COSMO COVID Social Mobility & Opportunities Study

Wave 1 June 2023

Technical Report

Kantar Public – Keith Bolling, Becky Hamlyn, Jonathan Kennett, Luke Taylor, David Xu

Supported by UK Research and Innovation

Partners The Sutton Trust, Centre for Longitudinal Studies, UCL Centre for Education Policy & Equalising Opportunities

In collaboration with Kantar Public

Contact

For queries about this report: jonathan.kennett@kantar.com

To contact the research team at UCL: cosmostudy@ucl.ac.uk

Authors

Kantar Public - Keith Bolling, Becky Hamlyn, Jonathan Kennett, Luke Taylor, David Xu.

How to cite this report

Bolling, K., Hamlyn, B., Kennett, J., Taylor, L., Xu, D. (2023) COVID Social Mobility & Opportunities study (COSMO): Wave 1 Technical Report. London: UCL and Sutton Trust.

This guide was published in May 2023 by the UCL Centre for Education Policy and Equalising Opportunities (CEPEO), UCL Centre for Longitudinal Studies (CLS), and Sutton Trust.

The UCL Centre for Education Policy and Equalising Opportunities (CEPEO) is a research centre that carries out cutting-edge research focused on equalising opportunities across the life course. Its work seeks ways to improve education policy and wider practices to achieve this goal. For more information, visit www.ucl.ac.uk/ioe/cepeo.

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Social Research Institute, University College London. It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ucl.ac.uk.

The Sutton Trust champions social mobility through programmes, research and policy influence. Since 1997 and under the leadership of founder Sir Peter Lampl, the Sutton Trust has worked to address low levels of social mobility in the UK. The Trust works to improve social mobility from birth to the workplace so that every young person – no matter who their parents are, what school they go to, or where they live – has the chance to succeed in life. For more information, visit www.suttontrust.com.

Kantar Public UK is an independent research agency that manages high quality cross-sectional and longitudinal research to inform the development of public policy, public service delivery and public communication on behalf of UK government and public sector organisations, academic institutions, and charities. For more information, visit www.kantarpublic.com/.

Please visit the study website for initial findings, questionnaires, and news and updates: www.cosmostudy.uk.

Contents

1 Introduction

The COVID Social Mobility and Opportunities (COSMO) study is a national longitudinal cohort study in England set up in 2021 to understand how young people's lives have been affected by the COVID-19 pandemic. This technical report provides an account of the design, development and methodology of the first wave of the survey (Wave 1) which was fielded between September 2021 and April 2022.

The COSMO study has been branded as 'Horizons' in all respondent-facing communications.

1.1 Background and objectives

The cohort of young people who were in Year 11 in the 2020/21 academic year (those born between September 2004 and August 2005) arguably suffered the most acute effects of COVID-19 on their schooling, experiencing severe disruption at a crucial transitional stage in their education.

The COSMO study seeks to generate high-quality evidence to answer the central research question of how the COVID-19 pandemic affects socio-economic inequalities in life chances, both in terms of short-term effects on educational attainment and well-being, and long-term educational and career outcomes for this cohort.

To achieve this aim, a representative sample of young people who were in Year 11 in the 2020/2021 academic year across England were invited to a baseline survey, with the intention of following them over time as they progress through the final stages of their education and into the labour market. Although the young person is the core unit of analysis, the study also sought to include a paired survey questionnaire with a parent or guardian[1] of the young person to complement the young person's data. To further enrich the data, the study also collected consent from young people for linkage to administrative data[2] on post-16 exam performance and other external data sources.

[1] Any parent or guardian of a sampled young person was eligible for this survey. "Parents/guardians" and "parents" are used interchangeably in this report.
[2] Consent was sought to link survey data to records from Department for Education (DfE), HM Revenue & Customs (HMRC), Department for Work & Pensions (DWP), Higher Education Access Tracker (HEAT) and the National Tutoring Programme (NTP)

The Wave 1 surveys of young people and their parents collect information on a wide range of factors including the socio-economic status of the young people and their households, their educational experiences and wellbeing during the COVID-19 pandemic, their experiences of returning to school, and catching up with disrupted learning.

COSMO is carried out by a collaboration of UCL Centre for Education Policy & Equalising Opportunities (CEPEO), the UCL Centre for Longitudinal Studies (CLS), the Sutton Trust, and Kantar Public who carried out the data collection for Wave 1. Kantar was supported by NatCen Social Research during the face-to-face stage of the fieldwork.

This first wave of the study was funded by UK Research and Innovation (through the Economic and Social Research Council) as part of its COVID-19 rapid response funding [ES/W001756/1]. In addition, the Sutton Trust funded an 'add on' to the main study, focusing on disadvantaged young people with high prior attainment. This is referred to as the Sutton Trust boost sample throughout this report.

The project is further supported by key stakeholders to ensure co-production of policy-relevant evidence including: the Department for Education (DfE), the Office for Students (OfS), Administrative Data Research (ADR UK), the Education Endowment Foundation (EEF), Transforming Access and Student Outcomes in Higher Education (TASO).

The study design and survey processes for COSMO were approved by the UCL IOE Research Ethics Committee.

1.2 Summary of methodology

This report provides a full account of the methodology. An overview is provided below while full details are provided in the relevant chapters.

The survey was initially set up to cover three audiences: young people, parents and school staff. However, despite significant attempts to recruit staff to a schools-based survey, it proved too difficult to secure their co-operation due to staff shortages and increased workloads faced by schools during the pandemic. Therefore, the schools survey element was eventually dropped (for further information on this stage, see Appendix A).

1.2.1 Survey sample

The household-based survey of young people and parents was based on the cohort of young people studying in Year 11 in the 2020/2021 academic year, that is those born between September 2004 and August 2005. The sample - which included both the main sample and the Sutton Trust boost - was mainly based on young people attending state schools in England during the 2020/21 academic year. In addition, a small supplementary

sample of students who were attending independent schools in England in the equivalent cohort was also included (see section 1.2.3 below for an overview of this stage).

For the state school sample, addresses of young people were selected from the National Pupil Database (NPD) which is maintained by the Department for Education (DfE). The sample was clustered by schools to improve the efficiency of face-to-face fieldwork at later stages. When drawing the NPD samples, pupils from more disadvantaged backgrounds were over-sampled within the original issue schools that were selected for the main study to improve the representation of these groups in the final sample. Overall, 35,719 students were included in the issued NPD sample (33,719 in the main sample and 2,000 in the boost sample).

For each sampled address, a parent of the young person living at the same address was also invited to take part in the survey.

1.2.2 Fieldwork

Original issue sample

Fieldwork ran between September 2021 and April 2022, at which stage the cohort of young people were in Year 12. The target questionnaire length was 30 minutes, for both young people and parents.

Throughout fieldwork, efforts were made to maximise the number of households where both the young person and a parent participated, as this provides a more complete picture of household characteristics. Within each household, only one parent was asked to complete the questionnaire, and any resident parent could choose to do this.

COSMO used a sequential mixed-mode design which comprised an initial online data collection phase followed by in-home interviewing for a proportion of non-responding households.

The online phase took place between September 2021 and November 2021, and this consisted of a launch mailing followed by up to 4 reminders. A subsequent face-to-face stage was included in the design which was planned to meet the following objectives: i) improve response rates among young people and parents; ii) help achieve overall target sample sizes; iii) help improve the sample balance; and iv) help increase the rate of complete households where only one eligible household member had already competed online.

The initial survey design assumed a targeted sub-sample of c.50% of non-responding households after the online phase would be issued face-to-face; the budget allowed for c. 9,480 households to be issued to face-to-face.

To achieve the optimal balance between the objectives outlined above, all partially complete households where only a parent or a young person (but not both) had

responded online were allocated to face-to-face, followed by households with the lowest response probabilities across the whole sample up to the maximum budgeted sample size.

However, the objectives for the face-to-face fieldwork were not fully met due to COVID-19 related challenges. Across the industry, face-to-face interviewing re-started in England gradually from October 2021.The original plan for COSMO was for face-to-face fieldwork to be conducted between November 2021 and March 2022, following the initial online fieldwork phase. However, this plan was adapted following further 'Plan B' COVID-19 restrictions introduced in December 2021 which meant that in-home face-to-face fieldwork had to be temporarily halted once again[3]. As an interim measure, Kantar Public carried out a 'knock to nudge' stage in February 2022 which involved interviewers knocking on doors to encourage young people and parents to complete online. Interviewers then returned to full in-home interviewing in March 2022, and the fieldwork timeline was extended until Easter 2022 to cover as much face-to-face fieldwork as possible within the overall more limited time available.

Reserve sample

However, the final number of face-to-face interviews achieved was still much lower than originally planned due to these unanticipated fieldwork disruptions and interviewer capacity issues. Therefore, to meet sample size targets for Wave 1, a decision was made in early 2022 to issue a fresh sample from a reserve sample of addresses which had been selected at the outset alongside the main sample of addresses. For the reserve sample, data were collected via online methods only between March 2022 and April 2022 (timing constraints meant that there was no possibility of a face-to-face follow-up for this group).

Incentivisation

Young people and their parents were offered a voucher conditional on survey completion to the value of either £10 or £20. Higher incentivisation was targeted at students and their parents expected to be from more disadvantaged backgrounds, based on information from the sampling frame, to help boost response amongst these groups.

A more detailed summary and timeline of the different fieldwork stages is provided below at Table 1.3.

1.2.3 Supplementary independent school sample

There is no pre-existing sample frame of students in independent schools and therefore this sample needed to be generated via a random sample of independent schools in England. Overall, 33 independent schools agreed to participate in the study and, following a within-school pupil selection stage, staff contacts at these schools sent survey

[3] https://www.gov.uk/government/news/prime-minister-confirms-move-to-plan-b-in-england

invitations to Year 12 students and their parents on Kantar's behalf. Due to logistics, independent school students were only contacted by web, and were not included as part of the face-to-face follow up. Given the different methods of sampling and fieldwork, the response rate among independent school students and their parents was much lower compared with students sampled via the NPD.

1.2.4 Achieved sample sizes and response rates

As noted above, overall, 35,719 students were included in the issued NPD sample. This comprised 33,719 in the main sample (22,719 original sample and 11,000 reserve) and 2,000 in the boost sample (1,600 original sample and 400 reserve).

The main and Sutton Trust boost achieved samples combined include 10,051 cases which comprise data from a matching young person and one of their parents, and a further 3,736 cases where the data only include a young person with no matching parent interview. This provides a total sample of n=13,787 young people. For the sample of young people without a paired parent interview, further attempts will be made to recruit a matching parent at Wave 2.

There were also 1,680 households (from the main and boost) where only the parent was successfully interviewed, with no matching young person interview. Although all surveyed cases have been included in the Wave 1 dataset, these cases have been given a zero-weight value in the dataset and will not be included at Wave 2. Only the 13,787 complete households or partial households where an interview was achieved with a young person will be included in the Wave 1 survey analysis and taken forward to Wave 2.

The total Wave 1 sample of 13,787 young people was made up of 13,113 young people sampled from the NPD (state school) sample and 674 from the independent school sample. Within the NPD sample of 13,113 young people, the following useable sample sizes were achieved: 12,154 main sample young person interviews and 959 young person interviews as part of the Sutton Trust boost.

The final response rates for the NPD sample are summarised below. It is worth noting that the need to issue the reserve sample meant that the final response rate to the study was lower than it would have been if we had only included the original issued sample only.

