Hidden variation in school performance tables: the difficulties in identifying mathematics departments that are effective for disadvantaged students.

Fay Baldry[1], Jenni Ingram[2], Andrea Pitt[3], Victoria Elliott[4]

[1]*University of Leicester,* [2]*University of Oxford,* [3]*University of Warwick,* [4]*University of York*

The achievement gap between disadvantaged students and others has received considerable attention from politicians and the press. School performance tables, which are used to rank schools, now include measures of this gap. Results from this study show that using alternative measures of the achievement gap would have a considerable impact on the position schools hold in these tables. Combining measures has allowed schools to be identified that are differentially effective for disadvantaged students in mathematics.

Keywords: achievement gap, disadvantage, performance tables

Introduction

This article explores different measures of the achievement gap between disadvantaged and non-disadvantaged students at the end of Key Stage 4 (KS4) in England. This is based on school level data, published by the Department for Education (DfE) in performance tables and relates to students aged 15 to 16 years of age. The possible implications the choice of measure has on judgments made about schools will be discussed, along with the potential for different statistical models to identify effective mathematics departments in schools.

Since their inception, performance tables have been the cornerstone of school accountability systems and have retained a high public profile (Burgess, Wilson & Worth, 2014). Whilst the tables for secondary schools have remained focussed on the General Certificate of Secondary Education (GCSE) exams taken by the end of KS4, successive governments have changed the benchmark indicators used to rank schools. One modification was the inclusion of mathematics and English in the headline figure of the percentage of students achieving 5 A*-C GCSE grades; these changes have influenced school practice (Goldstein, 2013).

Recently, the DfE has focussed both funding and accountability measures on students who meet their disadvantage criteria, that of eligibility for free school meals in the previous six years or being looked after/in care in the six months before the census date, and has published measures that have separate figures for disadvantage students and others. The government’s stated aim is to evaluate the impact of the Pupil Premium, which is the additional funding that the schools receive for these students, and to hold schools accountable for closing the gap between disadvantaged and non-disadvantaged students (DfE, 2013).

Data from the 2013 performance table, containing school level progress data for 2013, 2012 and 2011, was used in this study. Whilst pupil level data can be accessed by schools and other authorised parties, it is not made available to the public or the press. Judgements regarding school effectiveness are made by the media, parents and others using performance tables; it is this publically available data that is analysed here.

Literature review

Almost since their inception, the role of performance tables in evaluating school effectiveness has been questioned (Goldstein & Spiegelhalter, 1996). Objections have focussed on the interpretation of individual measures through to the uncertainties involved in using earlier cohort data to predict future results. There have also been arguments about using probabilistic models for the statistical analysis (Gorard, 2006).

The disparity in educational attainment between different groups of students, often referred to as the achievement gap, has been widely discussed (Goodman & Burton, 2012). One area of debate has been the choice of particular indices of educational attainment, particular when changes over time are considered (Gorard, 2006; Connolly, 2006). In addition to advocating the use of proportional change, Gorard offered three multiplicative indices, the segregation ratio, the cross product ratio and the achievement gap (see Appendix). However, he also demonstrated that the same raw scores could produce different values for each of these indices, and acknowledged that interpretation was not straightforward. In examining both pass and failure rates, Connolly (2006) argued that proportional change was as susceptible to misinterpretation as percentage point gaps, but for pass rates between 20 and 80 percent both would highlight similar trends over time. Outside this range he advocated the use of effect size calculated from the chi-squared statistic.

Strand (2011) argued that students' intake characteristics have a significant effect on attainment, and therefore threshold measures, such as the proportion of students attaining five A*-C GCSEs, explain little of school-level factors. A value added (VA) measure has been advocated as being more useful (Wilson & Piebalgo, 2008); it compares students' 'best eight' GCSE scores with the median of all students with the same KS2 average score, which is then aggregated at a school level. When comparing schools' VA scores, Goldstein and Spiegelhalter (1996) justified the use of confidence intervals by arguing that cohorts could be treated as a sample from a superpopulation of, say, any future group of students. When this approach was taken, the majority of schools appeared statistically indistinguishable from each other (Rowe, 2009).

