Attainment trends for primary grammar, punctuation and spelling (GPS), reading and maths

A nationwide analysis of state schools in England using assessments provided by RS Assessment from Hodder Education

November 2019

Katie Blainey, RS Assessment from Hodder Education
Timo Hannay, SchoolDash

RS ASSESSMENT FROM HODDER EDUCATION

SchoolDash UNDERSTAND.EDUCATION

Summary

Key findings of this paper include:

- The effects of gender differ considerably by subject. In grammar, punctuation and spelling (GAPS tests) girls, on average, perform better than boys at the start of primary school and maintain this advantage all the way to age 11. A similar pattern applies to reading (PiRA tests). In maths (PUMA tests), girls and boys initially show very similar performance, but by age 6-7 boys overtake girls and stay ahead until the end of primary school.
- Within each subject, there are often gender differences by topic. These are less evident in reading, where girls consistently outperform boys in all areas (comprehension, inference and LSP, or Language, Structure and Presentation). But in grammar, punctuation and spelling (GPS), gender differences tend to decline with age for grammar, spelling and vocabulary while increasing for punctuation. In maths, gender differences at the end of primary school are relatively high for number and measures, smaller for fractions and statistics, and virtually absent for operations and geometry.
- Having analysed the relative performance of summer-born pupils in reading and maths in our last white paper, we look here at GPS performance. Consistent with other subjects, this is lower for summer-born children and although the gap narrows as pupils get older, it persists until at least the end of primary school. These effects are broadly similar for all topics (grammar, punctuation, spelling and vocabulary).
- We also examine whether these gender and age gaps vary by type of school. The results indicate that differences tend to be linked to pupil cohorts rather than schools. This suggests that no particular groups of schools are more effective than any others in reducing the effects on attainment of pupil gender or season of birth.

Introduction

As in our previous study, we have analysed aggregate anonymous data from MARK (My Assessment and Reporting Kit), a free online marksheet and reporting service for customers of the Progress in Grammar, Punctuation and Spelling Assessment (GAPS), Progress in Reading Assessment (PiRA) and Progress in Understanding Mathematics Assessment (PUMA) primary tests provided by *RS Assessment from Hodder Education*. This paper updates and extends our 2018 analysis[1]. Unlike the DfE's statutory tests, which are currently used to assess each child twice (in Years 2 and 6), these assessments are typically sat every term, providing more detailed information about children's progress during their primary years. This paper provides a high-level analysis of over three million termly standardised test results produced between October 2015 and July 2019.

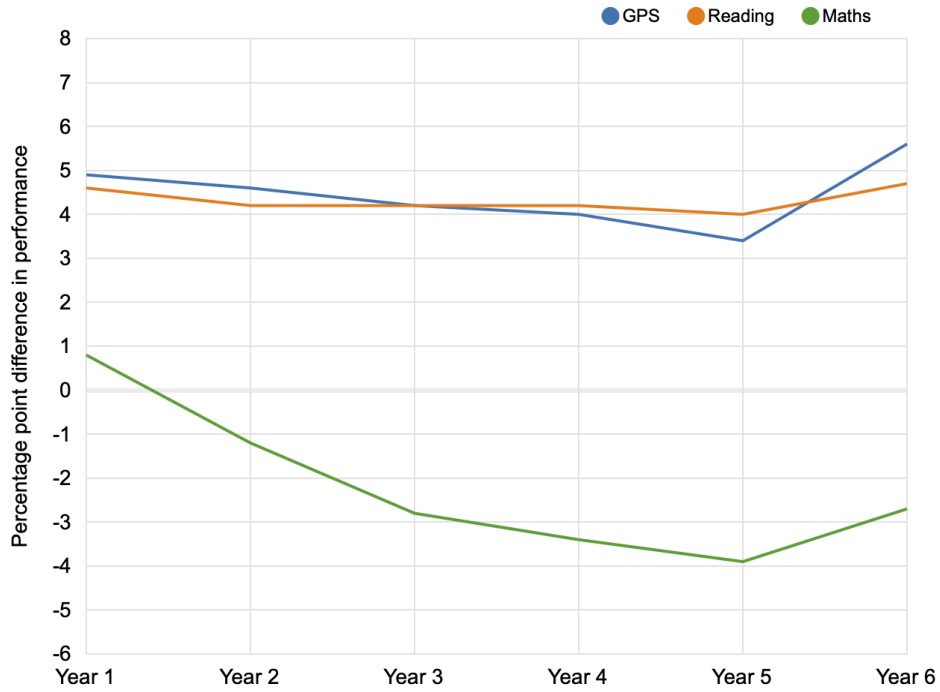
For further information about the kinds of schools included and an explanation of the test scoring system, please refer to the Appendix.

[1] See risingstars-uk.com/whitepaper2019

A comparison of boys and girls

Figure 1 shows the mean differences in test scores between girls and boys by year group and subject. In reading tests, girls achieve higher average scores than boys at the start of primary school and maintain this advantage through to Year 6. GPS tests show a similar trend. In maths tests, there is initially little if any difference between the average performance of boys and girls, but by Years 2-3 boys overtake girls and stay ahead until the end of primary school.