Table 1.1. Response rates for the NPD main and boost sample combined: from online and face-to-face stages

	Issued sample	Achieved sample	Response rate
Young people	35,719	13,113	36.7%
Parents	35,719	11,368	31.8%
Complete household	35,719	9,845	27.6%

Of the 13,787 young people surveyed (including independent school students), 13,445 completed online and 342 were interviewed in person. Of the 9,330 parents surveyed within complete households (including parents of independent school students), 8,918 completed online and 412 were interviewed in person.

As most of the survey responses were achieved from the online stage, a summary of the final online response rates for the NPD sample are summarised below. Since complete households would have been made up a mixture of online and face-to-face interviews, the table only shows response rates for young people and parents individually.

Table 1.2. Response rates for the NPD main and boost sample combined: online responses only

	Issued sample	Achieved sample	Response rate
Young people	35,719	12,771	35.8%
Parents	35,719	10,850	30.4%

The final achieved sample is of a high quality, as demonstrated by the fact that the achieved sample profile (with design weighting applied) was a good match to the population profile (see section 11.4).

1.3 Summary of fieldwork and timeline

A summary of the different stages of fieldwork and the associated timeline is provided in Table 1.3 below.

Table 1.3: Summary of survey stages and timeline

Fieldwork phase	**Sample subgroup**	**Dates**
Original issued sample		
Initial web survey launch letter	All original issued sample	22 September 2021
Web survey reminder 1 letter	All remaining non-responders	8 October 2021
Web survey reminder 2 letter	All remaining non-responders	20 October 2021
Email to young people in unpaired households to remind them to ask their parents to complete the survey	All young people respondents where a parent interview had not yet been achieved, and who had provided an email address	28 October 2021
Initial F2F stage (halted early due to further COVID-19 restrictions)	All non-responders selected for F2F stage	10 November - 10 December 2021
Break-off letter	Remaining non-responders from original sample not issued to initial F2F stage, who had broken off the web survey before reaching the threshold for a complete interview	17 November 2021
Web survey reminder 3A letter	Remaining non-responders from original sample not issued to initial F2F stage	13 December 2021
Web survey reminder 3B letter	Remaining non-responders from original sample issued to initial F2F stage who had not been contacted by F2F by this stage due to further COVID-19 restrictions	21 December 2021
Knock-to-nudge stage	All remaining non-responders from initial F2F allocation plus some additional YP and parents in new 'unpaired' households created from the previous web mailing (3A)	2 February – 6 March 2022
Return to full face-to-face	A subset of remaining non-responders from the F2F allocated sample above	7 March - 18 April 2022
Web survey reminder 4 letter (final mop-up web reminder)	All remaining unpaired parents/YP in households which had been allocated to F2F but were not in the event contacted during the F2F stage	8 April 2022

Table 1.3 (continued)		
Reserve sample		
Initial web launch letter	All reserve issued sample	17 March 2022
Web survey reminder 1 letter	All remaining non-responders	29 March 2022
Web survey reminder 2 letter	All remaining non-responders	7 April 2022
Independent school sample		
Recruitment of schools	All sampled independent schools	27 May 2021- July 2021
Web fieldwork phase	All students and their parents selected from the independent schools recruited for the study	October 2021-18 April 2022
All samples		
Fieldwork close	All samples	18 April 2022

2 Sample design: state schools

The target population for this study consists of all young people (and their parents/carers) in England studying in Year 11 (Y11) in the 2020/2021 academic year.

In this chapter we outline the sample design used for pupils that were in Y11 and educated in a state school during 2020/21. The sample design used for independent school pupils can be found in Chapter 7.

2.1 Summary of sample design

The sample was drawn from the DfE National Pupil Database (NPD). The key aspects of the state school sample design were as follows:

(1) A two-stage sample design

It was planned that COSMO would use a sequential mixed-mode design. This was anticipated to consist of an online data collection phase followed by face-to-face in-home interviewing[4].

To support this approach, it was necessary to cluster sampled pupils geographically. Clustering improves the efficiency of in-home interviewing by minimising interviewer travel time and so is substantially more cost-effective than an un-clustered approach. The first stage of sampling was to draw a random sample of schools (as Primary Sampling Units (PSUs)). At the second stage, a random sample of pupils was selected from each sampled school.

(2) Disproportionate sampling

There are some sub-groups which are of substantial scientific interest, but which have a relatively low population incidence. The sample was designed to oversample pupils from the following disadvantaged backgrounds:

- Those eligible for Free School Meals (FSM) at any point over the last six years
- From the six main minority ethnic groups (Indian, Pakistani, Bangladeshi, Black Caribbean, Black African, and Mixed)

[4] Although in the event the face-to-face stage was considerably more limited than initially planned due to COVID-19 related challenges (see section 6.2)

- Those that speak English as an Additional Language (EAL)

(3) A large overall sample size to provide precise estimates

It is critical that sample sizes remain sufficiently large at the later waves of the study (taking into account likely attrition at later waves). As such, it was decided to target c.11,100 interviews with state school pupils (with an overall target of c.12,000 including independent school pupils).

Full details on the sampling approach used for state school educated young people is provided below.

2.2 Sample frame and exclusions

The NPD was used to sample Year 11 pupils in state schools, as recorded in the Spring 2020/2021 pupil-level census[5]. The use of the NPD as a sampling frame for state schools was made possible through a Data Sharing Agreement[6] between UCL, Kantar Public and the DfE, following an application to DfE NPD team.

The Spring Census data extract provided consisted of 580,450 records. This file was de-duplicated to ensure that pupils only appeared once (using the NPD "RecordStatus" variable). Following this, we were left with 580,278 valid records.

For efficiency reasons (to provide workable interviewer assignments for the face-to-face phase) very small schools were excluded from the study. Special schools and establishments that offer Alternative Provision tend to be very small, and we therefore used different thresholds for these schools[7]. The thresholds used were:

- Alternative provision - we excluded any schools with <5 pupils in Year 11. This excluded 18.4% of these establishments but only 2.1% of their pupils.
- Special schools – we excluded any schools with <5 pupils in Year 11. This excluded 10.5% of special schools but only 2.1% of their pupils.
- Other state schools – we excluded establishments with <30 pupils in Y11. This excluded 8.6% of schools but only 0.8% of their pupils.

With these exclusions in place, we were left with 575,708 pupils at 3,881 establishments (there were 4,285 establishments before this exclusion). This represents 99.2% coverage of Year 11 pupils in state schools.

[5] The fieldwork timings (beginning in September 2021) did not allow the 2021/2022 NPD to be used for the state school sampling.
[6] DSAP number DS 00554.
[7] If we had used the same threshold for all schools, we would have excluded nearly all special schools and Alternative Provision.

It should be noted that these non-covered pupils remain part of the target population and the weighting design (outlined in Chapter 11) is designed to compensate for this non-coverage.

2.3 Sample design

The final sample design was informed by the population profile for the following variables: FSM eligibility, ethnic minority background, speaking EAL. The objective of the sample design was to boost these groups to allow for robust standalone analysis, while ensuring that this did not have too detrimental an impact on the precision of overall estimates.[8] The following tables show the profile of the population (for the three variables used for the disproportionate sample design).[9]

Table 2.1: Population counts

	N	%
Total	**575,708**	**100**
FSM eligibility (last six years)		
No	427,049	74.2
Yes	148,659	25.8
Ethnicity		
Indian	16,851	2.9
Pakistani	25,740	4.5
Bangladeshi	10,277	1.8
Black Caribbean	7,209	1.3
Black African	23,236	4.0
Mixed	31,945	5.5
White & other (incl. missing & refused)	460,450	80.0
English as an Additional Language		
EAL	97,471	16.9
Not EAL (or missing)	478,237	83.1

Based on this descriptive analysis, Kantar Public agreed with UCL and Sutton Trust to draw a sample with the following targets.

[8] Or too detrimental an impact on other sub-group analysis.
[9] This table is based on the population that remained after excluding small schools. Although it should be noted that the profile based on the 580,278 valid records is very similar.

- 46% of sampled individuals in the sample to have been eligible for FSM in the last six years (up from 25.8%)
- The six main minority ethnic groups to each be boosted up to c.5.5% of the sample:
  - Indian (from 2.9%)
  - Pakistani (from 4.5%)
  - Bangladeshi (from 1.8%)
  - Black Caribbean (from 1.3%)
  - Black African (from 4.0%)
  - Mixed (already 5.5% in the population)

We decided not to explicitly stratify by EAL as, by boosting the groups listed above, we would also slightly improve the representation of those that speak English as an Additional Language. Based on our calculations, we expected c.21% of the drawn sample to be part of this group (it is 17% of the population).

We estimated that with design weighting applied, the design effect due to this disproportionate sample design would be c.1.39.

2.4 Drawing the sample

Stage 1 - Drawing the PSUs (schools)

At stage one, we sampled 750 schools (460 for original issue and a reserve of 290) using a PPS (Probability Proportionate to Size) approach. The target number of pupils to sample from each school was set at 49[10]. To support the sample design, each pupil was classified into a single stratification category and provisional pupil-level sampling probabilities were calculated (as per table below).

Table 2.2: Calculating provisional sampling probabilities

	Population	Target to sample	Provisional sampling probability *p(provisional pupil)*
No FSM – Indian	14,820	1,778	0.120
No FSM – Pakistani	17,286	1,357	0.079
No FSM – Bangladeshi	6,069	1,194	0.197
No FSM - Black Caribbean	3,852	1,080	0.280
No FSM - Black African	13,339	1,160	0.087

[10] The mean value was 49, but some variation was necessary as some schools had fewer than 49 Y11 pupils.

No FSM - Mixed	20,740	1,324	0.064
No FSM - Other (incl. white, missing, refused)	350,943	11,952	0.034
FSM - Indian	2,031	244	0.120
FSM - Pakistani	8,454	664	0.079
FSM - Bangladeshi	4,208	828	0.197
FSM - Black Caribbean	3,357	941	0.280
FSM - Black African	9,897	861	0.087
FSM - Mixed	11,205	698	0.062
FSM - Other (incl. white, missing, refused)	109,507	12,670	0.116
Total	575,708	36,750	n/a

The provisional pupil-level sampling probabilities (provided in the table above) were then aggregated to the school-level (using URN[11]) to calculate each school's size measure for the PPS sampling procedure.

The Measure of Size for *school a* was calculated as follows:

$$MoS_a = \sum p(provisional\ pupil)_a$$

This size measure can be interpreted as the number of Y11 pupils at each school weighted by their value to the study (i.e., pupils in groups we want to oversample received a larger weight). This approach was used as it should allow the disproportionate sample design to be implemented while retaining more or less equal school-level sample sizes (some variation was still necessary due to some schools having fewer than 50 pupils).

The school sampling probability for *school a* was then calculated as follows:

$$p(School\ a\ selected) = \frac{MoS_a}{\sum MoS} \times 750$$

Prior to selection, schools were implicitly stratified using the following variables:

- Establishment type: Special / Alternative Provision / Other
- Admissions policy[12]: Selective / Non-selective / Missing or NA
- Region: the nine former Government Office Regions

[11] This is the "School Unique Reference Number" used by the Department for Education.
[12] It should be noted that this does not vary in Alternative Provision and Special schools. As such this stratification variable was only used for "other" types of establishment.

Finally, a systematic random sample of schools was drawn with school sampling probabilities as previously calculated. The sampled schools were then systematically allocated to the main (460 establishments) and reserve pools (290 establishments) at random. This allocation used the same implicit stratification as for the overall sample of schools.