From 2007 to 2011, contextual value added (CVA) scores were published in the tables, which were derived from multilevel models and had explanatory factors that included prior attainment, gender, ethnicity and deprivation indicators. The argument that this approach produces a more accurate measure of school performance has substantial support (Leckie & Goldstein, 2009; Kelly & Downey, 2010), although there is still some debate as to whether using probabilistic models is a valid approach to cohort data (Gorard, 2010) and if treating the residual errors as a value added measure is justified (Schagen, 2006). Even for those advocating a multilevel approach, the use of CVA scores to rank schools on their effectiveness has been challenged. Unidentified or oversimplified variables, limitations inherent in assessments and missing data are some of the issues acknowledged to contribute to the level of uncertainty in these models (Rowe, 2009). This has led to multilevel models being described as an effective mechanism to identify outliers, but ineffective as a means to judge individual schools (Leckie & Goldstein, 2009; Wilson & Piebalgo, 2008).

Method

Secondary schools from fifteen English local authority regions in the midlands were considered. They were included if they were non-selective and had three consecutive years of data in the 2013 performance tables, which gave a sample of 352 schools. Calculations were based on the progress measure, which is the proportion of students attaining the government's expected three levels of progress from KS2 to the end of KS4. This was used as it was the only measure that differentiated between disadvantaged students and others for mathematics. The analysis was repeated for each individual year's data and three-year totals, though, due to space limitations, illustrative results focussing on 2013 mathematics are reported here.

In order to consider how different measures could affect the interpretation of an achievement gap, five measures were calculated: raw percentage point gaps, segregation ratios, cross product ratios, achievement gaps (Gorard, 1999) and effect sizes from the chi-squared statistic. Some of the data was published in integer percentage form, producing calculations that were estimates; rounding errors were considered as part of the analysis.

Goldstein (2013) argued that multilevel modelling is the most suitable analytical approach for clustered data, such as students in schools, and there is evidence of regional variation even though many schools are now run independently of local authorities (Gibbons, McNally & Viarengo, 2012). Consequently, a two-level hierarchical logistic model was constructed, with school and local authority level data, using MLwiN version 2.29 (see Appendix). The output variables were the proportion of students making expected progress, with the model repeated for disadvantaged and non-disadvantaged groups across the different years. Explanatory variables at a school level found to be significant, and therefore retained, were the proportion of students classified as: English as an additional language, special educational needs, arriving with low prior attainment and boys. At a local authority level, a measure of urbanisation, calculated from Defra's (2009) rural/urban scores, was significant and therefore included. Checks indicated that the equations met the assumptions required by the model and the residuals were interpreted as a contextualised progress measure.

The relationships between progress measures and achievement gap indices were explored using measures of correlation, descriptive statistics and changes in rank, in order to identify patterns and outliers. Ranks were often used as the context was comparison of schools. The stability of school rankings across the three years was considered in terms of percentage change in position.

In order to identify effective mathematics departments, schools were filtered based on the progress measures and the achievement gap indices. From the multilevel model, the top 30 percent of schools had confidence intervals wholly above the midline. As such, 30 percent was the benchmark used to select schools under a two-stage process; schools ranked in this top bracket for both progress measures for disadvantage students and the achievement gap indices for 2013 data were chosen, then they were retained if similar calculations from the three-year totals demonstrated a level of consistency. Calculations were repeated for English and used to differentiate between those schools that were just effective in mathematics from those effective in both subjects. This offered the possibility of separating mathematics departmental effects from wider school factors.

Findings and discussion

Measures of the achievement gap between disadvantage students and others.

The relationships between the formulae for the five achievement indices meant that all these measures were highly correlated. However, there was considerable variation for some schools when these were converted into ranks and compared with the percentage point gap. For the segregation ratio and cross product, 20 percent of the 352 schools had ranks that differed by more than 35 places, whilst this was 30 percent of schools for the chi-squared effect size and 59 percent for Gorard's (1999) achievement gap.

