made quicker. For example, would getting examiners to give an intuitive score for each essay on a scale from 1–10, and then processing these scores using a Rasch model result in an equally useful set of student scores as traditional marking, but in a fraction of the time?

It is, of course, not necessarily true that the results shown here with respect to GCSE English essays would be repeated for other subjects. It is doubtless the case that certain types of student performances lend themselves more readily to CJ, whereas others are easier to mark. Nonetheless, the results here are important in understanding where the benefits of CJ derive from. Recognising that these are not solely caused by switching the way in which judgements are elicited, but also in the number of such judgements and how they are analysed, allows for a more nuanced comparison of the relative advantages of CJ and traditional marking. Failing to recognise these differences risks the two approaches never being compared on a like-for-like basis.

This article should not be taken as a criticism of the existing system for marking high-stakes examinations. Indeed, the examinations regulator in England, Ofqual, have themselves stated that "fundamentally, we believe this is a system that people can have confidence in" (Ofqual, 2014b, p.3). However, seeking for improvements in any system is worthwhile, and we hope that this article can provide a useful addition to existing research in terms of thinking about whether, and how, CJ might form a part of any improvements.

References

Benton, T. (2017). How much do I need to write to get top marks? *Research Matters: A Cambridge Assessment publication, 24*, 37–40.

Bramley, T. (2007). Paired comparison methods. In P. Newton, J.-A. Baird, H. Goldstein, H. Patrick & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards*. (pp.246–294). London: Qualifications and Curriculum Authority.

Bramley, T. (2015). *Investigating the reliability of adaptive comparative judgement*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Bramley, T., & Vitello, S. (2018). The effect of adaptivity on the reliability coefficient in adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, https://doi.org/10.1080/0969594X.2017.1418734.

Chalmers, R.P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48*(6), 1–29.

Child, S., Munro, J., & Benton, T. (2015). *An experimental investigation of the effects of mark scheme features on marking reliability*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Christodoulou, D. (2018). Comparative judgement – the next big revolution in assessment? *researchED*, *1*(1), 13–14.

Ofqual. (2014a). *Review of Double Marking Research*. Ofqual/14/5381. Coventry: Office of Qualifications and Examinations Regulation.

Ofqual. (2014b). *Review of Quality of Marking in Exams in A Levels, GCSEs and Other Academic Qualifications*. Ofqual/14/5379. Coventry: Office of Qualifications and Examinations Regulation.

Pollitt, A. (2004, June). *Let's stop marking exams*. Paper presented at the annual conference of the International Association for Educational Assessment, Philadelphia, USA.

Pollitt, A., & Murray, N. L. (1996). What raters really pay attention to. In M. Milanovic, & N. Saville (Eds.), *Performance testing, cognition and assessment. Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem* (pp.74–91). Cambridge, UK: Cambridge University Press.

Robitzsch, A. (2018). *sirt: Supplementary item response theory models*. R package version 2.5–45. Retrieved March 19, 2018 from https://CRAN.R-project.org/package=sirt

Steedle, J. T., & Ferrara, S. (2016). Evaluating Comparative Judgment as an Approach to Essay Scoring. *Applied Measurement in Education, 29*(3), 211–223.

How have students and schools performed on the Progress 8 performance measure?

Tim Gill Research Division

Introduction

In October 2013, the Department for Education (DfE) announced that new 'headline' performance measures for schools would be introduced to replace the previous measure of the percentage of students achieving five or more grades A* to C at General Certificate of Secondary Education (GCSE) Level, including English and Mathematics. The new measures (known as *Attainment 8* and *Progress 8*) are based on performance in a student's best eight subjects at GCSE (or equivalent), although with some restrictions; students are required to take the English Baccalaureate (EBacc[1]) qualifications in English and Mathematics, as well as at least three other EBacc qualifications. The remaining three slots can be filled either by other EBacc qualifications or by other approved, non-EBacc qualifications.

One of the reasons for the introduction of the new measures was concern that the previous measures penalised schools with a low-attaining intake. As Progress 8 is a value-added measure, it already accounts for the prior attainment of the student and should, in theory, no longer penalise these schools. The following method is used to calculate school-level Progress 8 scores:

- Calculate the Attainment 8 score for each student. This is the total points score for their eight highest scoring eligible qualifications. Points are based on the grade achieved (e.g., for GCSEs, points are on a 1–8 scale; 1 = G, 8 = A*[2]).

1. EBacc is the *English Baccalaureate*, a school performance measure which shows the proportion of pupils studying the 'core' academic subjects at KS4. Only specific qualifications (mainly GCSEs) are eligible for inclusion in the EBacc.

2. This example is for 'old' GCSEs. The scores for new GCSEs (9–1 grading scale) are slightly different.

- Compare this score with the mean Attainment 8 score for students with the same prior attainment (as measured by Average Fine Level on Key Stage 2 [KS2] tests[3]). The Progress 8 score for a student is the difference between the two, divided by 10 (this turns the score into a *per qualification* measure, as the English and Mathematics points are double-weighted in the calculation). This division is always by 10, even if a student takes fewer than 8 eligible qualifications.
- Calculate the mean Progress 8 score for all students in the school. This is the school's Progress 8 score.

By definition, the mean Progress 8 score for students with the same prior attainment is always zero. However, despite the value-added structure of Progress 8, there have been various criticisms levelled at the new measure. At the level of the school, there is evidence that it is biased towards Selective schools (e.g., Allen, 2016; Andrews, 2017). At student level, there is evidence that certain groups perform systematically better than others. These include female students, and those of Chinese ethnicity (Andrews, 2017), non-free school meals (non-FSM) students (Andrews, 2017; Sherrington, 2017), and non-Pupil Premium and EAL students (Thomson, 2017).

A further criticism of Progress 8 is that a few student-level outliers can have undue influence on the school-level score. Both Allen (2017) and Sherrington (2017) found that having just a handful of students who for one reason or another did not sit any eligible qualifications (and therefore achieved Attainment 8 scores of zero) can severely reduce a school's overall Progress 8 score. A recent UK Government policy document on school performance tables (DfE, 2017a) revealed that the DfE plans to consult with schools on this issue with a view to making changes to the methodology in future years.