Stage 2 – Sampling Pupils

For PSUs with more than 50 pupils (725 of the 750 schools selected at Stage 1), a PPS sample of 50 pupils[13] was drawn. The provisional within-school pupil sampling probability was calculated as:

$$p(pupil\ h\ at\ school\ a\ selected) = \frac{p(provisional\ pupil)_h}{\sum p(provisional\ pupil)_a} \times 50$$

In some instances, the within school pupil sampling probabilities exceeded 1. Where this was the case, these sampling probabilities were capped at 1 and the provisional within-school pupil sampling probabilities were automatically increased for all other pupils at the affected school to ensure a sample of 50 per school. This was done iteratively until no pupil had a within-school sampling probability of more than one. These values were then used as the final within-school pupil sampling probabilities.

For PSUs with fewer than 50 pupils (25 of the 750 schools selected at Stage 1), all Y11 pupils were selected for the study. For these pupils the within school pupil sampling probability is:

$$p(pupil\ h\ at\ school\ a\ selected) = 1$$

Prior to selection, pupils were implicitly stratified within each school using the following variables:

- Being eligible for FSM in the last 6 years: Yes / No
- Ethnic minority group: Indian / Bangladeshi / Pakistani / Black African / Black Caribbean / Mixed / Other
- Gender: Male / Female
- Special Educational Needs (SEN): Education, Health and Care Plan / SEN support / No Special Educational Need

A systematic random sample of pupils was then drawn using the final within-school pupil sampling probabilities[14]. In total, 36,994 pupils were sampled for the study. Within the

[13] This is above 49 to compensate for the 25 establishments where <50 could be sampled
[14] It should be noted that it was possible for more than one young person to be sampled from the same household (e.g., if there were twins / siblings). 235 of the individuals sampled for the main stage (out of 36,994) lived in a household where at least one other individual was sampled.

460 schools allocated to original issue, 22,719 pupils were sampled. Within the 290 schools allocated to the reserve, 14,275 pupils were sampled.

As will be discussed in Chapter 6, some reserve sampled ended up being issued into the field. Reserve sample was selected from all 290 reserve schools – a random systematic sample of 11,000 was selected from the available reserve cases in these PSUs (14,275). In the end, there were 3,275 reserve cases that were not issued into the field.

The overall NPD pupil sampling probability was calculated by multiplying the school sampling probability by the within-school pupil sampling probability (and accounting for only some of the reserve sample being issued into field):

$$p_{NPDstudent} = \frac{460}{750} * p(School\ a\ selected) * p(pupil\ h\ at\ school\ a\ selected) + \frac{290}{750} * p(School\ a\ selected) * p(pupil\ h\ at\ school\ a\ selected) * \frac{11{,}000}{14{,}275}$$

Checking the sample selection

Once the sample was selected, checks were carried out to ensure that the selected sample matched the intended sample design.

As shown in the table below, the sample drawn for the study closely matches the individual-level targets set. The only minor exception is for Black Caribbean which make up c.5% of the selected sample rather than the c.5.5% targeted.

This discrepancy was because these pupils were heavily clustered in a small number of schools. Representation of this group could potentially have been increased to 5.5% of the sample by allowing substantial variation[15] in the number of pupils sampled from each school. On balance it was agreed that it was preferable to minimise variation in school sample sizes and to continue with this sample design.

[15] In some schools, more than 80 pupils would need to be sampled to support this design.

Table 2.3: Checking sample selection vs targets					
	Target	Original + Reserve (750 PSUs)		Original only (460 PSUs)	
	%	n	%	n	%
Total	-	**36,994**	-	**22,719**	-
No FSM (last 6y)	54%	20,167	54.5%	12,315	54.2%
FSM (last 6y)	46%	16,827	45.5%	10,404	45.8%
Indian	5.5%	2,046	5.5%	1,226	5.4%
Pakistani	5.5%	1,993	5.4%	1,353	6.0%
Bangladeshi	5.5%	1,944	5.3%	1,267	5.6%
Black Caribbean	5.5%	1,777	4.8%	1,104	4.9%
Black African	5.5%	2,067	5.6%	1,292	5.7%
Mixed	5.5%	2,062	5.6%	1,259	5.5%
Other	67%	25,105	67.9%	15,218	67.0%
EAL	21%	7,939	21.5%	5,056	22.3%
Not EAL	79%	29,055	78.5%	17,663	77.7%

2.5 Sutton Trust Boost sample

The Sutton Trust boost sample was drawn after the main study sample was selected. The boost sample was drawn from the 460 schools selected as original issue for the main study (using the process described above). No Sutton Trust boost cases were sampled from the reserve schools.

The definition of pupils included in the boost sample was as follows:

- Eligible for FSM in last 6 years AND
- In the top 33% in the combined reading, maths, and GPS (Grammar, Punctuation and Spelling) KS2 score (the score weighted as follows: maths 50%, reading 25% & GPS 25%)

We aimed to interview c.1,060 pupils as part of the boost.

In the original issue sample selected for the main study there were 22,719 pupils (within the 460 schools selected as original issue). Of these young people, 1,976 were part of Sutton Trust's population of interest (as defined above).

Within these original issue PSUs, there were a further 2,868 pupils that were eligible for the Sutton Trust boost, that is part of this population of interest, that had not been selected for the main study. From these pupils, a further random sample of 2,000 were selected for the Sutton Trust boost (1,600 original issue and 400 reserve). As further explained in Chapter 6, all of the reserve sample which were part of the Sutton Trust boost were issued into the field.

2.6 Issued sample size and targets

In total, we aimed to achieve c.11,000 interviews for the main sample and a further c.1,060 interviews for the Sutton Trust boost. For the main study, 33,719 addresses were issued into field (22,719 original issue and 11,000 reserve). For the boost, 2,000 addresses were issued into field (1,600 original issue and 400 reserve). The following useable sample sizes were achieved: 12,154 young person interviews as part of the main sample and 959 young person interviews as part of the Sutton Trust boost.

3 Questionnaire development

This chapter describes the questionnaire content, questionnaire development, the approach for collecting consent for data linkage, and provision of contact information for future waves.

3.1 Questionnaire content

Two questionnaires[16] were developed:

- Young person questionnaire
- Parent questionnaire

The overarching aim of COSMO is to provide a representative data resource to support research into how the COVID-19 pandemic has affected the life chances of pupils with different characteristics, in terms of short-term effects on educational attainment and wellbeing, and long-term educational and career outcomes. The main unit of analysis is young people. However, data was also collected from a parent in the same household to complement the data collected from young people. The parent interview served to enrich the household data with information on socio-economic background and with direct reports of parents' experiences during the pandemic.

All questions for the Young Person and Parent questionnaires were designed to work in both web and face-to-face modes (see section 3.2.3 below). For the web survey, the entire questionnaire was self-completed online. For the face-to-face survey, the more sensitive questions were administered as self-completion (CASI) which respondents completed via the interviewer's tablet.

The full questionnaires, annotated with variable names, are provided on the COSMO study website[17] A summary of the content for each questionnaire is provided below.

3.1.1 Young Person questionnaire content

A core focus of the questionnaire was on disruptions to education due to the COVID-19 pandemic, covering the two major UK lockdowns that led to school closure (Lockdown 1: from April to July 2020, and Lockdown 3: from January to March 2021) as well as the time

[16] A third questionnaire was also developed for a schools-based survey. However, as detailed in Appendix A, this survey was eventually dropped given challenges associated with fieldwork during the pandemic.
[17] https://cosmostudy.uk/about/study-design-and-data-collection

in between when most schools were open (September to December 2020), and when young people returned to school after Lockdown 3 (March 2021 to July 2021).

To manage interview length and reduce burden on young people, two sections of the questionnaire were asked to random half samples. The first random half sample (Module A) received Section I: Cancelled Assessments, while the other random half sample (Module B) received Section K: Extra-Curricular Activities Pre- and Post-Pandemic.

All young people were asked for their consent to link some administrative data to their records (see section 3.3 below).

A summary of the content is provided below in Table 3.1.

Table 3.1: Summary of questionnaire coverage for young people

Section	Topics
A: Introduction, verification and opening demographics	• Survey Introduction • Verification of NPD sample details (name, address) • Demographics including sex, gender and date of birth
B: Household grid	• Number of other household members, and their age, gender and relationship to the young person
C: Current status	• Current status (all activities) and main status • Details of jobs, apprenticeships, training courses • If not in education, employment or training (NEET): whether looking for work and reasons associated with this status
D: Qualifications studying	• Place of study/training • Number and types of academic and vocational qualifications working towards, and subjects studied
E: Education during lockdown 1/Year 10 (April-July 2020)	• Whether attended school in person (e.g. because a parent was key worker) • Time spent on schoolwork • Remote learning and provision, and contact with teachers/tutors • Access to devices for online learning • Problems related to studying during this period
F. Education during lockdown 3/Year 11 (January-March 2021)	• As for Lockdown 1 (see above)
G. Education during Year 11 when schools were open (September-December 2020 and March-July 2021)	• Reasons for non-attendance during this period and whether affected by school closures or bubble closures • Estimated total absence during this period
H: Catch up	• Provision of catch-up activities • Perceived impact of pandemic on education and motivation
I: Cancelled assessments (asked to a random half sample)	• Experience of teacher assessments in place of exams • Whether exam results were worse or better than expected and impact on future plans • Intentions to re-sit GCSEs
J. Education and career aspirations	• Perceived likelihood of applying to or getting into university and reasons for not planning university • Planned university courses for those considering this • Most likely activity in two years' time • Plans for vocational qualifications in the future • Attitudes towards future life (importance of having a job/career, raising a family, etc.) • Changes in future educational and career plans because of the pandemic • Participation in activities about careers advice and informal careers advice (family, friends etc.)

	• Sutton Trust Boost sample questions: Awareness of educational access and support programs, and whether has applied to them

K. Extra-curricular activities pre and post-pandemic (asked to a random half sample)	• Participation in extra-curricular activities in Year 10 and whether organised by school or outside of school
	• Participation in extra-curricular activities in Year 11 after schools re-opened and whether organised by school or outside of school
	• Extra-curricular activities asked about included:
	  ○ Sports and exercise
	  ○ Other clubs (arts, crafts, music, drama, etc.)
	  ○ Classes associated with church/religion
	  ○ Voluntary or community work
	  ○ Activities that involved overnight stays (such as Duke of Edinburgh)

L. Attitudes to education (including motivation)	• Attitudinal questions measuring locus of control
M. Health and wellbeing (CASI)	• Experience of COVID-19 including long COVID • Experience of major life events since start of pandemic • Mental health and wellbeing scales (Rosenberg scale, GHQ-12, GAD2, PHQ-2) • Life satisfaction • Self-assessed general health
N. Friends, peers and family support (CASI)	• Peer support and social provisions sale • Experience of bullying and harassment including cyber harassment and discrimination • Perception of school support for mental health • Caring responsibilities • Ethnicity (see note* below this table)
O. Health Related Behaviours (CASI)	• Cigarette and e-cigarette use • Drug use • Sleep habits • Physical exercise • Self-harm

P. Linkage	• Linkage consent asked to link records from:
	  ○ Department for Education (DfE)
	  ○ Education Endowment Foundation
	  ○ Higher Education Access Tracker
	  ○ Department for Work and Pensions (DWP)
	  ○ HM Revenue and Customs (HMRC)

Q. Recontact, signposts and closing screens	• Updating of young person's contact details for future waves, signposting to sources of support and closing

* Due to a questionnaire error, based on incorrect assumptions about the use of information provided on the sample frame, the young person's ethnicity was not asked to the original NPD sample (as it was erroneously thought this could be accessed via NPD linkage). This error was corrected for the issue of reserve sample. To help fill in the missing data for original issue NPD sample, young people were asked to supply information on ethnicity as part of the between-wave keeping in touch exercise that was administered in August 2022 (see section 6.8).

