Department for Education

Analysing family circumstances and education

Increasing our understanding of ordinary working families

April 2017

Contents

Table of figures

Summary

1.   The Government has set out its Plan for Britain, which aims to deliver the right deal for Britain abroad and a better deal for ordinary working people at home. Central to this plan is the education system. Currently, analysis of education data focuses on the educational experience of children captured within the Department's measures of disadvantage. This means we have very limited understanding of the experiences of children in families with modest incomes compared to the experiences of children in the wealthiest 10% of families. By knowing more about the experience of these families in the education system, we can ensure we have the necessary insight to better inform policy making for ordinary working families - just as we do for disadvantaged children and children with special educational needs, and to prevent discrimination based on gender and ethnicity.

2.   This technical consultation sets out the work we have done to link pupil information with parental income for all pupils; the impact of housing costs on how we measure this; our proposals for establishing and maintaining the database required to deliver this; and the privacy and data access controls that we have put in place to ensure the highest standards of privacy and ethics controls. We are consulting on the next steps in developing the database and on whether to officially refer to the group on modest incomes as ordinary working families.

3.   The approach set out is designed in relation to DfE as it relies on the current measure of disadvantage that is specific to the education system. This measure considers those who are either in receipt of free school meals, which are qualified for through a number of primarily out of work benefits, or through receipt of pupil premium, which is qualified for through being in receipt of free school meals within the last six years or being a looked after child.

4.   In undertaking this, we have worked across Government to share data, in line with current legislation and with full regard to the privacy impacts of the work, to understand family circumstances of pupils in England. The working paper sets out:

- The data matching and linking process undertaken to ensure we can link pupils to their parents and the parents' income, as well as the further work needed in this area.
- An analysis of attainment by equivalised gross income decile before housing costs.
- An analysis of the impact of adjusting for housing costs using a newly developed methodology to test whether the conclusions remain robust.
- Methodology to define income groups, including for those below median income but not disadvantaged – and consults on whether to officially refer to this group as ordinary working families. This is to provide statistics and research - no individual will be identified or labelled as being from an ordinary working family through this.

- How we can develop the database in the future to better understand the issues explored in this paper.
- The need to balance the availability for research of these data with the rights of citizens on how these data will be used and their guarantee of privacy. These data will only be used for research purposes and cannot legally be used to identify individuals. This is set out in law and the Government has no plans to change this.

5.   The work presented here shows the progress to date and indicative findings from the analysis which are starting to emerge but **caution should be taken in drawing definitive conclusions from these findings until we have completed further work.**

Testing the data linking

6.   Input received through the promised discussions as part of the Schools that Work for Everyone process made clear that a good starting point for linking pupil information with parental income would be HMRC and DWP income data. Utilising the data sharing powers under the Education and Skills Act 2008, a test database linking pupils and parents' income has been created containing pupil details from the National Pupil Database and household equivalised income (see figure 1) covering all sources of income for those on tax credits and earnings for those not in receipt of tax credits. The matching was successful, particularly for those on tax credits.

7.   While this has proved that we can successfully link parental income and pupil attainment, it is clear that data on self-employed income will be necessary to improve the accuracy of the dataset. Additionally, more insight could be drawn from analysing changes over time through developing a time-series.

8.   We have also undertaken analysis of the impact of housing costs on family income to test the robustness of our analysis. To do this we have adjusted family income by the average housing costs (rents and mortgages) for the local authority district where the family live and then equivalised the income. (Full methodology details are in Annex B.)

9.   Ideally, we would have housing cost data below local authority district level but this was not possible at this time. We will consider how to develop this method further.

10.   We would welcome views on:

- expanding the number of years for analysis to provide time-series and improve our ability to understand the links between family circumstances and education.
- developing the database to include more sources of income, for example from self-employment.
- developing and improving the methodology for adjusting for housing costs.

The analysis so far

11.   For the purposes of this document, we have undertaken analysis by income decile and also summarised for those both above and below the median who are not eligible for pupil premium. The median income for families has been set using the nearest decile to the median equivalised household income of the households with children in our database where we have valid income. This median has been used in preference to survey estimates at this stage due to the lack of some self-employed income data.

12.   For identifying those in lower income families after housing costs, we have utilised data from the Family Resources Survey over 13 years. We have taken the average family costs on rents and mortgages and inflated these for each year to 2015/16 prices. We have then deducted these from gross family income before equivalising. While this gives an adjustment for housing costs based on the area the family live in, it does not reflect the actual housing costs of each individual family. At present, we have not identified data sources which would allow a complete adjustment for housing costs.

13.   Overall, the adjustment of housing costs shows a similar picture as before housing costs. In our analysis of area of residence of pupils in families below median income who are not eligible for pupil premium there is a strong correlation between the percentage of such pupils in each local authority district both before and after housing costs. The housing costs adjustment increases the percentage of pupils in London and other inner city areas who are below median income. The trends in attainment are consistent both before and after housing costs are accounted for so while the numbers change once accounting for housing costs, the overall trends are unchanged.

Ordinary working families

14.   There is no official definition of an ordinary working family. However, the Government frequently refers to a group who are not entitled to pupil premium but earning modest incomes as ordinary working families. This document has set out an approach to looking at this group through better understanding of those families who are below median household equivalised income but who are not eligible for pupil premium. This approach shows there are around a third of children in these families.

15.   Unlike eligibility for pupil premium or free school meals which derives directly from Government policy, this is necessarily an analytical approach intended to provide a basis for a clearer analysis of educational outcomes for ordinary working families. We would welcome your views on the statistical analysis and methodology outlined here for looking at this group, and how to refer to them in DfE publications.

Privacy and ethics

16.   A privacy notice has been published alongside this document for the work carried out so far and full privacy impact assessments will be undertaken for the proposed expanded database as well.

17.   The Education and Skill Act 2008 restricts use of this database for research purposes only and there is no proposal to change this.

18.   On 16 February, I set out that our system for data access was under review and I will be reporting on this in early May. While it is our intention to make the database available for research purposes once it is of sufficient quality, we will need to make sure access is in line with highest modern standards of security and privacy protection. We will be working with the Office for National Statistics and Administrative Data Research Network to develop options.

Next steps

19.   We would welcome views by 30th June 2017 on

- Expanding the number of years for analysis to provide time-series and improve our ability to understand the links between family circumstances and education.
- Developing the database to include more sources of income, for example from self-employment.
- How to develop and improve the methodology for adjusting for housing costs.
- The statistical analysis and methodology outlined here for looking at ordinary working families, and how to refer to them in DfE publications.

20.   We will be discussing the method further with users and considering views on the methods set out here to consider the way forward for our analytical publications. Views can be provided by e-mailing hop.statistics@education.gov.uk by the 30th June 2017.

Acknowledgements

21.   This work shows the Government Statistical Service coming together to legally utilise the data available across Government to improve the statistical evidence base available for policy and research. It could not have been achieved without analysts from across DWP, HMRC and ONS working with DfE to deliver this analysis.

Iain Bell

Head of Profession for Statistics, DfE

Data linking

22.   Current measures of disadvantage in education rely solely upon receipt of free schools meals or pupil premium eligibility. As set out in the Schools that Work for Everyone consultation, the Government wants to identify pupils that are not captured by these measures but are whose families are nevertheless just managing to get by. To achieve this, the DfE needs a more nuanced view on household incomes to identify those with incomes above the current thresholds but who are still facing socio-economic barriers to their education.

23.   We committed to discuss our proposals with the education sector and academics to help our thinking in developing these measures. During these discussions we received three messages consistently:

- Data from HMRC and DWP on income would be an obvious starting point.
- We need to think about housing costs.
- Reassurance that these data will not be used to identify individuals.

24.   To address this, DfE, DWP and HMRC have undertaken a feasibility study to determine if it is possible to link personal information relating to children currently in education with employment and benefit claims information about their parents/guardians. The purpose is to research and identify how educational attainment varies with income, and the challenges faced by pupils from lower income families that are not currently identified as disadvantaged.

25.   This data linking has been undertaken utilising the data sharing powers under the Education and Skills Act 2008, as amended by the Small Business, Enterprise and Employment Act 2015. The legal aspects, and measures taken to ensure privacy is respected and protected, are explored further in the final section.

26.   The following information was linked:

- personal characteristics of the pupils including: age, sex, ethnicity, primary language, location, school level information, attainment and progress measures;
- tax credit information via Child Benefit claimed by parents/guardians;
- employee tax data on parent/guardian income;
- Housing Benefit; and
- other benefits.

27.   We are only linking the information relating to pupils who were in the state-funded education system as of the academic year 2015/16.

28.   As this process has been experimental, with a purpose of determining feasibility of such a dataset, the quality of the match has been investigated. While broadly successful, there are some important issues with the robustness of the output which we will be

seeking to overcome as the data is developed further. The successes and challenges are detailed in the following sections.

Methodology

29.   The methodology used to produce the dataset is set out in the steps below and illustrated in figure 1:

- DfE isolated the identifying characteristics for pupils in state-funded education from the Spring Census 2015/16, stored in the National Pupil Database (NPD).
- These records were securely transferred to analysts in DWP for the matching process.
- DWP used the identifying information, such as name, sex, address and date of birth to match Child Benefit records, which include information about the child in question and their parents/guardians. This allows for a link to be made between a child and their parents/guardians. The same process is performed with housing benefit record where a Child Benefit link is not possible.
  - These records provide National Insurance numbers for the primary claimant and any partner they have, allowing for matching to a broad range of tax and benefit records.
  - Of the 7.6 million records for pupils on the NPD, we are able to match 97% (7.4 million) to Child Benefit records.
- Following identification of the parents/guardians, DWP combine these with employee income tax records provided by HMRC. These are then used to construct a gross household income measure.
  - HMRC record gross household income for the purpose of administering tax credits. Using this value is the preferred way of determining household income as this covers a wide range of income streams. 60% of NPD records have been successfully matched to tax credit records, which is in line with the proportion of the population in receipt of tax credits. This equates to just under 4.6 million pupils, of which 2.7 million are of primary age (reception to year 6, 60%) and 1.9 million are of secondary age (year 7 to year 13, 40%).
  - Employee income tax records only cover income relevant to the collection of tax. As such, other sources (most notably self-employment income) are not recorded, resulting in underestimates for household income in some cases. This is used where tax credits data is unavailable, and covers a further 23%.

10

- While self-employment income records exist technical and timing constraints have prevented their inclusion at this time. These will be included in future iterations.
- DWP calculate equivalised household incomes, taking into account the benefits received by each adult, and then adjust the income based on the total number of people in the household – this process is described in Annex A. DWP then place these equivalised household incomes into the national deciles of household income[1], preventing precise incomes from being known by those analysing the data to further protect privacy. The data was then stripped of all other identifying information and securely transferred back to DfE for analysis.
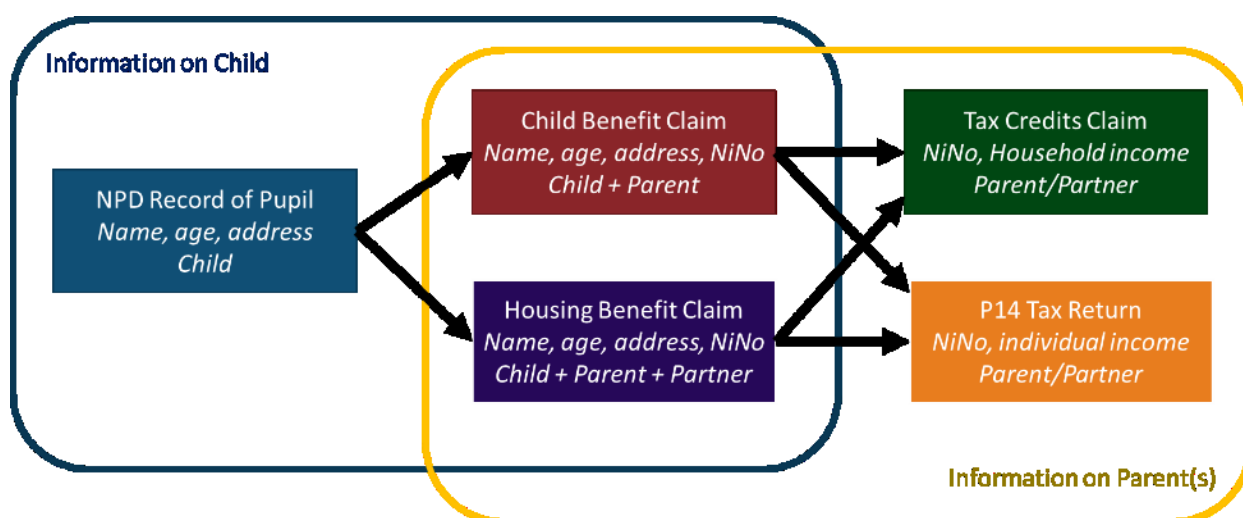
Figure 1: Diagram of data matching to construct household income

30. The process of analysing the newly created dataset is discussed in the next section.

Robustness of income information

31. While a high match rate has been achieved, and there is no evidence of substantial issues with the matching methodology or process, analysis of the data broken down in this manner suggests that there are issues arising from the two ways that we derive household incomes – via tax credit information or via employee tax information.

32. The median household income for those with children is reported in the Living Cost and Food Survey as £24,900, which closely aligns with the fourth decile of all households' income, at £25,400. 70% of pupils within our dataset come from families with

---

[1] Office for National Statistics (ONS): Gross household income by income decile group, UK, 2014

income below this level, suggesting that the work to complete the incomes from non-earnings sources will be important in improving the accuracy of the matching.

33.   This issue likely derives from multiple sources:

- As discussed further below, in some cases this data does not capture all of a household's family members. Additional adults beyond the primary adults are not included.
- Where gross household income is determined from employee tax information, we are missing income from self-employment, creating reduced household incomes. While some cleaning of the data has been done to account for this, cases such as those where one parent is self-employed and their partner has low income employment cannot presently be fully accounted for. This means that some caution should be undertaken when looking at results for those in the lowest income decile who are not classified as disadvantaged.

34.   For this reason, we have set the median income throughout this document as the **median household income measurable in our dataset**, as being the most accurate representation of the group from available data sources. This means as more data become available we can refine the median estimate and ensure as consistent approach as possible. We expect as the additional data is added, the median of the dataset will align more closely with other estimates of equivalised median household income.

35.   **The data and threshold used to define households as below the median income in this analysis should therefore be treated as provisional. Caution should be taken in drawing definitive conclusions from these findings until we have completed further work.**

36.   These limitations impact on the characteristic and attainment figures derived from the data so far to some extent. Figures 2 and 3 show Attainment 8 and key stage 2 Reading scores for each income group, using just tax credits or just employee tax information to derive the income groups. As can be seen, when employee tax information is used to derive household incomes, attainment is consistently higher. However, the scale of the difference, at a national level, does not affect the trends seen as discussed later in this document. So while this means some caution should be exercised in using the analysis, the trends and patterns appear to be robust.
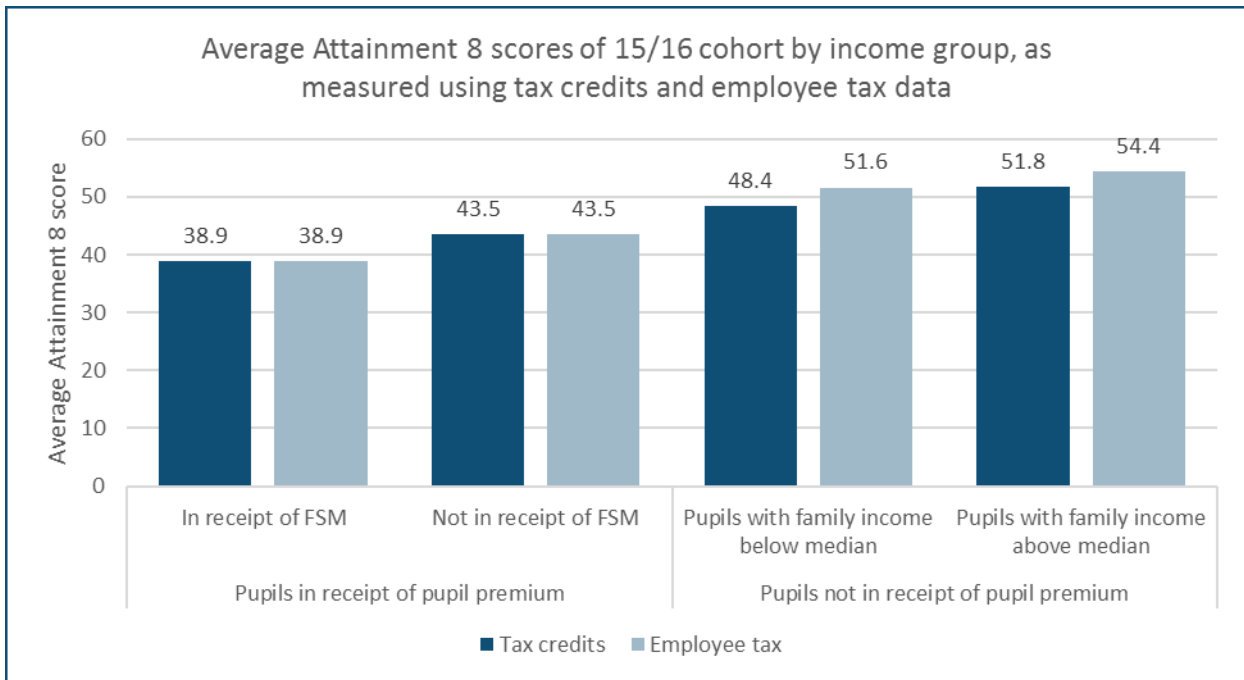
Figure 2: Attainment 8 scores by income group as derived from tax credits or employee tax information
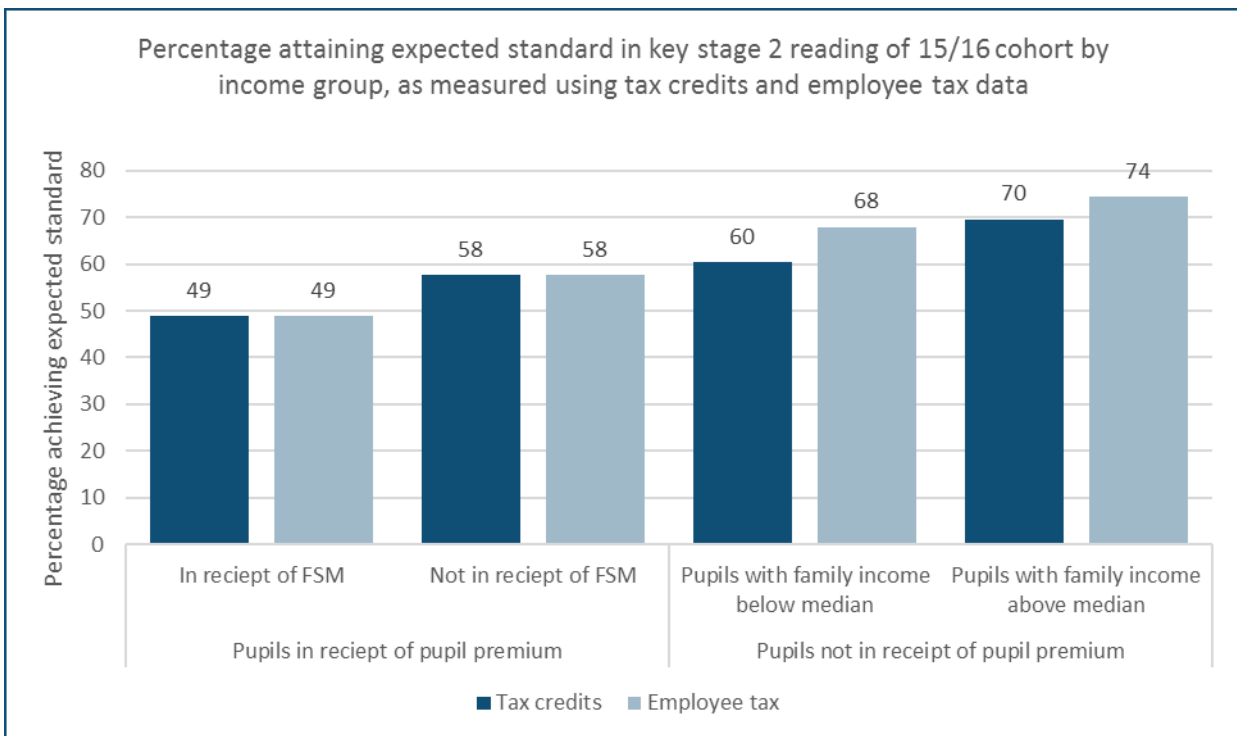
Figure 3: Percentage achieving expected standard in key stage 2 reading by income group as derived from tax credits or employee tax information

37.   Particular caution should be taken when considering small groups, such as an individual school's pupils. At this stage we have not done so, as statistics produced on smaller populations will be more sensitive to errors introduced through these limitations.

38.   It should be noted some of the difference between the results when deriving income through each method are to be expected, and does not wholly represent

13

methodological issues. Where employee tax data is used, this covers the entire population, excluding the small percentage we were unable to match. However, where tax credits data is used, only those in receipt of tax credits are considered. It is to be expected that, for families in the same income band, those families in receipt of tax credits have genuinely different circumstances to those not in receipt of them.