The purpose of this research was to delve deeper into the relationship between Progress 8 scores and various student- and school-level factors. Prior research tended to focus on basic differences in Progress 8 mean scores between groups. This article presents similar analyses on more recent data and builds on this with a more detailed analysis, including a linear regression model to infer which factors were most important in determining scores at student level.

Data and methods

We used data from the 2015/16 academic year. We were interested in both the school- and student-level Progress 8 scores. Students' Progress 8 scores were taken directly from the National Pupil Database (NPD), which is administered by the DfE. The NPD includes examination results for all students in all qualifications and subjects in schools and colleges in England, as well as student and school background characteristics such as gender, ethnicity and level of income-related deprivation. Data on school-level Progress 8 scores was downloaded from a DfE website (https://www.compare-school-performance.service.gov.uk/download-data).

We used school classification information from *Edubase* (the DfE's register of educational establishments), which classifies schools by their school 'type' and by their selection policy. We excluded data on Independent schools and Special schools because they are not subject to the same accountability measures as State schools. We also excluded data on further education (FE) and sixth form colleges because, although some offer GCSEs or other qualifications to students in Key Stage 4 (KS4), this is a relatively rare occurrence.

The geographical region of each school was downloaded from a UK Government website (https://get-information-schools.service.gov.uk).

We undertook a descriptive analysis of Progress 8 scores for different groups of students and schools, followed by statistical modelling of Progress 8 scores at student level. We categorised students by a number of different background characteristics recorded in the NPD:

- *FSM status:* We classified students by whether or not they had claimed for free school meals (FSM) in any of the past six years.
- *SEN status:* We classified students with Special Educational Needs (SEN) by the categories used in the NPD. These were (in order of the amount of extra support needed, from low to high): SEN support; Statement of SEN; or Education, Health and Care Plan (EHCP).
- *Ethnicity:* The NPD categorised students into one of seven ethnic groups: White; Asian; Black; Chinese; Mixed; Other; or Unknown. Chinese students were in a category of their own due to a well-known tendency to perform very well compared to other Asian students.
- *EAL status:* The English as an Additional Language (EAL) classification in the NPD was into one of three categories: English; Other; or Unclassified. It should be noted (see Strand, Malmberg & Hall, 2015) that the definition of EAL in the NPD only accounted for whether the student was exposed to an additional language in their home or community. It did not actually tell us their level of proficiency in English.
- *School type:* We classified students by their school type, taken from *Edubase*: Academy (Comprehensive); Academy (Modern); Academy (Selective); Comprehensive; Secondary Modern; and Grammar.
- *School gender:* Girls'; Boys'; or Mixed.
- *Region:* South East; London; North West; East of England; West Midlands; South West; Yorkshire and the Humber; East Midlands; or North East.

Additionally, we investigated the relationship between two non-categorical variables and Progress 8 scores. These were:

- *Prior attainment:* Measured by students' KS2 Average Fine Level. This was not included in the descriptive analysis because it is already accounted for in the calculation of Progress 8, and therefore average Progress 8 scores do not vary with levels of prior attainment. However, it was still included as a factor in the regression model.
- *Income-related deprivation:* A measure of deprivation commonly used in analyses of student performance is the *Income Deprivation Affecting Children Index* (IDACI). This measures the percentage of children in the area where the student resides who live in income-deprived families. As such, it cannot tell whether or not the students themselves are income deprived[4].

For the school-level analysis, we calculated various measures to indicate the composition of students in each school, as the types of students attending a school were likely to have a significant impact on the Progress 8 score for that school. They may also have an impact on the Progress 8 scores for individual students. We calculated the following school-level variables (using the data in the NPD on students at the end of KS4):

3. Average Fine Levels are derived from the marks achieved on KS2 tests in Mathematics and English.
4. For further information on IDACI calculation, including definitions of children, families, and income deprivation, see https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-technical-report

- *Mean prior attainment (KS2 Average Fine Level):* This was calculated as the mean of the prior attainment of each student in the school who was at the end of KS4 in 2015/16.
- *Mean level of implied deprivation* (as measured by IDACI).
- *Percentage of students who claimed for FSM in at least one of the past six years.*
- *Percentage of students with SEN:* For the purpose of the school-level analysis, the three SEN categories were combined to create a binary variable (student with SEN or not).
- *Percentage of White students.*
- *Percentage of Black students.*
- *Percentage of Asian students.*
- *Percentage of students with English as an additional language (EAL).*

Schools were also classified by their type (according to the *Edubase* categories), by their gender and by the region.

Descriptive analyses

These analyses consisted of the calculation of means and standard deviations of Progress 8 scores for students and schools in each category (for the categorical variables). For the non-categorical variables, we present correlations and scatter plots of the relationship with the Progress 8 score (at either student or school level).

Regression analysis

The purpose of the regression model was to be able to investigate the effect of each factor in turn, whilst controlling for all other factors. As such, it provides a more nuanced approach than the descriptive data analysis and can determine which were the most important factors in accounting for variance in Progress 8 scores. It can also highlight any significant interaction terms, where the effect of one variable is different at different levels of another variable.

We fitted a statistical model to determine which of the factors described above were most important in predicting the student-level performance on Progress 8. A multilevel model was used to account for the clustering of students within schools. The general form of the model was as follows:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots \beta_k x_{kij} + \mu_j + \varepsilon_{ij}$$

Where y_{ij} is the Progress 8 score for student i in school j, x_{1ij} to x_{kij} are the independent variables, β_1 to, β_k are the regression coefficients, μ_j is a school effect (this is technically known as the Level 2 random effect and allows the model to account for the clustering of pupils within schools) and ε_{ij} is the residual difference between a student's predicted and actual Progress 8 score.

Results

Descriptive analysis of students' performance on Progress 8

This section presents the results of descriptive analyses of student-level Progress 8 scores, by various background factors. This summary is similar to that provided by the DfE (see https://www.gov.uk/government/statistics/revised-gcse-and-equivalent-results-in-england-2015-to-2016), although their analysis included data from Special schools and FE colleges, while ours does not.