The asymmetry of the segregation ratio (Connolly, 2006) could be considered an advantage in so far as it promotes schools with higher attainment up the ranks when compared to schools with the same percentage point gap (Table 1). School A and school B have identical percentage point gaps, but when segregation ratios are considered, school A has a higher ranking, whereas school B's ranking drops.

Table 1: Comparison of schools with similar ranks (mathematics 2013 data).

	Percentage of students making expected progress		Percentage point gap		Segregation Ratio	
	Disadvantaged	Non-disadvantaged	Index	Rank	Index	Rank
School A	75	89	-14	115	0.96	95
School B	41	55	-14	115	0.75	177
School C	58	78	-20	197	0.74	176

Regardless of which measure was used it would still be difficult to judge schools on one index alone. School B and school C have very similar segregation ratios, with school B's percentage point gap considerably smaller, yet the attainment of students, including disadvantaged students, is higher in school C. This highlights a possible tension between equity and effectiveness (Strand, 2010). Progress measures for disadvantaged students do correlate with achievement gap indices (R^2=0.546 for percentage point gap 2013 data). However, as Figure 1 illustrates, when these relationships are considered at an individual school level, associating good levels of progress with small achievement gaps would be misleading for many schools.

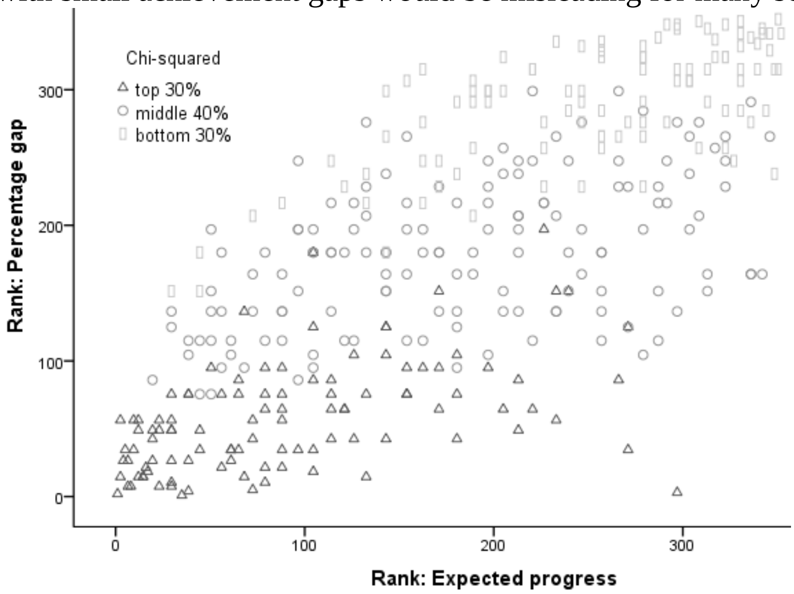

Figure 1: Disadvantaged students' progress against percentage point gap (mathematics, 2013 data)

When the three years of consecutive data was analysed, the progress measures and achievement gap indices indicated that a majority of schools did not hold stable positions relative to other schools over time. For the achievement indices, half of the 352 schools changed rank by over 88 places, regardless of which measure was used. A similar pattern emerged when the proportion of students making expected progress was considered, with over half the schools changing rank by at least 76 places.

Identification of schools

Filtering of the schools using a 30 percent benchmark across progress and achievement gap measures identified 32 schools for mathematics, representing 9 percent of the total. Equivalent calculations for English identified 36 schools, of which 12 were also selected for mathematics. Follow-up studies are planned in two of these schools, one from the mathematics only group and the other from the overlapping mathematics and English group.

The differences between mathematics and English extended beyond the selection criteria. Whilst not exploited in this study, a comparison of coefficients in a bivariate version of the multilevel model indicated that the explanatory factors had differing effects on mathematics and English. Also, equivalent measures calculated for the two subjects were positively correlated. For example, the correlation coefficient for proportion of deprived students making expected progress for 2013 was $R^2=0.609$. However, this did represent substantial difference for many schools; half of the 352 schools had a gap between mathematics and English ranks greater than 56 places, and the 10 percent of schools with the greatest difference had gaps in excess of 147 places. This pattern was repeated across the other measures.