39.   To mitigate this problem, an assumption has been made that a family should be considered within the above median income group where they:

- do not have any recorded employee tax income;
- do not receive any benefits income; and
- do not receive tax credits.

40.   This assumption has been made on the basis that they are very likely to be from a family with above median income if they do not receive tax credits or any benefit income, and because their attainment and progress characteristics better align with that group. This covers 832,000 pupils.

41.   In future, this will be addressed by linking additional income information to build a more complete picture of household income as the data is developed. The effect this has on the above discrepancy will be closely examined.

Adjusting for housing costs

42.   There is no dataset that exists which has every families' housing costs. So, we have developed a methodology in partnership with DWP and ONS to get as local an indicator of housing costs possible and adjust family incomes by this. The method is as follows:

- We have taken 13 years of data on rental and mortgage costs from the Family Resources Survey and inflated each year's data to 2015/16 prices to align with the analysis year. This amount of data was used in order to enable more local analysis.

- We have then created an average housing cost for each local authority district in England.

- We have then subtracted this from each family's income. We then equivalised the family income on the after housing costs basis.

43.   This approach has some weaknesses – in that it does not obviously adjust for every individual's experience and in most local authority districts there will be variation of housing costs. However, it clearly will adjust for the large differences that exist between local authority districts and help ensure the analysis is robust to these.

44.   Further details on this methodology are set out in Annex B.

Other known methodological limitations and data issues

45.   In addition to the above issue, there are a number of other smaller issues which are known, and which should be considered when interpreting the analysis. These are discussed below, including their likely effect and how they may be addressed in future.

The data only identifies up to two parents or guardians and young people under 16 or aged 16 to 20 and still in education or training	Due to the way data are recorded for the payment of benefits and collection of taxes, only these household members are referenced. Adult children or grandparents, for example, which may be part of the household would not be identified. Depending on specific circumstances, this could lead to either over-estimation or under-estimation of equivalised household income. Further work, comparing results to survey data, is needed to establish the impact of this issue but early estimates suggest it is small. This is a result of how the administrative data used is recorded, and it is unlikely significant improvements can be achieved with further development.
Pupils and family characteristics are taken as a snapshot, and may not account for changing circumstances.	Characteristics of pupils are taken from the Spring 15/16 School Census, while characteristics of their broader household rely on circumstances being recorded as part of tax and benefit records. While many characteristics are unlikely to regularly change, some (such as family composition, home address or school) may change within the year. These may, in turn, be important for educational attainment or financial situations. Future developments will seek to bring in additional years of data, (i.e. more cohorts of pupils) which will help track long term changes in family structures. However, it will not be possible to fully track changes to circumstances in more dynamic families, as these are not recorded through the administrative data sources.
The matching process uses fuzzy matching, and creates some small anomalies	The matching process uses combinations of names, date of birth, addresses and similar factors to attempt to match to Child Benefit records, though a process known as 'fuzzy matching'. Not all factors are able to

| | be matched in every case, and multiple matches may be found of varying qualities. The data goes through a complex cleaning process to isolate the best matches but there may still be incorrect results – the scale of this is difficult to estimate but analysis suggests it is less than 1%.

The cause of this has not yet been traced but likely arises from complex cases such as twins, who share many of their identifying characteristics (e.g. address, surname, date of birth).

Further investigation of the matching process and refinement to it will seek to alleviate these issues |
|---|---|
| **There are 832,000 pupils in families who do not receive Tax credits or other benefits, and are recorded as having no income. Following investigation, they are being treated as having above median gross household income.** | The families appear to receive no income of any kind, which is unlikely at this scale. However, the lack of a tax credit record means that we are unable to account of a number of income streams, such as money earned via being self-employed. As such, we believe this group primarily consists of people earning sufficient money through other streams to be above the tax credits threshold, and as such fit with our more affluent pupils. This is further supported by other characteristic information of this group, which mirrors pupils from more affluent families. |

Future developments

46.   As discussed above, there are a number of areas that will be targeted for future developments of the dataset. These have been derived following consultation with expert groups during development and the gaps that have become apparent during our analysis. The main areas are set out below:

Further income information	As discussed extensively above, it has become apparent that the limitations on the income information included at this stage have caused substantial gaps in the data. We will work with HMRC and DWP to expand this, and develop a more reliable metric for household income.
Matching across multiple years	For this feasibility study, a single year of data was linked. Expanding this to include further years, and particularly being able to link those years, would allow for explorations of not only how circumstances and outcomes have changed over time for income groups, but also an exploration of the journeys of individual families. However, these data will not be used to identify individual children or families. This may allow for better identification of those who are at risk of sliding into disadvantage or understanding of factors which allow for pupils to perform well despite their socio-economic disadvantages.
Linking to further education data sources	Inclusion of the Independent Learner Records for Further Education, Higher Education Statistics Agency data and the Longitudinal Education Outcomes dataset would substantially enrich the data source, as it would develop a complete view for those in education beyond the age of 16.

Analysis of results by family equivalised income decile

Analysis of findings before housing costs accounted for

47.   In order to help understand the circumstances of all children, we have analysed the dataset by income decile after taking out children who are disadvantaged (that is pupils who have been in receipt of free school meals at any stage in the last six years, or have been looked after by their local authority or adopted from care). It is well known that family size is higher in households with lower incomes, and the process of equivalisation further exaggerates this. As such, the distribution of the pupil population is not surprising. However, the high number and percentage in the lowest income decile will also partly reflect families where they are not on tax credits and we do not have full self-employed income (as discussed at paragraph 29).
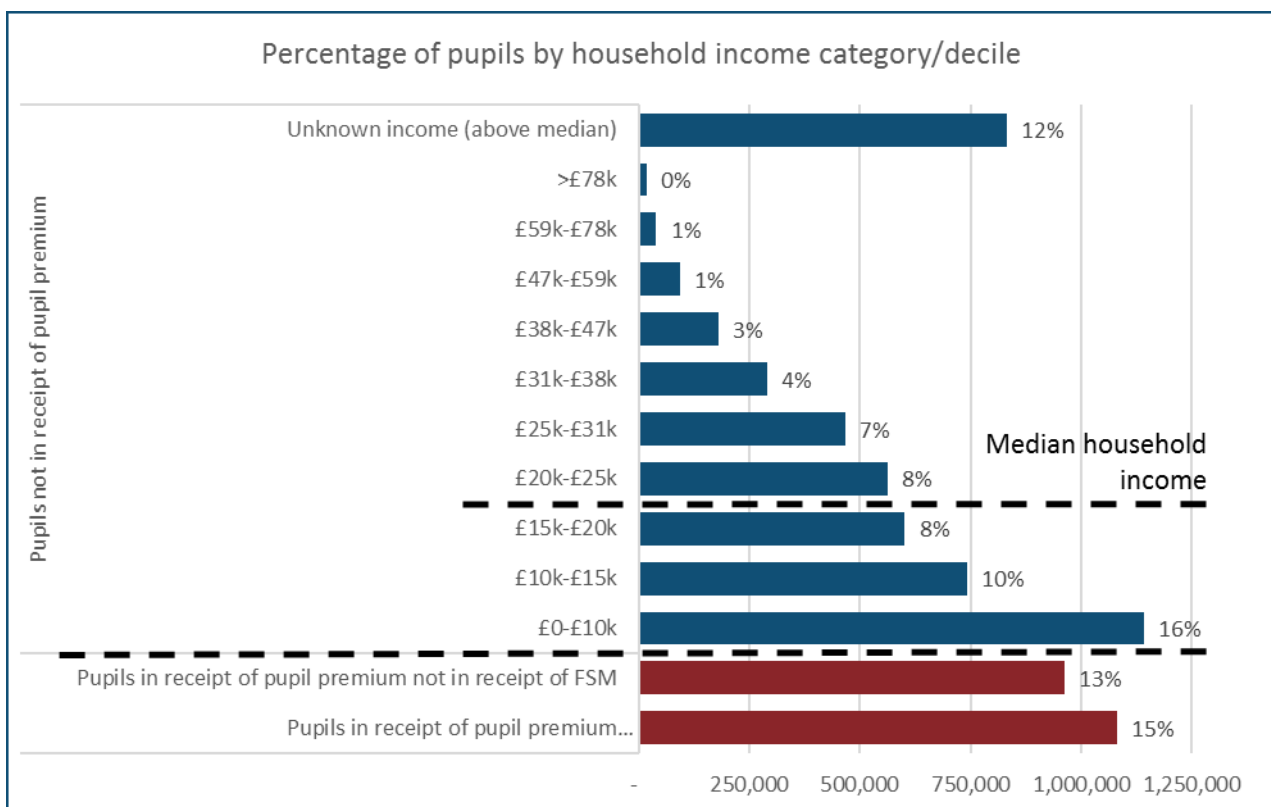

Figure 4: Distribution of pupils by household income

Those in the category 'other above median, income unknown' indicates the approximately 832,000 pupils who have characteristics consistent with being in households with incomes above the median, but for whom our data does not record any income. This is discussed further in the methodology section.

Key stage 2 results

48.   Figure 5 below shows the percentage of pupils achieving the expected standard in (all of) reading, writing and mathematics at the end of key stage 2 in 2016. This shows that attainment at age 11 rises as household equivalised income rises. The exception is between the lowest decile and the second lowest where the income grouping may be affected by the lack of information on all income sources.

49.   Nationally, 53% of pupils in state-funded schools achieved this standard. All income groups not currently classed as disadvantaged were above the national average.
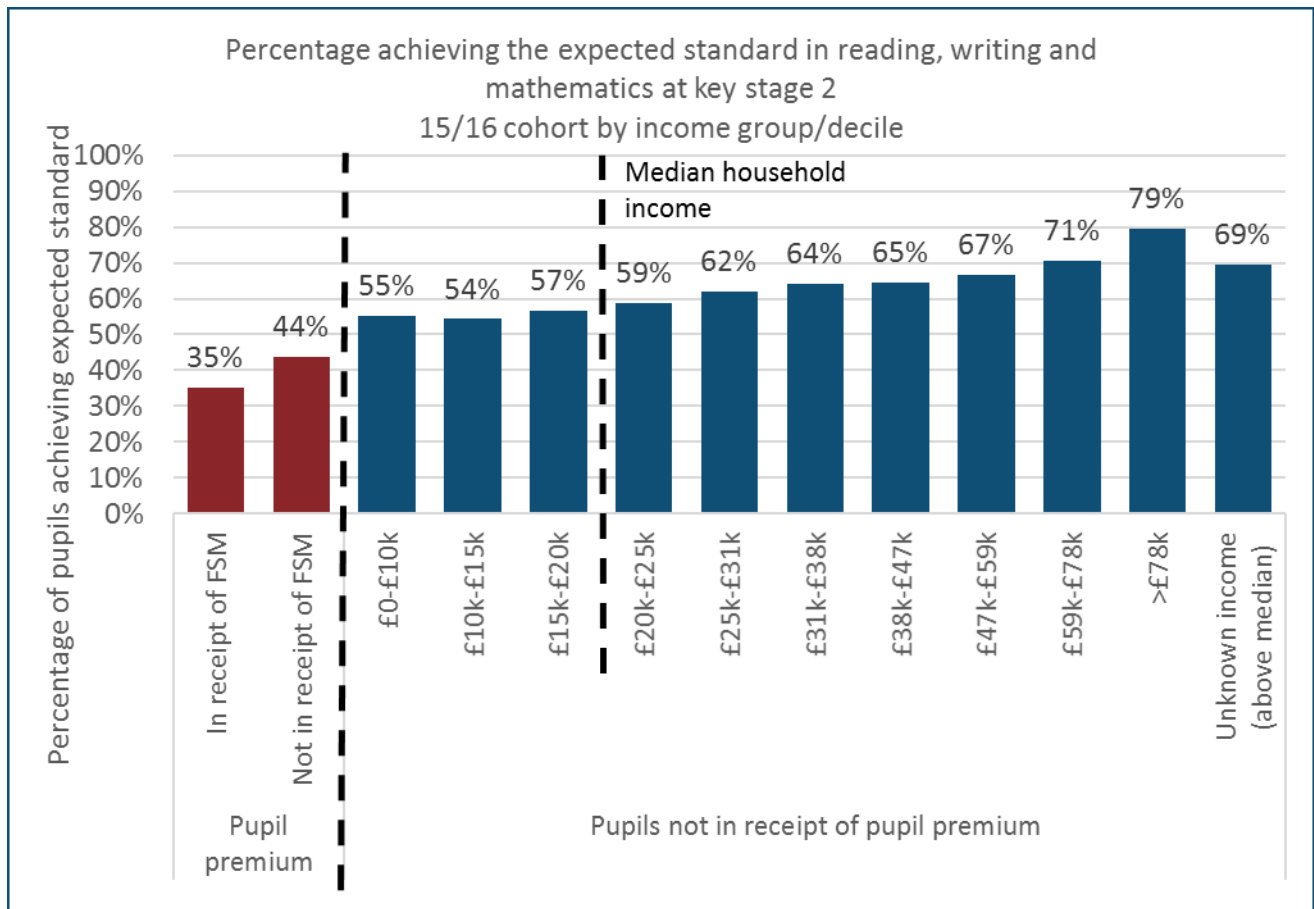

Figure 5: Key stage 2 pupils achieving expected standard by income

50.   Figure 6 below shows progress made to the end of key stage 2 in 2016, which is measured using a value added method that compares pupils to others with similar levels of attainment at the end of key stage 1.
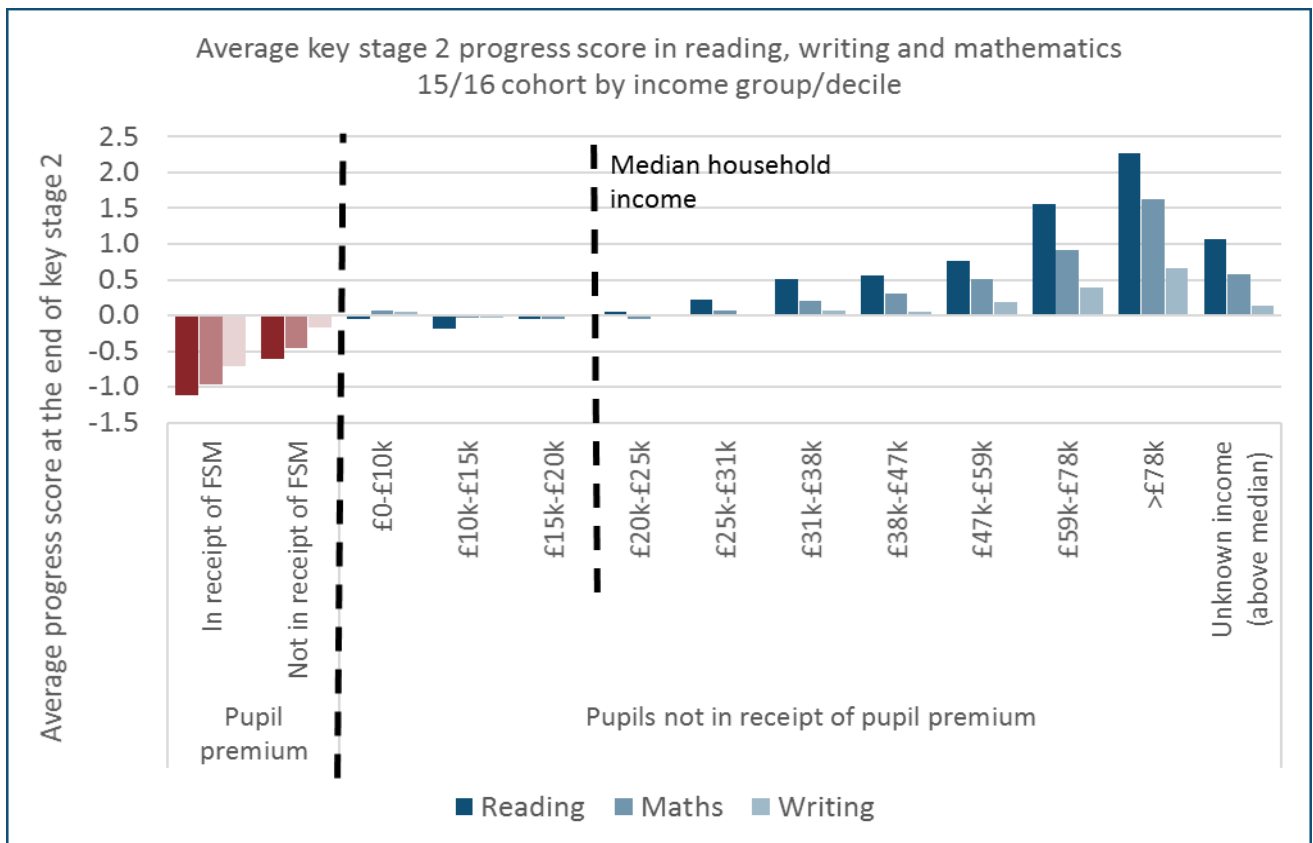
Figure 6: Key stage 2 progress by income

51.   All three subjects show the same pattern, with negative scores for disadvantaged pupils indicating that they make less progress than the national average, and progressively more positive values with increasing income showing that higher progress made relative to their peers. The lower income groups are similar to the national average in each of reading, writing and maths with similar progress made by non-disadvantaged groups with income up to £31,000.

52.   There is much more variation by household income in the progress made at primary for reading than for writing. It is unclear if this reflects a real difference in their relationship with income, or is affected by differences in assessments.

Key stage 4 results

53.   Figures 7, 8 and 9 show the key stage 4 results for three attainment measures: Attainment 8 (achievement across a suite of 8 GCSE and other subjects); entry to and achievement of good passes at A*-C in the English Baccalaureate (approved English, mathematics, sciences, languages and history or geography GCSEs); and achievement of 5+ A*-C grades including English and mathematics GCSEs.

54.   The patterns are very similar across this basket of measures and show the same patterns as at key stage 2. Children eligible for pupil premium have lower attainment than all other families and are below the national average. With the exception of the same

caveat on the two lowest deciles, the analysis again shows that attainment and progress increase with family income.

55.   There are bigger increases seen towards the highest income groups, although there are smaller numbers of children in these groups. This is particularly evident for the English Baccalaureate.
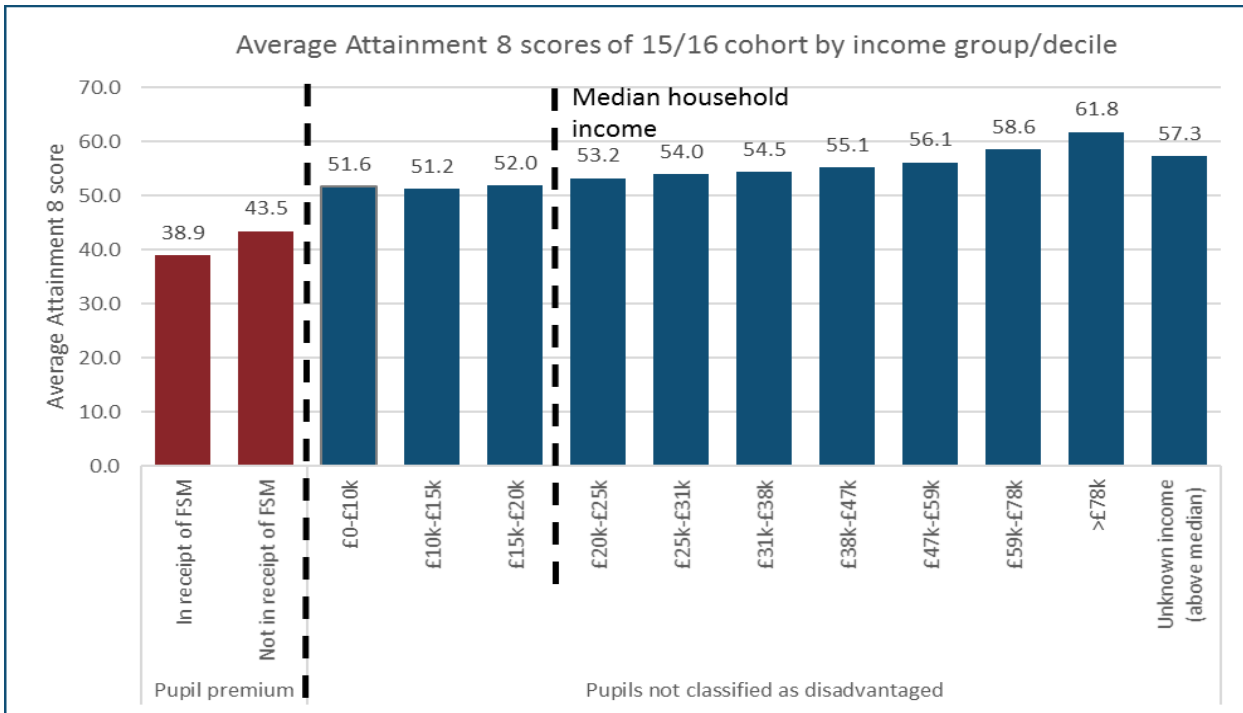