Table 1: Distribution of Progress 8 scores, by student characteristics

		No. of students[5]	*Mean P8*	SD *P8*
Gender	Female	248,547	0.13	1.02
	Male	253,257	-0.12	1.08
FSM	No	367,843	0.12	0.95
	Yes	133,406	-0.31	1.24
SEN status	None	435,912	0.06	1.00
	SEN support	55,397	-0.38	1.32
	Statement	8,019	-0.31	1.31
	EHCP	2,044	-0.49	1.40
Ethnic group	Asian	45,726	0.34	0.98
	Black	23,688	0.21	1.04
	Chinese	1,584	0.70	0.83
	Mixed	20,858	-0.01	1.12
	White	399,119	-0.06	1.05
	Other	6,196	0.50	1.06
	Unknown	4,236	-0.05	1.16
Language	English	436,739	-0.06	1.05
	Other	63,842	0.42	1.03
	Unclassified	826	0.02	1.18
School type	Academy (Comprehensive)	290,343	0.01	1.06
	Academy (Selective)	17,467	0.33	0.72
	Academy (Modern)	12,251	-0.03	1.06
	Comprehensive	172,900	-0.04	1.08
	Grammar	3,005	0.32	0.75
	Secondary Modern	5,838	-0.14	1.10
School gender	Mixed	451,353	-0.03	1.07
	Boys'	20,096	0.15	0.96
	Girls'	30,355	0.31	0.95
Region	East Midlands	44,612	-0.11	1.06
	East of England	57,473	0.05	1.03
	London	68,467	0.20	1.08
	North East	24,552	-0.11	1.04
	North West	69,862	-0.12	1.08
	South East	79,486	0.05	1.04
	South West	49,264	-0.02	1.04
	West Midlands	56,281	-0.04	1.03
	Yorkshire and the Humber	51,753	-0.01	1.07
	All	**501,804**	**0.00**	**1.06**

Table 1 presents the number of students in each category, and the mean and standard deviation (*SD*) of Progress 8 (P8) scores, by each of the student-level factors.

This shows a small, but noteworthy, difference between the mean scores for girls (0.13) and boys (-0.12). This means that on average, girls made the equivalent of a quarter of a grade more progress per subject than boys.

Around a quarter of students were in the FSM category and, on average, they had substantially lower Progress 8 scores (-0.31) than students in the non-FSM category (0.12).

There were around 13% of students categorised as having some level of SEN, of which most were in the SEN support category (11%). Only 1.6% of students had a Statement of SEN and 0.4% had an EHCP. It is clear that students with no SEN had a much higher mean Progress 8 score than any of the SEN students. Students with EHCP had the lowest mean Progress 8 score, followed by those with SEN support and those with a Statement.

5. Totals by characteristic may not always add up to the total number of students included, due to missing data.

It is interesting that students with a Statement (i.e., a higher level of special needs) made better progress on average than those receiving SEN support (lower level).

White students made up about 80% of the population, with 9% Asian and around 5% Black. Only about 0.3% of students were Chinese. The results show that Chinese students had the highest scores on average, followed by those in the 'Other' category and Asian students. The lowest mean was for White students (-0.06). Thus, all other ethnicities made more progress than Whites. However, this analysis takes no account of other factors, such as deprivation levels, which may be more important in determining Progress 8 scores.

The results in Table 1 show that around 87% of students were English speakers. Non-English speakers had a much higher mean (0.42) than English speakers (-0.06).

In terms of school gender, students in both types of single-sex schools had higher Progress 8 scores on average than those in mixed schools. This is likely to be due to a high proportion of single-sex schools also being selective, and students in these schools made more progress on average (see Table 1). Students in girls' schools made more progress than those in boys' schools, which may be partly because girls made more progress than boys (see Table 1).

Finally, the differences in mean Progress 8 scores between regions were generally quite small. However, students in London had a notably higher mean (0.20) than any other region. The regions with the lowest means were North West (-0.12), East Midlands, and North East (both -0.11). This analysis takes no account of the background characteristics of the students in each region, which may be more important in determining Progress 8 scores.

Figure 1 displays a plot of the relationship between the implied deprivation experienced by students (as measured by IDACI) and their Progress 8 score. This shows a small, negative relationship between the two measures (correlation = -0.14), indicating that more deprived students had lower Progress 8 scores on average.

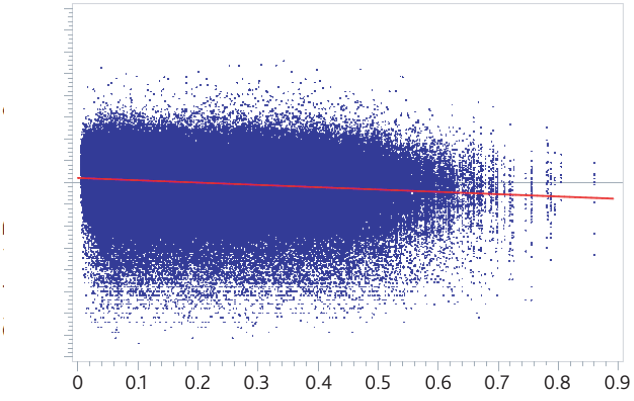

Figure 1: Relationship between student IDACI and Progress 8 score

Descriptive analysis of schools' performance on Progress 8

Table 2 presents descriptive analyses of Progress 8 (P8) scores by various school-level factors (school type, school gender, and school region).

The mean scores by school type were very similar to the mean scores in the student-level analysis (see Table 1). Schools classified as Selective academies or Grammar schools had higher mean Progress 8 scores on average (0.33, compared with negative mean scores for all other school types). The lowest mean was for non-academy Secondary Modern schools (-0.14). There were very few Grammar schools or Selective academies with Progress 8 scores below zero. In contrast, the maximum Progress 8 scores for Secondary Modern schools was just 0.59.