3.1.2 Parent questionnaire content

The main focus of the parent/guardian questionnaire was to complement the information obtained from young people and to provide more context on household demographics. Questions included but were not limited to parents' level of education, working status throughout the pandemic, occupation, income, and ethnicity, all of which provide important background information on young people. A *household reference person* approach was used when collecting information for occupation coding, so that this measure would be less dependent on the responding parent/guardian. Sections on parenting and parents' attitudes to education also help to contextualise young people's experiences.

There were also questions on parents' experiences over the course of the pandemic, particularly around COVID-19 related disruptions to education, home learning and tuition, as well as impacts on household finances, and family life.

Parents'/guardians' own health and wellbeing, including their COVID-19 infection and vaccination status are also covered.

A summary of the parent questionnaire is provided below.

Table 3.2: Summary of questionnaire coverage for parents

Section	Topics
A. Introduction and verification checks	• Survey Introduction • Verification of NPD sample details (name of young person, address) • Demographics including gender, age, relationship to young person, marital status and which parents live at the same address as the young person
B. Attitudes to Education	• Engagement with school reports received by young person • Parental aspirations and views on likelihood of whether young person will go to university (and reasons if not) • Parental attitudes to education
C. Parenting, home learning, tuition & catch-up	• Level of parenting control (e.g. knowing where they are, setting times for being back home) and how close they feel to the young person • Parental help provided during Lockdown 1 and Lockdown 3 • Non-attendance of young person at school due to COVID -19 • Attitudes to remote learning during the pandemic • Use of private tutors over the course of the pandemic and lockdowns and any other additional COVID-related expenditure (laptops etc.) • Perception of impact of pandemic on young person's progress • Contact with school on issues relating to COVID-19
D. Working status across the pandemic	• Main status of parent before the pandemic • Work history covering from before the beginning of the pandemic until survey date (each unique status and date they ended)

	• Whether parent was classified as a key or critical worker during the pandemic • Location of work (home, office, hybrid etc.) • Whether parent experienced any changes related to their working status over the course of the pandemic (whether furloughed, whether took a pay cut, etc.) • Details of partner's current status where applicable
E. Parental tenure, HRP and occupational details	• Tenure • Establishment of Household Reference Person (HRP) • For the HRP: Details of last job and questions to establish NS-SEC coding.
F. Parental education	• Highest academic and/or vocational qualification • Partner's highest academic and/or vocational qualification
G. Parental income	• Sources of income and estimate of income collected in bands
H. COVID History and vaccination (CASI)	• Vaccination status • Whether needed to self-isolate and number of times
I. Pandemic impact on family life (CASI)	• Effects of the pandemic on certain aspects of life (sleep, smoking, hours of work, amount of money spent, etc.) in Lockdown 1, and in Lockdown 3 • Whether the household experienced major life events since the beginning of the pandemic (loss of a job, death of someone close, moving, etc.)
J. Parent health and wellbeing (CASI)	• Mental health and wellbeing scales (GHQ-12, GAD2, PHQ-2) • Life satisfaction • Self-assessed general health
K. Disadvantage (CASI)	• Comparison of current financial situation to pre-pandemic • Whether fallen behind on rent or mortgage since the beginning of the pandemic • Self-assessment of financial situation • Financial hardship including rent/mortgage arrears, issues with housing, food poverty and use of food banks • YP's eligibility for free school meals
L. Closing demographics	• Ethnicity, country of birth, religion, internet connection
M. Contact details, signposting and closing screens	• Name and contact information for parent, whether parent lives in the same address as the YP and updating of either if necessary for future waves, signposting to sources of support and closing

3.2 Content development and testing

3.2.1 Approach

The COSMO questionnaires were developed by UCL, Sutton Trust and Kantar Public working in collaboration. To help inform development of the questionnaire, meetings were

held with relevant stakeholders, and input was received from researchers, governmental organisations and funders.

In developing the questionnaires, other relevant surveys were consulted, and pre-existing questions were used or adapted where possible to build on prior experience and ensure comparability. Other surveys consulted included, but were not limited to, the Longitudinal Survey of Young People in England: Cohort 2 (LSYPE 2, also known as "Our Future"), Next Steps (also known as LSYPE 1), the Millennium Cohort Study, CLS COVID-19 surveys on national longitudinal cohort studies, and Understanding Society. A number of new questions were also developed.

The questionnaires were developed during the period May to August 2021 over the following stages:

- UCL provided initial drafts of questionnaires, and these were developed further in discussion with Kantar and wider stakeholders.
- Cognitive testing was conducted via online video conferencing to test the understanding of the questions among the relevant age group and among parents of young people in this age group.
- The content was finalised based on results of cognitive testing and further discussion with UCL.

It is worth noting that because COSMO Wave 1 was funded by the UKRI COVID-19 rapid response fund, and needed to be in the field as quickly as possible to collect accurate information on the experiences of young people about the pandemic, the project had very tight timescales. These timescales did not allow for a pilot stage to test questionnaire flow, fieldwork processes and interview length. Therefore, a small number of informal pilot interviews were carried out by the research team using personal networks to ensure the questionnaire worked well, and to derive approximate timing estimates.

As there was a range of potential topics of interest, there was a need to prioritise content to ensure that the questionnaire did not exceed the target interview length of 30 minutes.

3.2.2 Cognitive testing

Cognitive testing describes a process of testing survey questions to ensure that they are clear, relevant and comprehensible, and that they meet the intended measurement objectives. Researchers present questions to respondents and then use a mixture of pre-prepared and reactive probes to explore how participants decide on an answer. This provides valuable insight into how the questions are being interpreted and can help improve question wording.

Cognitive testing was carried out in June and July over two rounds. Testing focussed mainly on newly drafted questions which had not been used in previous studies or where we anticipated specific recall or comprehension issues in the context of the COSMO cohort.

Respondents were recruited via an external recruitment agency according to quotas on gender and region, and with further quotas to ensure minimum numbers of young people i) attending an independent school and ii) in receipt of free school meals. Cognitive interviews took place via an online video conferencing (Zoom). In total, 21 interviews were conducted over two rounds: 15 interviews with young people in the relevant age cohort and 6 interviews with parents of this age cohort. All respondents were provided with a £40 voucher payment to thank them for their participation.

3.2.3 Designing questions for multiple modes

Due to the sequential mixed-mode design, questions were designed to be compatible for both online and face-to-face presentation. Kantar Public's questionnaire template adopts a 'mobile first' approach for designing online survey questions and optimises question format by mode. When using survey data collected via multiple modes, it is important to consider how this may affect analyses. "Mode effects" are generally taken to mean differences in observed responses to survey items which are due solely to the mode of data collection.

The questions were designed to be presented in both modes using a set of rules for how different types of questions should be presented by mode. To help reduce mode effects, attempts were made when designing the questionnaire to ensure that the online questionnaire was as similar as possible to the face-to-face approach. Examples included using showcards for the face-to-face data collection and making all "Don't Know" codes explicit in both modes (these were included on showcards). Additionally, the use of a self-completion (CASI) section within the face-to-face interview for more sensitive items helped to mitigate against social desirability bias which can be associated with interviewer-led modes.

Nevertheless, despite attempts to align questions across mode, mode effects will be unavoidable as the two approaches can never be truly identical. Some examples of why measurement may still vary between modes:

- Face-to-face interviewers can provide motivation or clarification when required; this cannot truly be replicated online.
- People who would not disclose sensitive personal information or socially undesirable opinions/behaviours to an interviewer (for example drug use or self-harm) may be more willing to provide this information online. As noted above, this was mitigated by placing the more sensitive questions in a self-completion CASI section during the

face-to-face interview, although it is still possible that the presence of an interviewer created some mode effects associated with these more sensitive questions[18].

3.2.4 Use of showcards in the face-to-face interview

As noted above, showcards were used at many questions to provide a comparable question presentation across online and face-to-face modes. Due to COVID-19 protocols, we adopted a different way of administering showcards for COSMO. Respondents with a smartphone were given the opportunity to scan a showcard QR code which took them to a webpage on the survey website where they could view all the showcards on their phone. As an alternative, interviewers also had available a pack of standard showcards which were laminated and easy to wipe clean between interviews.

3.3 Data linkage

Young people were asked for their consent to link administrative data to their survey data, held by a variety of organisations:

- Education records, held by the DfE, including the National Pupil Database (NPD) and Individualised Learner Records (ILR) - covering achievement in school and further education as well as details about the school, college or training centre young people attended.
- Records about young people's enrolment in the National Tutoring Programme, held by the Education Endowment Foundation.
- Records covering students' progression from school into Higher Education and beyond, held by the Higher Education Access Tracker (HEAT).
- Information on benefit and employment programs, kept by Department for Work and Pensions (DWP).
- Information on employment, earnings, tax credits, occupational pensions and National Insurance Contributions, kept by HM Revenue and Customs (HMRC).

Taken together, consent to linkage to NPD, ILR, DWP and HMRC records allows for linkage to the UK Government's combined Longitudinal Educational Outcomes (LEO) dataset, which is based on a combination of these administrative datasets. The procedures for explaining and obtaining these consents from young people were approved by the UCL

[18] In a small proportion of in-home interviews, the interviewer conducted the CASI section by reading out the responses because the respondent was not able or willing to complete this section by themselves due to reading/eyesight or language issues. The self-completion was administered by the interviewer in 64 in-home parent interviews and 14 in-home young person interviews.

IOE Research Ethics Committee.

3.3.1 Data linkage consent process

When young people were invited to participate in COSMO, they were sent a leaflet which included information about data linkage consent, and here it was emphasised that this was entirely their choice. Moreover, in the respondent-facing website, there was a separate page on data linkage, where young people could access some frequently asked questions (FAQs) on this. These FAQs made clear how the linkage process worked, which data holders they would be asked about, and the purpose of data linkage. The webpage also emphasised that they may choose to consent to some rather than all linkages, that they can complete the survey without consenting to any of them, and young people were also informed about issues like data retention and the ability to request a withdrawal of their consent.

As the young people were over the age of 16 at the time of the interview, there was no parental consent necessary for data linkage. However, on the website, it was emphasised that young people could discuss this with their parents if they wished to do so, and parents also received a copy of the survey leaflet which outlined this process.

Within the survey, at the beginning of the consent module, young people were informed of the steps of data linkage, that information on them will be collected on an ongoing basis unless they told the study team to stop, and that they could change their permissions at any time.

The proportions of young people who consented to the various linkage requests are presented in section 8.4. No data linkage consents were collected from parents.

3.4 Provision of contact information for future waves

It was important to collect further contact information to allow future re-contact of both young people and parents via a range of channels. The approach differed according to the following types of sampled individuals.

Young people selected from NPD sample: At the beginning of the interview, participating young people sampled via the NPD were asked to confirm and, if necessary, update the name and address from the NPD record as well as the school year (expected to be year 12, though a few may have repeated or skipped a year). In addition, students who said that they were now at a different school to the one attended in Year 11 (the school recorded in the NPD sample) were asked to provide the details of their current school or college attended in Year 12.

The final module of the interview asked young people to provide further contact details to allow further re-contact via multiple channels at later waves: mobile number, landline number and email address.

Young people attending independent schools: For this sample of students, we had no prior information about name and address as the sample was generated via schools rather than pre-selected (see Chapter 7). Therefore, for this group, we needed to collect all contact information in the questionnaire. In the opening section of the questionnaire, students were asked to provide details of the school they attended in Year 11. Information about the school they were attending in Year 12 was already known as they had been sampled from one of the participating schools in the independent school sample.