Figure 1: Mean reading, GPS and maths performance of girls relative to boys



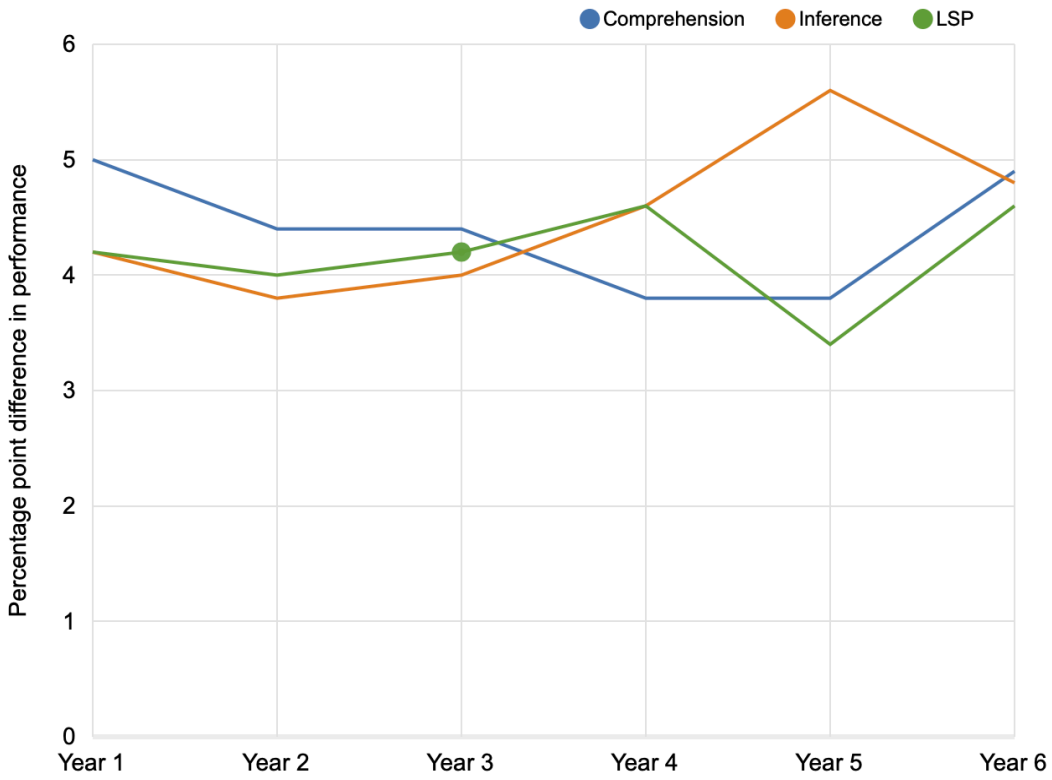
Sample sizes are in the range 27,000-272,000 tests per data point.

It is important to emphasise that the variation within each gender is far larger than the differences in mean values between the two groups. To put it another way, knowing the gender of a child on its own provides almost no indication of their likely performance in any of these tests. Nevertheless, understanding the scale and time course of those differences that do exist is a potentially important route to understanding the effectiveness and equality of the primary education system as a whole.

As well as these differences between subjects, there are also interesting variations between topics within the same subject. Figure 2 shows mean differences in test scores by reading topic. In this case there is relatively little variation between comprehension, inference and LSP (language, structure and presentation), which each show roughly constant gender gaps of 4-5 percentage points across all year groups. This is consistent with the results we presented last time using data gathered up to July 2018.

Professor Clare Wood, Nottingham University: *"For me, this flags the importance of developing resources for the early years, as children are entering school with these differences in ability, and school instruction as it stands is not effective in closing the gender gap, but is effective in ensuring that equivalent progress is made, regardless of gender of pupil. One observation is that perhaps we over-encourage non-fiction reading for boys both in the early years and at school, especially in relation to leisure reading, as there is some evidence that fiction reading is associated with reading outcomes."*

Figure 2: Reading topic performance of girls relative to boys

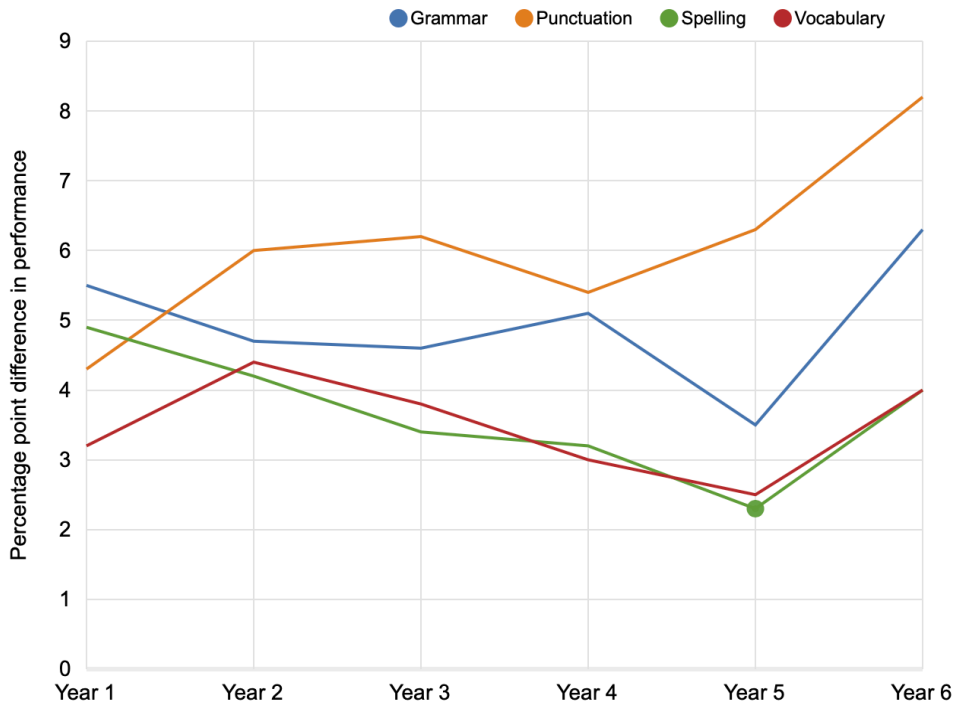


Sample sizes are in the range 130,000-272,000 tests per data point.

In contrast, Grammar, Punctuation and Spelling (GPS) topics show divergent behaviour with respect to gender, as shown in Figure 3. The gender gaps for grammar, spelling and vocabulary all tend to decline between Years 1 and 5 while that for punctuation tends to increase. All topics appear to show relative increases in the performance of girls during Year 6, though we would urge caution in interpreting this particular result as sample sizes for Year 6 are smaller than for other year groups. It will therefore be interesting to see if this effect persists in future analyses with even larger sample sizes.

Professor Clare Wood, Nottingham University: *"This data suggests that instructional approaches to grammar, spelling and vocabulary are effective at reducing gender gap, but more work is needed to improve instructional approaches for punctuation in Key Stage 1 and Key Stage 2."*