Table 2: Distribution of mean Progress 8 scores, by school characteristics

		No. of schools	*Mean P8*	*SD P8*	*Min P8*	*Max P8*
School type	Academy (Comprehensive)	1,745	-0.03	0.41	-2.51	1.37
	Academy (Selective)	139	0.33	0.20	-0.24	0.75
	Academy (Modern)	79	-0.03	0.32	-1.00	0.59
	Comprehensive	1,053	-0.05	0.34	-1.36	1.08
	Grammar	23	0.33	0.21	-0.06	0.81
	Secondary Modern	43	-0.14	0.35	-0.83	0.54
School gender	Mixed	2,722	-0.05	0.38	-2.51	1.31
	Boys'	151	0.16	0.32	-0.60	1.15
	Girls'	209	0.31	0.26	-0.49	1.37
Region	East Midlands	268	-0.14	0.37	-1.23	0.93
	East of England	345	0.02	0.41	-2.32	0.75
	London	429	0.19	0.35	-1.71	1.14
	North East	147	-0.14	0.35	-1.08	0.69
	North West	445	-0.15	0.42	-2.51	1.37
	South East	473	0.04	0.34	-1.14	1.08
	South West	307	-0.04	0.33	-1.19	0.91
	West Midlands	371	-0.05	0.37	-1.79	1.31
	Yorkshire and the Humber	295	-0.02	0.36	-2.09	0.93
All		**3,082**	**-0.02**	**0.38**	**-2.51**	**1.37**

In terms of school gender, girls' schools had the highest mean Progress 8 score, followed by boys' schools. The difference between the average Progress 8 score in girls' schools and in mixed schools (0.36) was equivalent to more than one third of a grade. These results were very similar to the results for students attending single-sex or mixed schools (see Table 1), but it is worth noting that there were very few girls' schools with a negative Progress 8 score.

Schools in London had the highest mean Progress 8 scores by some distance (0.19). The only other regions with a positive mean were South East (0.04) and East of England (0.02). The regions with the lowest mean were North West (-0.15), East Midlands, and North East (both -0.14). The 'London effect' is a well-researched phenomenon that has been attributed to a number of different factors (see Blandon, Greaves, Gregg, MacMillan & Sibieta, 2015), and the results here suggest it is present in terms of progression as well as attainment.

The remaining school-level factors that we investigated were continuous variables, which are better analysed through correlation coefficients and scatter plots. Table 3 presents the correlation coefficients between Progress 8 scores and the value of each of these

Table 3: Correlations between school-level factors and Progress 8 scores

School factor	*Correlation coefficient*
Mean prior attainment	0.33
FSM students (%)	-0.28
SEN students (%)	-0.21
EAL students (%)	0.23
Mean IDACI	-0.27
White students (%)	-0.23
Asian students (%)	0.18
Chinese students (%)	0.17
Black students (%)	0.14

variables at school level. Figure 2 plots the data for four of these factors (prior attainment, percentage FSM, percentage SEN, and percentage EAL).

To give an idea of what the prior attainment scale means, a student who only just achieved Level 4 in English and Mathematics (the minimum expected level in each subject) would have an Average Fine Level equal to 4.0. Figure 2 shows a distinct positive relationship between the two, suggesting that schools with higher prior attaining students also tended to get higher Progress 8 scores on average. The correlation between the two measures was 0.33. The slope of the line of best fit was 0.48, meaning that an increase in KS2 mean of 1 led to (on average) an increase in Progress 8 score of 0.48. Note that there appeared to be two separate groups of schools, with the main body of schools having a KS2 mean score of between 4 and 5, and a smaller group with a KS2 mean score of between 5 and 5.5. Almost all of this second group were Selective schools, which were able to select exclusively high-attaining students.

There was a clear negative relationship between the school percentage of FSM students and Progress 8 scores (correlation = -0.28). However, we can see from Figure 2 that there were many schools with high percentages of FSM students and positive Progress 8 scores, and many with low percentages of FSM students and negative Progress 8 scores.

There was also a significant negative relationship between the percentage of students with SEN in a school and Progress 8 score, with a correlation of -0.21. In other words, schools with higher percentages of SEN students tended to have lower Progress 8 scores. This is not a surprising finding, as students with special needs made less progress on average than other students (see Table 1).

There was a significant positive relationship between the percentage of EAL students and Progress 8 scores (correlation = 0.23). Schools with higher percentages of EAL students were more likely to have higher Progress 8 scores than those with low percentages of EAL. We expected this result, given that EAL students made considerably more progress on average than students with English as their first language (see Table 1).

Table 3 also shows that the correlation between the mean level of implied deprivation (as measured by IDACI) experienced by students attending the school and the school Progress 8 score was -0.27. Given the negative relationship between disadvantage and Progress 8 score at an individual level, it is not surprising that the same relationship was also visible at school level.

The ethnic make-up of schools was also considered as a possible influence on Progress 8 scores. The correlations in Table 3 show that in each case there was only a relatively weak relationship between the percentage of the ethnic group and Progress 8 mean scores. The correlation was negative between the percentage of White students and Progress 8 score (-0.23) whilst correlations were positive between the percentages of Asian, Chinese, or Black students and Progress 8 score (0.18, 0.17, and 0.14 respectively). Again, these results reflect the student-level results in Table 1, which show more progress on average made by Asian, Chinese or Black students, compared with White students.

Regression analysis of student-level Progress 8 scores

Table 4 presents the results of the modelling of school-level Progress 8 scores. Model 1 included no predictors, just an intercept, to assess the amount of variance in achievement between schools. From the random effects part of the table, we calculate that schools accounted for around

