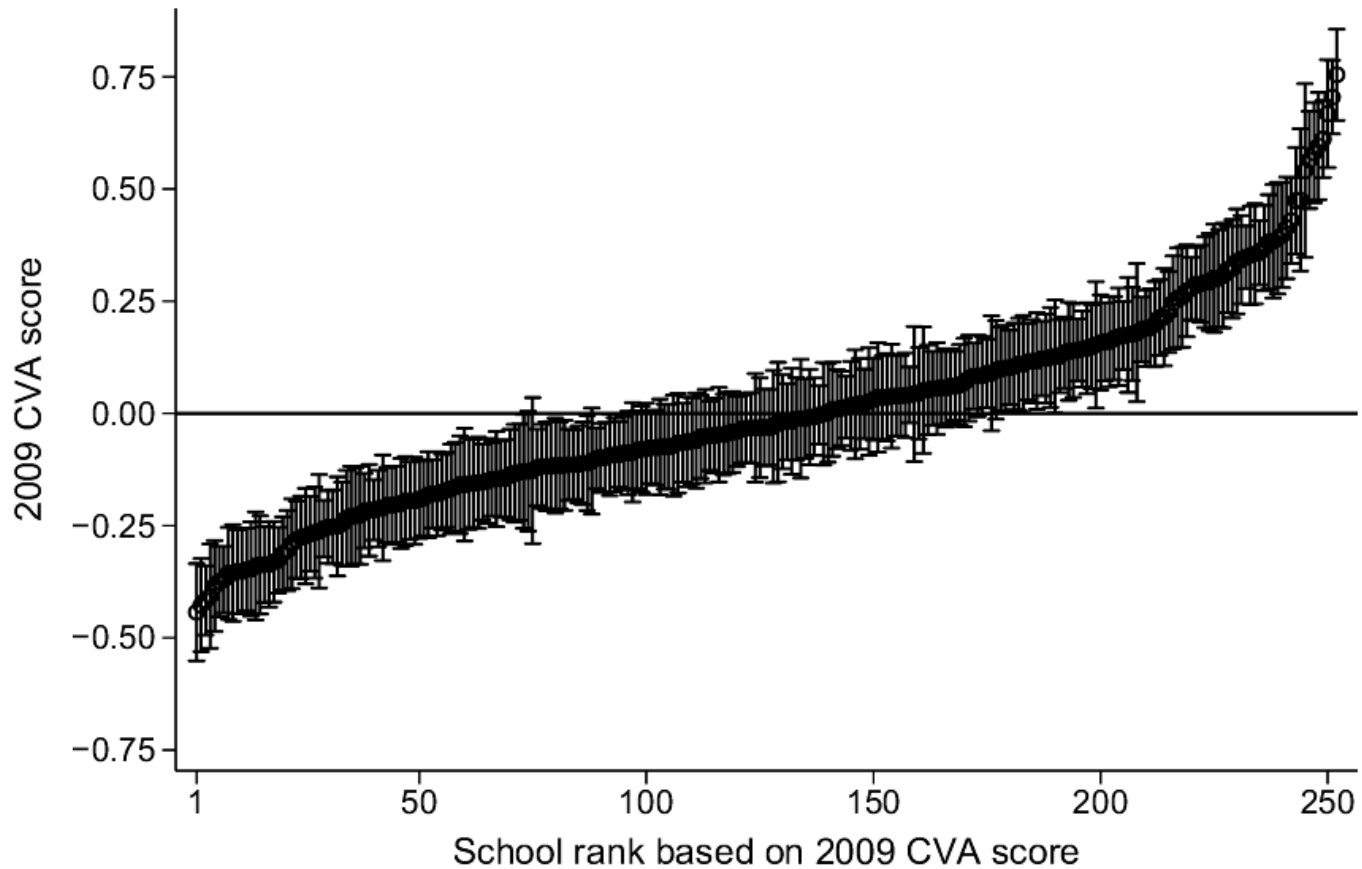
$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}$$
$$u_j \sim \text{N}(0, \sigma_u^2), \quad e_{ij} \sim \text{N}(0, \sigma_e^2)$$

- y_{ij} is the total age 16 GCSE score for student i in school j
- x_{ij} is their average age 11 English and maths score
- u_j is the **CVA school effect** for secondary school j
- e_{ij} is the student level random effect or residual