In the final module, young people were then asked to provide their full name and address as well as further contact details such as telephone numbers, email addresses and mobile phone numbers.

Parents of young people in NPD sample: At the beginning of the interview, parents of young people sampled via the NPD were asked to confirm and, if necessary, update the name, address of the selected child from the NPD record.

The NPD record does not contain details of parents and therefore, in the final module, parents were asked to provide their full name, and also address if not the same as the one recorded in the NPD record. We also checked if the parent lived at the same address as the young person, and collected an alternative address if relevant. As with young people, parents were also asked to provide further contact details where these were available, to allow re-contact via other channels at later waves: mobile number, landline number and email address.

Parents of young people attending independent schools: Collection of data was similar to those who had a child in the NPD sample but in addition parents were asked to provide the name and address of their child as a verification.

Further contact detail updates, as well as collection of missing ethnicity information, was collected in a separate between-wave 'Keeping in touch' (KIT) mailing conducted in July/August 2022. Details of this are covered in section 6.8.

4 Ethics and informed consent

This chapter describes the processes applied to ensure ethics approval, survey completion based on informed consent, and the approach taken to asking sensitive questions and safeguarding.

4.1 Ethics committee approval

The study design and survey processes for COSMO were approved by the UCL IOE Research Ethics Committee. This application covered sampling, incentive approach, data linkage consents, participant information, privacy notice, signposting to sources of support, survey mode, questionnaires, and other relevant dimensions of the study.

4.2 Consent

Participation was based on the principle of informed consent. As all sampled young people were over the age of 16 there was no formal requirement for parental consent, although parents were given information about the young person survey in addition to information about the parent survey.

Information about the survey was available via several channels:

- The survey leaflet which was included in the first mailing to both young people and parents
- The advance survey letter and later reminder mailings (which did not include a leaflet but instead a set of FAQs were included on the back of the letters).
- The survey website which included participant FAQs, Privacy Notice, information about data linkage, background to the survey, sources of help and support.

Across this correspondence respondents were informed of the following:

- Background to the study and confirmation of research ethics approval
- Who the study was funded and managed by
- How young people (and parents) were selected for the study
- Overview of survey topic coverage
- That the decision to take part in the survey was completely voluntary
- That they had a right to decline to answer certain questions or withdraw from the study at any time, either completely or just from one wave
- That we planned to contact them again in the future

- That the information they provide would be treated confidentially and in accordance with relevant legal frameworks (Data Protection Act and GDPR)
- Information and data linkage (for more detail on see this section 3.3)
- Information about how to complete the survey and claim their voucher
- Sources of support relevant to more sensitive issues covered in the survey
- Signposting to the survey privacy policy

At the start of the survey, respondents were asked to confirm that they had read the information contained in the leaflet and that they were happy to take part before starting the survey:

> "Welcome to the Horizons survey. This survey is being conducted on behalf of University College London and the Sutton Trust with support from the Department for Education.
>
> This is the first year of this survey, and we would like to thank you for taking part.
>
> Please confirm by clicking on the button below that you have read the leaflet which was sent with your invitation letter and that you are happy to take part. Then click the (>) button to continue."
>
> *Help text [AFTER Privacy Policy]: The data controller for this project will be University College London (UCL). The data protection privacy notice for this project, as well as general privacy notices for UCL, can be found on the Survey Privacy Information section on the Horizons Study website at www.horizonsresearch.co.uk/. Anonymised data will be made available to researchers through the UK Data Service or similar organisations.*

4.3 Sensitive issues and safeguarding

For the web survey, the entire questionnaire was self-completed online. For the face-to-face survey, the more sensitive questions were administered as self-completion (CASI) which respondents completed via the interviewer's tablet.

Sensitive questions included in the CASI module for the young person questionnaire included: mental health and wellbeing, physical health, bullying and harassment, major life events such as death in the family, risky behaviours such as smoking, alcohol and drug use, self-harm and suicide.

Sensitive questions included in the CASI module for the parent questionnaire included: mental health and wellbeing, pandemic impact on family life, major life events such as divorce or death in the family, financial hardship.

A safeguarding approach was built into the script for young people who indicated that that may have self-harmed or attempted suicide in the previous 12 months. Where a respondent answered 'yes', 'don't know' or 'prefer not to say' to either of these questions they were immediately directed to a targeted screen which provided information on where to seek further help including contact details for the Samaritans.

More generally, all respondents had access to a range of relevant support sources which were included on the survey leaflet and on the survey website. These were also signposted at the end of the questionnaire.

5 Respondent engagement

This chapter covers our strategy to respondent engagement, which was designed to help maximise response rates, provide informed consent at all stages, and to assist respondents with participation in the survey.

5.1 Respondent materials and branding

The study was known as 'Horizons' for participants, rather than COSMO. A specialist agency was commissioned to develop a logo for the survey and to design an engaging survey leaflet. The Horizons branding and visuals were then used across all survey materials and the survey website.

The following respondent-facing materials and resources were developed over the course of the study:

- Survey website
- Participant information sheets
- Privacy notice
- Survey invitation and reminder letters
- Survey leaflet
- Sources of help and support
- Survey helpline

5.2 Website, privacy notice and participant information sheets

A website was set up at http://www.horizonsresearch.co.uk which contained the following resources:

- A homepage
- Participant information sheets set up as a set of website FAQs (separate pages were available for young people, parents and schools[19])
- Survey privacy notice

[19] The website contained participant information for school staff taking part in the schools' survey which covered some of the same material but tailored to the schools' audience. However, as noted elsewhere, this survey was ultimately dropped to fieldwork challenges during the pandemic (see Appendix A).

- Information about data linkages asked about during the survey
- Background information about the survey and its objectives
- Sources of help and support
- A 'Contact us' page with details on how to access the survey helpline (see 5.4 below) or to contact the research team at UCL

In order to access the survey, NPD participants were directed to a landing page at www.horizonsresearch.co.uk/survey. They would then enter their log-in and password to be taken to the survey.

Visitors to the main website purposely could not access this landing page to avoid confusing independent school participants who were using the website to find out further information about the study. The independent school participants accessed the survey via a school specific open link that was sent to young people and their parents directly, so they did not use the same log-in process. For more information on the fieldwork process for independent school pupils, see section 7.3.

5.3 Respondent invitation letters and survey leaflet

The following relates to the NPD (state school sample). The procedures for contacting students (and their parents) at independent schools was different and is covered in more detail in Chapter 7.

The mailing strategy was based on the following sequence of letters. More information about the content of the letters and survey leaflet is provided in section 6.4, while a timeline is provided in section 1.3.

Young people and parents received separate different letters. At Web reminders 1 and 2, adapted versions of letters were used for 'break-offs' that is those who had started the survey but not completed it[20].

In line with best practice, wording varied across mailings to emphasise different messages.

The original sample received up to four letters (reminders 3A and 3B were sent to different sub-samples):

- Launch letter invite and survey leaflet
- Reminder 1 letter with FAQs on the reverse
- Reminder 2 letter with FAQs on the reverse
- Bespoke letters to target break-offs with FAQs on the reverse

[20] In practice, this included all those who had not reached the threshold designated as a useable break-off (see section 10.1)

- Reminder 3 with FAQs on the reverse; two variations were sent to two different subgroups of non-responders:
  - Reminder 3A letter - this was sent to remaining non-responders from the original sample who had not been allocated to the face-to-face stage
  - Reminder 3B letter - sent to all remaining households which were due to be contacted F2F but did not end up being contacted by an interviewer (see sections 1.2.2 and 6.2 for more explanation on this)
- Reminder 4 letter with FAQs on the reverse - final 'mop-up' to all non-responders who had not been contacted F2F, but this time targeted only at young people or parents in partial households, in order to increase the rate of full households rather than create new unpaired households at this late stage

The reserve sample, which was issued later, was sent a launch letter invite followed by two reminders.

As noted in section 6.4, all correspondence to young people and their parents in the original sample were sent in separate envelopes to maximise response rates among young people. As names were not available from the sampling frame for parents, they were addressed to 'Parent of the [NPD child]'. However, when issuing the reserve sample, both young person and parent letters were sent in the same envelope addressed to the 'Parent of [NPD child]'. This was because at this stage we were especially focused on ensuring a high rate of matched households and reducing the number of returns where only a young person or parent in the household participated.

Copies of survey materials for the NPD sample are included in Appendix B, and copies of survey materials for the independent school sample are include in Appendix C.

5.4 Respondent helpline

A survey helpline hosted by a specialist team at Kantar Public was set up to deal with queries about the survey. The helpline could be accessed by participants by Freephone or by email. Most queries were dealt with directly by the helpline team although where necessary queries were forwarded to the research team. UCL also forwarded any queries that they received to ensure that Kantar resolved these issues too.

5.5 Assistance for respondents with low levels of English language skills

Respondent-facing materials were not translated into minority languages. However, at the face-to-face stage, interviewers were able to use another member of the household to assist with interpretation in cases where a respondent struggled with English language,

either because English was not their first language, or because of learning/literacy difficulties.

6 Fieldwork

This chapter covers the fieldwork procedures for the NPD (state school) sample. Fieldwork procedures for the supplementary independent school sample were different and this is detailed in Chapter 7.

6.1 Overview of original fieldwork plan and timing

Fieldwork for Wave 1 was conducted between 22 September 2021 and 18 April 2022. All fieldwork was conducted by Kantar Public, with support from NatCen during the face-to-face stage of the study.

The original intention was that the study should use a sequential multi-mode design involving online data collection initially, followed by face-to-face data collection for a targeted sub-set of non-responding households. The original intention was that a targeted sub-sample of c.50% of non-responding households after the online phase would be issued face-to-face. The criteria used for allocating non-responding households for face-to-face issue after the online phase is outlined in section 6.3.

6.2 Changes to fieldwork protocol due to the impact of COVID-19

The original fieldwork plan was for the online phase of the survey to close at the start of November 2021 after a second targeted reminder and for face-to-face fieldwork to be conducted between November 2021 and March 2022. The online survey was closed on 1st November for all households allocated to the face-to-face sample although it was kept open for households not being issued face to face.

All in-home interviewing in the UK was paused from March 2020 until September 2021 due to the COVID-19 pandemic, and so COSMO was one of the first studies to go back into face-to-face fieldwork after this lengthy break across the entire industry. This had implications in terms of the Kantar Public and NatCen interviewer panels not being at full strength due to interviewers leaving the industry during the lockdown period, many existing interviewers being reluctant to return to work due to the ongoing circulation of the virus, and the need for interviewers who were willing to work to follow strict COVID-19

protocols on the doorstep and in households. In the event, although the in-home survey phase was launched in the field on 10th November 2021, fieldwork had to be halted again on 10th December 2021 due to the emergence of the Omicron variant and the re-introduction of some 'Plan B' restrictions[21]. All in-home interviewing on the study remained on hold for three months until the start of March 2022.

Since the nature of the study was time critical it was not possible to simply delay fieldwork indefinitely and so the lost fieldwork time could not be fully recovered. The original fieldwork period was extended by about three weeks from the end of March 2022 until mid-April 2022, but this was an absolute deadline which could not be extended again. Key considerations included the need to collect the data during Year 12 school year and to finish fieldwork before end-of-year examinations which typically take place from May onwards.