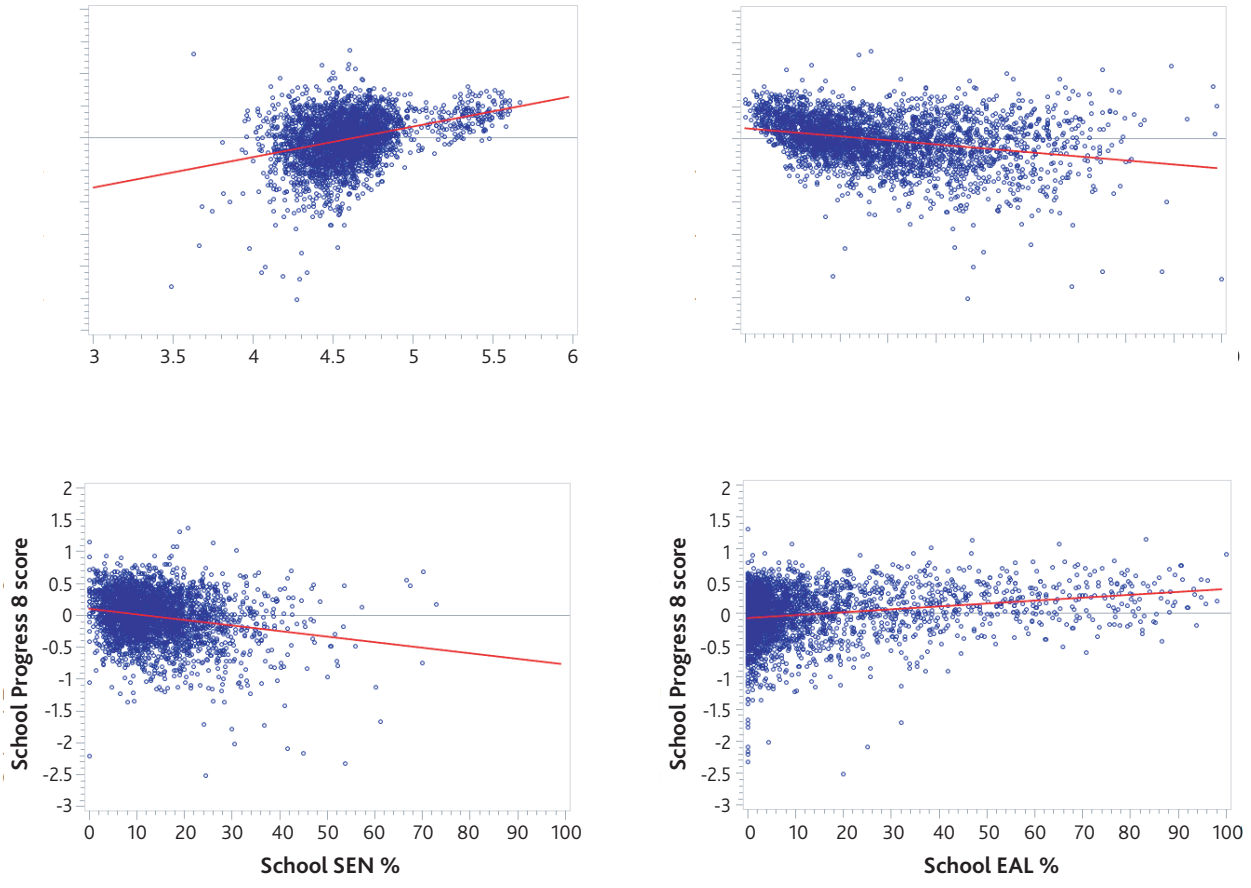

Figure 2: Relationship between school-level variables and Progress 8 score

11.5 per cent of the variance in Progress 8 scores[6]. This is a substantial proportion and suggests that the use of a multilevel model was justified. Model 2 then included all of the predictor variables, whether at the student or school level, with statistical significance ($p<0.05$) indicated in the table by bold type. A further model (results not shown in the table) included any statistically significant interactions between fixed effects[7]. Some of the results from this final model are presented and discussed later in this article.

The size of each coefficient represents the change in Progress 8 associated with a particular category compared with the base category (for the categorical variables). As Progress 8 is essentially a mean grade measure, the coefficients represent the change in mean grade associated with a particular category. For continuous variables, the coefficients represent the change in mean grade associated with a unit increase in that variable.

Main effects

The results of Model 2 included a significant effect of prior attainment, which is surprising since the calculation of Progress 8 already accounts for this. The size of the effect (-0.23) suggests that an increase of one KS2 level for a student was associated with a fall in Progress 8 of between one fifth and one quarter of a grade. The existence of this effect was thought to be due to the presence of two different prior attainment effects, which cancel each other out overall (so that the mean Progress 8 score at each level of prior attainment is zero): firstly, a between-school effect, where schools with higher performing intakes tended to display higher Progress 8 scores (see Figure 2 and also the *School KS2 mean coefficient*); and secondly, a within-school effect where, within any given individual school, candidates with the highest prior attainment were likely to be those with the lowest Progress 8 scores[8].

This means that, for example, a high prior attaining student will tend to make more progress on average if they are in a school with other high prior attaining students than if they are amongst students with low prior attainment. Equally, a low prior attaining student will have higher Progress 8 scores on average if they are in a school with high-attaining students. If we compare high prior attaining and low prior attaining students in the *same* school, on average, the students with low prior attainment will have slightly higher Progress 8 scores.

The results also showed that females made more progress than males (0.25 of a grade on average). More disadvantaged students tended to make less progress: being an FSM student was associated with a significantly lower Progress 8 score (by 0.39 of a grade); and an increase in the level of implied deprivation was significantly associated with a fall in Progress 8 score (by 0.09, for each increase in IDACI of 0.1). Furthermore, even after accounting for these effects at individual level, schools having a larger proportion of their students receiving free school meals were associated with lower Progress 8 scores. However, this effect was very small – an increase in FSM students of 20 per cent was associated with a fall in Progress 8 of 0.04. In terms of SEN, each SEN

Table 4: Student Progress 8 score - Regression coefficients

Note: Standard errors in brackets; bold type indicates statistical significance at $p<0.05$.