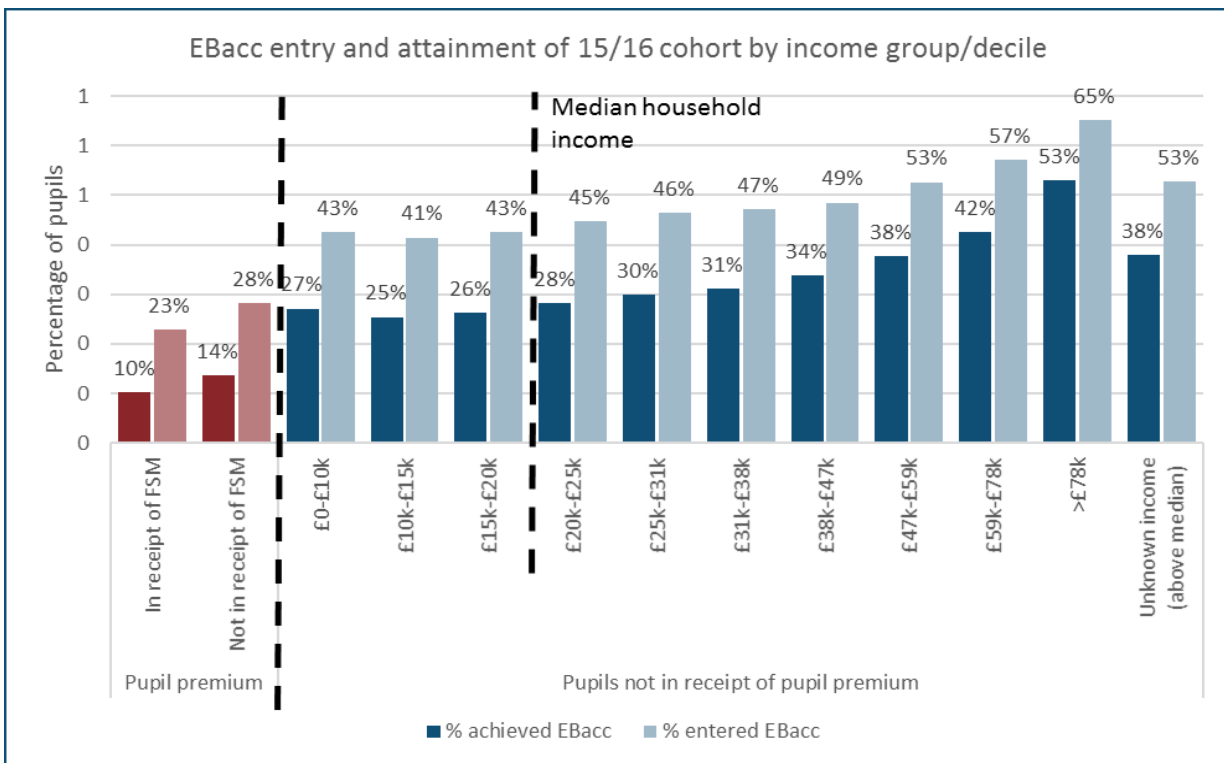**Figure 7: Attainment 8 scores by income**

Figure 8: EBacc entry and attainment by income
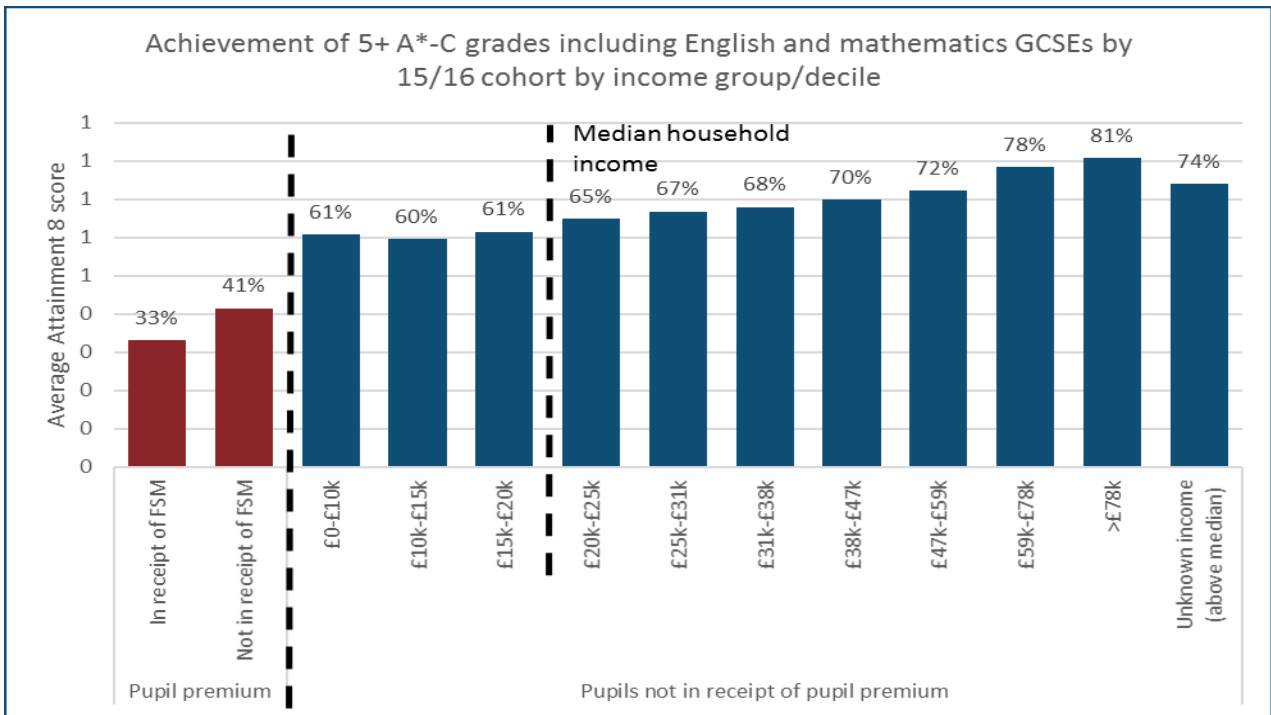
Figure 9: Achievement of 5+ A*-C grades including English and mathematics GCSEs by income

56.   Progress 8 is a value added measure which compares Attainment 8 scores of pupils with similar prior attainment at the end of primary school. Figure 10 below shows that, as at primary, pupils from disadvantaged groups have negative scores on average, showing below average progress. Average Progress 8 scores increase with higher household income, and again, show larger increases for the very highest income groups.

Figure 10: Progress 8 scores by income

22

Geography and income
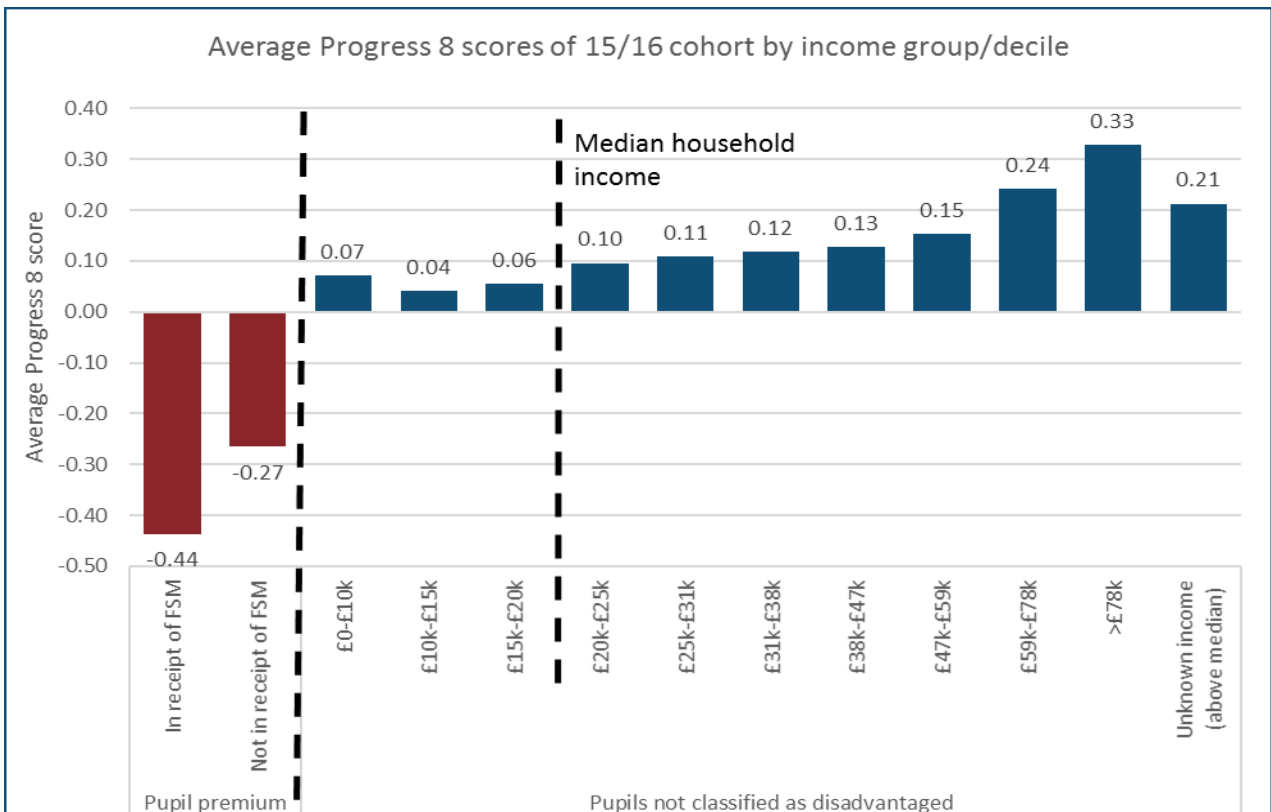
57.   The results presented above show detailed analysis before housing costs are introduced. We have first looked at children in families below median equivalised household income who are not classified as disadvantaged, and how this compares to what we already know about the location of pupils in receipt of pupil premium.

58.   As can be seen in the maps below, the lower income families are not evenly distributed throughout the country, and pupils from these families are focused in different areas to those already captured within disadvantage measures.

59.   Those pupils captured by our current disadvantage measures are found in the highest proportions in urban centres. Figure 13 shows this for London, and the same patterns can be seen in other larger cities, such as Birmingham and Manchester.

60.   The location of children below median income does also depend on their income levels. Those on the lowest incomes, shown in figure 12, are in their highest proportions in local authorities outside urban centres in the North and East of England. Unlike those eligible for pupil premium funding, they are not as commonly found in city centres – this is most apparent in London, where there are lower proportions across most local authority districts within the M25 than outside it.

Figure 11: Density of pupils eligible for pupil premium funding by local authority district

Darker shading indicates higher proportion of pupils in area eligible for pupil premium.
See accompanying ODS tables for underlying data.

Figure 12: Density of pupils in below median income families by local authority district

Darker shading indicates higher proportion of pupils in area in families below median income (but not considered disadvantaged). See accompanying ODS tables for underlying data.
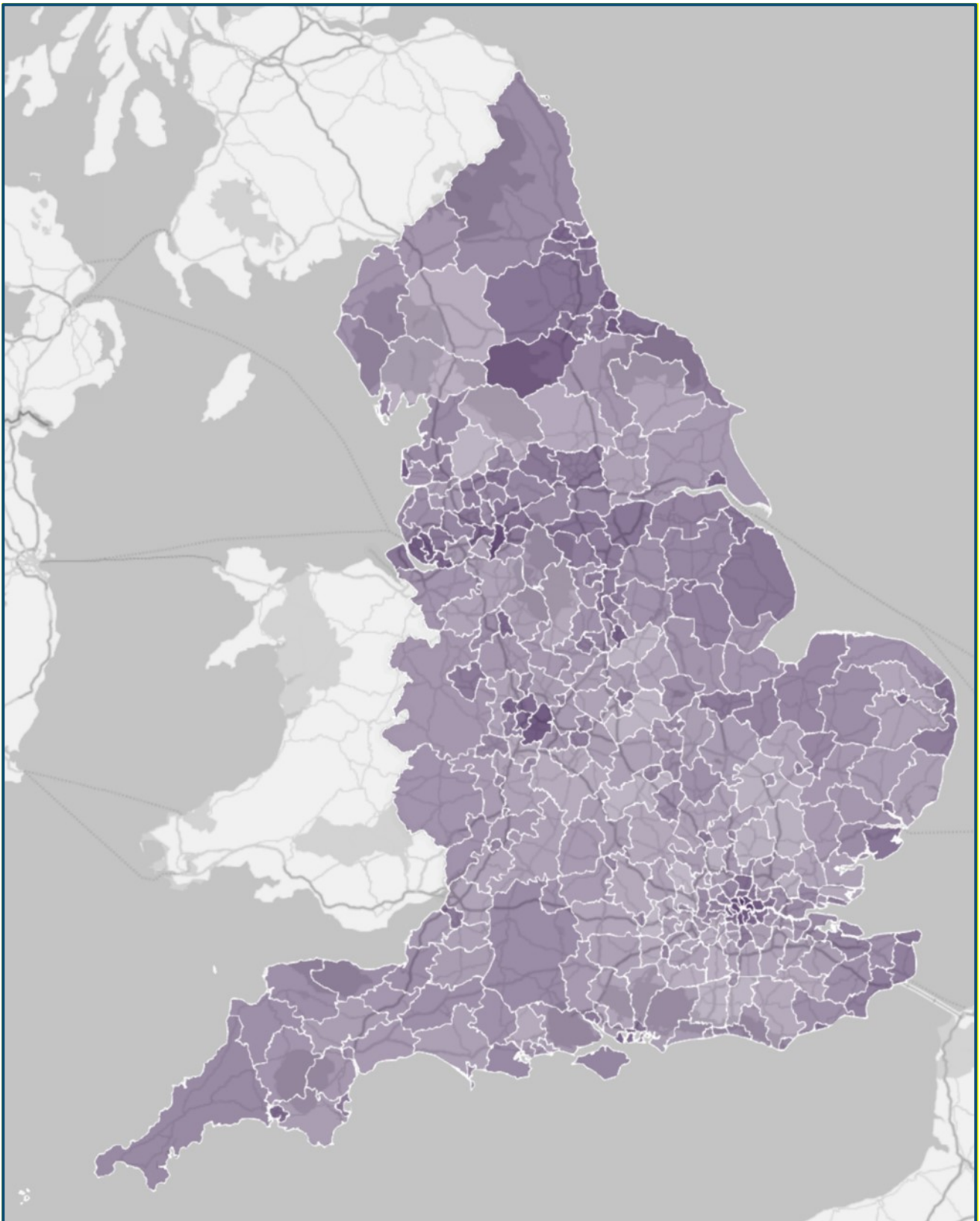
Figure 13: Density of pupils eligible for pupil premium in London by LA district

Darker shading indicates higher proportion of pupils in area eligible for pupil premium. See [accompanying ODS tables](#) for underlying data.
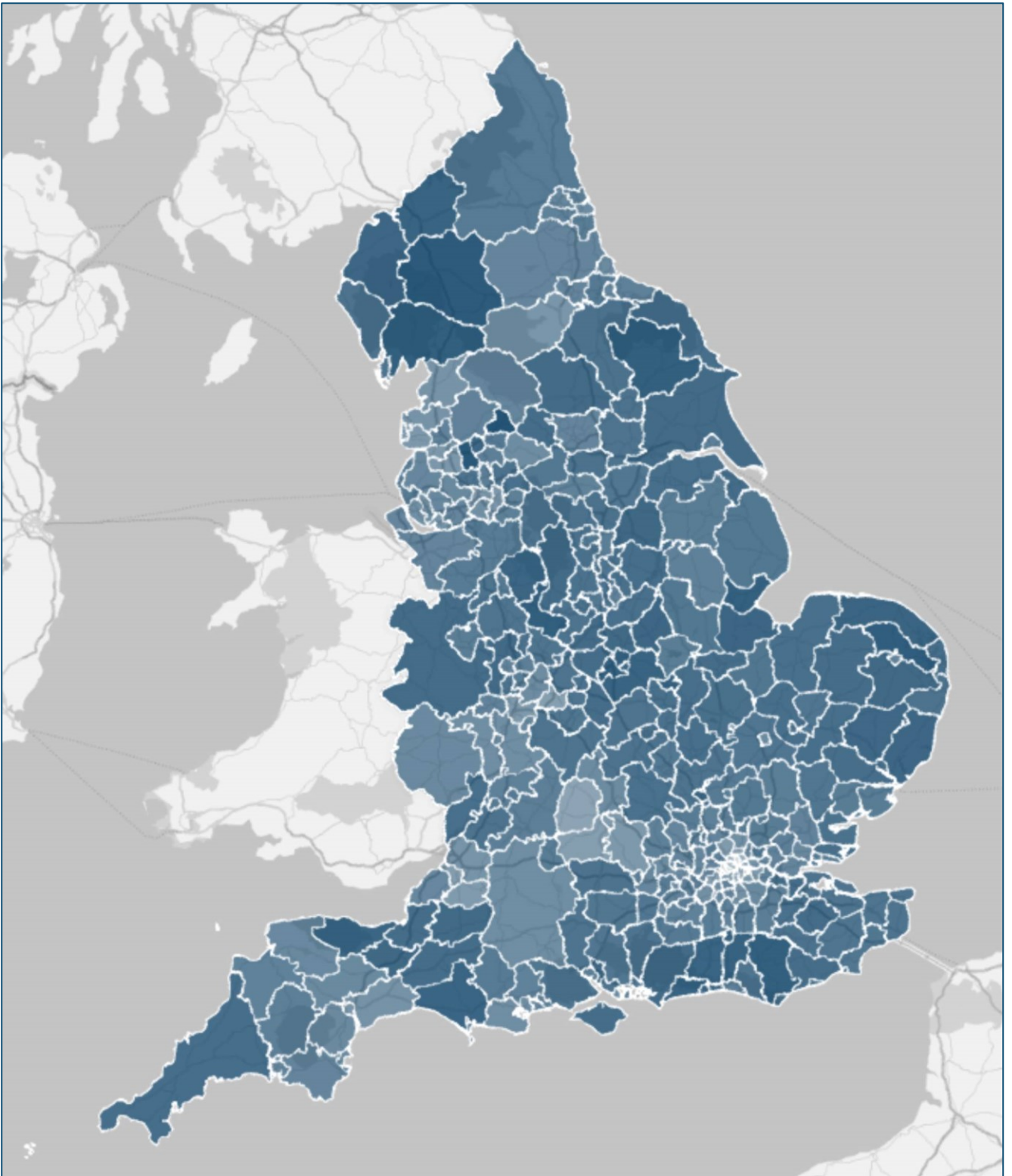

Figure 14: Density of pupils in below median income families in London by LA district

Darker shading indicates higher proportion of pupils in area in families below median income (but not considered disadvantaged). See [accompanying ODS tables](#) for underlying data.
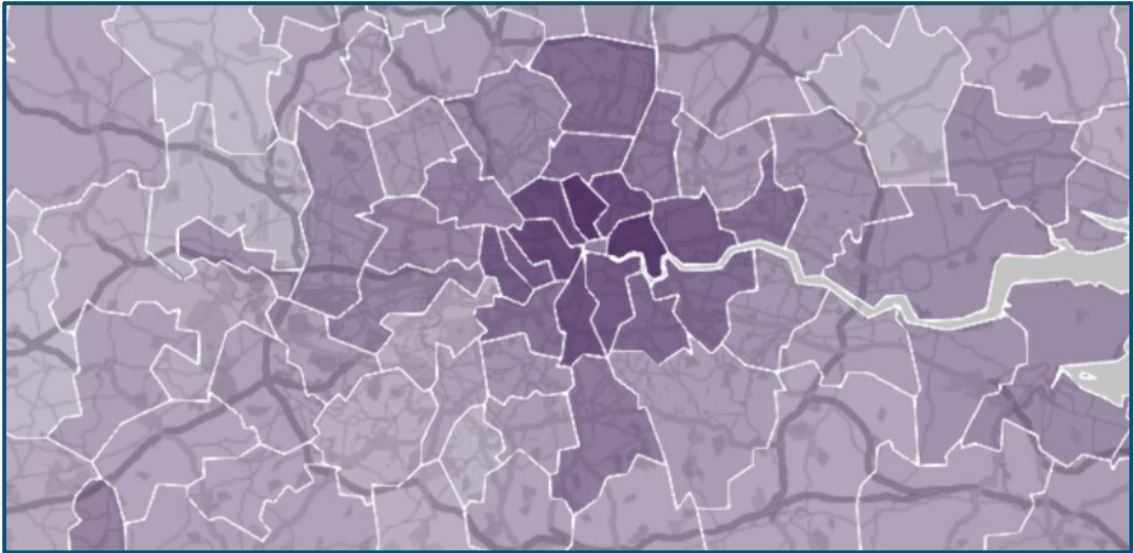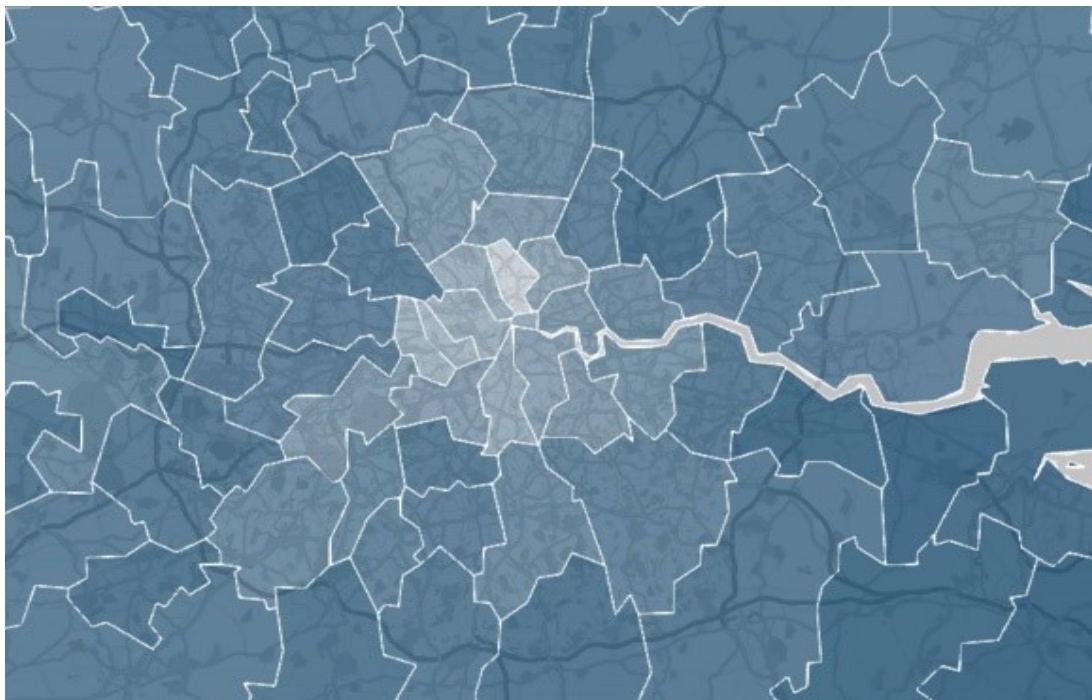
Opportunity areas

61.   The DfE has designated 12 social mobility 'cold spots' as opportunity areas across the country. This seeks to support the Department's goal to ensure all children can access high-quality education at every stage of life. This will be done through the creation of local partnerships with early years providers, schools, colleges, universities, businesses, charities and local authorities.