Conclusion

Schools do appear to be differentially effective for groups of students as well as for different subjects, something which is rarely highlighted in the public discussions of performance table data. With the reported long tail of low attainment in mathematics (Hansen and Vignoles, 2005), gaining an understanding of the factors that influence achievement for different groups could offer insights into how to provide effective mathematics education for disadvantage students. Further research using pupil-level data to construct multivariate multilevel models could be of benefit in this context.

The most recent performance tables contain over 370 variables, with political agendas and the media determining which form the headline figures. Closing the achievement gap has been a recent government focus, but it is too early to know which, if any, of the measures involving disadvantaged students will be prioritised, or how the possibly competing outcomes of the achievement gap and attainment will be judged. Our findings support the notion that no single measure would be sufficient to capture the complexities of school effectiveness and all are too susceptible to variation to rank schools with any degree of certainty. A focus on simply ‘closing the gap’ could mask both high and low progress for disadvantaged students.

References

Burgess, S., Wilson, D., & Worth, J. (2013). A natural experiment in school accountability: The impact of school performance information on pupil progress. *Journal of Public Economics*, *106*, 57-67.

Connolly, P. (2006). Keeping a sense of proportion but losing all perspective: a critique of Gorard's notion of the ‘politician's error’. *British Journal of Educational Studies*, *54*(1), 73-88.

Defra (2005) LA classification dataset [accessed 18/12/2013] http://archive.defra.gov.uk/evidence/statistics/rural/rural-definition.htm

DfE (2013). *Schools and College Performance Tables: Statement of Intent 2013.* Department for Education.

Gibbons, S., McNally, S., & Viarengo, M. (2012). *Does additional spending help urban schools? An evaluation using boundary discontinuities* (No. 6281). Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit.

Goldstein, H. (2013). Evaluating educational changes: a statistical perspective. *Ensaio: Avaliação e Políticas Públicas em Educação* (AHEAD),

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)*, 385-443.

Goodman, R., & Burton, D. (2012). What is the nature of the achievement gap, why does it persist and are government goals sufficient to create social justice in the education system? *Education 3-13*, *40*(5), 500-514.

Gorard, S. (1999). Examining the paradox of achievement gaps. *Social Research Update*, *26*, 1-4.

——— (2006). Yet another perspective: a response to Connolly. *British Journal of Educational Studies*, *54*(4), 471-475.

——— (2010). Serious doubts about school effectiveness. *British Educational Research Journal*, *36*(5), 745-766.

Hansen, K., & Vignoles, A. (2005). The United Kingdom education system in a comparative context. In S. Machin & A. Vignoles (Eds.), *What is the Good of Education?* (pp, 13-35). Princeton, NJ/Woodstock: Princeton University Press.

Kelly, A., & Downey, C. (2010). Value-added measures for schools in England: looking inside the 'black box'of complex metrics. *Educational Assessment, Evaluation and Accountability*, *22*(3), 181-198.

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*(4), 835-851.

Rowe, K. J. (2009). Assessment, League Tables and School Effectiveness: Consider the Issues and 'Let's Get Real'!. *The Journal of Educational Enquiry*, *1*(1), 73-98.

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement*, *21*(3), 289-314.

——— (2011). The limits of social class in explaining ethnic gaps in educational attainment. *British Educational Research Journal*, *37*(2), 197-229.

Wilson, D., & Piebalga, A. (2007). *Accurate Performance Measures But Meaningless Ranking Exercise?: An Analysis of the English School League Tables*. University of Bristol, Centre for Market and Public Organisation.

Appendix

Progress	Dis-advantaged	Non-disadvantaged
Expected	*a*	*b*
Not expected	*c*	*d*

Segregation	Achievement gap (Gorard)	Cross
$\frac{\frac{a}{a+c}}{\frac{b}{b+d}}$	$\frac{a-b}{a+b} - \frac{(a+c)-(b+d)}{a+c+b+d}$	$\frac{ad}{bc}$

Basic model for the proportion of pupils making expected progress:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 X_{ij} + \beta_2 X_{2j} + v_j$$

i: school level *j*: local authority level