Data

- National Pupil Database (NPD)
- We focus on the 2009 GCSE cohort
- We analyse a 10% random sample of schools

School effects for the 2009 cohort



- ~60% of schools are significantly different from the overall average

School effects for the 2016 cohort

- Will the same significant differences remain in 2016?
- We must factor in the additional uncertainty that arises from predicting seven years into the future
- We use a **multivariate response version** of the CVA model for eight cohorts of students to do this

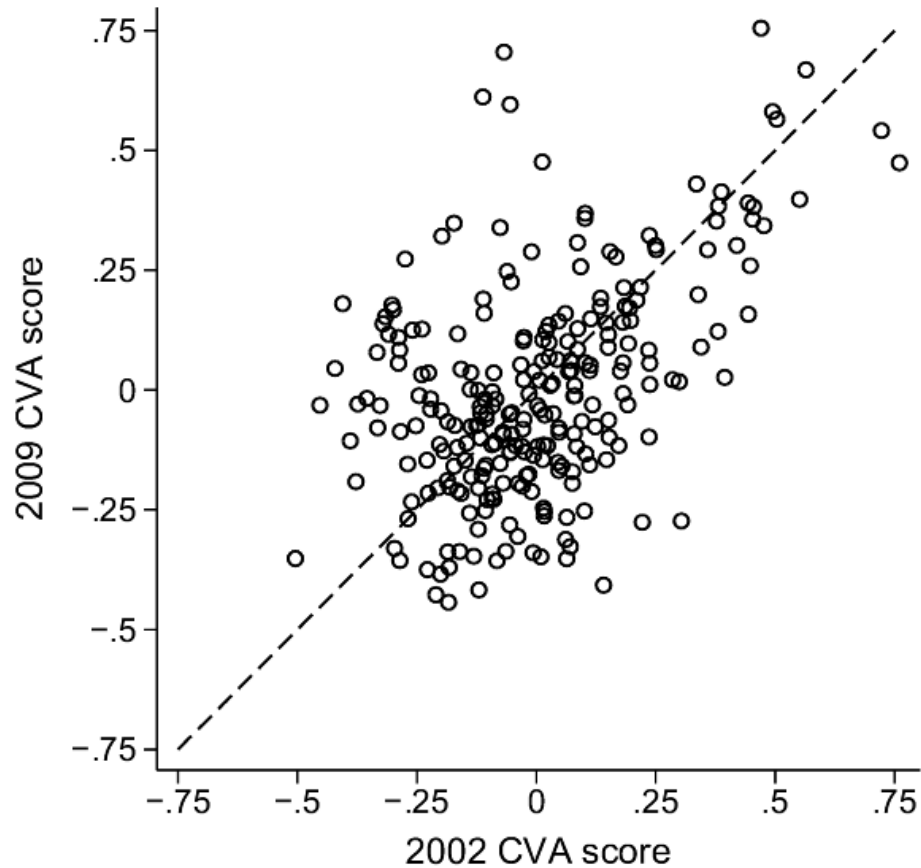
Multivariate response model for all eight cohorts: 2002-2009

- These correlations measure the **stability of school effects** over time

	2002	2003	2004	2005	2006	2007	2008	2009
2002	1.00							
2003	0.90	1.00						
2004	0.82	0.90	1.00					
2005	0.75	0.82	0.90	1.00				
2006	0.69	0.75	0.82	0.90	1.00			
2007	0.62	0.69	0.75	0.82	0.90	1.00		
2008	0.58	0.62	0.69	0.75	0.82	0.90	1.00	
2009	0.55	0.58	0.62	0.69	0.75	0.82	0.90	1.00