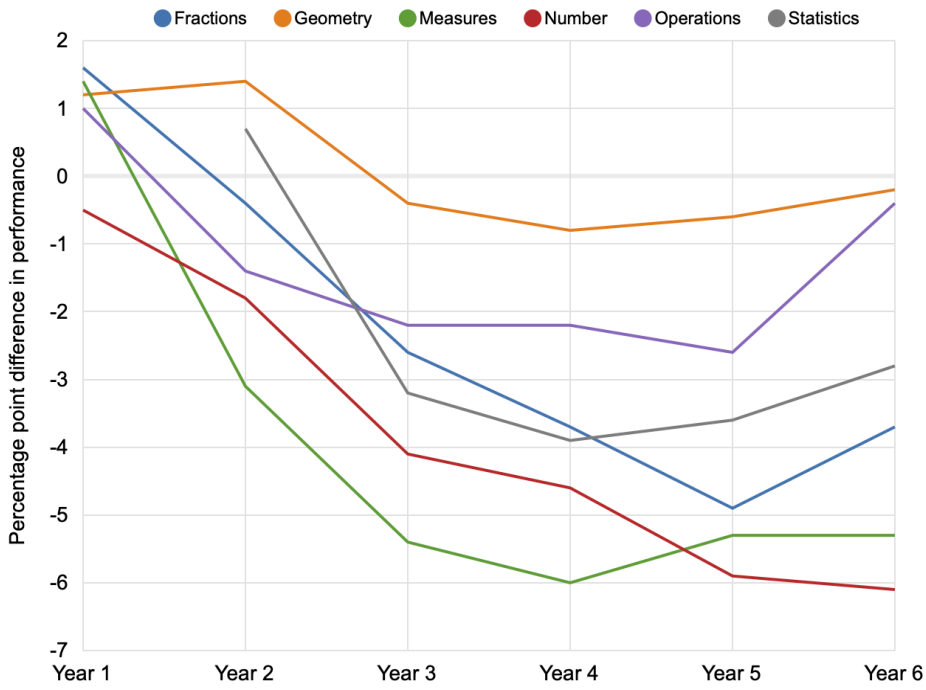
Figure 3: GPS topic performance of girls relative to boys



Sample sizes are in the range 27,000-63,000 tests per data point.

As shown in Figure 4 (and as we also saw in our previous study), the trends displayed by maths topics also vary. By the end of primary school, the gender gaps for number and measures are relatively large, those for fractions and statistics are smaller, and those for operations and geometry are virtually absent.

Figure 4: Maths topic performance of girls relative to boys



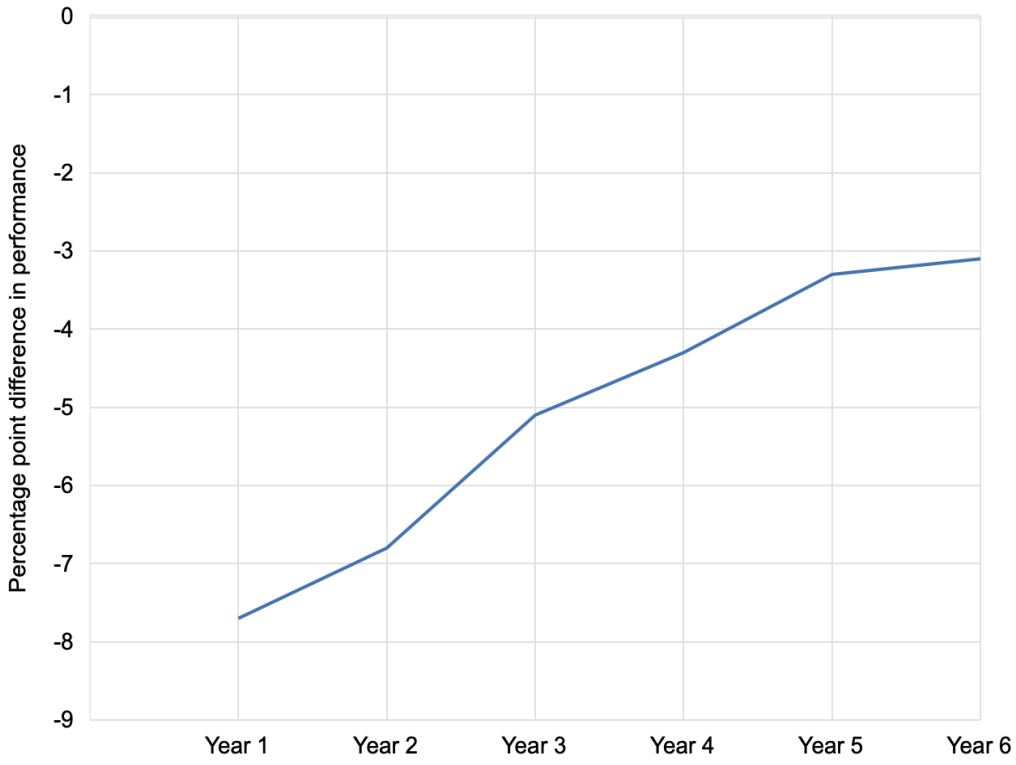
Sample sizes are in the range 123,000-261,000 tests per data point.

Grammar, Punctuation and Spelling performance

We previously saw that summer-born pupils tend to show lower average performance in maths and reading, and that this difference declines as pupils get older. Figure 5 shows the same analysis for GPS. Here, too, the gap is relatively large – almost 8 percentage points – in Year 1 and declines to about 3 percentage points by Year 6.

Professor Clare Wood, Nottingham University: *"This is interesting as the data does suggest that summer-born pupils are at a developmental disadvantage at the start of school but primary school-based approaches to instruction are being effective at supporting the children to catch up. However, arguably the rate of improvement is still too slow, and there is scope to do more here in relation to supporting their development across the curriculum."*

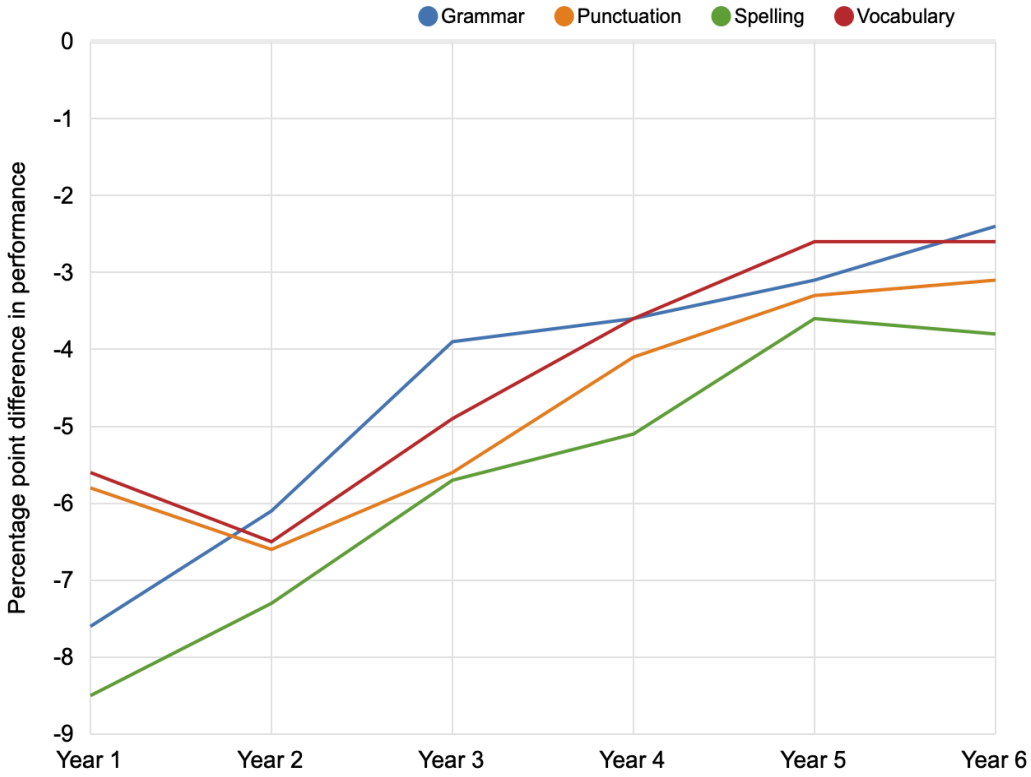
Figure 5: GPS performance of summer-born pupils relative to other pupils