62.   As is shown below, almost all of these areas see higher proportions of pupils from disadvantaged and below median income families than the national average. This underscores the Department's focus on these areas, and their need for additional support so that these pupils can achieve better standards of living than their parents.
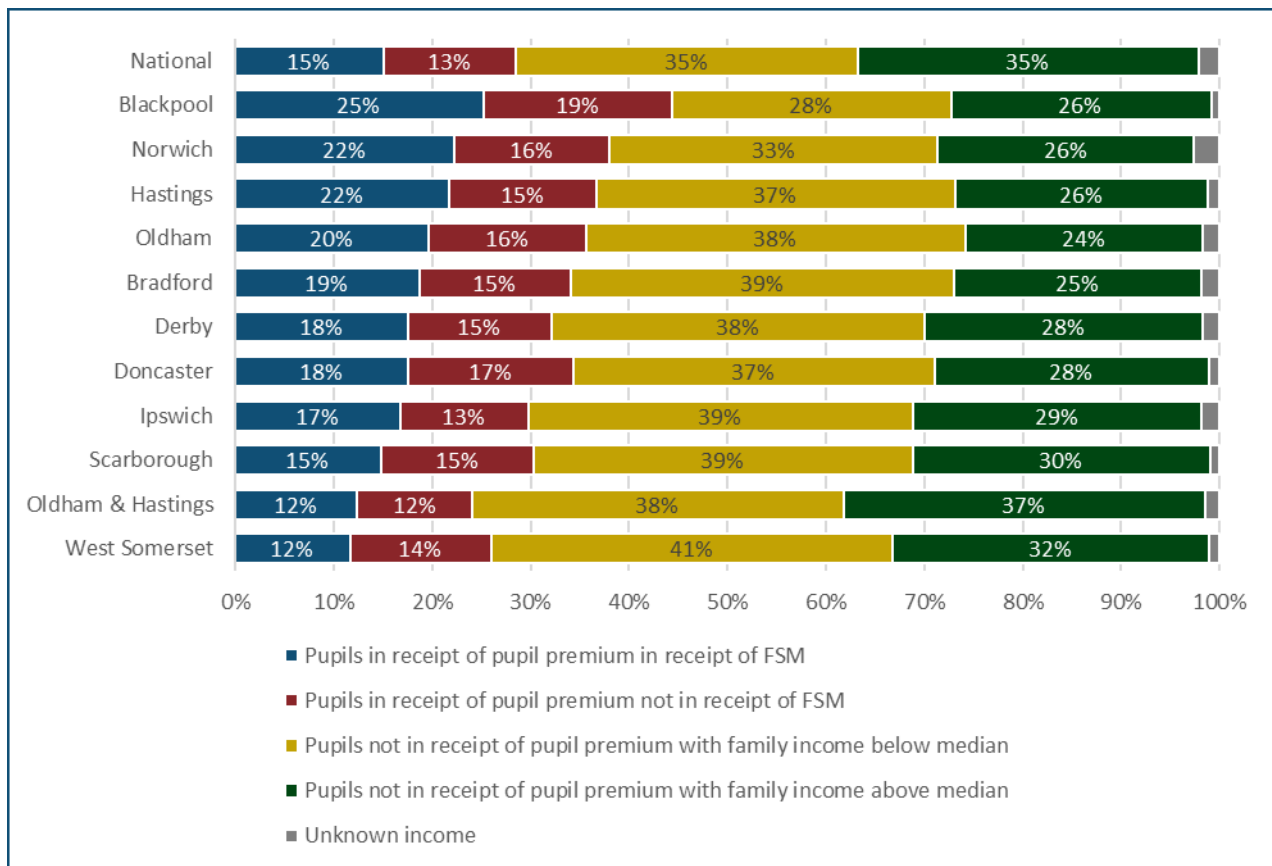

Figure 15: Proportion of pupils by income group in opportunity areas

Impact of housing costs

63.   The results by location show the need to fully understand the impact of housing costs on this analysis. In order to do this we have undertaken an area-based adjustment methodology as set out in paragraph 42 and Annex B. This ensures that each family's income is adjusted for the average housing costs for the local authority district in which they reside.

64.   This achieved the expected results after adjusting for average area housing costs. Under this measure 62% of pupils are in families below median household income, and 31% are below the median household income and not eligible for pupil premium. These proportions are similar to, though slightly lower than, those before housing costs.

65.   Looking at the results by local authority district, there is a high correlation between percentage of pupils below median income and not eligible for pupil premium both before and after housing costs. Each area saw a lowering of the percentage but less so in 32 local authority districts that are predominantly in London and metropolitan areas. This is the expected feature of accounting for the highest housing costs in London and other urban centres.

66.   The changes in proportions of below median income families are shown nationally and around London in figures 16 and 17, where darker areas are those with higher relative proportions and lighter areas with lower relative proportions. This further illustrates the shift in toward London and shifts away from the Home Counties, particularly surrounding the capital.
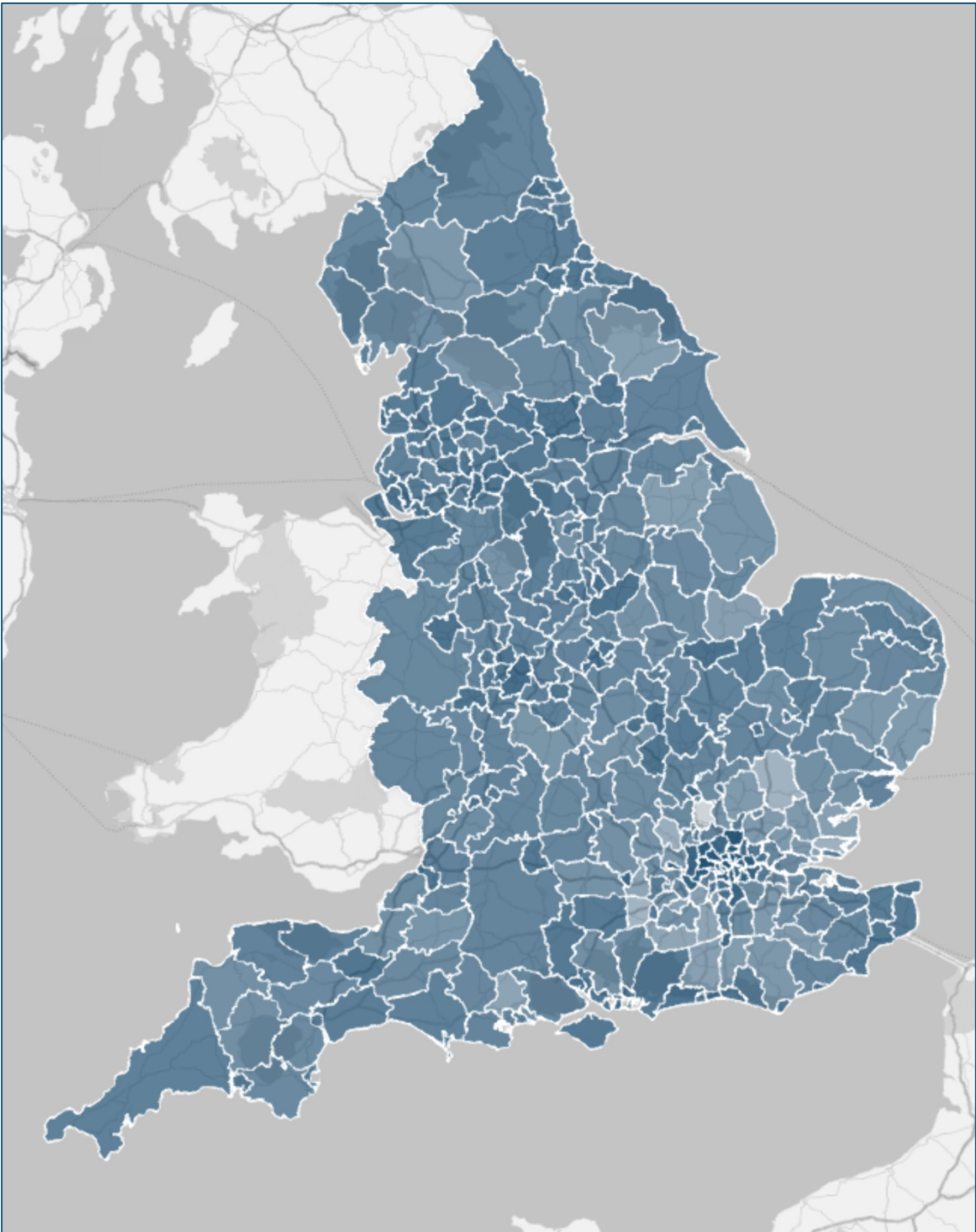
Figure 16: Map of change in the proportion of below median income families after housing costs

Darker shading indicates a greater absolute change (percentage point difference) in proportion of pupils in area in families below median income when housing costs are taken into account. See accompanying ODS tables for underlying data.
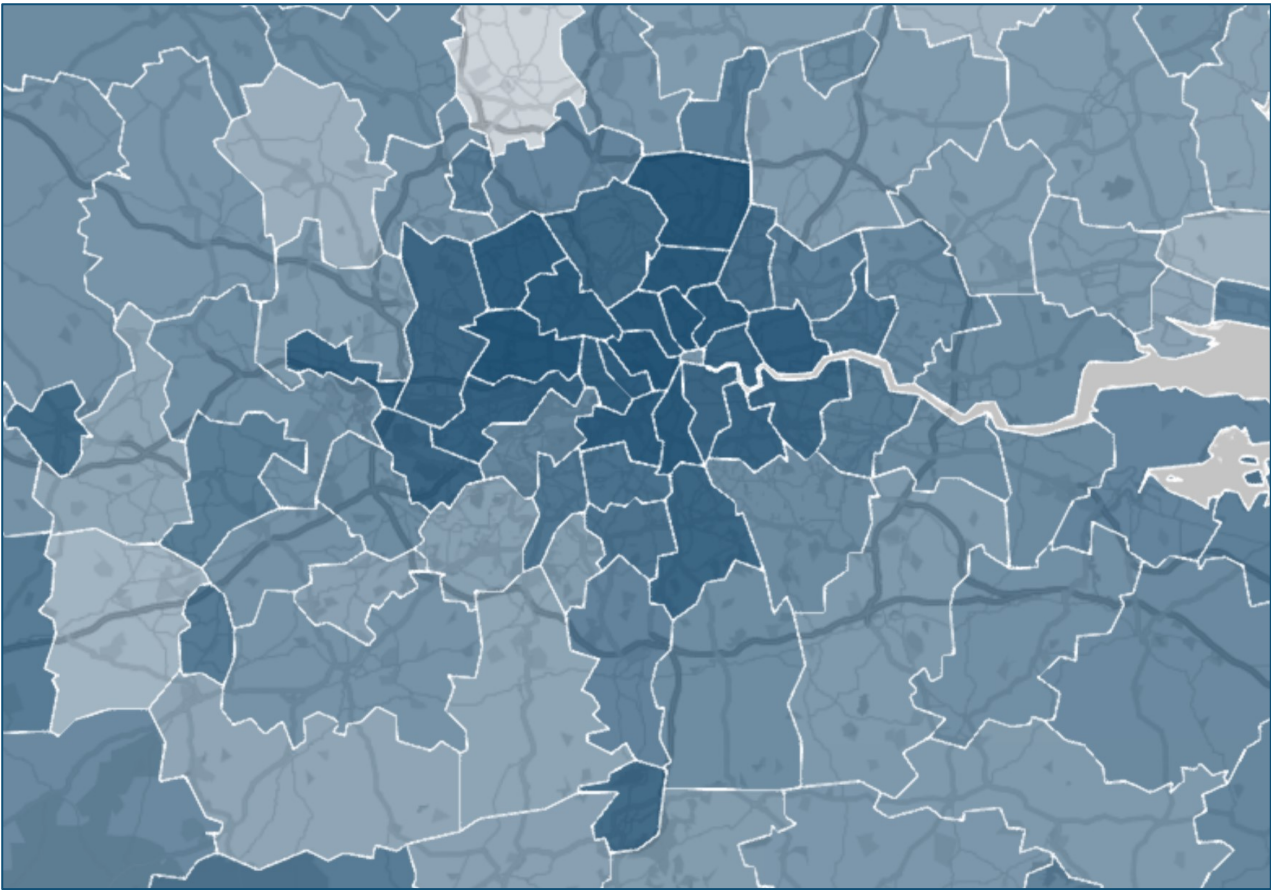
Figure 17: Map of London showing change in the proportion of below median income families after housing costs

Darker shading indicates a greater absolute change (percentage point difference) in proportion of pupils in area in families below median income when housing costs are taken into account. See accompanying ODS tables for underlying data.

In order to explore the sensitivity of the analysis to housing costs, we looked at key stage 4 Attainment 8 and Progress 8 results both before and after housing costs, for pupils eligible for pupil premium, those below median income not eligible for pupil premium and above median income. The results show that at a national level the analysis by income level is very similar, as seen in figures 18 and 19.

Figure 18: Average Attainment 8 score before and after housing costs

Figure 19: Average Progress 8 score before and after housing costs

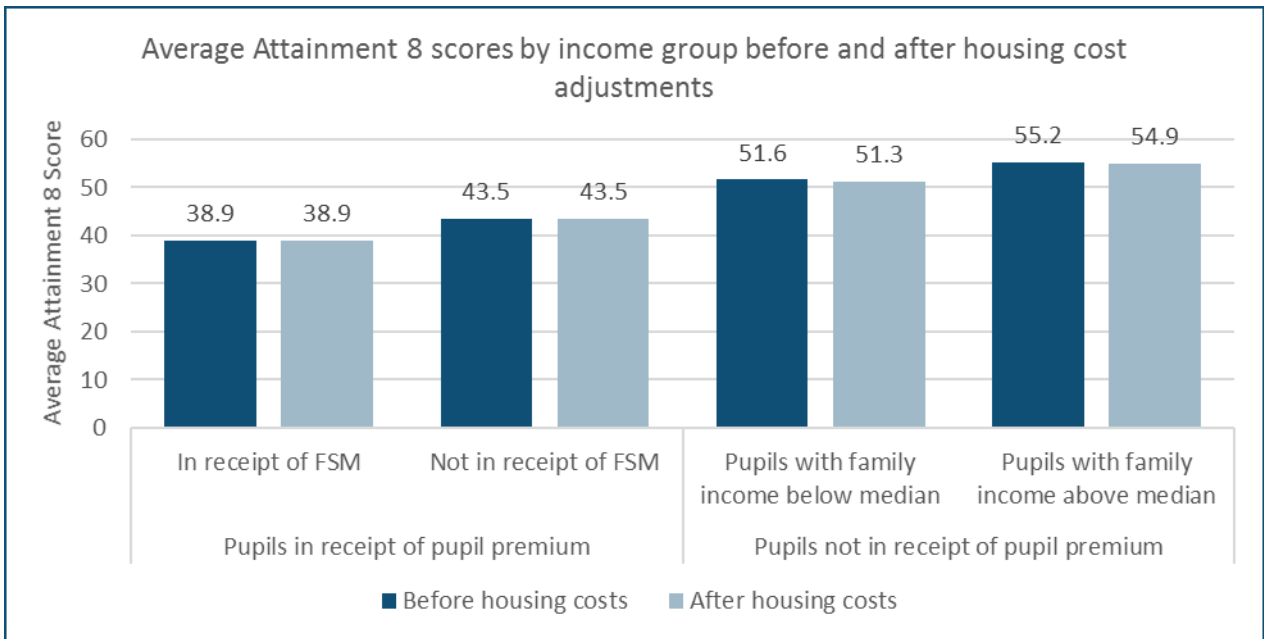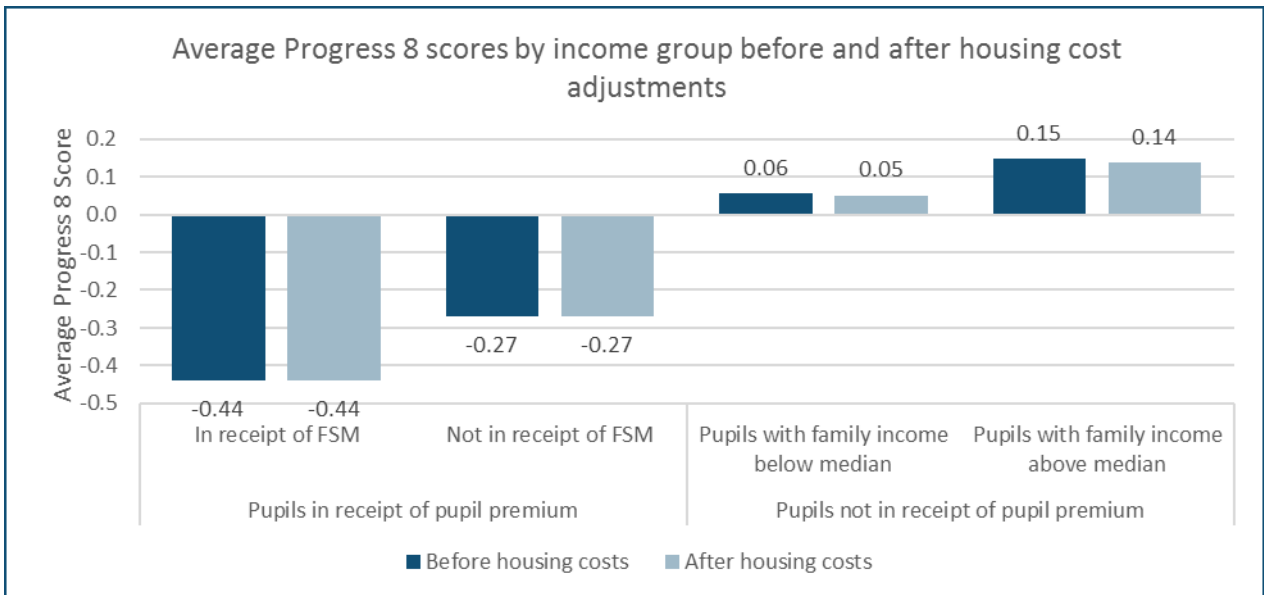67.   This suggests that although the following section which focuses on analyses before adjusting for housing costs may show some changes if housing costs were taken into account, its findings will remain broadly accurate.

Detailed analysis of results before housing costs

68.   A working definition has been set using the nearest decile to the median equivalised household income of the households with children in our database with a valid income. This has been used in preference to survey estimates at this stage due to the lack of some self-employed income; this is discussed further in the data linking section.

69.   Additionally, it should be noted that in the following analysis we have excluded pupils in Year 12 and 13. This is because the dataset only covers those in state-funded schools, and does not include those in other institutions such as sixth-form and other further education colleges. As such, our data on those in post-16 education has a substantial bias at this time. We have also excluded a further 0.02% of pupils in the dataset with implausible income data.

70.   The following income groupings have been used in this paper:

- pupils in receipt of pupil premium;
    - pupils in receipt of FSM;
    - pupils not in receipt of FSM;
- pupils not in receipt of pupil premium;
    - pupils in families above median income (with equivalised income greater than £20,000);
    - pupils in families below median income (with equivalised income less than £20,000); and
- pupils with unknown incomes.

71.   The median point of £20,000 is based on equivalised income, as explained in Annex A. For a two parent family with two teenage children, this would equate to an unadjusted household income of around £33,000. For a lone parent family with one young child, it would equate to an unadjusted household income of around £17,000.

72.   This estimate of the median household income is lower than that estimated by other sources and is based on households included in our dataset. This lower estimate reflects the incomplete data on income in our current dataset and is likely to be adjusted in further work on this topic.