Fixed effects		*Model 1*	*Model 2*
Intercept		-0.012 (0.014)	0.027 (0.022)
KS2 Average Fine Level			**-0.226 (0.002)**
Gender	Male		
	Female		**0.249 (0.003)**
FSM	No		
	Yes		**-0.393 (0.003)**
IDACI			**-0.933 (0.013)**
SEN status	None		
	SEN support		**-0.480 (0.005)**
	Statement		**-0.503 (0.011)**
	EHCP		**-0.691 (0.022)**
Ethnic group	White		
	Asian		**0.188 (0.007)**
	Black		**0.232 (0.007)**
	Chinese		**0.428 (0.024)**
	Mixed		**0.068 (0.007)**
	Other		**0.318 (0.013)**
	Unknown		-0.029 (0.016)
EAL	No		
	Yes		**0.386 (0.006)**
	Unknown		**0.091 (0.034)**
School type	Academy (Comprehensive)		
	Academy (Selective)		**0.095 (0.040)**
	Academy (Modern)		-0.051 (0.036)
	Comprehensive		**-0.039 (0.012)**
	Grammar		0.066 (0.070)
	Secondary Modern		**-0.113 (0.048)**
School gender	Mixed		
	Boys'		**0.112 (0.028)**
	Girls'		0.022 (0.024)
School KS2 mean			**0.232 (0.043)**
FSM percentage			**-0.002 (0.001)**
IDACI mean			0.165 (0.144)
SEN percentage			**0.002 (0.001)**
EAL percentage			**0.003 (0.001)**
Asian percentage			**-0.004 (0.001)**
Black percentage			0.000 (0.002)
White percentage			-0.001 (0.001)
Region	London		
	East Midlands		**-0.163 (0.030)**
	East of England		-0.047 (0.029)
	North East		-0.021 (0.037)
	North West		**-0.117 (0.028)**
	South East		**-0.060 (0.028)**
	South West		**-0.080 (0.030)**
	West Midlands		**-0.064 (0.028)**
	Yorkshire and the Humber		-0.015 (0.030)
Residual variances			
Level 1 (Pupil level)		**1.008**	**0.895**
Level 2 (School level)		**0.131**	**0.085**

6. As calculated by the intraclass correlation coefficient (ICC). ICC = school variance/(school variance + error variance) = 0.131/(0.131+1.008) = 0.115.
7. A full table of results, including interaction effects, can be found in the conference paper version of this article, available on the Cambridge Assessment website: http://www.cambridgeassessment.org.uk/our-research/all-published-resources/conference-papers/
8. This within-school effect was confirmed by grouping candidates within each school into five groups based on their prior attainment. Each successively higher prior attaining group had a lower mean Progress 8 score (0.05, 0.04, 0.01, -0.02, -and 0.08).

status was associated with less progression, with the biggest effect being for those with EHCP (-0.691).

Compared to White students, all other ethnic groups were associated with significantly more progress, even after accounting for other factors, such as implied deprivation. The largest effect was for Chinese students (0.43 of a grade), followed by students from an Other ethnic background (0.32), Black students (0.23), and Asian students (0.19). Having English as a second language was associated with better progress (by 0.39 of a grade), compared with students who spoke English as their first language. There were some significant school type effects, with being in a Selective academy associated with significantly higher Progress 8 scores (by 0.10 of a grade) and being in a non-academy Secondary Modern associated with significantly lower Progress 8 scores (by 0.11 of a grade), compared with being in a Comprehensive academy. There was also a statistically significant negative effect of attending a Comprehensive school, but this was very small. It is interesting that the effect of attending a Grammar school (either an academy or not) was much reduced compared to that found in the descriptive analysis (see Table 1). In fact, attending a non-academy Selective school was not significantly different from a Comprehensive academy. This is probably because the effect seen in the descriptive analysis was more due to attending a school with a high mean prior attainment than to attending a Selective school.

Students in single-sex boys' schools made significantly more progress than those in mixed schools, even after accounting for the fact that these schools tended to have a higher attaining intake and were more likely to be selective than mixed gender schools. However, whilst the descriptive analysis showed that students in girls' schools made the most progress on average, there was no significant difference between girls' school students and mixed school students in the statistical model. This is likely to be because the effects of an individual's gender and of school mean prior attainment were more important in determining Progress 8 scores.

There were several statistically significant school-level effects, but these were very small. These included the percentages of SEN, EAL and Asian students in a school. The coefficients for these effects ranged from -0.004 to 0.003. An effect of -0.004 meant that an increase of 20 per cent was associated with a reduction in the Progress 8 score of just 0.08.

Finally, there were significant effects of the geographical region of the school. Compared with London, being in any another region was associated with lower Progress 8 scores. However, the effects were smaller than those seen in the descriptive analysis (and were not significant for some regions). Even so, there is evidence that the London effect was present, even after accounting for other factors.

Interaction effects

There were a number of significant interaction effects in the final model (results not shown in Table 4). A few of the more interesting effects are presented in this section. For the complete set of results, please refer to the full paper on the Cambridge Assessment website. Using the values of the coefficients from the final model, it was possible to calculate predicted Progress 8 scores for different groups of students, and therefore demonstrate the effects of these interactions. These are shown in Figures 3 to 7.

Gender x SEN status

Figure 3 presents predicted Progress 8 scores for male and female students, with different levels of SEN status[9]. This shows that female students with no SEN had a substantially higher predicted score than males, but this gap was smaller for students with SEN. For those with a Statement or EHCP, there was almost no difference in predicted scores between males and females.

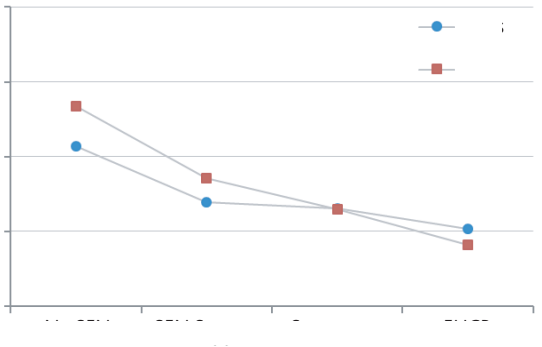

Figure 3: Predicted Progress 8 scores for male and female students with different levels of SEN

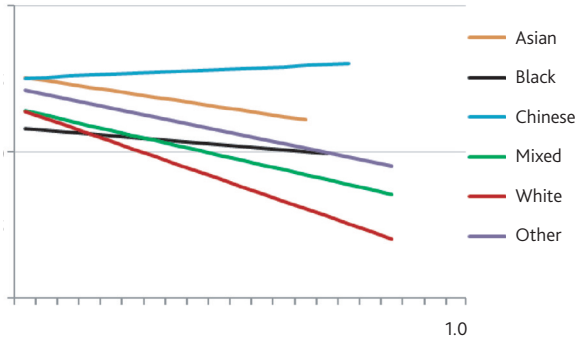

Figure 4: Predicted Progress 8 scores for students of different ethnicity, by IDACI score