Sample sizes are in the range 27,000-63,000 tests per data point.

As shown in Figure 6, the scale and trajectory of these differences are broadly consistent across different topics (grammar, punctuation, spelling and vocabulary).

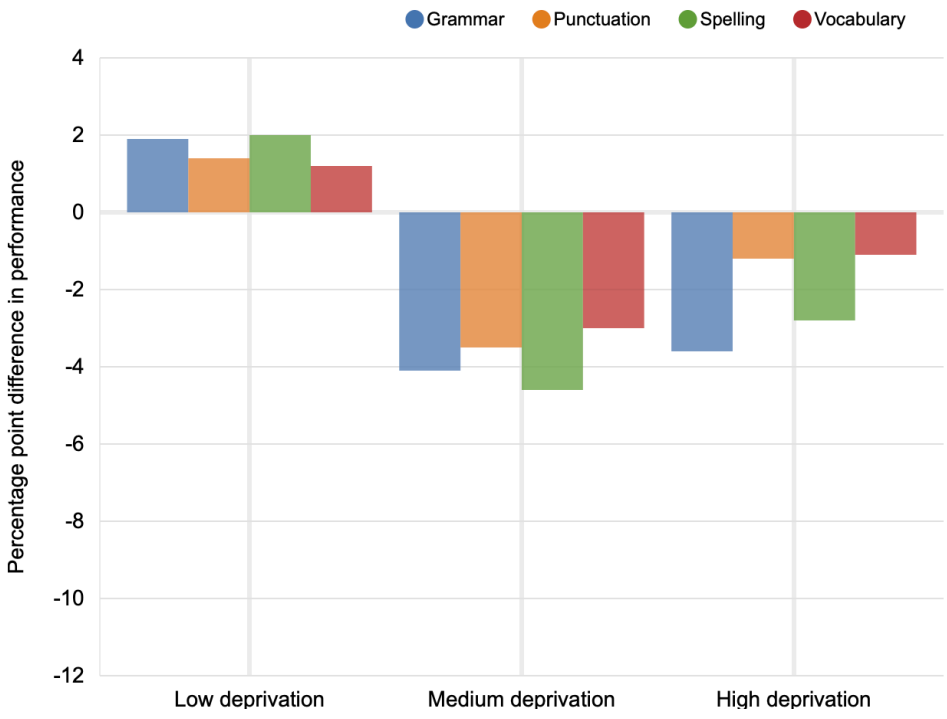
Figure 6: GPS performance of summer-born pupils relative to other pupils, by topic



Sample sizes are in the range 27,000-63,000 tests per data point.

Our 2018 study also showed clear trends in reading and maths performance depending on levels of poverty. This is also the case for GPS. Figure 7 shows that low-deprivation schools (<20% of pupils eligible for free school meals) had substantially higher average performance in all GPS topics than those with medium (>=20%) or high deprivation (>35%).

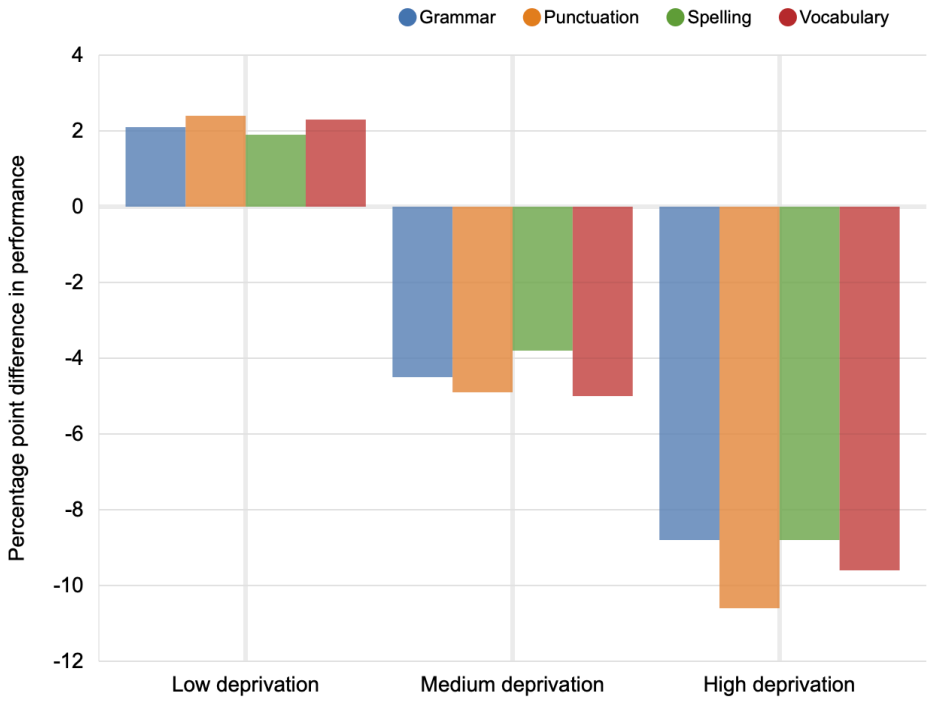
Figure 7: Relative GPS performance in Year 1 by school deprivation level



Sample sizes are in the range 4,000-32,000 tests per data point.