73.   Applying these definitions to the data gives the breakdown seen in figure 21, overleaf.
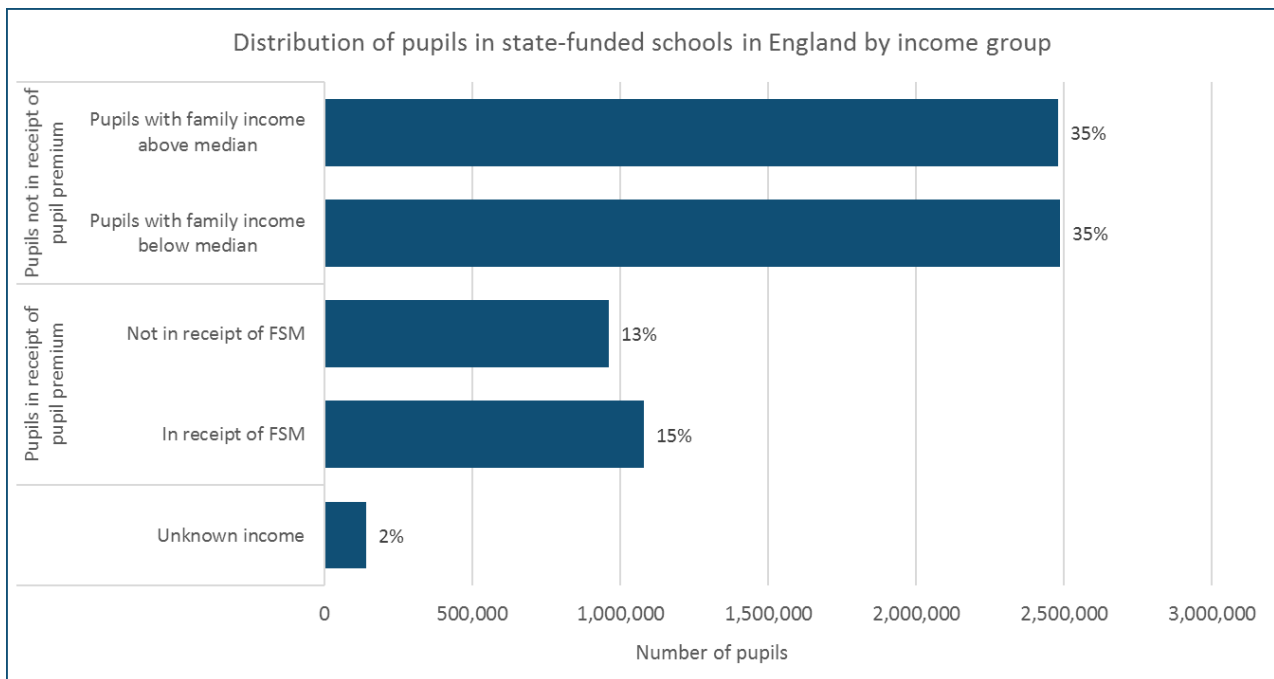
Figure 21: Distribution of pupils in state-funded education in England by income

74.   35% of all pupils (2.5 million) in the NPD fall are below the median income, but are not already considered disadvantaged by the DfE. Together with the 1.1 million pupils in receipt of free school meals and 1.0 million others currently receiving pupil premium funding, these groups cover 63% of all pupils in the NPD. This is above 50% for three reasons:

a) the median is defined for families, and lower income families contain, on average, more children;
b) some disadvantaged families not in receipt of free school meals will now be earning above median income; and
c) our available income data is in deciles and we have taken the nearest decile which includes some families with incomes slightly higher than the median in the dataset.

Family characteristics

75.   Families in lower income groups more often have their first child earlier, are younger and include more children, when compared to their more affluent peers.

76.   Families already captured by our disadvantage measures most commonly have their first child during the parents' early twenties, as seen in figure 22. This is a decade younger than other income groups. Other below median income families have their first child earlier on average than families with above median income, but still are most likely to have that child in their early thirties.
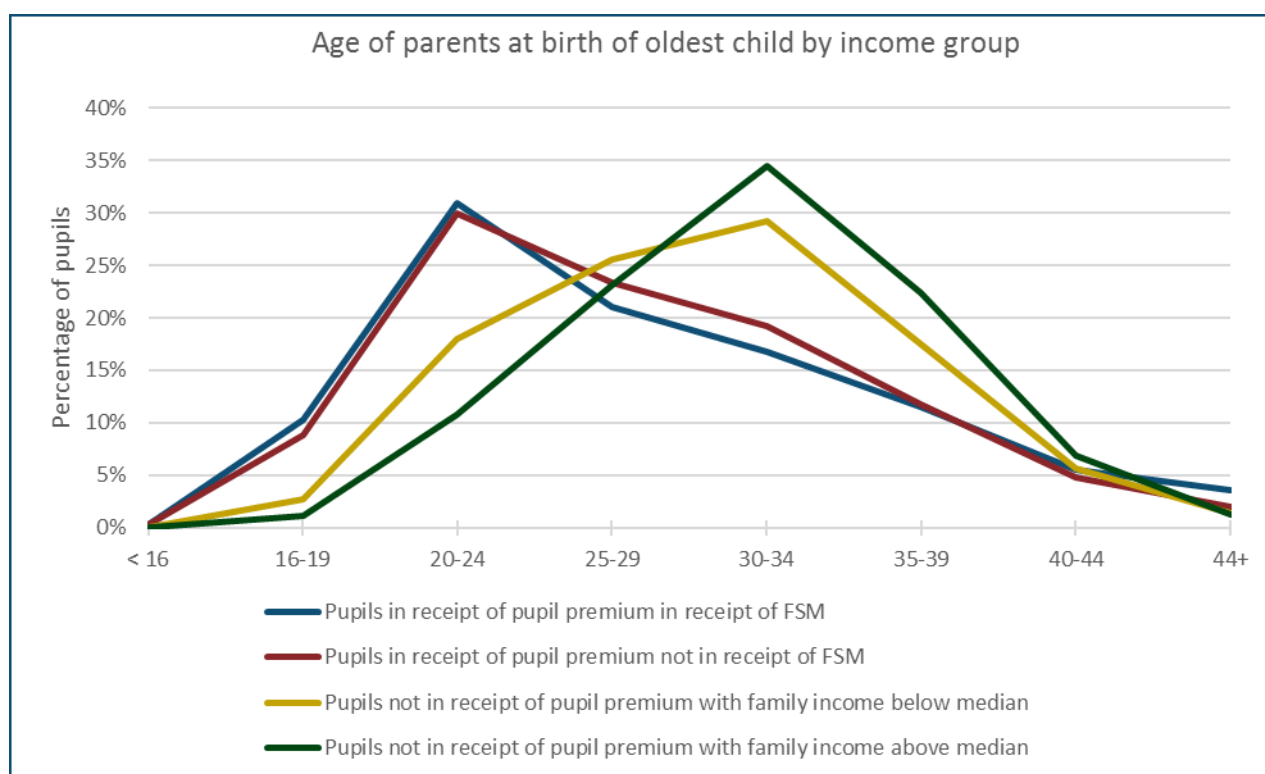
Figure 22: Age of parents at birth of oldest child

77.   In addition, pupils from below median income families are more likely to be younger than those with above median incomes but older than disadvantaged pupils.

78.   Lower income families are more likely to include more children. This is partially explained by the methodology – larger families see their effective income lowered by the equivalisation process (see Annex A) as they require a larger household income for the same standard of life. As such, a two parent family with a household income of £25,000 might not be below the median equivalised household income with one child, but would with two.

Ethnicity and language

79.   Figure 23 shows that there are progressively more pupils from White backgrounds in families above median income compared to those below median or disadvantaged. This is reversed for pupils from most other ethnic backgrounds, with the exception of Asian families which constitute a bigger percentage of below median income families than they do of disadvantaged or above median income families.
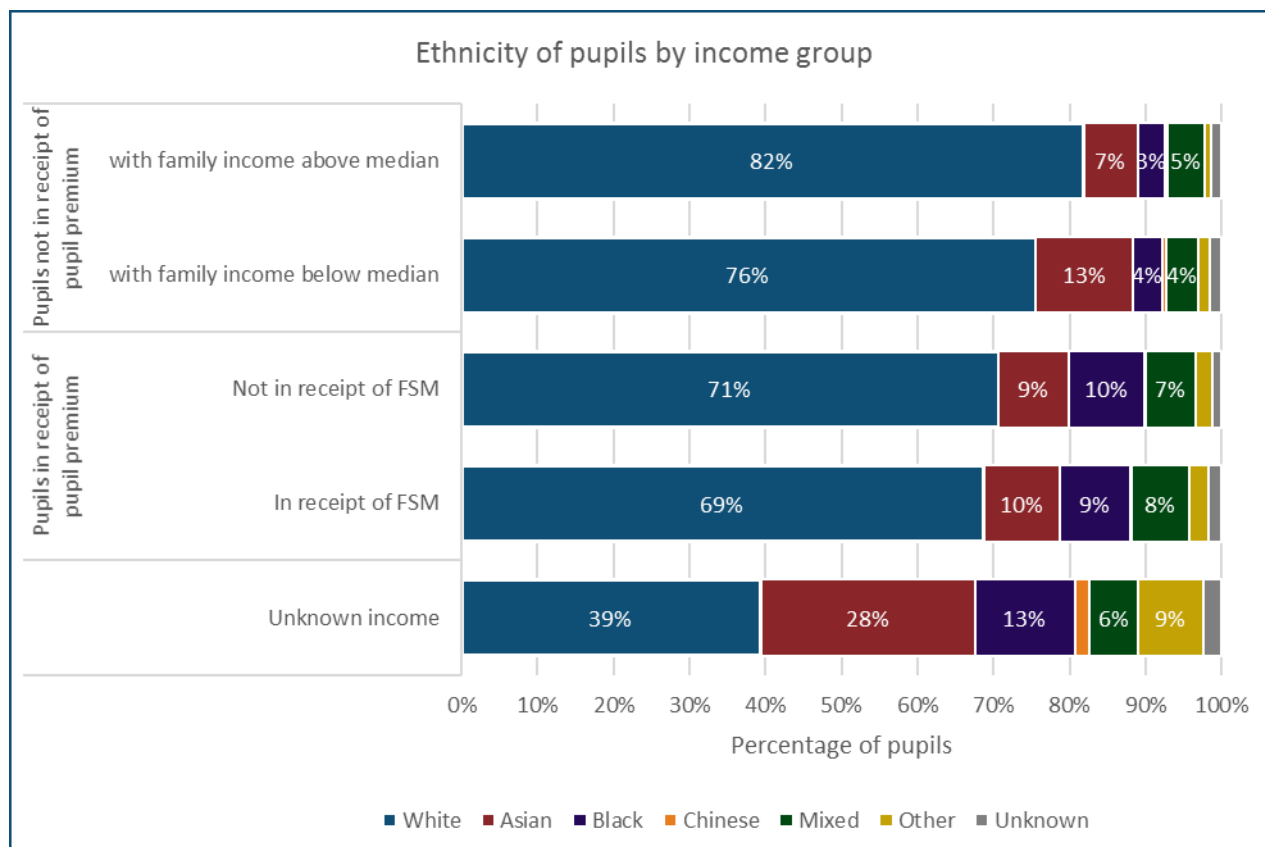

Figure 23: Ethnicity of pupils by income group

80.   Those with unknown incomes are the most unusual group, with four in ten being White, compared to more than eight in ten nationally. This group also shows high levels of having English as an additional language. This would be consistent with this group being largely from a recent immigrant background, particularly non-European. The attainment characteristics of this group, with low attainment in their earlier school years but large progress in later years, would also be consistent with this. This would also explain why making a match to Child Benefit records has not been possible, as access to benefits such as this is heavily restricted for those who have recently arrived from non-EEA countries.

81.   Similarly, the proportions of pupils with English as an additional language are higher among lower income families. There is a particularly notable difference between one and two parent families: pupils from two parent families are substantially more likely to have English as an additional language than pupils from lone parent families across most income groups, especially in below median income families.
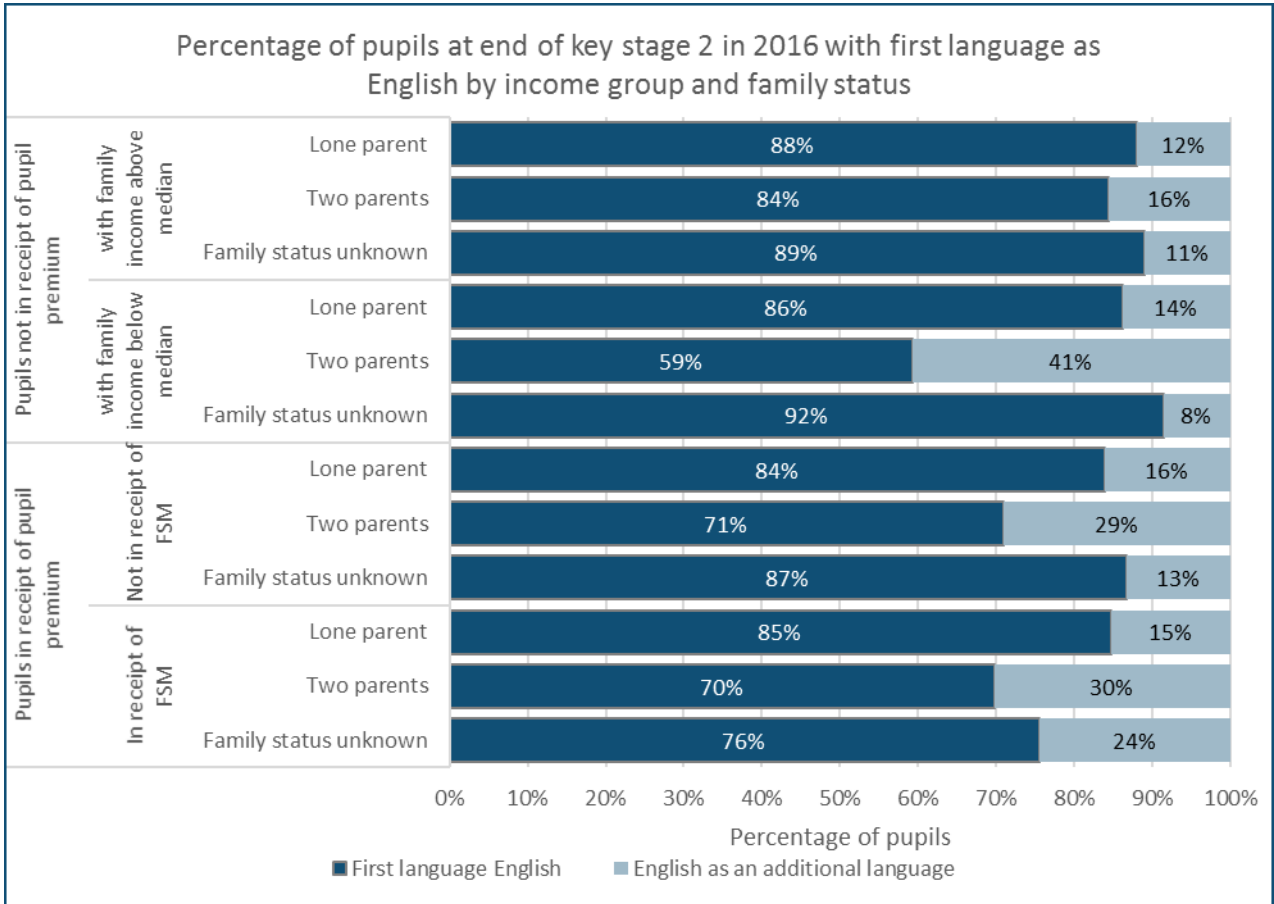
Figure 24: Percentage of pupils at the end of key stage 2 in 2016 with first language as English by family income and family status

Schools attended

82.   The types of school attended by those across the income range vary significantly. This has been examined across both markers of school quality, such as Ofsted ratings and attainment, and by selection criteria, such as grammar schools and faith schools.

83.   Across all income groups, 61% to 62% of pupils in each group attend a school that is rated 'Good' by Ofsted. However, there are larger differences in the proportions attending 'Outstanding' and 'Requires improvement' schools, with more affluent pupils more likely to attend a high performing school and less likely to attend a low performing school.

Percentage of Pupils attending Schools by Ofsted Rating and Income Group
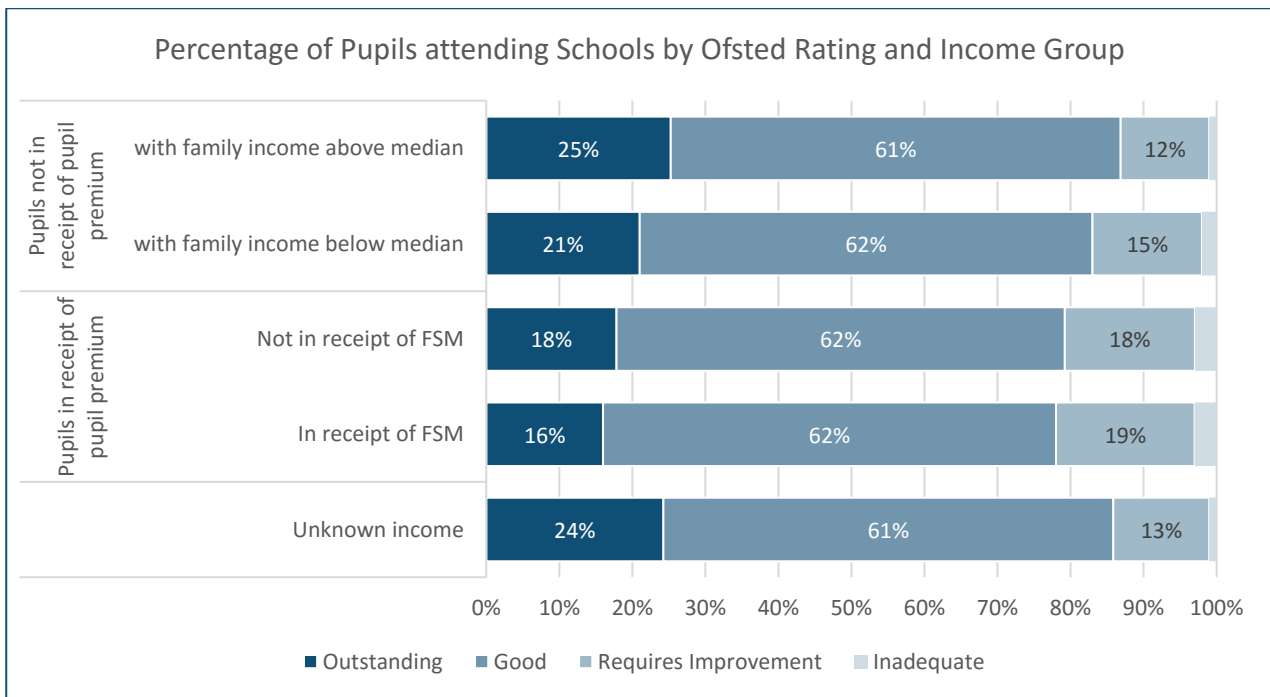
Figure 25: Proportion of pupils in each income group by Ofsted rating of the school they attend

84.   Figure 26 shows the family circumstances of pupils in non-selective schools, as defined in DfE's recent publication on school performance and houses prices analysis, for both primary and secondary schools. This defines the attainment decile of schools based on data between 2012 and 2014, using the proportions attaining level 4 and level 5 at key stage 2 and attaining 5 A*-C grades at GCSE including English and maths.

85.   The charts show that disadvantaged pupils are significantly less likely to be in higher performing non-selective schools and that below median income families, who are not currently considered disadvantaged, are present in similar proportions across all schools.