- The **seven-cohort apart correlation** is just 0.55

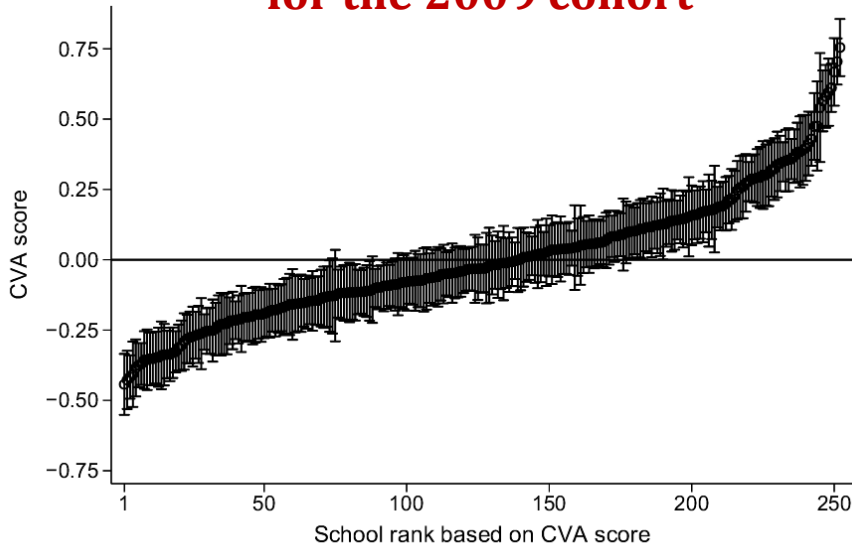
School effects for the 2009 cohort vs. the 2002 cohort



- The correlation of 0.55 implies a substantial reordering of schools
- The government implicitly assume that there is no reordering

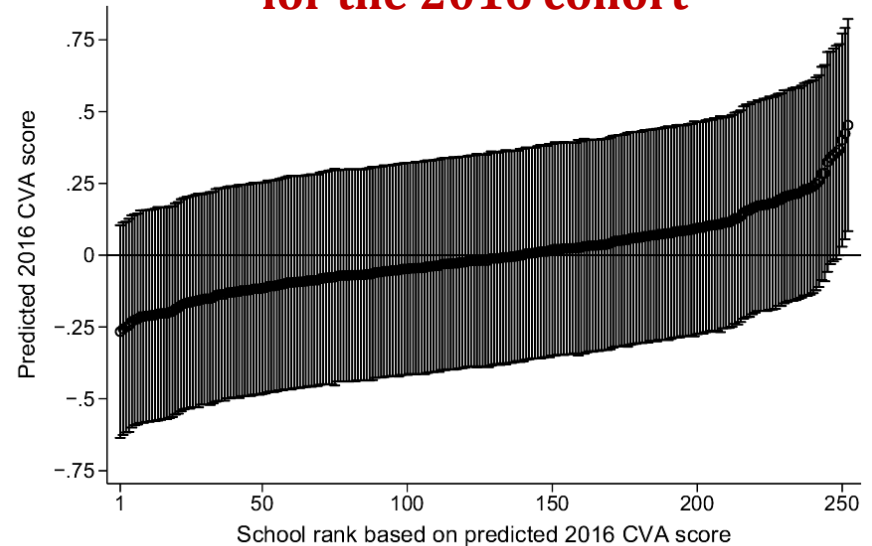
Comparison of the school effects for the 2009 and 2016 cohorts

School effects for the 2009 cohort



More appropriate for inferences about school accountability

Predicted school effects for the 2016 cohort



More appropriate for inferences about school choice

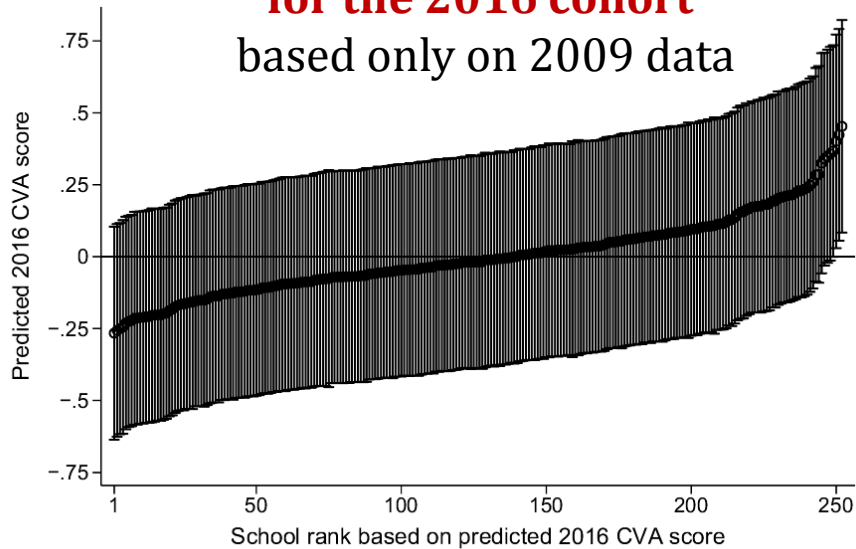
- Different users want different things from league tables

Making more precise predictions?