By Year 6 these deprivation-related differences widen, as shown in Figure 8.

Figure 8: Relative GPS performance in Year 6 by school deprivation level

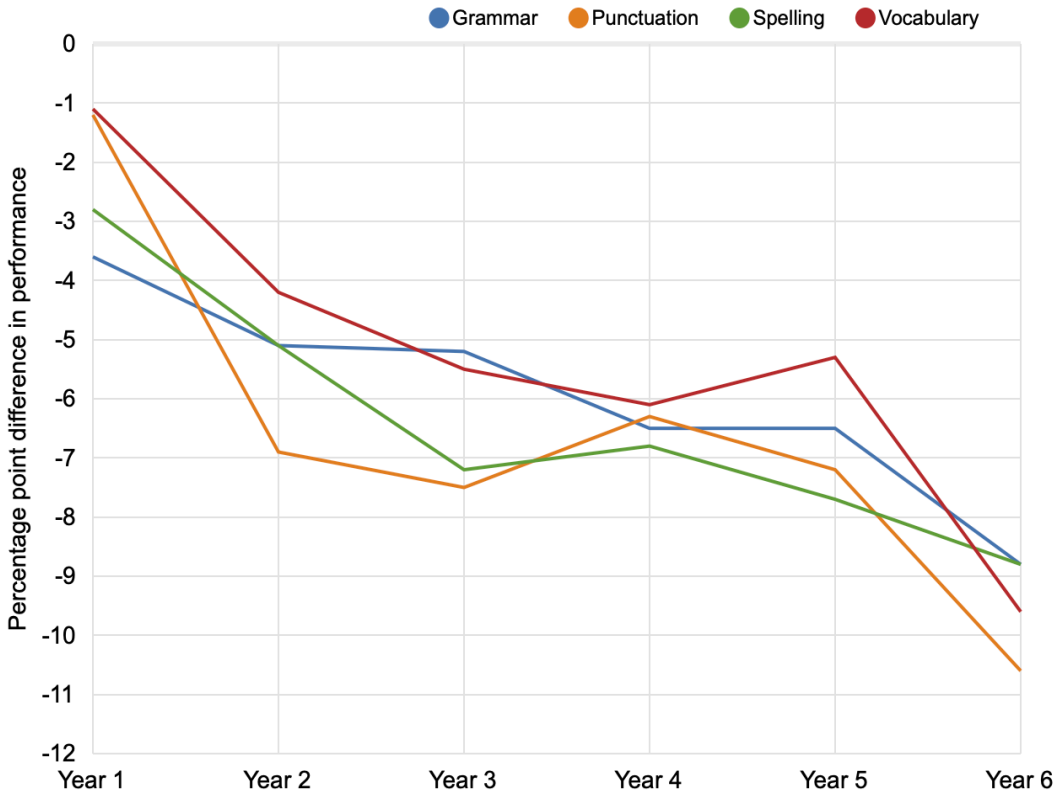


Sample sizes are in the range 2,000-20,000 tests per data point.

Focusing just on high-deprivation schools, the differences between Years 1 and 6, seen above, are part of a general trend in which the gap increases as pupils progress through their primary years. This is shown in Figure 9.

Professor Clare Wood, Nottingham University: *"This indicates we do not yet have an effective process for reversing the achievement gap for children in high-deprivation schools, and this needs to be addressed as a matter of urgency. Recent research on the impact of Pupil Premium funding suggests that it may be benefiting these pupils, but so much will depend on how this resource is deployed at school level, and the extent to which schools engage with advice on how to best support children from disadvantaged backgrounds. In relation to reading, the lack of school library provision in primary schools may be contributing to this pattern of results, as children from high need backgrounds may rely on access to such spaces and resources more heavily than other children."*

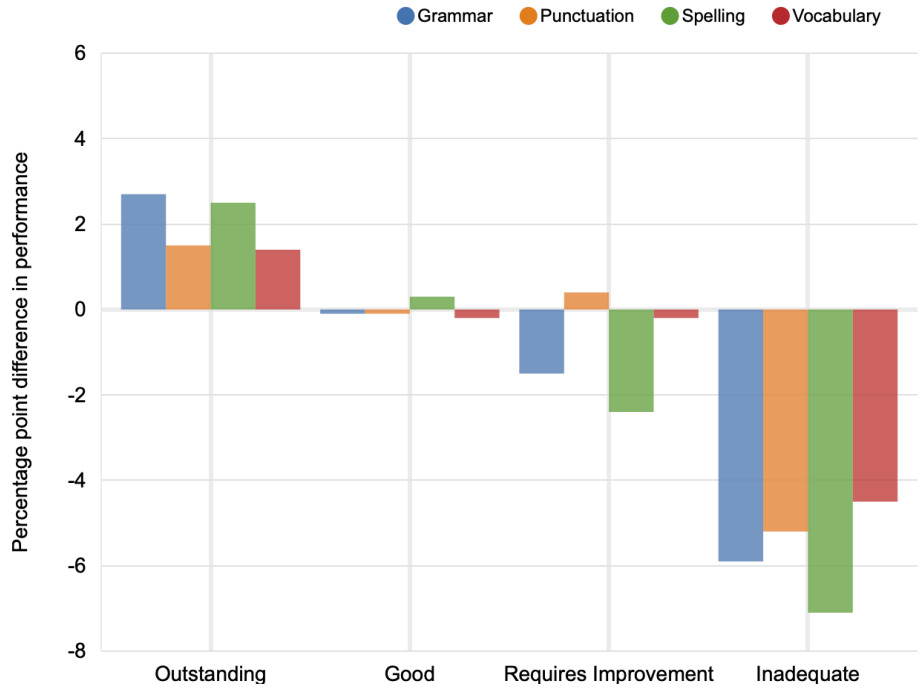
Figure 9: GPS performance in high-deprivation schools relative to all schools, by year group



Sample sizes are in the range 2,000-43,000 tests per data point.

Average GPS performance also varies by Ofsted rating, as shown in Figure 10, with 'Outstanding' schools showing the highest mean scores for Year 1 pupils and 'Inadequate' schools the lowest.