In an attempt to compensate for the lost fieldwork time, some changes were made to the original fieldwork protocol as follows:

- The online survey was re-opened and a third reminder (3B) was sent to all non-responding individuals during December 2021.
- Kantar Public implemented a 'knock to nudge' (KTN) stage from the start of February 2022 until the resumption of in-home interviewing. This involved interviewers' visiting selected non-responding households which had been allocated to face-to-face to encourage young people and parents to complete the survey online and did not involve the interviewer entering respondents' homes at any point. However, as noted above, a variety of factors meant that the available interviewer resource was limited during this period which meant that only a relatively small proportion of sample was worked in this way. Additionally, the allocation of households to KTN was not done on any systematic basis but rather was based on available interviewer resource during this period.

Fieldwork based on in-home interviewing resumed on 7th March 2022 and continued for six weeks until 18th April 2022. However, due to the ongoing issue with interviewer capacity during this time it was not possible to allocate and cover all the sample that was originally allocated for face-to-face issue.

Based on the suspension of face-to-face fieldwork for almost three months, ongoing interviewer capacity issues, and a reluctance of participants to let interviewers into their home, it became clear that even when face-to-face fieldwork resumed in early March 2022, this would not be sufficient to achieve close to the target number of interviews.

As a result, a decision was taken in early 2022 to issue a second batch of sample drawn from a reserve sample which had been selected at the outset of the study. Given the available time, it was agreed that this sample of households would only involve online

[21] https://www.gov.uk/government/news/prime-minister-confirms-move-to-plan-b-in-england

data collection and fewer reminder mailings would be sent compared with the original sample. While the issuing of this reserve sample enabled the target number of achieved interviews to be met it meant that the final response rate to the study was lower than anticipated.

Further details about the online and face-to-face contact procedures that were implemented can be found in sections 6.4 and 6.5.

6.3 Allocation of sample to face-to-face after online stage

COSMO was set up as a sequential mixed mode study with an initial online phase followed by a face-to-face phase. Under the original design the face-to-face phase had four objectives: (i) to help maximise the study response rate; (ii) to help achieve the overall target sample sizes and target sample sizes for key sub-groups; (iii) to help equalise response rates between sub-groups to ensure a more balanced sample profile; and iv) to help increase the rate of complete households where only one eligible household member had already competed online. However, the changes to the fieldwork plan necessitated by the COVID-19 pandemic meant that the number of cases worked in the face-to-face phase was far less than anticipated and so these objectives were not fully met.

Based on these objectives - as well as budgetary considerations – the initial survey design assumed that a targeted sub-sample of c.50% of non-responding households after the online phase would be issued face-to-face. Based on the estimated online response rate and the likely face-to-face response rate, the budget allowed for c. 9,480 households to be issued.

In deciding how to select non-responding households for allocation to face-to-face, the aim was to use an approach which achieved the optimal balance between the objectives outlined above. Three separate allocation approaches were considered and tested:

- Using a regression non-response model with participation in the initial online phase of the survey as the dependent variable and sample frame variables as predictors to calculate estimated response probabilities for each young person[22]. Households would then be allocated to face-to-face issue by prioritising those with the lowest response probabilities *within schools*.
- Using the same non-response model as above but prioritising households with the lowest response probabilities across the *whole sample*.

[22] Variables used in the model FSM, Ethnicity, EAL, Gender, region, IDACI quintiles, SEN status, KS2 reading score (tertiles), KS2 maths score (tertiles), KS2 GPS score (tertiles) and school band (Special/AP/Other)

Not using a model-based approach but simply selecting cases at random for face-to-face issue based on their explicit cell stratification (FSM and ethnicity) with a view to achieving set target sample sizes in each cell.

Since each approach has its pros and cons, a preliminary sample was drawn using each of the three approaches outlined above based on interim data taken near the end of the online fieldwork period. The resulting profile of the three selected samples were compared to assess which yielded the best sample profile and to consider the practical issues and implications of implementing each sample in the field.

Following this preliminary analysis and review stage it was agreed to tweak the non-response model slightly and to use the approach of prioritising households with the lowest response probabilities *across the whole sample*.

One slightly unexpected outcome from the online phase was the relatively high proportion of partial household responses received – where the young person or the parent had responded, but not both. After discussion it was agreed that since complete households were the priority, all partial households at the end of the online phase should be issued face-to-face stage. As a result, the final allocation process was as follows:

- All partial households at the end of the online phase were automatically allocated for face-to-face issue.
- Following this allocation, completely non-responding households were sampled by selecting households with the lowest response probabilities *across the whole sample* up to a total of 9,465[23].

In the end, the actual number of households initially issued was 9,430 since the sample allocation was done a few days before the online fieldwork ended and so late online responders and opt outs were removed from the sample. This sample consisted of 5,808 completely non-responding households (which was 62% of all non-responding households issued); 2,743 partially responding households where the parent was the non-responder, and 879 partially responding households where the young person was the non-responder.

However, because of the changes to the fieldwork protocol outlined in section 6.2, the face-to-face sample allocation was re-worked before the start of Knock-to-Nudge fieldwork in February 2022. The additional online reminder sent in December 2021 to those originally allocated to face-to-face issue (reminder 3B) meant that some individuals were taken out of the sample because they had in the meantime completed the survey online (or opted out). At the same time, the extra online reminder sent to those not originally allocated to face-to-face issue (reminder 3A) created new partial households which were added into the face-to-face sample.

[23] To account for break-off cases non-responding households were defined as those which failed to reach the threshold designated as a useable break-off (see section 10.1 for a definition of a useable break-off).

As a result of the above, 918 households were taken out of the initial sample allocation, while 424 new households were added. This means that in total 9,854 households were allocated for face-to-face issue at some point during fieldwork: 5,673 completely non-responding households (which was 57% of all non-responding households); 3,118 partially responding households where the parent was the non-responder; and 1,063 partially responding households where the young person was the non-responder.

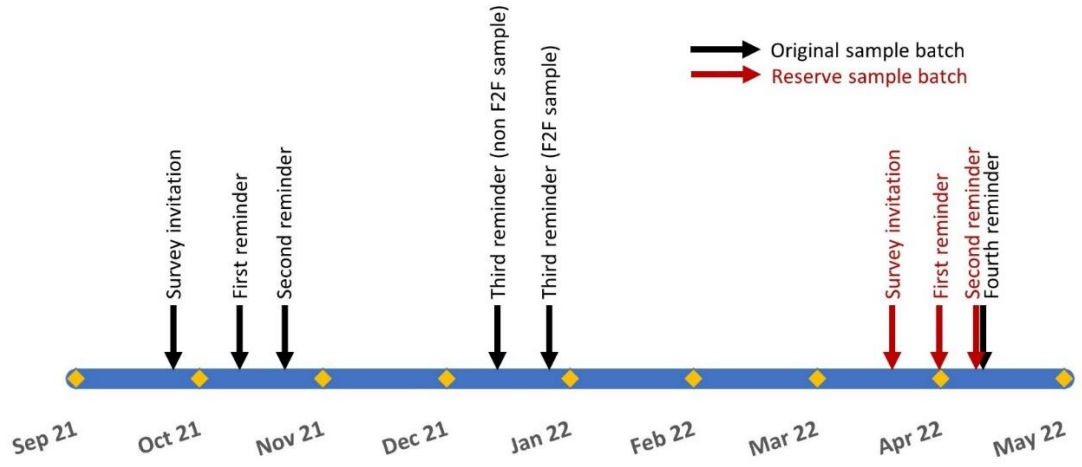
6.4 Contact procedures for the online phase

The full timeline for schedule of contact with participants is covered in Table 1.3.

Due to the ongoing pandemic the original fieldwork plan and fieldwork timings had to be reviewed and adapted several times during the study as outlined in section 6.2. This meant that the contact procedures for the online phase were more extensive than was originally planned both in terms of the reminder strategy and the fact that two separate batches of sample were released at different times. Given the more limited time available for fieldwork, the reserve sample had by necessity a more limited contact strategy compared with the original sample.

Figure 6.1 summarises the contact strategy implemented for the online survey based on the fieldwork timeline. The contact strategy for the independent school sample is discussed in Section 7.3.

Figure 6.1 Contact procedures for online phase



Initial survey invitation

All households selected as part of the NPD sample were sent an initial survey invitation letter (see Appendix B, section 2). Letters were branded with the Horizons study logo and sent in plain white envelopes. The invitation letter contained:

- A brief description of the study which emphasised the fact that it was a new study set up specifically to see how COVID-19 has affected young people's lives.
- The URL of the survey website (horizonsresearch.co.uk) and username and password details for logging in (a QR code was also included on the letters).
- Details of who was leading the study and who was carrying out the data collection.
- Assurances of confidentiality
- The fact that a £10/£20 voucher was available as a thank you for completing the survey.

Additionally, the initial mailing also contained a survey leaflet which provided more information about the study (see Appendix B, section 1) including:

- More details about the content of the survey and who was carrying it out.
- An explanation of how individuals were chosen for the survey and its voluntary nature.
- Further information about confidentiality, data protection and data linkage.
- Directions to the survey website and the survey helpline or email inbox if the participant had any questions or concerns.
- Other sources of advice and support about the topics covered in the survey.

Although the content of the launch letter and leaflet sent to the original sample and the reserve sample were very similar there was one important difference. With the original sample, invitations were sent to young people and parents in separate envelopes: the letter to the young person was addressed to them by name, while the letter to the parent was addressed to "the parent of [named young person] . However, with the reserve sample the invitations were sent in the same envelope addressed to 'the parent of [named young person selected from the NPD]'. The reserve sample launch mailing contained a letter for the parent, a separate letter for the young person with their name on it, and one copy of the survey leaflet. This change was made because of the number of partial households – where either a young person or a parent within a household responded but not both – which the original sample had yielded. It was hypothesised that this approach might be better at achieving complete household returns (see sections 8.1, 8.2 for more details).

Planned reminder mailings

The original fieldwork plan had involved sending up to two reminders before the start of face-to-face fieldwork. All non-responding individuals were sent a first reminder letter approximately two weeks after the initial invitation and a second reminder after another two weeks (see Appendix B, sections 3 and 4). The spacing between the dispatch of reminders was slightly shorter for the reserve sample compared with the original sample due to the need to finish fieldwork by a set date. Slightly adapted versions of the letter

were sent to those who had entered the survey (termed a 'break-off') but not completed it, encouraging them to go online again and complete the survey to the end.

The reminder mailings did not contain another copy of the survey leaflet. Instead, the back of the letter contained 'Commonly Asked Questions' which summarised some of the information included in the survey leaflet.

The difference between the original and reserve sample mailings for the invitation letter (where the reserve sample included invitations in the same envelope) was replicated for the reminders, except where the parent had responded but the young person had not. In this case the letter was addressed directly to the young person.

Additional reminders

Following the changes to the face-to-face fieldwork outlined in section 6.2, additional targeted reminders were sent in an attempt to boost response to the online survey as much as possible due to the challenges associated with face-to-face fieldwork. These additional reminders were only sent to non-responders in the original sample as follows:

- A third reminder was sent following the suspension of face-to-face fieldwork in December 2021. This was sent to all non-responding individuals in the original sample, except for any households which had been contacted during the short face-to-face fieldwork window. This reminder made explicit reference to the fact that the period for the online survey had been extended due to ongoing restrictions. Reminder letters sent to households which had been allocated for face-to-face issue contained an additional sentence noting that an interviewer might call at the door in the new year. Beyond this, there were two variations of Reminder 3 sent to two different subgroups of non-responders:
  - Reminder 3A was sent to all non-respondents who had not ever been allocated to face-to-face (mailing sent 13 December)
  - Reminder 3B was sent to all non-respondents who had initially been allocated to face-to-face but had not yet been contacted face-to-face by this stage due to further COVID-19 restrictions. This subgroup was therefore re-opened on the web platform and sent a further web reminder (mailing sent 21 December)
- Due to the lower response seen among parents, a highly targeted email reminder was sent to young people who had completed the survey (and provided their email address) but whose parent had not. This email stressed the importance of getting the views of both young people and parents and asked them if they could remind their parents to complete the survey. This was done to try and boost the number of complete households in the survey.
- A final reminder (Reminder 4) was sent in April 2022 as a 'mop-up' towards the end of fieldwork. Since achieving a complete household interview (young person and parent) was considered the most important priority, this final reminder was targeted only at young people or parents in partial households which had not been contacted during the face-to-face fieldwork phase. No additional reminders were sent to any

young people or parents in households that had not responded at all, as at this late stage we did not want to create new unpaired households.