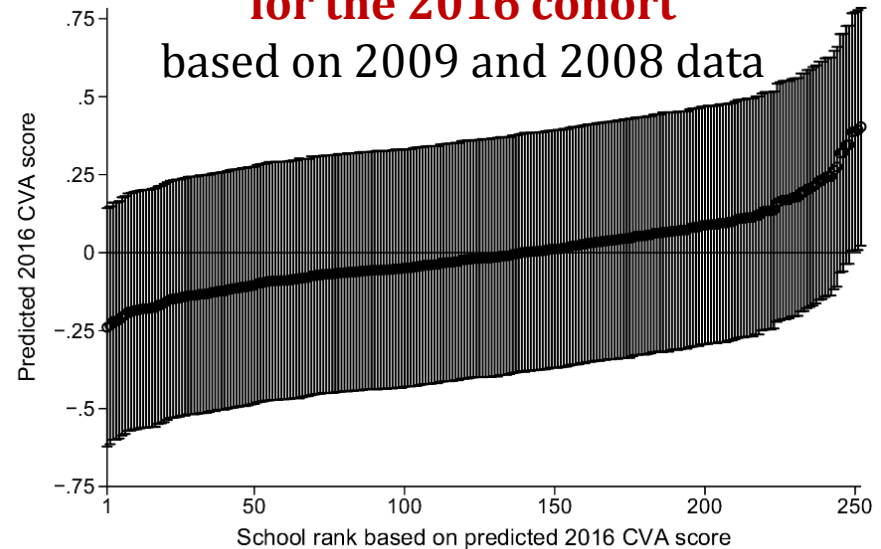
- We have used 2009 data to predict 2016 performance
- What about using data from 2008, 2007,...?
 - Note that earlier cohorts will add increasingly less information

Making more precise predictions?

**Predicted school effects
for the 2016 cohort
based only on 2009 data**



**Predicted school effects
for the 2016 cohort
based on 2009 and 2008 data**



- There is no visible improvement in the precision of the predictions

Limitation 2

Should we adjust for school level variables?