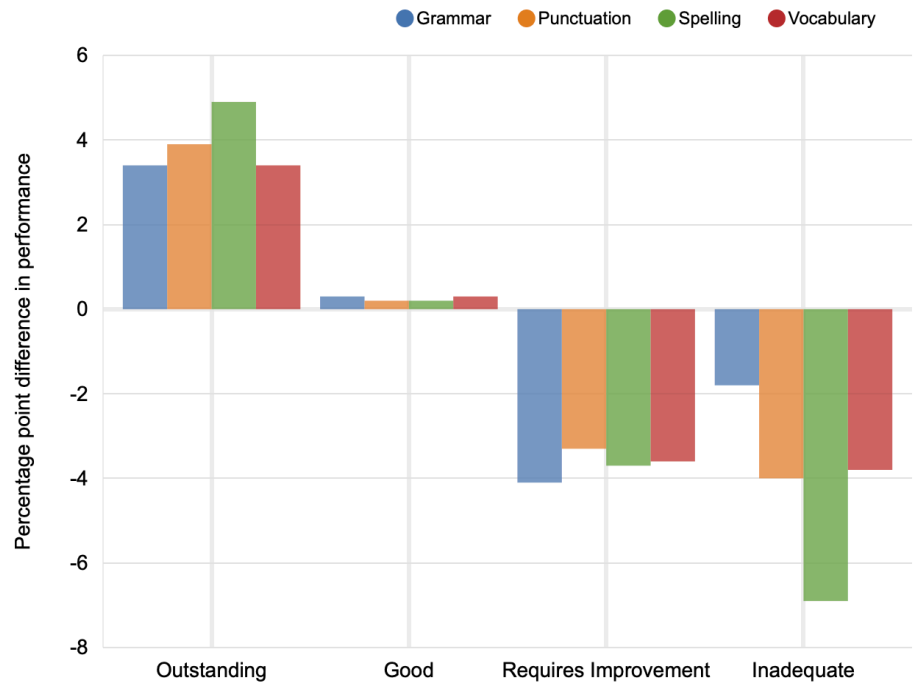
Figure 10: Relative GPS performance in Year 1 by school Ofsted rating



Sample sizes are in the range 1,000-29,000 tests per data point.

Figure 11 provides results for Year 6, by which time the differences are even more pronounced. Relative performance at schools rated 'Requires Improvement' seems to decline more than at 'Inadequate' schools, but this should be interpreted with caution as 'Inadequate' schools are relatively rare and sample sizes correspondingly small.

Figure 11: Relative GPS performance in Year 6 by school Ofsted rating



Sample sizes are in the range 1,000-18,000 tests per data point.

Variations in gender and age gaps

The data described above, along with our previous analysis, demonstrate clear national differences in the average performance of boys and girls, as well as between summer-born pupils and their classmates born at other times of the year. However, these mean values hide a great deal of variation, which raises the interesting question of whether there are any schools at which these gaps are consistently smaller (or larger) than the overall national trend.

To explore this further we looked at whether particular schools showed propensities to larger or smaller gender gaps. There were no systematic differences by common school groupings such as size, deprivation level or Ofsted rating, but this does not exclude the possibility that certain schools are nevertheless able to maintain a narrow gap. Figure 12 analyses this by showing the relationship between the gender gaps in two consecutive academic years: 2017-18 (horizontal axis) and 2018-19 (vertical axis). Each dot represents a school. The orange dots compare the same cohorts – specifically, Year 5 pupils in 2017-18 and Year 6 pupils in 2018-19. There is a positive relationship between these metrics (R2=0.53), which indicates that cohorts with small (or large) gender gaps in one academic year tend to maintain similar gaps even after the pupils have moved up into the next year group.

For comparison, the blue dots in Figure 12 show data for Year 6 pupils in both academic years, i.e. two different cohorts of pupils. Here there is no correlation (R2=0.00), which indicates that schools with small (or large) gender gaps in any particular year group do not tend to maintain this difference from one year to the next as new pupil cohorts flow through the system.

Figure 12: Reading gender gap

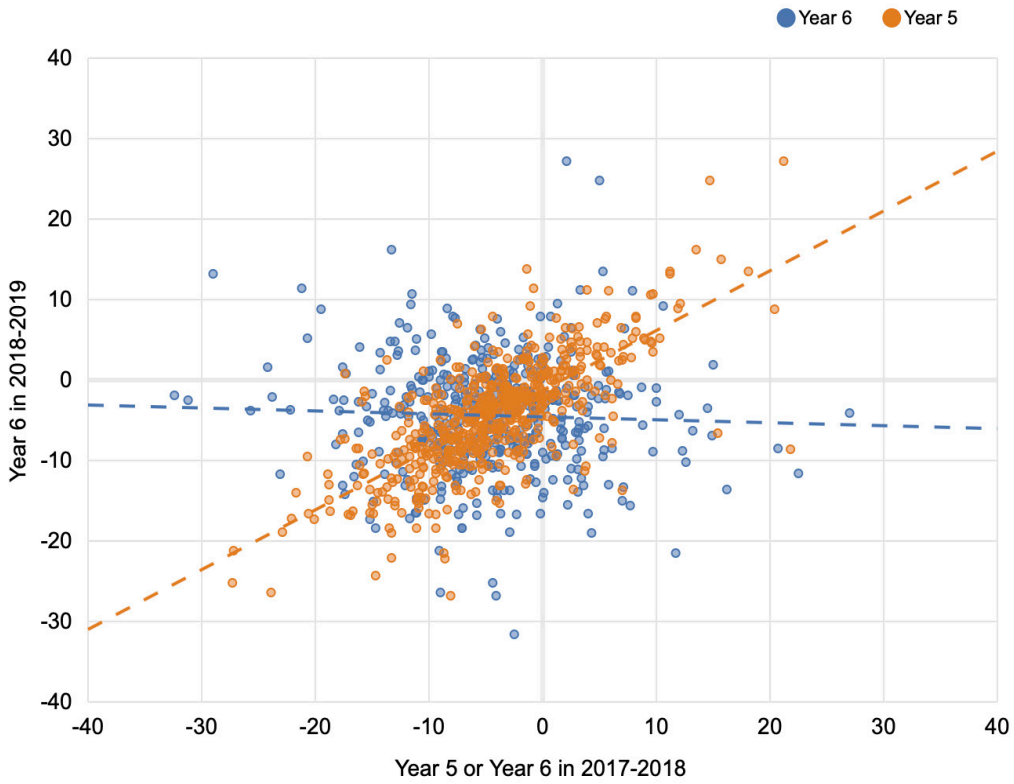
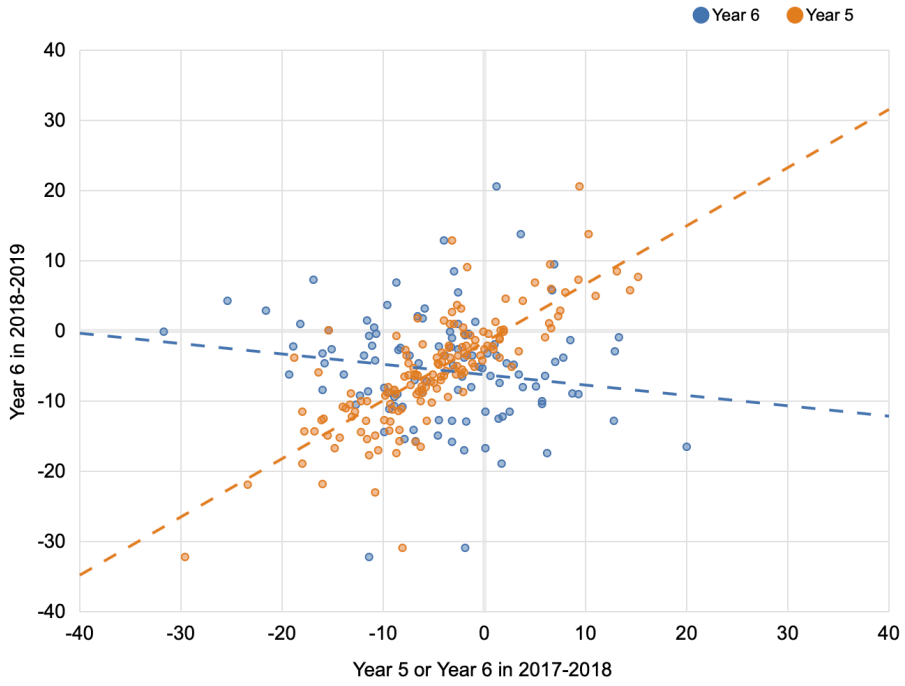