6.5 Face-to-face contact procedures

As noted in section 6.2 the fieldwork procedures changed during Wave 1 in response to the ongoing COVID-19 pandemic. In particular, the plans for face-to-face fieldwork had to be amended several times due to the emergence of the Omicron virus, continued lockdown restrictions, and ongoing challenges with interviewer resources which affected the whole research industry. Only the original sample batch had any face-to-face fieldwork element; time restrictions meant that the reserve sample batch only involved online fieldwork.

Table 6.1 shows that face-to-face fieldwork was divided into three distinct phases.

Table 6.1: Stages of face-to-face fieldwork

Phase 1	In-home interviewing	10th November – 10th December 2021
Phase 2	Knock to nudge	2nd February – 6th March 2022
Phase 3	In-home interviewing	7th March – 18th April 2022

For the in-home interviewing phases, the interviewer had several tasks all of which had to be carried out with strict adherence to a fieldwork protocol designed to ensure interviews could be conducted safely during the pandemic[24]. These tasks were:

- To make contact with the household and confirm that the named young person and their parent or guardian were living at the address.
- If the young person was living elsewhere to attempt to get a new address for them.
- Where contact was successful to collect as many contact details as possible (telephone numbers, email addresses) to make any future contact easier.
- To conduct an interview with either the young person and/or the parent depending upon the status of the household.

All sample was managed via an electronic contact sheet (ECS) which enabled interviewers to complete all these tasks. All addresses were issued at a household level but with corresponding records for each individual in the household. This enabled interviewers to know from the sample management system who they needed to interview in each household. If someone in the household (either young person or parent) had already completed an online interview a specific outcome code (code 970) was used so that the interviewer was aware of this in advance.

[24] Fieldwork safety guidance was issued by the Market Research Society for all research agencies working in the field and Kantar Public developed its own fieldwork protocols within the framework of the MRS guidance.

For the knock-to-nudge stage the interviewer tasks were largely the same as above, except no in-home interviews were conducted. Instead, this involved interviewers' visiting selected non-responding households to encourage young people and parents to complete the survey online. All face-to-face visits were conducted using a fieldwork protocol designed to ensure socially distanced COVID-secure contact. Interviewers were able to provide households with more information about the survey, answer any questions, and also provide the log in details and passwords for the survey. Interviewers were also asked to collect contact details wherever possible to allow for further follow up by telephone. Interviewers recorded where households agreed to do the survey online after their visit, although it was understood that not everyone who agreed on the doorstep would convert into a complete interview.

6.6 Interviewer briefings

All face-to-face interviewers attended a project briefing before starting work on the study. Briefings were administered via pre-recorded videos delivered by members of the research team. Given the uncertainties over fieldwork timings it was decided that using a pre-recorded briefing would give more flexibility in terms of when and how briefings were conducted. Briefings were split into two shorter sessions lasting about an hour in total. Following the pre-recorded session, interviewers were then invited to an interactive session held on MS Teams led by members of the field and research teams. This provided a chance for interviewers to clarify any points or ask further questions.

The briefings covered the following:

- Background to the project and its purpose
- Overview of the sample and the multi-mode approach
- Online contact procedures and fieldwork materials
- Making contact and COVID-19 safety protocols
- Introducing the survey on the doorstep and answering FAQs
- Fieldwork procedures and tips for gaining a high response rate
- Content of the questionnaires, including self-completion sections and dealing with sensitive topics
- Signposting sources of help and support
- Data linkage consent
- Field administration

At the end of viewing each pre-recorded session interviewers were asked to complete a quiz which covered the main points of the briefing. They were also required to complete a practice interview for each survey before starting their assignment.

As well as the briefing slides, interviewers were also provided with a comprehensive set of project instructions for the study and a range of survey materials including copies of

survey letters and leaflets, calling cards, and COVID-19 leaflets. An additional interviewer note was provided for interviewers undertaking knock to nudge fieldwork.

6.7 Incentives

Incentives were used to encourage participation in the survey and boost response rates. Young people and their parents were offered either a £10 or £20 gift voucher upon completion of the survey. On the online survey participants could claim their e-voucher immediately by clicking on a link which took them directly to the website of the incentive provider. If they requested this, they were also sent an email containing a voucher code and details of how to claim the voucher. On the face-to-face survey participants were sent an email within 7-10 days of completing the survey which contained similar details. Participants who did not have an email or did not want to provide one were sent a gift card by post.

Differential incentivisation was used to help ensure a good representation of students and their parents from more disadvantaged backgrounds, who are typically less likely to respond to surveys. The higher voucher value was targeted at those attending a school with the highest rates of pupils eligible for free school meals. It was decided to offer differential incentives at the *school* level, rather than the *individual* level (e.g. by FSM status), to eliminate any risk that a pupil invited to take part in the study discovered they were being offered a different incentive value from other pupils at the same school. In all cases parents received an incentive to the same value as the young person in their household.

Overall, 17% of NPD sampled households were offered the higher incentive. However, the proportion of households offered the higher incentive varied considerably by stratum: for example, 32% of Bangladeshi households and 27% of Black Caribbean households were offered the higher incentive, while 22% of households eligible for free school meals were offered the higher incentive compared with 8% of households which were not eligible.

6.8 Keeping in touch exercises after Wave 1 fieldwork

Two 'Keeping in touch' (KIT) exercises were implemented between the period from the end of Wave 1 fieldwork (April 2022) to before the start of Wave 2 (October 2022).

6.8.1 KIT mailing to collect missing ethnicity data

As noted in section 3.1.1, the young person's ethnicity was erroneously not asked to the original NPD sample (the error was corrected for the reserve sample issue and

respondents interviewed by face-to-face towards the end of fieldwork). The first KIT exercise was primarily to collect missing ethnicity data from the original sample, although we also used this an opportunity to collect updated contact details.

A letter was sent all young people in the original issue achieved Wave 1 sample (NPD and independent: n=10,858) asking them to complete a short online form to ensure that the contact details we held for them were up to date, and to ask them a few additional questions. The questionnaire included questions to confirm if their name, address, email and landline/mobile numbers were up to date and to collect new contact information for anyone who had planned to move home in the next six months.

The KIT 'survey' was accessed via the survey website used for the main survey but directed to a 'keeping in touch' landing page. As for the main survey, participants could access the KIT 'survey' by using a unique set of log in details. We prepared a separate set of frequently asked questions for this exercise which we made available on the survey website throughout the KIT fieldwork.

To help encourage response one reminder was sent (either SMS, email or letter depending on what contact information we had available) and everyone who took part received a £5 e-voucher which they could redeem at the end of the KIT exercise.

The KIT 'survey' launched on 4 July and closed on 9 September 2022. In total, 5,229 responses were achieved, representing a response rate of 48.2%. This enabled us to update the Wave 1 data with the ethnicity from an additional 5,145 participants after excluding those who said 'don't know' or refused' in the KIT survey.

Documents related to the KIT exercise are included in Appendix D.

6.8.2 Pre-notification mailing to share findings from Wave 1

All young people and parents in the sample for Wave 2[25] were sent a pre-notification mailing shortly before the main launch in early October 2022. This comprised a cover letter and A5 postcard which included some visually engaging headline findings from Wave 1.

Prenotification mailings were sent in separate envelopes for young people and parents, and the cover letter was tailored depending on if the parent also took part in Wave 1 or if they were a new parent we were hoping to interview for the first time at Wave 2. See Appendix D for copies of prenotification correspondence.

The aim of this mailing was to act as an engagement tool and to remind families of the study in advance of the main Wave 2 launch which followed a week or so later.

[25] That is all young people and parents in all households where a young person had been interviewed at wave 1, and excluding anyone who had opted out the panel

7 Independent schools

The procedures for sampling and conducting fieldwork among independent school students and their parents differed from the main NPD (state school) sample and this chapter details the specific processes put in place for this sample.

Fieldwork documents for this stage of the study are included in Appendix C.

7.1 Independent school sample design

As set out in Chapter 2, the sampling frame used for state schools covered the 2020/21 academic year. For state schools, we therefore sampled pupils that were in Year 11 (Y11) in 2020/21, but the survey was conducted during the 2021/22 academic year when they were in Year 12 (Y12).

If we had asked independent schools to sample using pupil lists from 2020/21 (consistent with the state school approach), there would have been non-coverage as schools would not have been able to invite all sampled pupils (those that have left their school at the end of Y11). As a result, it was decided to ask independent schools to sample pupils attending Y12 in 2021/22.

Using the 2021/22 academic year for independent and 2020/21 academic year for state schools has two implications:

- Children who were in an independent school in Year 11 but moved to a state school for Year 12 are not covered by the study. In theory, this group is part of the target population but, because it is missing from both sample frames, there is no way to weight the data to compensate for this non-coverage (thought to be <1%).
- Potentially, some children will appear in both sample frames: specifically, those that moved from a state school in Year 11 to an independent school in Year 12. These respondents were identified retrospectively (via data collected in the survey questionnaire) and the weighting compensated for this (see Chapter 11).

7.1.1 Sample frame and exclusions

The DfE register of schools in England (GIAS accessed on 23rd March 2021) was used to identify the population of independent schools. This is a publicly available database.

Schools were eligible if all four of the following conditions were satisfied:

- The school was located in England
  (*GOR (name)* =/= "Wales")
- The school was an independent school
  (*TypeOfEstablishment (name)* = "Other independent school" or "Other independent special school")
- The school was open
  (*EstablishmentStatusname* = "Open"
- The school covered Year 12
  (*StatutoryHighAge* > 16)

In total, 1,112 independent schools were identified as eligible to be sampled (396 independent special schools and 716 independent schools).

For efficiency reasons, it was agreed that small independent schools would be excluded from the study. The GIAS database provides the number of pupils across all year groups (*NumberOfPupils*). To estimate the number of students in Y12 at each school - the total number of pupils attending the school was divided by the number of year groups (inferred from the range between the *StatutoryLowAge* and the *StatutoryHighAge* provided for each school).

Independent special schools tend to be extremely small. We therefore used different thresholds for independent schools and independent special schools.

- We excluded independent schools that we estimated had fewer than 25 pupils in Y12. Once this exclusion was in place, we were left with 435 schools (61% of the original total of 716). However, we estimated that these 435 schools educate 91% of Y12 students that attended an independent school in 2021/22.
- For independent special schools - we excluded schools that we estimated had fewer than 2 pupils in Y12. This left 235 schools out of 396 (59%). We estimated that these 235 schools educated 89% of Y12 pupils that attended an independent special school in 2021/2022.

7.1.2 Drawing the sample of independent schools

A systematic random PPES (Probability Proportionate to Estimated Size) sample of 240 schools was drawn. School sampling probabilities were proportionate to the estimated number of Y12 pupils in the school (inferred from GIAS data as outlined in the previous section). There were two explicit strata: (i) independent schools and (ii) independent special schools.