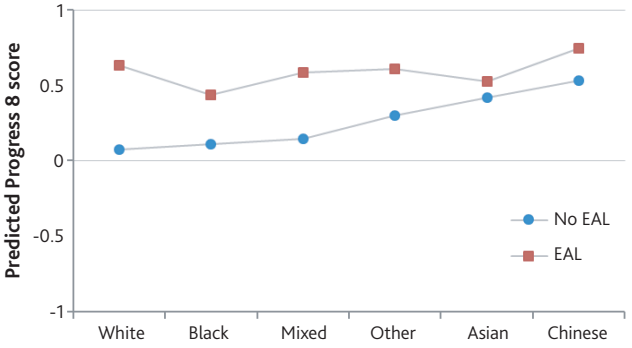

Figure 5: Predicted Progress 8 scores for students of different ethnicity, by EAL status

IDACI x Ethnic group

Figure 4 shows that the overall pattern was of increased IDACI associated with lower Progress 8 scores. However, for Chinese students this effect was reversed, so that increased IDACI was associated with very slightly higher Progress 8 scores. In terms of the other ethnicities, the effect of IDACI was largest for White students and was very small for Black students.

9. For this and all subsequent graphs of interactions, the predicted scores are for students in the baseline categories for all other categorical variables and with values of the continuous variables equal to the mean.

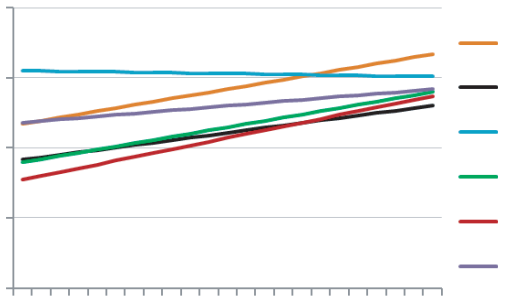

Figure 6: Predicted Progress 8 scores for students from different ethnic background, by school KS2

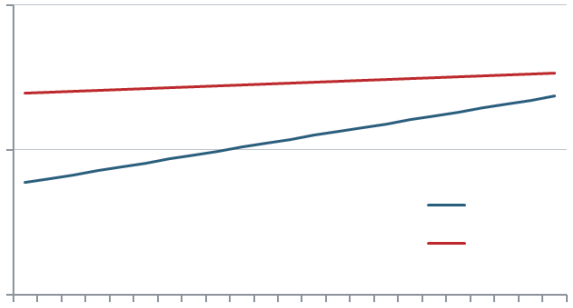

Figure 7: Predicted Progress 8 scores for students with different EAL status, by school KS2 mean

EAL x Ethnic group

Figure 5 presents the predicted Progress 8 scores for students from different ethnicities, by their EAL status. This shows that the effect of having EAL was different for each ethnicity. The effect was significantly larger for White students than any other ethnicity, and was smallest for Asian students.

Ethnic group x School KS2 mean

The effect of school prior attainment for students from different ethnic backgrounds is shown in Figure 6. This demonstrates that students from Chinese, Black, or Other backgrounds all had scores that were less sensitive to changes in school-level prior attainment, compared with White students. The biggest difference was for Chinese students, for whom school prior attainment had essentially no effect on their predicted Progress 8 score.

EAL x School KS2 mean

Figure 7 presents the predicted Progress 8 scores for students attending schools with different levels of mean prior attainment, by their EAL status. This shows that the effect of school-level prior attainment was much less for students with EAL, than for English speakers.

Discussion

Many of the findings presented in this article match those of previous research that investigated progress in secondary schools amongst different groups of students. We found that female students made on average a quarter of a grade (per qualification) more progress than males. This advantage in terms of progress in secondary school for female students was similar to that found in previous research (e.g., Sammons, 1995; Burgess, McConnell, Propper & Wilson, 2004). One possible explanation for this is the fact that, in the past, GCSEs had a substantial proportion of coursework or controlled assessment, which tend to favour girls. More recent research (Bramley, Vidal Rodeiro, & Vitello, 2015) has shown that girls also outperform boys in written examinations at GCSE, but by not as much as in coursework. It will be interesting to see whether the introduction of reformed GCSEs (most of which will no longer have coursework) leads to the gender gap closing somewhat.

We also found that students eligible for free school meals made less progress than other students, by around 0.4 of a grade on average. Students experiencing higher levels of implied deprivation also made less progress. Previous studies (e.g., Sammons, 1995; Strand, 2014; Sammons et al., 2014) found a similar negative relationship between progress and higher levels of disadvantage. One proposed reason was that more disadvantaged students have less support at home and lower educational aspirations.

The next finding was that all ethnic groups made more progress than White students, even after accounting for other factors, such as implied deprivation. The largest effect was for Chinese students (around 0.4 of a grade on average). Black and Asian students made about 0.2 of a grade more progress on average. This fits in with the findings from previous research (e.g., Sammons, 1995; Wilson, Burgess and Briggs, 2011; Strand, 2014), which suggests that the importance of high aspirations in many minority ethnic families and communities is the most likely explanation for this better progress.

We also found that students with EAL made about 0.4 of a grade more progress than students with English as their first language. Again, this matches the findings of previous research (e.g., Strand et al., 2015). However, part of this effect may be due to a change in the true EAL status over time. The NPD defines EAL students as those exposed to an additional language in their home or community, taking no account of their actual proficiency in English. We know that all of the students included in the analysis have been in the country for at least five years, because we have access to their KS2 test results. It may be that for many of these students, their knowledge of English will have improved significantly in that time and this may explain why their progress was so much greater.

Many of the school-level effects from the regression models were very small, but there were three variables (all closely linked) which were important. In the descriptive analysis we found that students in Selective schools or in single-sex schools made considerably more progress on average. However, in the regression model some of those differences disappeared, or became much smaller. In particular, the fact that students in schools with a higher attaining intake (as measured by KS2 mean) made more progress led to a big fall in the effect of attending a Selective school. Even so, some school type effects were still present after accounting for school mean KS2, with students in Selective academies performing better by about one tenth of a grade, and those in non-academy Secondary Moderns performing worse (by 0.11 of a grade) than students in Comprehensive academies. The girls' school advantage disappeared completely in the statistical model, whilst the boys' school advantage was still present.