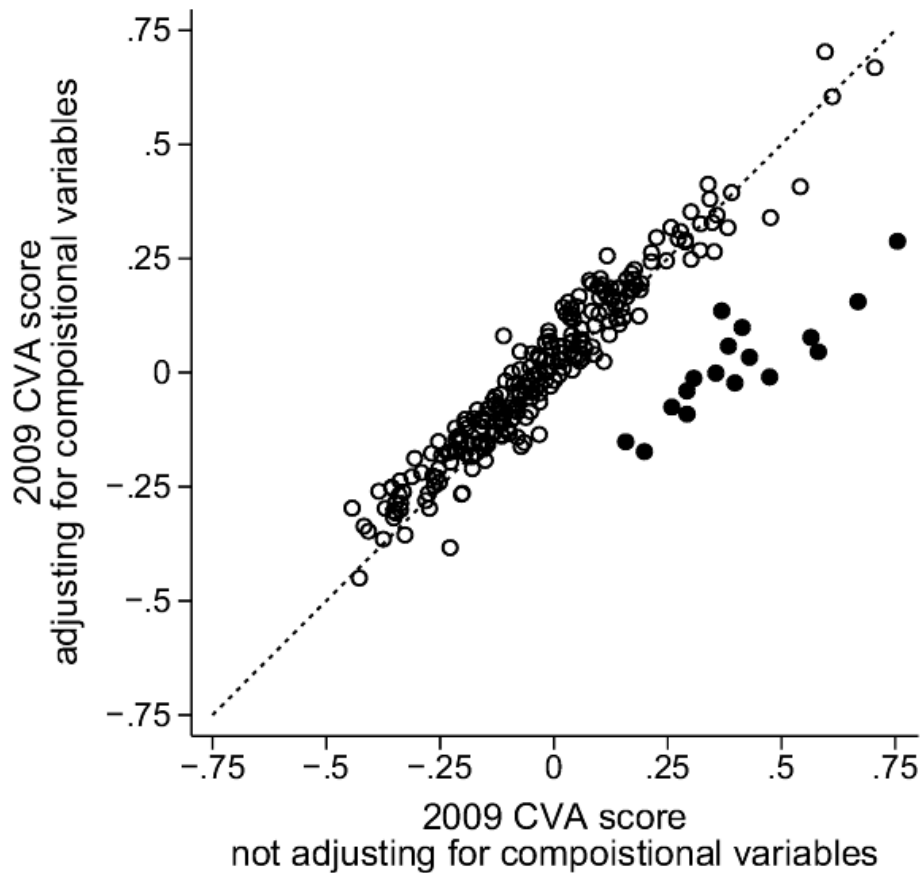
Adjusting for school level variables

- The government adjust for two school level variables
 - **School mean** of age 11 intake achievement
 - **School spread** of age 11 intake achievement
 - There is a positive effect of having a high achieving and homogenous intake
 - These are school composition variables that aim to measure peer group effects
- Including these variables removes peer group effects from schools' measured performances

Different users want different things from league tables

- **For choosing a school:**
 - No adjustment should be made
 - Parents are interested in how much better their child will do in one school than another
 - Peer-group effects are part of the difference between schools which is of interest
- **For holding schools accountable:**
 - This adjustment should be made
 - Schools should not be held accountable for factors outside their control
 - The government is interested in disentangling schools' policies and practices from their context and peer groups (this is ambitious!)

Adjusting and not adjusting for school compositional variables



Adjusting for the positive effects of having a high achieving and homogenous intake lowers the rankings of selective schools

However, selective schools' rankings will be lowered by too great an extent if selective schools are effective in their own right

Being a selective school is confounded with having a high achieving and homogenous intake

Other statistical limitations

Other statistical limitations

- At GCSE, students take **different combinations of subjects**
- Schools will be **differentially effective** for different types of students and for different responses
- **Student mobility** between schools is not recognised
- Students with **missing data** are listwise deleted
- Little is known about the **inter-rater reliability** of the tests
- For school choice, CIs for **multiple comparisons** are needed

Some broader limitations

- Huge financial cost to implement
- Teaching time is taken up with the administrative burden of the tests
- The range of knowledge and skills that tests assess is very narrow
- Stress caused by over-testing turns children off education

Conclusions

Conclusions

- School league tables ignore the uncertainty in using current performance as a guide to future performance
 - Adjusting for this uncertainty reduces the number of schools that can be separated to almost none
- For school choice, don't adjust for school-level factors, since this is part of the effect that parents are interested in
 - Adjusting for school achievement composition pushes selective schools down the league tables

Conclusions (cont.)

- We have shown that CVA scores contain very little information for choosing schools
 - This is just one more argument against their publication
 - However, the government insist that they are here to stay
 - In which case, strong health warnings are required
 - They should never be the sole basis of high-stakes decisions
- There is still an accountability role for CVA scores
 - But should only be used sensitively by experts
 - Can be used as a monitoring and screening device
 - However, it is not clear how to adjust for school compositional variables that are correlated with school policies and practices

Conclusions (cont.)

- The issues we have discussed are also relevant for primary school, post-16 schools and university league tables
 - Small size of primary schools makes estimated school effects even more imprecise
 - Universities are even harder to compare than schools due to lack of common curriculum and tests
- They are also relevant to other countries which publish school league tables to inform choice
 - Australia, Canada and the US

References

- Goldstein, H. and Leckie, G. (2008) School League Tables: What can they really tell us? *Significance*, 5, 62-64.
- Leckie, G. and Goldstein, H. (2009) The limitations of using school league tables to inform school choice, *Journal of the Royal Statistical Society: Series A*, 172, 835-851.
- Leckie, G. and 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*, 174, Part 3, *Forthcoming*.
- <http://www.dcsf.gov.uk/performancetables/>
- <http://www.bristol.ac.uk/cmpo/plugin/>