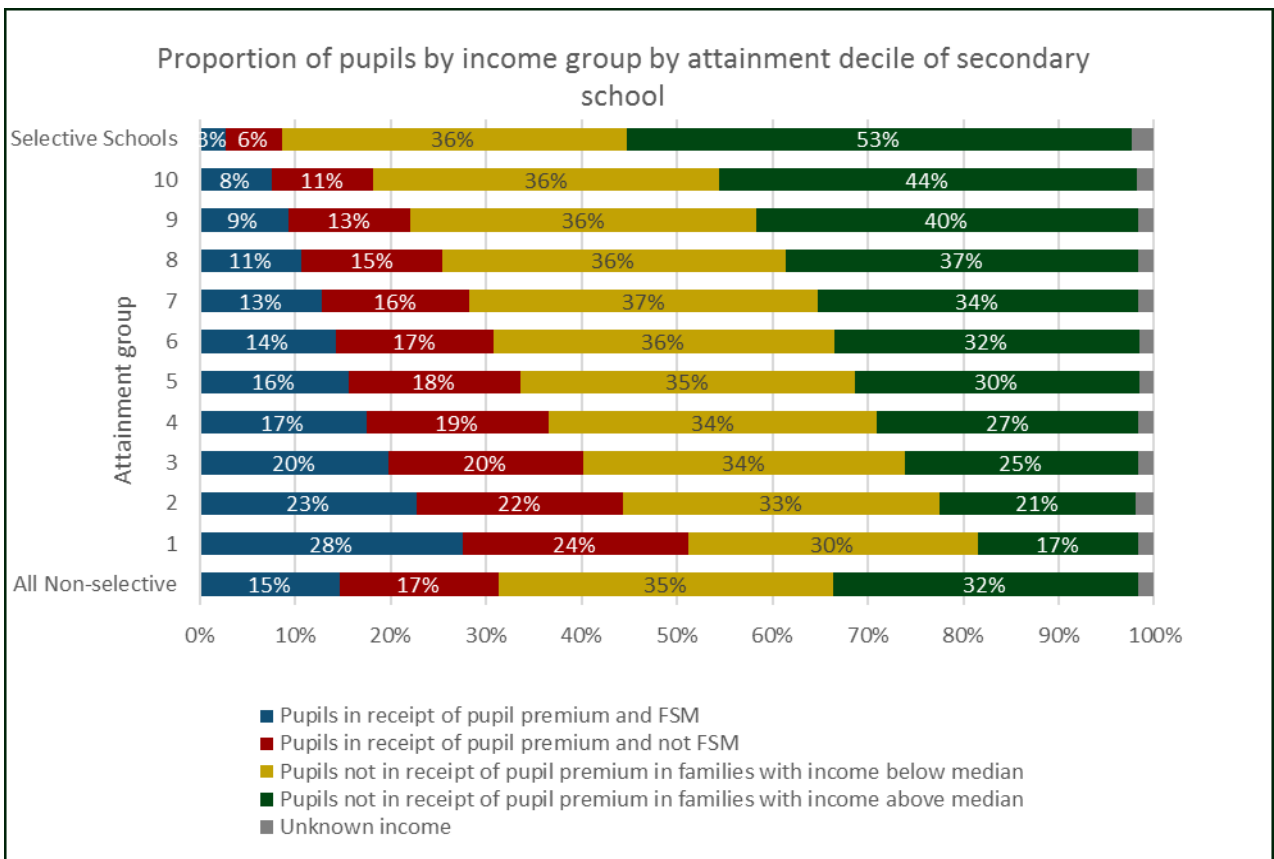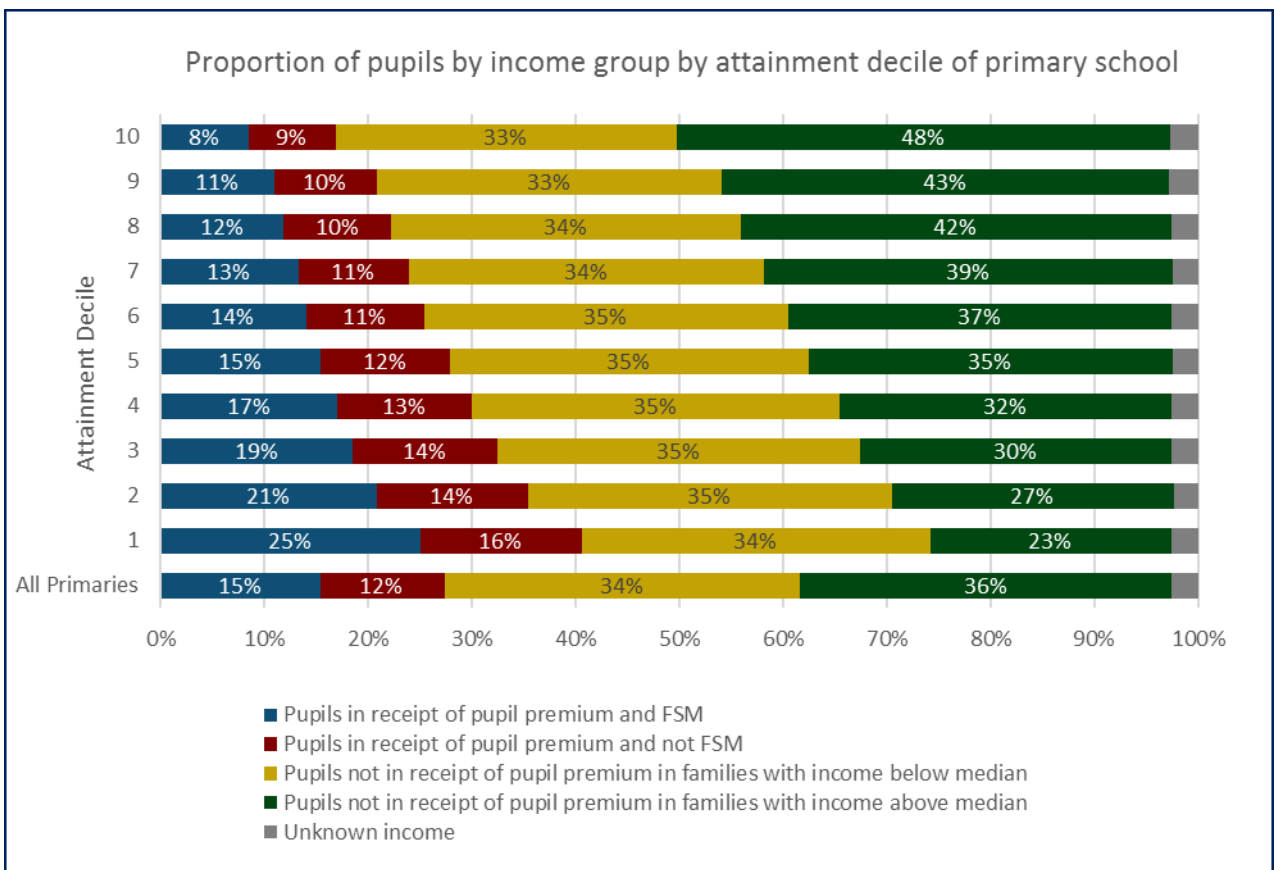**Figure 26: Proportion of pupils by income group by attainment decile of non-selective secondary schools**

Figure 27: Proportion of pupils by income group by attainment decile of primary school

38

86.   As is seen when considering national attainment, disadvantaged pupils achieve, on average, lower attainment scores. As such, those schools with lower proportions of disadvantaged pupils would be expected to have higher overall attainment even when all other factors are constant.

87.   It is not possible to determine from this analysis if these results are more influenced by this effect, or if high performing schools not admitting as many disadvantaged pupils. Other measures of school quality, such as Ofsted ratings, suggest that some variation due to the latter is present.

88.   Faith schools also see a similar, if smaller, under-representation of pupils from families in receipt of pupil premium at primary school level, as seen in figure 28. This does not continue into secondary, where proportions are largely consistent between faith and non-faith schools.
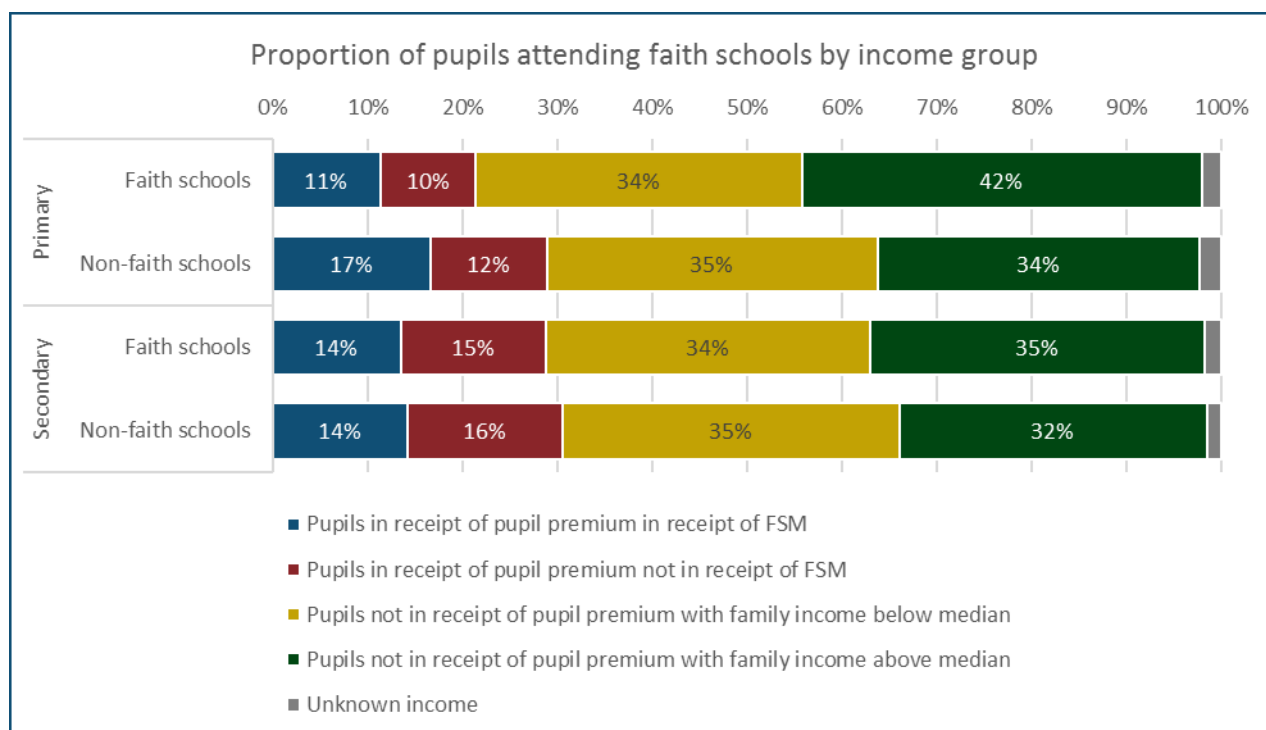
Figure 28: Proportion of pupils attending faith schools by income group

89.   For grammar schools, the key differences are again amongst disadvantaged children and those from families with above median incomes. Here the variation is at its largest, with less than one in ten pupils in selective schools from disadvantaged backgrounds and over half from the most affluent group. The percentage of children at selective schools from below median income families, who are not considered disadvantaged, (36%) is almost the same as the percentage for non-selective schools (35%). This is shown in figure 29.

Proportion of pupils attending grammar and non-grammar schools by income group

Percentage of pupils

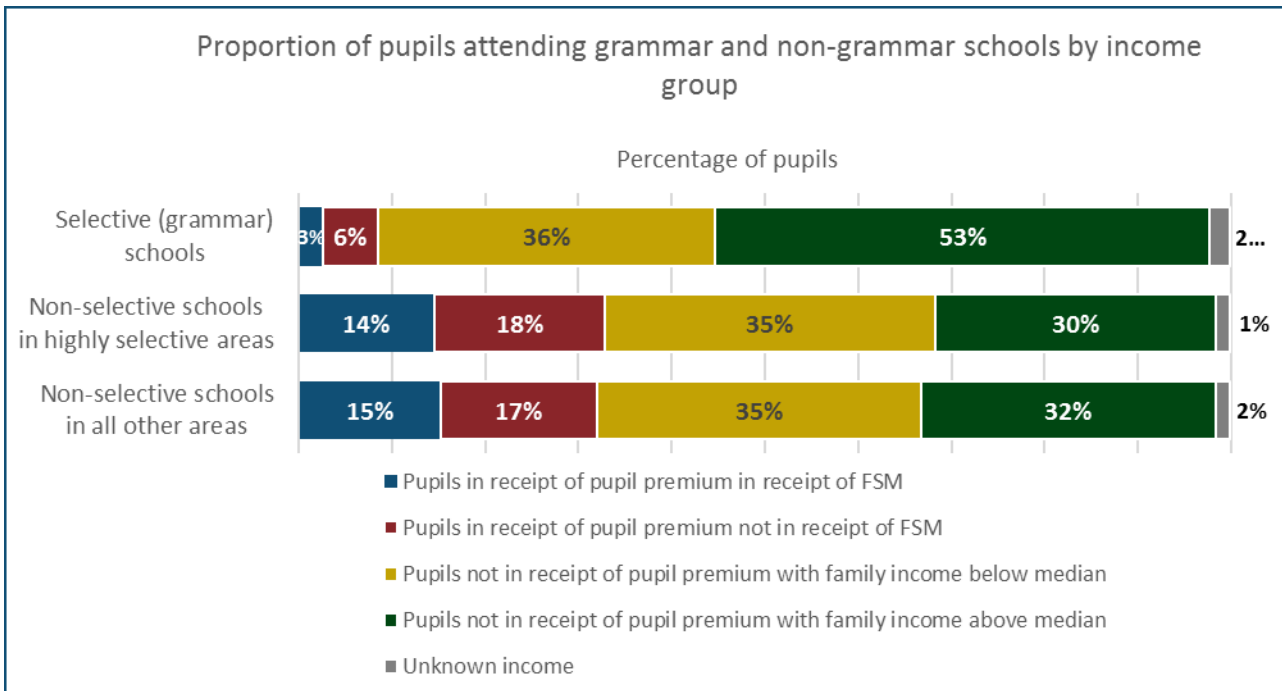|	Pupils in receipt of pupil premium in receipt of FSM	Pupils in receipt of pupil premium not in receipt of FSM	Pupils not in receipt of pupil premium with family income below median	Pupils not in receipt of pupil premium with family income above median	Unknown income
Selective (grammar) schools	3%	6%	36%	53%	2…
Non-selective schools in highly selective areas	14%	18%	35%	30%	1%
Non-selective schools in all other areas	15%	17%	35%	32%	2%

Figure 29: Proportion of pupils attending grammar and non-grammar schools by income group

Progress and attainment

90.   The dataset shows clearly that attainment and progress increase as families' incomes increase. However, the largest gaps are between those DfE already supports through pupil premium funding and free school meals and those above this criteria. Those from families with below median incomes have worse outcomes than those families above the median threshold, but are substantially closer to their performance than those already regarded as disadvantaged. This can be seen at key stage 4 in figure 30.

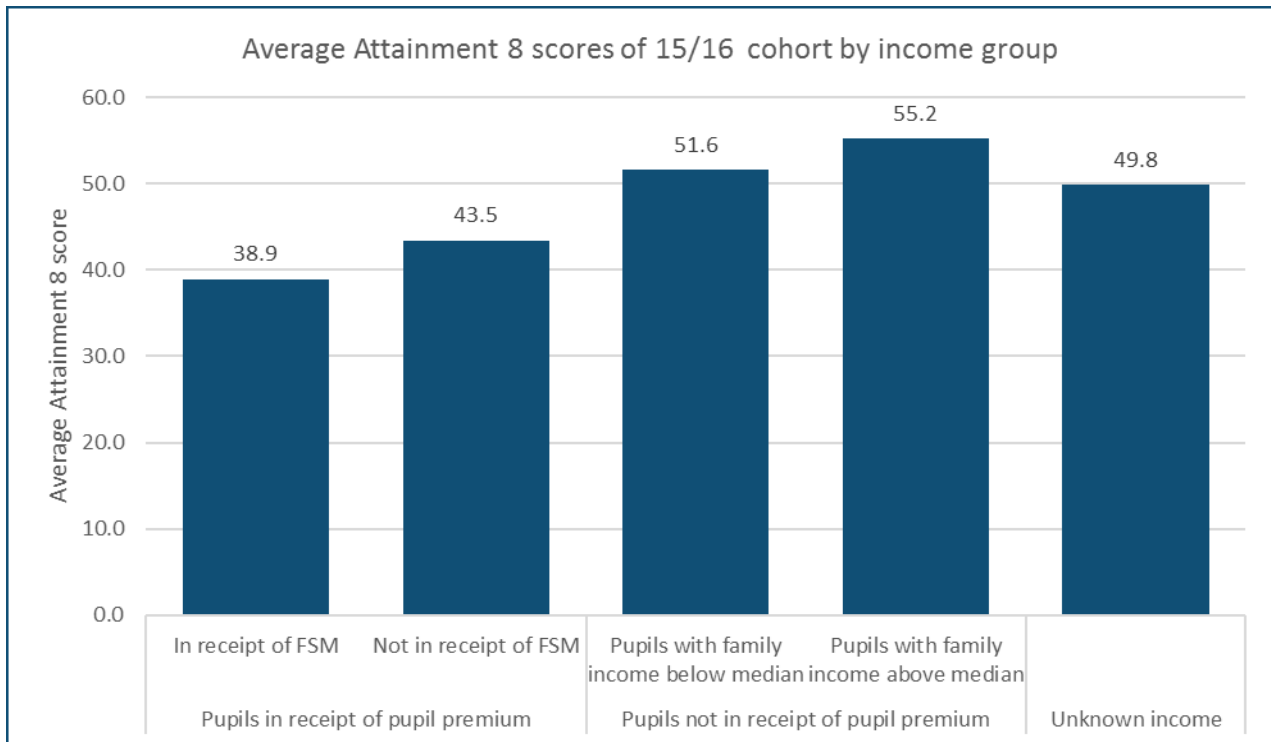

Figure 30: Average Attainment 8 scores by income group

91.   Looking at the former headline measure for attainment, the percentage achieving at least five A*-C to C grades at GCSE, including English and maths, the same pattern is present.
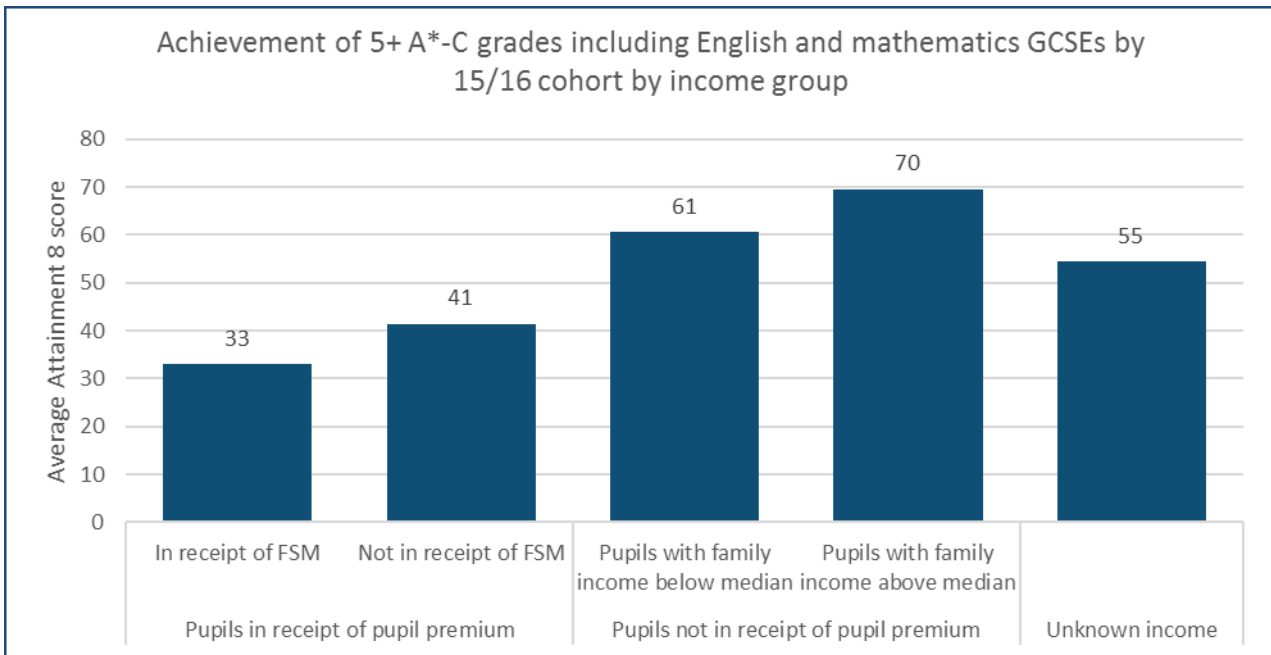
Figure 31: Achievement of 5+ A*-C grades including English and mathematics GCSEs by income group

Progress for all groups not currently regarded as disadvantaged is average or above, as indicated by having a progress score of above zero in figure 32. As highlighted in the data linking section, whilst the average Progress 8 may fall as the data quality is improved, it is expected that those families who are below median income will continue to make average, or close to average, progress. This compares to the substantially below average progress made by those in receipt of pupil premium support.
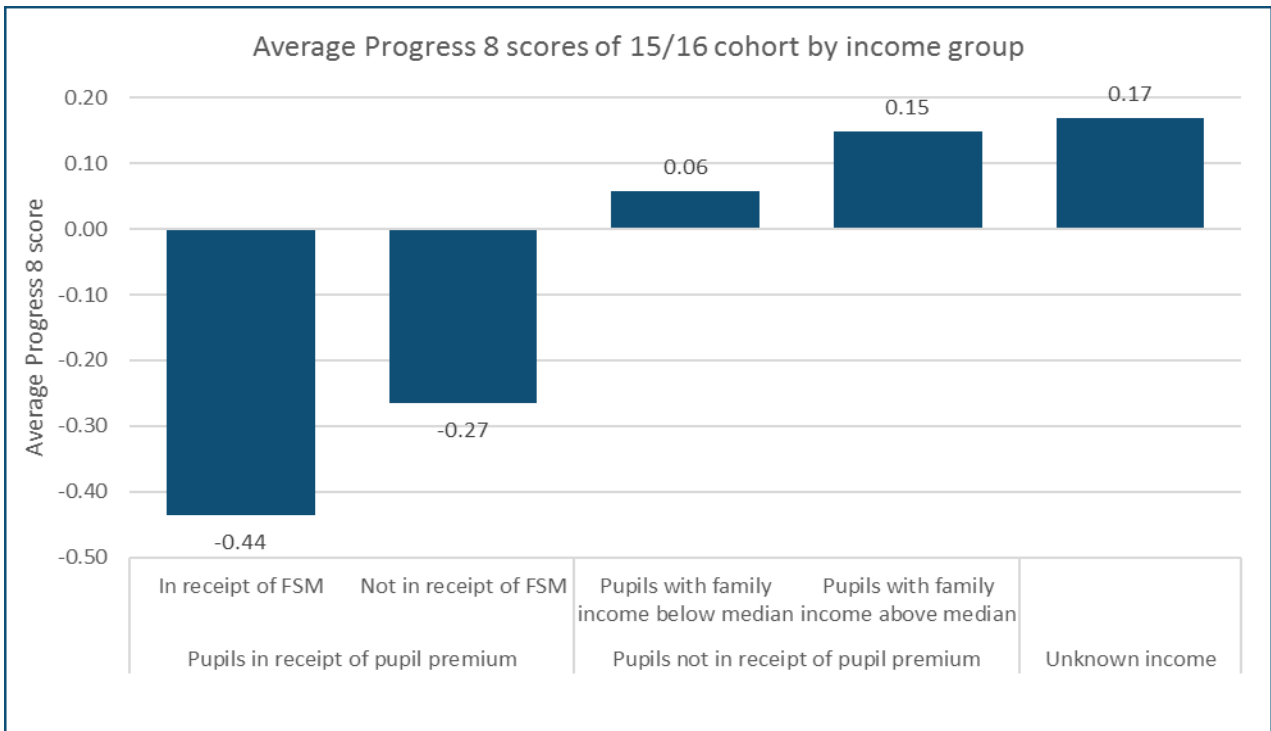