This suggests that boys perform better in single-sex schools than in mixed schools, but girls do not, after accounting for other factors.

The region in which students attended school also had a significant effect on Progress 8 scores, with lower mean scores for all regions compared with London. However, the effects seen in the regression model were smaller than those in the descriptive analysis (and were not significant for some regions). For example, the North East had the second lowest mean Progress 8 score (-0.14) and was also highlighted in a recent DfE report (DfE, 2017b) as having the highest percentage of schools below the 'floor standard'. However, this effect was not significant once other factors were taken into account. Despite these caveats, there was still evidence that the London effect was present, even after accounting for other factors.

Some interesting interaction effects were also present. The advantage for girls was much smaller (or disappeared completely) if they had any SEN. The implied deprivation effect was greatest for White students, compared with other ethnic backgrounds. Figure 4 showed that White students from more deprived backgrounds made the least progress on average. This confirms the findings of previous research (e.g., Strand, 2014), and suggests that the importance of high aspirations in non-White ethnic backgrounds extends to those from poorer families.

Perhaps of most importance was the finding that, despite the measure taking account of prior attainment at the student level, Progress 8 systematically penalised schools with a lower performing intake. This pattern was present for both Selective and non-Selective schools, which suggests that the difference seen between school types was more due to the prior attainment of the students than to the type of school per se. This confirms previous findings (e.g., Yang & Woodhouse, 2001) that students make more progress in schools with a higher attaining intake. It is not entirely clear why this would be the case, but Thomson (2015) has suggested three possible explanations: students in higher attaining schools receive more help at home; competition amongst students may drive up attainment; and higher attaining schools may be more effective because they can recruit better teachers. It may also be the case that low prior attaining students receive a boost in high-attaining schools due to the positive influence of higher attaining peers.

Overall, this analysis has shown that Progress 8 scores are strongly related to a plethora of school and student factors, most of which are beyond the control of teachers. Therefore, it is questionable that Progress 8 should be used as a measure for comparing schools, or for bringing them to the attention of The Office for Standards in Education, Children's Services and Skills (Ofsted) if they are below the floor standard. It could be argued that a fairer way of judging schools would be to take account of some of these factors when calculating school performance measures. Whilst debate over different ways of measuring school effectiveness is likely to continue (e.g., Allen, 2015; or Allen, Burgess & Mayo, 2018), an awareness of the potential limitations of any such measures is crucial to their interpretation.

References

Allen, R. (2015, May 18). *We cannot compare the effectiveness of schools with different types of intakes*. [Blog]. Retrieved from: https://educationdatalab.org.uk/2015/05/we-cannot-compare-the-effectiveness-of-schools-with-different-types-of-intakes/

Allen, R. (2016, September 16). *Progress 8 is too favourable to grammar schools and understates secondary modern achievement*. [Blog]. Retrieved from: https://educationdatalab.org.uk/2016/09/progress-8-is-too-favourable-to-grammar-schools-and-understates-secondary-modern-achievement/

Allen, R. (2017, January 20). *Outliers in Progress 8*. [Blog]. Retrieved from: https://educationdatalab.org.uk/2017/01/outliers-in-progress-8/#_ftn1

Allen, R., Burgess, S., & Mayo, J. (2018). The teacher labour market, teacher turnover and disadvantaged schools: new evidence for England. *Education Economics, 26*(1), 4–23.

Andrews, J. (2017). *Analysis: The introduction of Progress 8*. EPI: London. Retrieved from: https://epi.org.uk/analysis/introduction-progress-8/#

Blandon, J., Greaves, E., Gregg, P., Macmillan, L., & Sibieta, L. (2015). *Understanding the improved performance of disadvantaged pupils in London. Social policy in a Cold Climate, Working Paper 21*. Centre for the Analysis of Social Exclusion, LSE: London.

Bramley, T., Vidal Rodeiro, C. L., & Vitello, S. (2015). *Gender differences in GCSE*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Burgess, S., McConnell, B., Propper, C., & Wilson, D. (2004). Girls rock, boys roll: An analysis of the age 14–16 gender gap in English schools. *Scottish Journal of Political Economy, 51*(3), 209–229.

DfE. (2017a). *2017 School and College Performance Tables: Statement of Intent*. Department for Education: London.

DfE. (2017b). *2017 Revised GCSE and equivalent results in England, 2015 to 2016*. Department for Education: London.

Sammons, P. (1995). Gender, Ethnic and Socio-economic Differences in Attainment and Progress: a longitudinal analysis of student achievement over 9 years. *British Educational Research Journal, 21*(4), 465–485.

Sammons, P., Sylva, K., Melhuish, E., Siraj, I., Taggart, B., Toth, K., & Smees, R. (2014). *Influences on students' GCSE attainment and progress at age 1: Effective Pre-School, Primary & Secondary Education Project (EPPSE) Research Brief*. Department for Education: London.

Sherrington, T. (2017, February 24). *More issues with Progress 8*. [Blog]. Retrieved from: https://teacherhead.com/2017/02/24/more-issues-with-progress-8/

Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: intersectionality and 'getting it' for the white working class. *Research Papers in Education, 29*(2), 131–171.

Strand, S., Malmberg, L., & Hall, J. (2015). *English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database*. University of Oxford Department of Education: Oxford.

Thomson, D. (2015, May 5). *Why do pupils at schools with the most able intakes tend to make the most progress?* [Blog]. Retrieved from: https://educationdatalab.org.uk/2015/05/why-do-pupils-at-schools-with-the-most-able-intakes-tend-to-make-the-most-progress/

Thomson, D. (2017, March 2). *Putting Progress 8 in context*. [Blog]. Retrieved from: https://educationdatalab.org.uk/2017/03/putting-progress-8-in-context/

Wilson, D., Burgess, S., & Briggs, A. (2011). The dynamics of school attainment of England's ethnic minorities. *Journal of Population Economics, 24*(2), 681–700.

Yang, M., & Woodhouse, G. (2001). Progress from GCSE to A and AS Level: Institutional and gender differences, and trends over time. *British Educational Research Journal, 27*(3), 245–267.