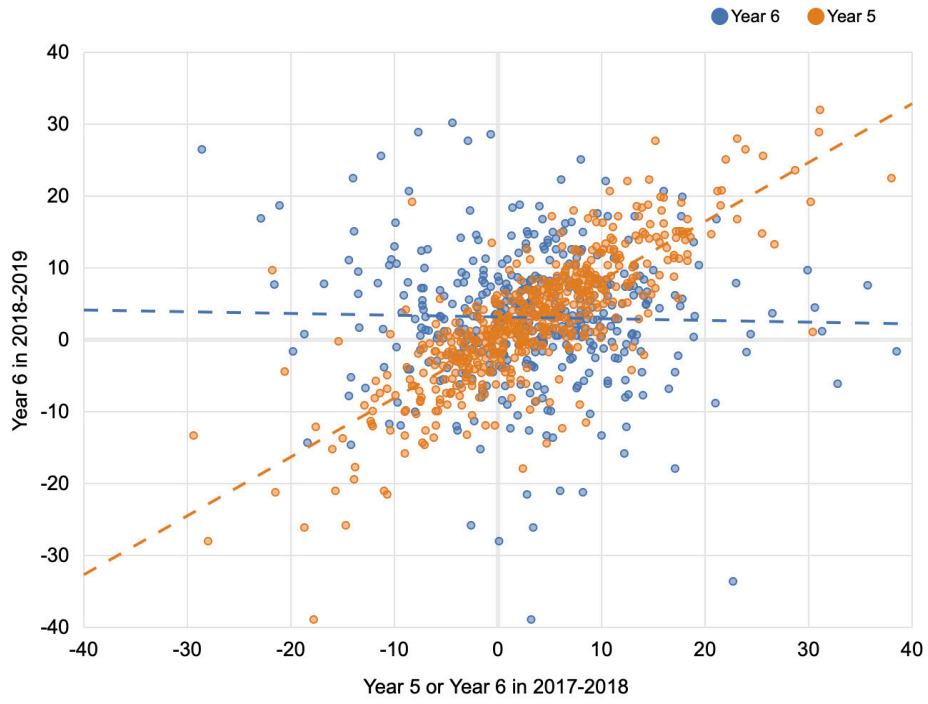
Figure 13 shows the same analysis for GPS, with similar results.

Figure 13: GPS gender gap



And the same holds true for maths, as shown in Figure 14.

Figure 14: Maths gender gap



Taken together, these results suggest that main driver of the variations in the gender gap are differences between individual pupil cohorts rather than differences between schools.

Figures 15, 16 and 17 shown similar analyses for the gap between summer-born and other pupils for reading, GPS and maths, respectively. Once again, variations in the age gap appear to be down to differences between pupil cohorts rather than schools.