Figure 32: Average Progress 8 scores by income group

42

92.   Looking further at the subjects entered, there is substantial variation in the proportions of pupils entering and achieving the EBacc at key stage 4, as is seen in figure 33.
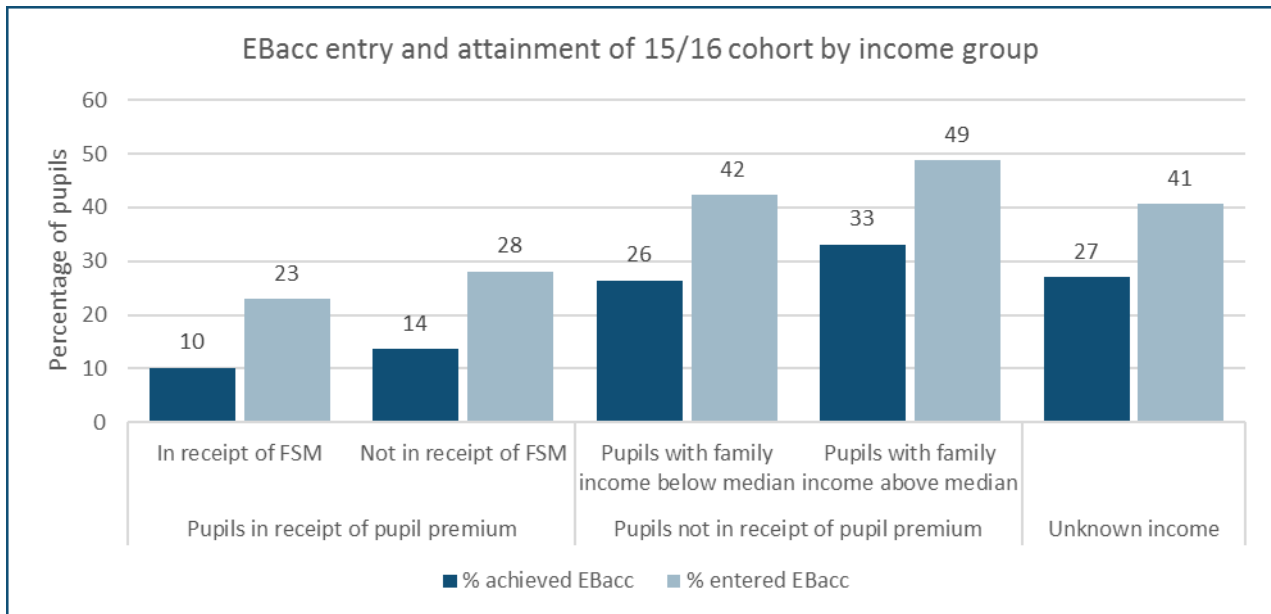
Figure 33: EBacc entry and attainment by income group

93.   There is a higher percentage of pupils from families earning above the median achieving the EBacc than there are pupils eligible for pupil premium entering it. Pupils from families who are not classified as disadvantaged but below the median income are less likely to achieve or enter the EBacc than their wealthier peers, but are still around twice as likely to achieve the EBacc as those eligible for pupil premium.

94.   A similar pattern is seen at key stage 2. However, here the gap between those above and below the median point is larger, with those families below sitting more at a mid-point between the disadvantaged and higher income groups.
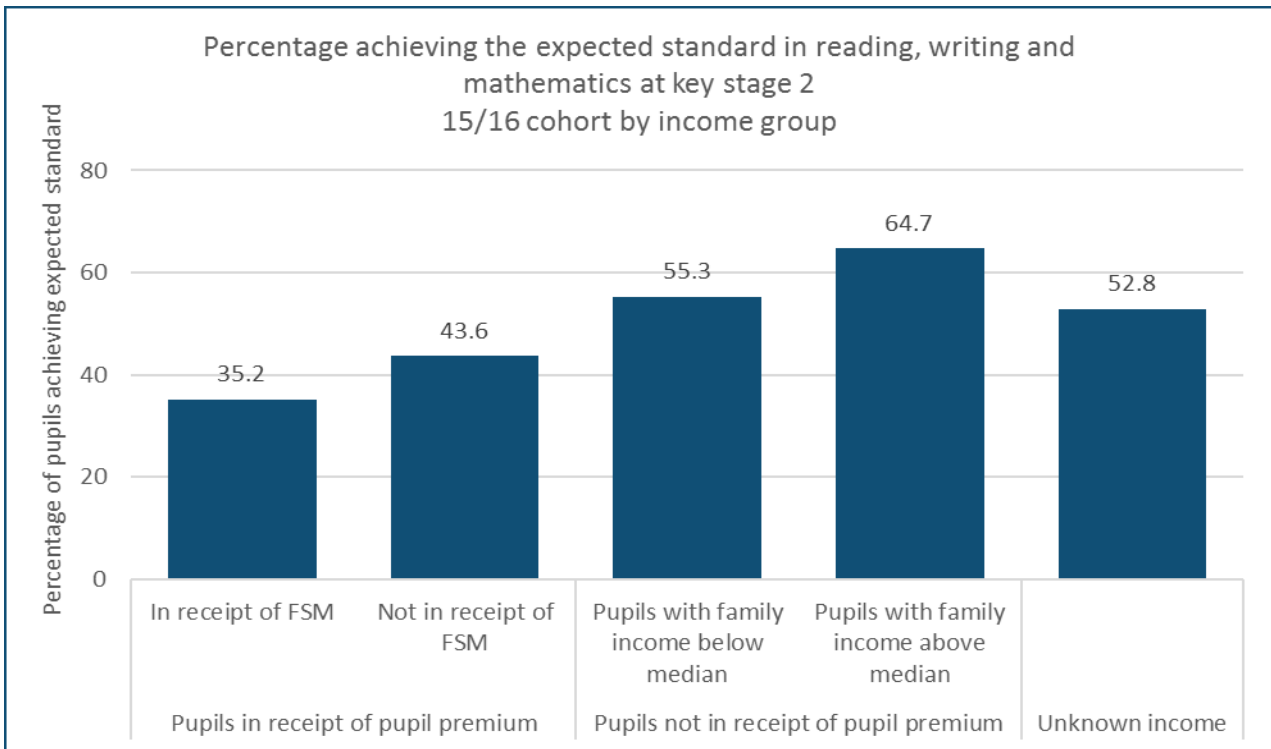
Figure 34: Percentage achieving the expected standard in reading, writing and mathematics at key stage 2 by income group

95.   The effect of family size on key stage 2 results has also been examined, as can be seen in figure 35. This analysis only considers eldest children to control for the effect of birth order, which has a similar but separate effect. These show results for the reading assessment, but similar patterns are seen among the other assessments and for children of different birth order.

Figure 35: Percentage of eldest children in the 15/16 cohort meeting the expected standard in reading at key stage 2 by family size and income group

96.   Across the lowest income bands, eldest children in larger families have lower attainment than those in smaller families. The exception to this, across all groups, a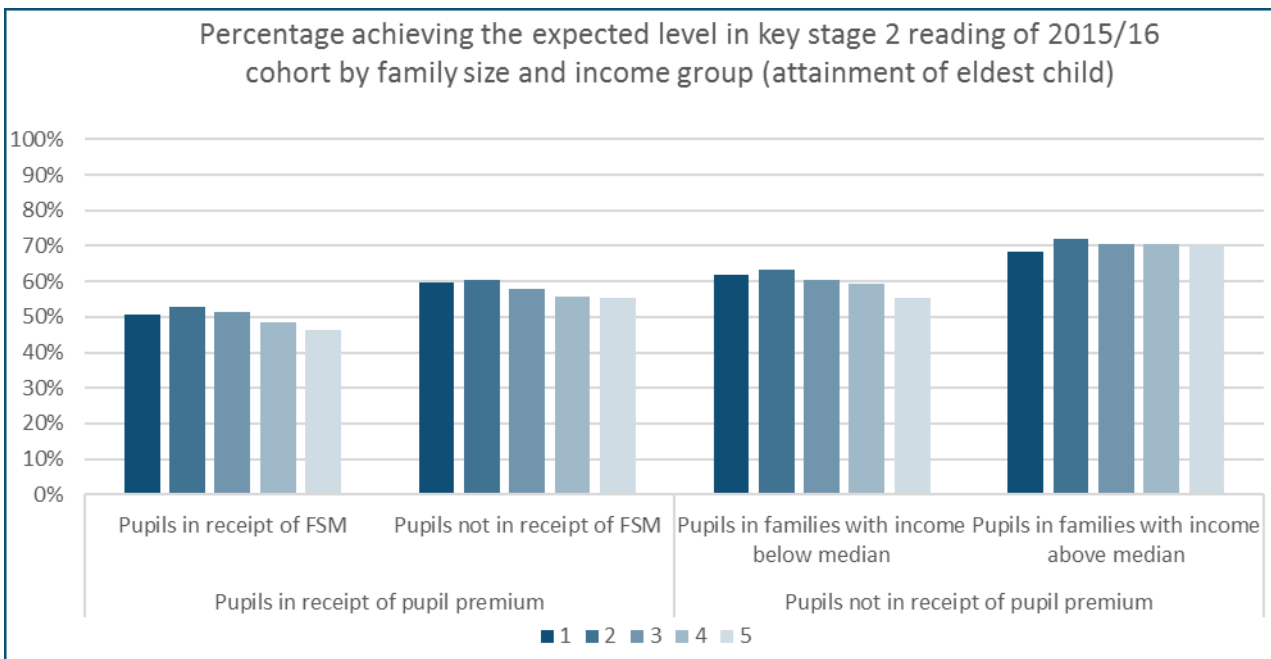re lone children, who perform worse than children with one other sibling. As will be discussed further below, less affluent families are also more likely to be larger. However, this pattern does not hold for more affluent families, where larger families perform as well as smaller families.

97.   It is only possible to reliably look at the effects of family size at key stage 2, as at key stage 4 it is much more likely the pupil will have older siblings who are not picked up in this dataset because they have already left school. This would result in incorrect identification of family size.

Participation in Higher Education

98.   As this feasibility study has been conducted with the most recent cohort of pupils, it is not possible to directly see the proportions from each income group who go on to participate in higher education.

99.   However, the destinations of pupils is of clear interest. As such, the proportions of pupils by income group at local authority district level have been compared to the participation in higher education, as determined through the POLAR dataset[2]. This shows the percentage of the eighteen year old cohort who proceed to higher education aged eighteen or nineteen.

100.   Moderate correlation has been found at the two ends of the income group scale. Areas with high proportions of those above the median have higher proportions participating in higher education, and areas with high proportions of those eligible for pupil premium funding have lower proportions participating in higher education. These trends can be seen in figures 36 and 37.

---

[2] Higher Education Funding Council for England (HEFCE): POLAR – Participation of Local Areas
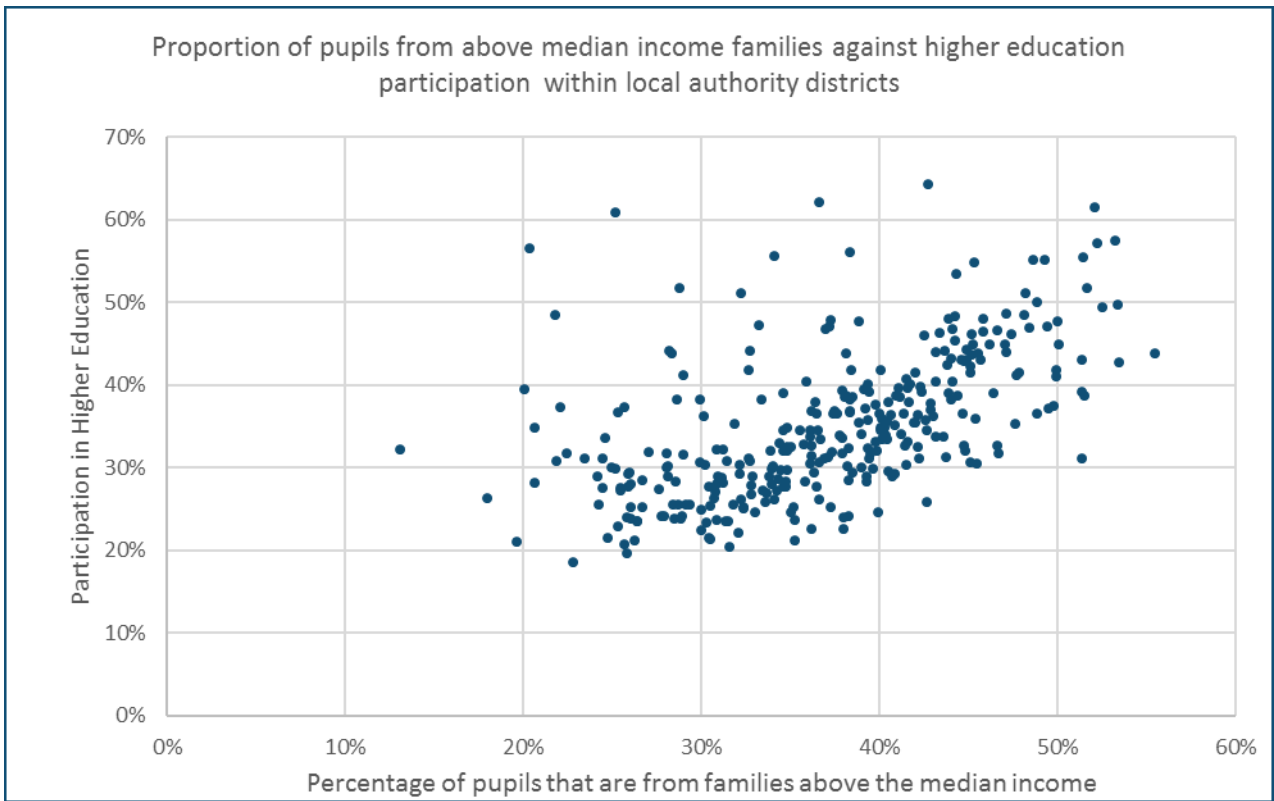
Figure 36: Proportion of pupils from above median income families against higher education participation within Local Authority districts
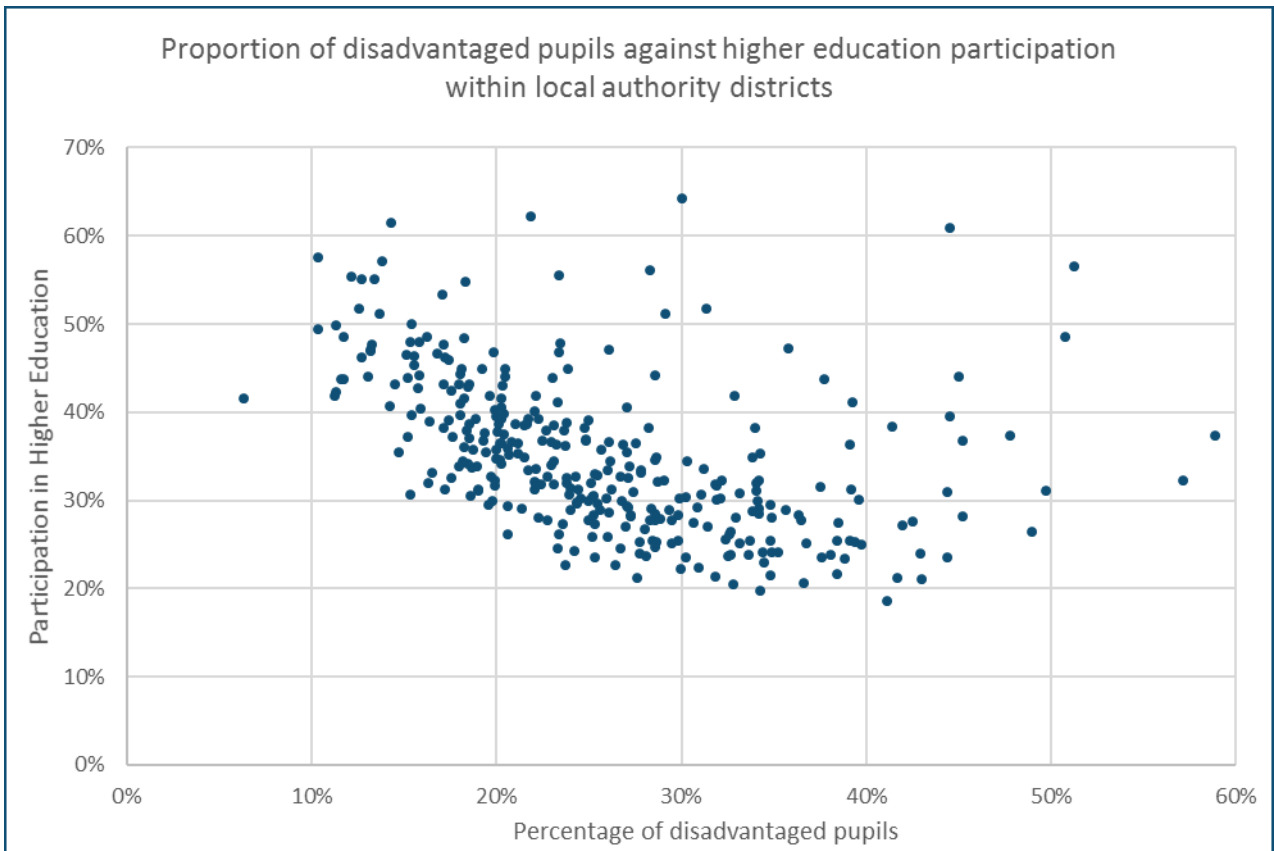

Figure 37: Proportion of disadvantaged pupils against higher education participation within Local Authority districts

101.   As can be seen in figure 38, those areas with higher proportions of families below the median income but not classified as disadvantaged, when not considering other groups, do not show a clear trend. This may be because in some cases lower proportions of these families indicate higher proportions of those above the median or those caught by disadvantage measures.
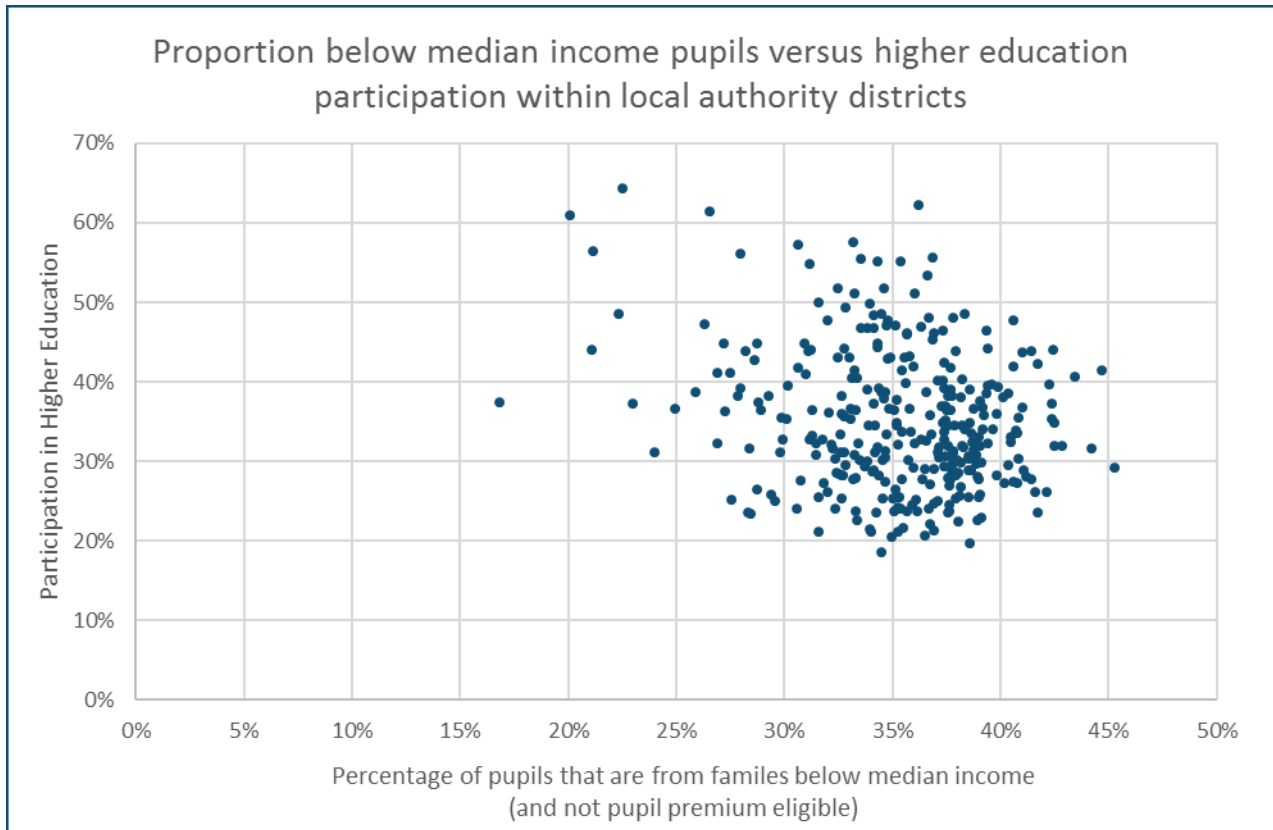
Figure 38: Proportion of below median income pupils against higher education participation within local authority districts

102.   Grouping below median income families with either their wealthier peers or disadvantaged pupils simply gives the inverse of the trends already discussed. The percentage of disadvantaged pupils combined with below median income families is simply 100% minus the percentage of pupils above the median income. In a similar way, the percentage of pupils in below median income families and above median families is equivalent to just those in disadvantaged families.

Lone parent families

103.   Our ability to identify pupils in lone parent households is limited to those for whom we have tax credit information. As such, there is a sizeable proportion of families of unknown family status, particularly in higher income groups.

104.   As can be seen in figure 39, lower income families are more likely to be lone parent families. While the proportion of unknown families among higher income groups is much higher, this trend holds true among those where the status is known. It is unlikely the unknown families are disproportionately lone parent families, as we would expect lone parent families to more easily qualify for tax credits.
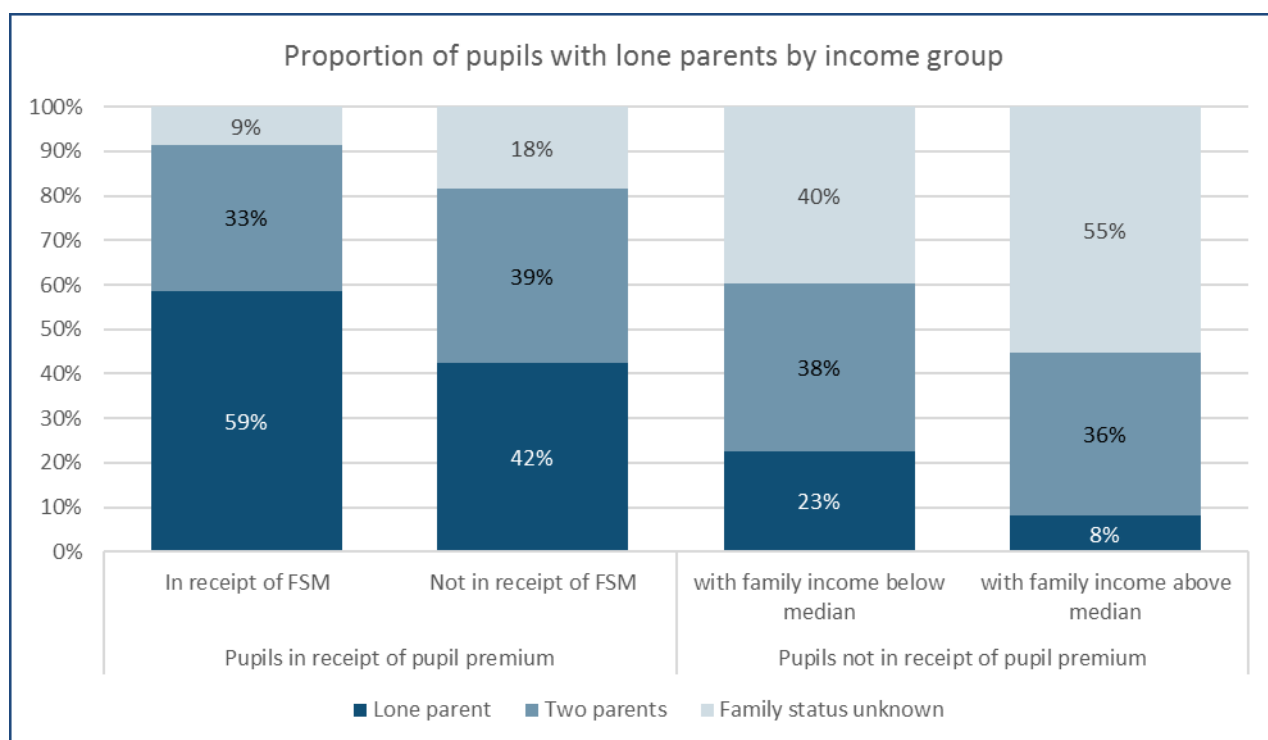

Figure 39: Proportion of pupils with lone parents by income group

105.   Understanding the income of lone parent families develops our understanding of attainment among pupils from these families substantially. Previous evidence showed children in lone parent families consistently underperforming those in two parent families. Performance at key stage 4, as shown in figures 40, is in line with this, with pupils in lone parent families consistently having lower attainment than those in two parent families. However, the gap is small in most cases, and much of the overall gap can be accounted for by the high proportion of lone parent families in lower income groups.

106.   The attainment among the unknown families group is higher in most cases, which may lessen this effect. However, even if all of these families are assumed to be two parent households, the attainment gap between pupils from lone parent families and two parent families shrinks considerably once income level is controlled for.
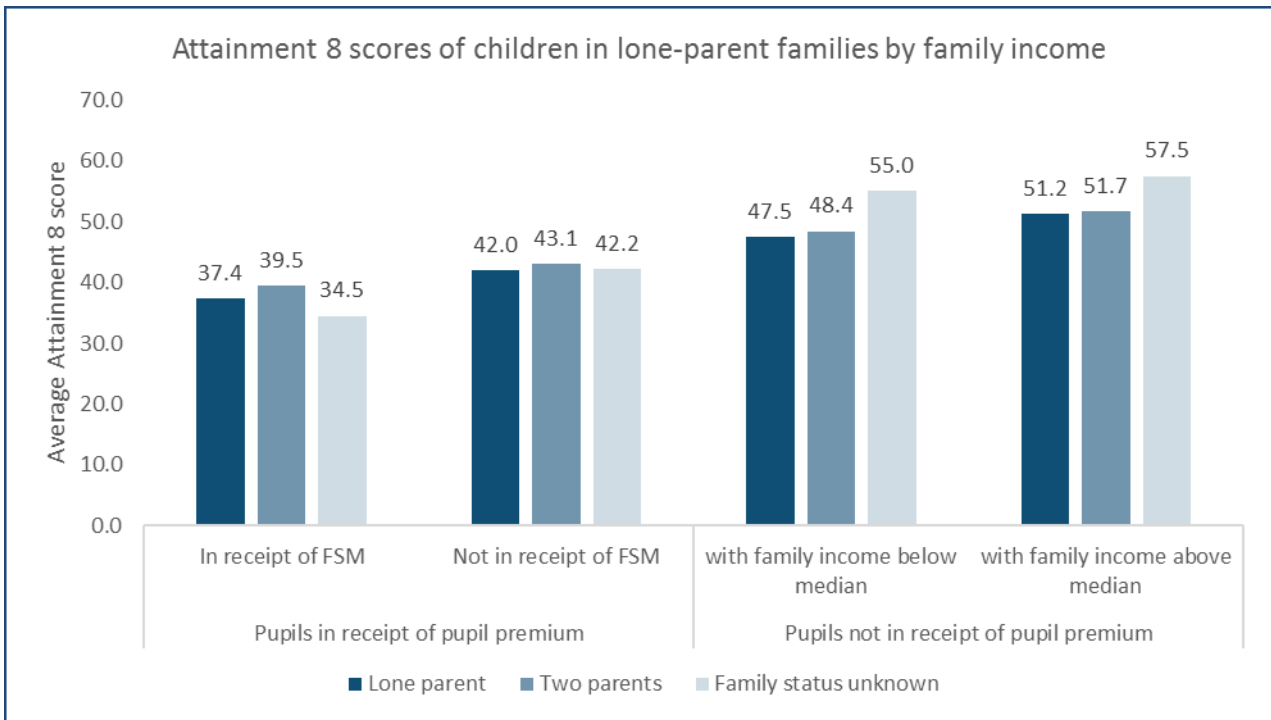
Figure 40: Attainment 8 scores of children in lone parent families by family income

107.   This point is clearer still at key stage 2, as seen in figure 41. Across all income groups, children from lone parent families slightly outperform those from two parent families. Again, the margin is small but the consistency when income is controlled for is notable given previous suggestions that being in a lone parent family was an inherent educational disadvantage.

Figure 41: Percentage achieving the expected standard in key stage 2 reading in lone parent families by family income

108.   While further investigation is warranted, particularly of unknown family status, this initial finding suggests that once income is adjusted for pupils in lone parent families perform slightly better than peers in two-parent families at key stage 2, but make slightly less progress into key stage 4.

Ordinary working families

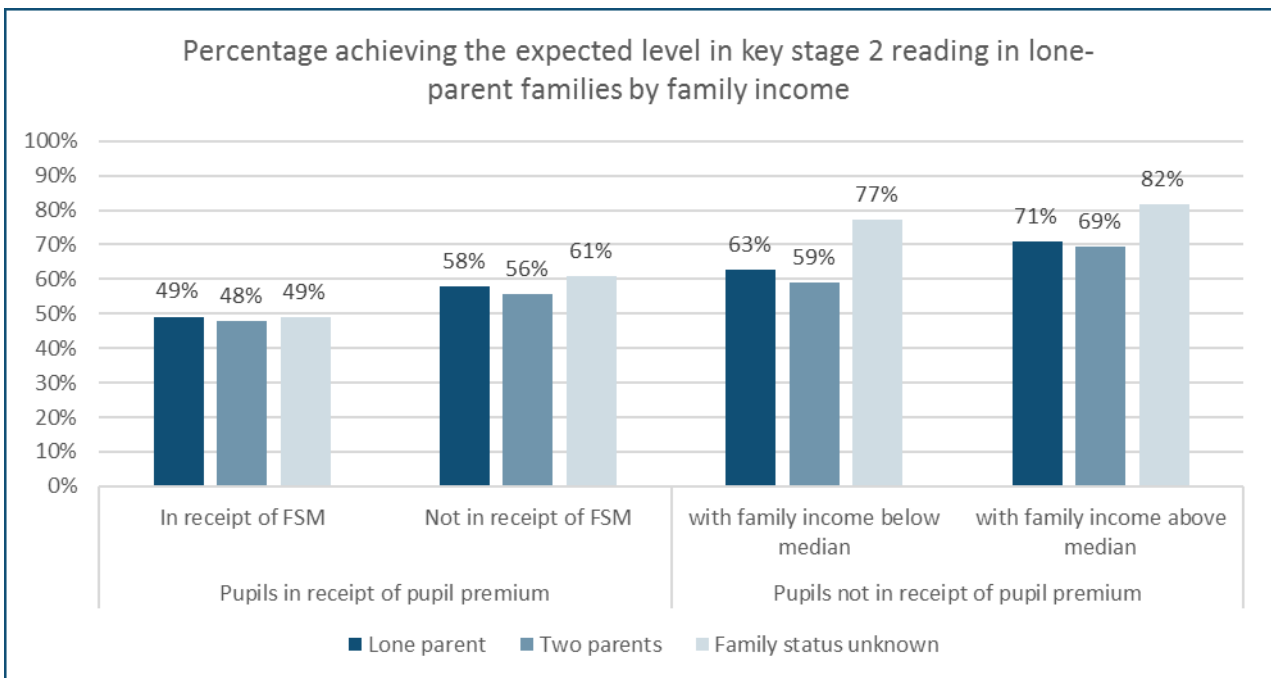109.   Within Department for Education statistics we have terms and definitions which are not used in other Departments. For example, the term disadvantaged is used to refer to those eligible for pupil premium in DfE statistics but not elsewhere.

110.   There is no official definition of an ordinary working family. However, the Government frequently refers to a group who are not entitled to pupil premium but earning modest incomes as ordinary working families. This document has set out an approach to looking at this group through better understanding of those families who are below median household equivalised income but who are not eligible for pupil premium. This approach shows there are around a third of children in these families.

111.   Unlike eligibility for pupil premium or free school meals which derives directly from Government policy, this is necessarily an analytical approach intended to provide a basis for a clearer analysis of educational outcomes for ordinary working families. We would your views on the statistical analysis and methodology outlined here for looking at this group, and how to refer to them in DfE publications.

Privacy controls and access

112.   The Department for Education takes very seriously the responsibility to handle data securely and appropriately. As part of this a privacy notice has been published, alongside this document, for the work carried out so far. In addition to this, full privacy impact assessments will be undertaken for the expanded database as well.

113.   Permitted Uses of the National Pupil Database are set out in the Education Act, section 537A. Sharing of data is permitted where for the purpose of promoting the education or well-being children in England are:-

- conducting research or analysis;
- producing statistics; or
- providing information, advice or guidance.

114.   Users must demonstrate the above, along with compliance with the Data Protection Act, proportionality of use (i.e. only asking for data needed to meet the objective) along with satisfying strict security standards.

115.   The matching of National Pupil Database Information with HMRC and DWP data is more specifically covered within The Education and Skills Act, as amended by The Small Business, Enterprise and Employment Act. The conditions of that act restricts use of the data to being for evaluating the effectiveness of education provision and assessment of policy related to that education and there is no proposal to change this. This database can therefore only be used for specified research purposes.

116.   The law also requires that, as far as is practical, the data must be anonymised to those who are carrying out the analysis, and identifying individuals for any operational purpose is prohibited. Access to names, date of births and addresses is restricted to government employees and contractors who need access in order to link together the information. The linked data produced and made accessible to analysts for research purposes has names, dates of birth and addresses, along with any other identifying information removed.

117.   All government employees and contractors must agree to our terms and conditions before accessing the linked information for research purposes, and measures are in place to prevent any such employees or contractors accessing other sensitive data sources which could potentially allow identification. We will monitor use of the linked information and if users fail to meet these conditions, access will be removed and we may consider further action.

118.   On 16 February, the Head of Profession for Statistics at DfE set out that our system for data access was under review and we will be reporting on this in early May. While it is our intention to make the database available for research purposes once it is of sufficient quality, we will need to make sure access is in line with highest modern
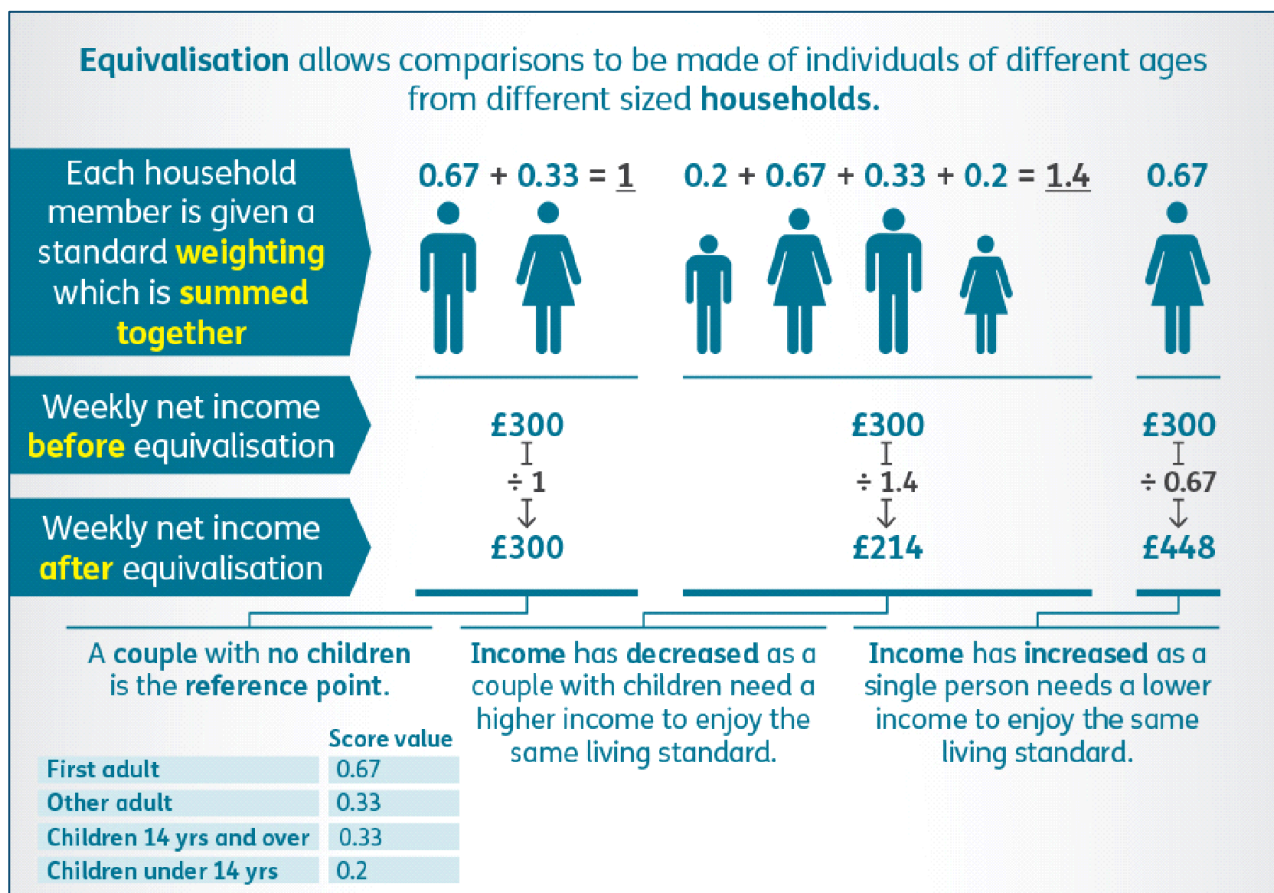
standards of security and privacy protection. We will be working with the Office for National Statistics and Administrative Data Research Network to develop options.

Annex A - Equivalising Household Income

Once household incomes, including benefits, have been determined through data matching, they undergo a process of equivalisation. Equivalisation adjusts incomes for household size and composition, taking an adult couple with no children as the reference point. For example, the process of equivalisation would adjust the income of a single person upwards, so their income can be compared directly to the standard of living for a couple.

The process used applies a weighting to each family member, using the Modified OECD Equivalence Scale, which is commonly used in government statistics for this purpose.

The process is demonstrated in the following graphic:

Equivalisation allows comparisons to be made of individuals of different ages from different sized **households.**

Each household member is given a standard **weighting** which is **summed together**

$0.67 + 0.33 = \underline{1}$ $0.2 + 0.67 + 0.33 + 0.2 = \underline{1.4}$ 0.67

Weekly net income **before** equivalisation

£300   £300   £300
÷ 1   ÷ 1.4   ÷ 0.67

Weekly net income **after** equivalisation

£300   £214   £448

A **couple** with **no children** is the **reference point.**

Income has **decreased** as a couple with children need a higher income to enjoy the same living standard.

Income has **increased** as a single person needs a lower income to enjoy the same living standard.

	Score value
First adult	0.67
Other adult	0.33
Children 14 yrs and over	0.33
Children under 14 yrs	0.2

Originally published in: Households below average income (HBAI) statistics on gov.uk

Annex B – Calculating Local Authority Housing Costs

There is no dataset that exists which has every family's housing costs. So, we have developed a methodology in partnership with DWP and ONS to get as local an indicator of housing costs possible and adjust family incomes by this. The method in detail is as follows:

We have taken 13 years of data on rental (both public and private sector) and mortgage costs (based on repayment mortgages, including both capital repayments and interest payments) from the Family Resources Survey. This covers the period 2003/04 (before which mortgages were classified differently) and 2015/16 (latest data). This amount of data was used in order to enable more local analysis. Households who own their property outright, who are buying with a non-repayment mortgage, or are part-renting and part-buying are excluded from the calculation. This means each household included has either a rental amount or a mortgage payment but not both.

In order to align with the analysis year, each housing cost in a survey year has been inflated by a constant amount calculated as the ratio of the UK average (median) 2015/16 housing costs to the survey year's overall average (median) housing cost using data from the FRS to ensure consistency.

We have then created an average (median) housing cost for each local authority district in England using the 13 years of data. The median was used in preference to the mean so outliers did not affect the calculation. This was then subtracted this from each family's income. We then equivalised the family income on the after housing costs basis to account for variations in family size. We validated this analysis by comparing the ranking of Local Authority areas in terms of private rents with a publicly available Valuation Office Agency results, as well as seeing what impact using census 2011 tenure mix weights instead of survey weights made.

This approach has some weaknesses – in that it does not obviously adjust for every individual's experience and in all local authority districts there will be variation of housing costs. While the estimates have been adjusted for non-response, it is the case that for some local authority districts the sample size is small even after combining 13 years of data and that the stratified cluster sampling will also affect the coverage of local authorities. However, it clearly will adjust for the large differences that exist between local authority districts and help ensure the analysis is robust to these.

Follow us on Twitter: @educationgovuk

Like us on Facebook: facebook.com/educationgovuk