Figure 15: Reading summer-born gap

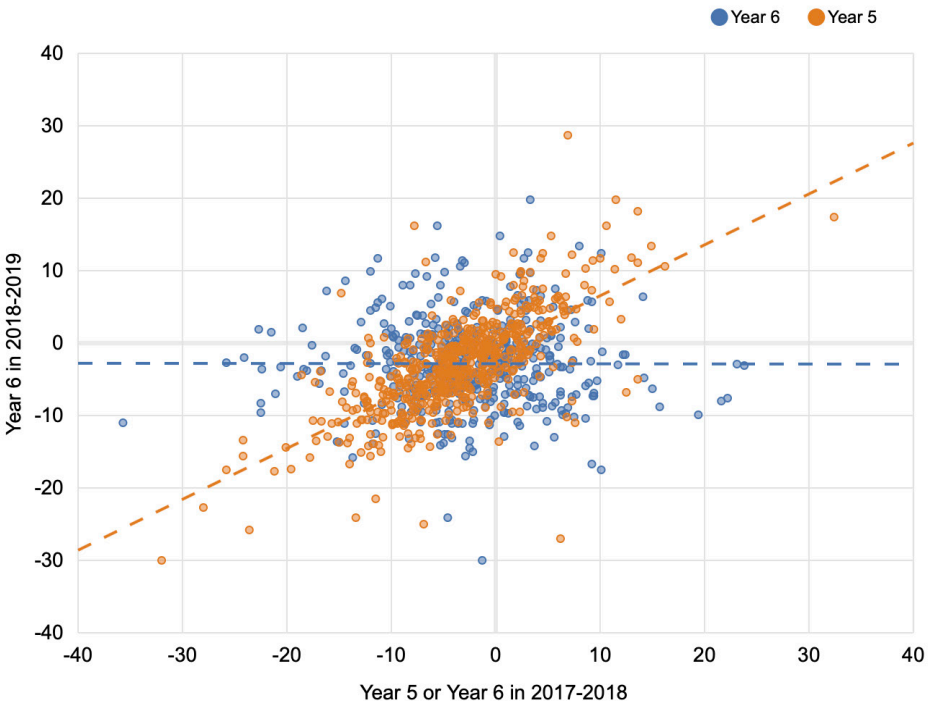


Figure 16: GPS summer-born gap

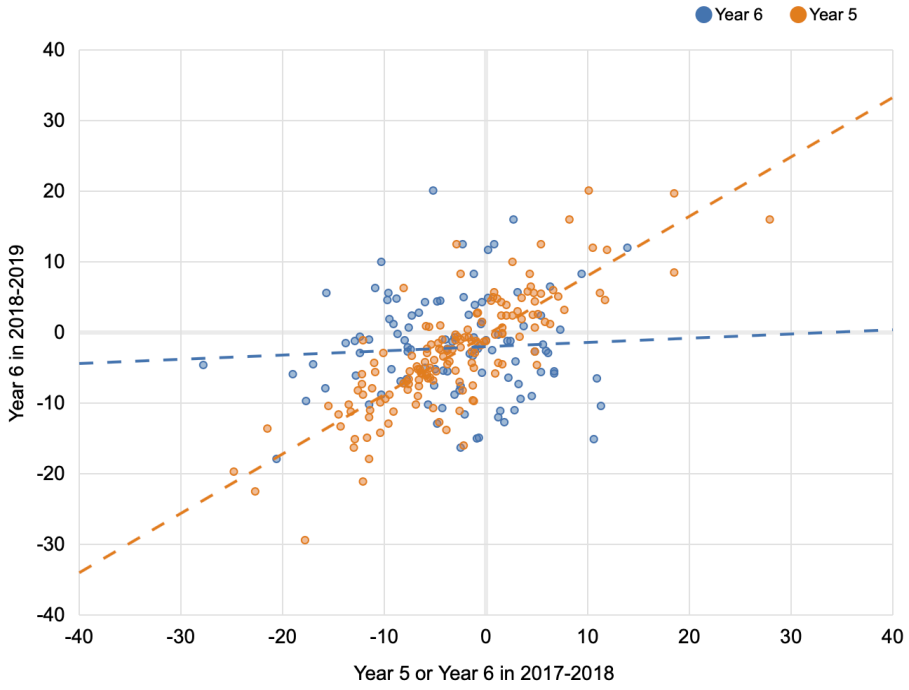
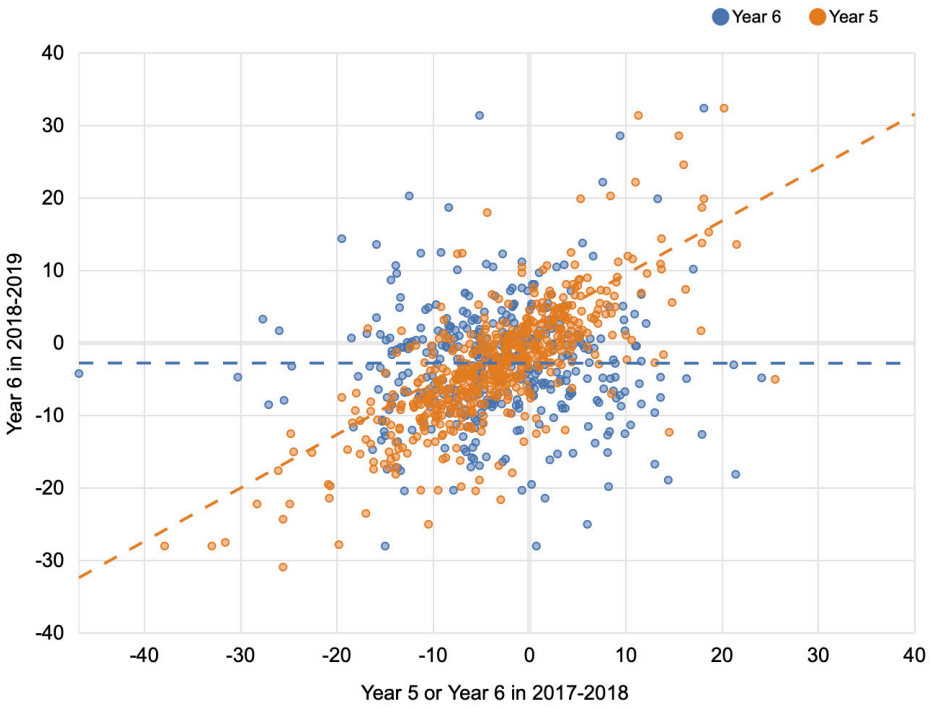


Figure 17: Maths summer-born gap



These results, while interesting, are arguably somewhat disappointing. If there had been clear examples of groups of schools with consistently smaller gender or age gaps, then it may have been possible to study them and learn how to reduce such differences. As things stand, however, these gaps appear to be a persistent aspect of the primary education system, at least in England, and there are no obvious signs of any particular subset of schools that is able to reduce them more effectively than any other.

Appendix

In order to protect the confidentiality of the institutions and individuals concerned, results have been analysed and presented in an anonymised aggregate form.

Represented schools

This analysis is limited to state primary schools in England. All regions and major school types are included, and the sample is broadly representative, albeit with some biases. For example, Yorkshire and The Humber is slightly over-represented while the North East is slightly under-represented; schools with Ofsted ratings of 'Inadequate' and 'Requires Improvement' are over-represented relative to 'Outstanding' schools; large schools are over-represented compared to small schools; and sponsor-led academies and over-represented relative to other school types. The most over-represented of these groups contains 79% more schools than would be expected if all groups were represented equally; the most under-represented contains 45% fewer schools.

Tests and scores

The data used in this report comes from standardised, termly tests: GAPS, PiRA and PUMA. The tests were taken from October 2015 to July 2019 inclusive and entered in to MARK, a free marksheet and reporting service. The termly tests are marked by teachers using a robust mark scheme, and raw scores are converted to standardised scores automatically in MARK. Only mainstream state schools in England have been included. We have analysed only results from fully completed tests, with non-zero scores, sat at the correct time of year, by a pupil within the correct age range.

The results presented are based on normalised scores, in which the overall mean score across all pupils for each test is set to zero. Any individual pupil, or group of pupils, may therefore have a score that is positive (higher than mean performance), negative (lower than mean performance) or zero.