The sampling probability (p_a) for **independent schools** was calculated as:

$p_a = (n_{ai}/n_i) * 228$

Where:

n_{ai} = estimated number of Y12 pupils in independent school a

n_i = estimated number of Y12 pupils in all (eligible) independent schools in England

The sampling probability (p_a) for **independent special schools** was calculated as:

$p_a = (n_{as}/n_s) * 12$

Where:

n_{as} = estimated number of Y12 pupils in independent special school a

n_s = estimated number of Y12 pupils in all (eligible) independent special schools in England

Before a systematic random PPES sample of schools was drawn, within each stratum schools were sorted by:

- Region
- Whether they are mixed or single sex
- Whether they have boarders

This helped ensure that the sampled schools were representative of all eligible schools in terms of these factors.

Selected schools were then randomly allocated to original issue (120 schools – 114 independent and 6 independent special schools) and reserve (the remainder). However, as set out in section 7.1.3 below, due to difficulties engaging schools the whole of the reserve ended up being issued (all 240 schools). As such, the school level sampling probabilities (p_a) did not need to be adjusted to reflect whether schools were allocated to original issue or reserve.

7.1.3 Issued sample size and achieved sample size against targets

We aimed for c.40 independent schools participate in the study and to achieve c.1,000 interviews (c. 25 per school). In total, all 240 schools selected were issued into field (all original issue and all reserve), 35 schools initially agreed to co-operate, and 33 of these schools issued survey invitations (see section 7.2 below for details of recruitment).

In the event, and as detailed further in section 7.1.4 below, the following numbers of online interviews were achieved with students and parents from the independent sector:

- 674 young people

- 363 parents, of which 206 were included in the final dataset as they were in a complete household which also contained a young person

7.1.4 Pupil-level sampling

The 35 initially co-operating schools were asked to distribute the survey invitations to pupils and their parents/guardians on Kantar Public's behalf; as noted in section 7.2 below, in the event 33 schools sent invitations to schools.

As the aim was to achieve approximately 1,000 achieved interviews with independent pupils and a matched sample of 1,000 parents, we asked schools to sample 60 pupils with the expectation of a c. 50% response rate among each group (35 x 30=1,050) although in the event the within-school response rate was lower than this (see section 8.3 for details of independent school response rates).

All co-operating schools were sent detailed sampling instructions which stressed the need for students to be selected randomly and systematically. To help ease burden on schools, schools were asked to select whole forms/tutor groups rather than select a random sample across the whole year group.

Where schools had fewer than 60 pupils in the Y12 group, they were asked to invite all their Y12 pupils. For larger schools, Kantar Public worked with them to randomly select an appropriate number of forms to invite to the study (with the aim of inviting around 60 pupils). The selection process[26] was as follows:

- Schools were asked to provide a list of forms/tutors with the number of Y12 students in each. These were then sorted alpha-numerically. Schools were not required to pass on any personal information relating to students or form tutor leaders.
- Schools were asked to exclude students with an international home address where practical (these tended to be boarding students).
- The research team at Kantar Public then systematically selected a random selection of forms which resulted in a sample of approximately 60 pupils (if an exact number was not possible, we selected a sample of forms designed to achieve slightly more rather than slightly fewer than 60 students).
- Kantar Public than communicated with schools to tell them which forms should be sent email invitations (invitations to be sent to all students in these forms and their parents after excluding any international students)
- As the independent school survey was administered via an open link rather than a unique link (see section 7.3 below) we stressed to schools the importance of only communicating with selected students and not to advertise the survey on any open forum such as a student intranet or newsletter.

[26] See Appendix D, section 4

For any schools that did not have clearly defined forms, a suitable alternative approach was determined.

As noted, cooperating schools provided information on the total number of forms in Y12 (f_{ai}) and the number of forms invited to participate in the study ($f_{aSampled}$). This was used to calculate the within-school pupil sampling probability($p_{s|a}$):

$$p_{s|a} = f_{aSampled} / f_{ai}$$

Finally, the pupil sampling probability was calculated by multiplying the school sampling probability by the within-school pupil sampling probability:

$$p_{independentStudent} = p_a * p_{s|a}$$

7.2 Recruitment of independent schools

Independent school recruitment took place from May to July 2021, and this was undertaken by a specialist recruiter with experience of the school sector.

Contact details of schools were obtained from the DfE register for schools although this did not include email addresses. Therefore, a manual exercise was undertaken to collect these from the appropriate school websites to augment the sample. Where possible, an email address for the head teacher, head of sixth form and pastoral lead were collected to increase the potential number of decision-makers at the school who could be contacted directly.

At the outset, we communicated with the Independent Schools Council (ISC) and independent school Heads Associations including: Girls' Schools Association (GSA), Headmasters' and Headmistresses' Conference (HMC), Independent Schools Association (ISA) and the Society of Heads. All of these organisations agreed to support the study, and this support was mentioned in all correspondence with independent schools. We also received permission from these bodies to use their logos in all correspondence.

Recruitment then began by contacting the selected schools by letter and email addressed to the headteacher to introduce them to the study and to let them know that we would be in touch to discuss whether they would be interested in taking part. This correspondence highlighted the study objectives and what would be required of them, stressing that we only needed their agreement in principle at this stage as survey invitations would not need to be issued until September 2021.

A follow-up telephone call was then made a few days later to schools where we had either positive contact or no contact, addressing any queries schools had about the research. All details from emails and telephone calls were securely recorded in a sample log.

Calls and emails continued until we were able to speak to the headteacher or an alternative decision maker, and until a final decision could be recorded. Where independent schools agreed to participate in the study, they provided a nominated contact for Kantar Public to liaise with to administer the surveys from September 2021 onwards.

The contact strategy further included a reminder letter and reminder email sent a few weeks after the original mailing. While attempting to recruit from the original sample of 120 schools it became clear that reaching the headteacher was problematic as letters often did not reach them promptly and school receptionists were sometimes reluctant to put calls through to them. Therefore, the reminder strategy was adjusted so letters and emails were sent to both the headteacher and head of sixth form/pastoral lead to broaden our reach at the telephone follow up stage. The correspondence to the headteacher was amended to inform them that we would also be contracting the head of sixth form or pastoral lead about the study, while the letter to the head of sixth form/pastoral lead stated that we had also written to the head teacher so there were aware of this. In the majority of cases where schools agreed to participate, agreement was secured as a result of schools contacting Kantar Public via email after correspondence eventually reached an appropriate contact.

Given these challenges in recruitment, we decided to also use the reserve sample of schools (120 additional schools on top of the initial sample of 120 schools), to maximise our changes of securing participation with schools.

Initially 47 schools signalled their intent to participate in the study, although 12 of these subsequently declined by the end of the Summer holidays, leaving 35 schools in the sample. A further two schools who supplied information about form numbers did not in the event send out any email invitations, so the final number of participating independent schools was 33.

7.3 Fieldwork approach

Independent school students and their parents were only contacted to take part in the survey online. As we had no information about their personal information in advance, it would not have been possible to conduct any face-to-face follow-ups with these samples.

In September 2021, the 35 schools who had initially agreed to participate were contacted again to request information regarding the form names and number of pupils in each form for sampling purposes. As detailed in section 7.1.4, school forms/tutor groups were randomly drawn, and this information was fed back to the nominated contact at each school along with school-specific survey links to send to the selected pupils and to a

parent or guardian. Each school received two school-specific survey links, one for young people and one for their parents, which allowed progress to be tracked at school level.

This was a key difference between the NPD and independent survey administration. For the NPD sample, we had contact details of NPD students in advance; therefore NPD students and their parents accessed the survey via the website using unique log in details. As we did not have contact details of independent school students in advance, the survey for this group was accessed via a school-specific open link sent by the school (schools were provided with a template email which they could use or adapt) and contact details of students and their parents were collected within the questionnaire. This led to a manual matching process at the data editing stage to link independent school students with their parents to create household-level data (see section 10.3). Editing was also required to remove a small number of cases where the student (and their parent) was not in the correct survey cohort[27] or where there were duplicates (see section 10.1).

Survey fieldwork for the independent school sample took place from October 2021, with the majority of survey invites issued in the autumn term (September to December 2021). Progress was monitored on a regular basis to see which schools had sent out the survey invitations and also to monitor any unusual response patterns. For example, an exceptionally high return rate may have suggested that all pupils had been invited to participate (rather than just those in sampled forms). Monitoring progress enabled us to contact specific schools to reconfirm that the survey links had been sent out and to ask them to send email reminders to boost response rates where interim response was looking low. As noted in section 7.2 above, in the event responses from at least one pupil or parent were achieved from 33 schools.

Although independent students and their parents did not access the survey via the survey landing page used by NPD students and parents, they were still able to access the website for further information such as FAQs, the privacy notice and details of the helpline.

All fieldwork documents relating to the independent school sample can be found in Appendix C.

[27] In some cases we suspected that form groups may have contained a mixture of Year 12 and Year 13 students which may have led to some Year 13 students receiving the survey link – any such cases were detected and removed as part of the survey cleaning process (see section 10.1)

7.4 Questionnaire differences (independent vs NPD sample)

The questionnaire was in the most part the same for the NPD (state school) sample and the independent school sample, although there were a few differences relating to logistical issues as follows:

Young people

- While state school respondents were asked to confirm their details as supplied in the NPD to verify that there were correct, independent school pupils were asked to enter their name/address details as we did not already have these.
- Independent school pupils were asked to confirm that they attended the school they had been sampled from and were screened out if not. They were also asked to provide their form or tutor group so we could validate this against the selected sample of forms (see section 10.1).
- Independent school pupils were asked directly for their ethnicity.[28]
- While state school participants received their incentive codes as soon as they completed their survey, incentive codes for independent school participants were sent out in four batches by email (or post if no email address was recorded) to allow for validation of the completed surveys and to ensure participants did not receive multiple incentives in the small number of cases where duplicate entries were detected (see section 10.1).

Parents

- While state school parents were asked to confirm their child's name and address, independent school parents were asked to enter their child's full name since we did not already have this data (this was required to aid the household matching process, see section 10.3).
- Parents were asked to confirm that their child attended the school linked to the sample record and were screened out if not.
- As for independent school students, parents received incentive codes by email or post after survey completion to allow for an interim process of validation.

[28] As discussed in section 3.1.1, this was originally omitted from the original issue NPD sample as it was erroneously thought this would be obtained by NPD linkage

8 Survey response

This chapter covers the number of achieved cases in different sample types, the response rates associated with each sample type, data linkage consent rates and rates of survey break-offs. The key sample types covered are: sample source (NPD sample or independent school sample); sample batch (original or reserve); main or boost sample (NPD only); and data collection mode (online or face to face). Unless otherwise stated all numbers quoted refer only to valid (useable) cases left on the datafiles after the removal of cases for quality control reasons and those defined as unusable partials (see section 10.1 for the definition of the threshold criteria used to define a useable case).

8.1 Summary of achieved interviews

A total of 13,787 valid young person interviews and 11,731 valid parent interviews were achieved at Wave 1. At a household level this represented 10,051 paired interviews where an interview was achieved with both the young person and one of their parents – termed a complete household - and a further 5,416 unpaired interviews where either a young person or a parent interview was achieved but not both – termed a partial household. This means that at least one interview was achieved at 15,467 separate households.

Although both paired and unpaired interviews have been included on the Wave 1 dataset, only complete households or partial households where an interview was achieved with a young person will be taken forward to Wave 2: a total of 13,787 households. Although included in the Wave 1 dataset, partial households where only a parent was interviewed (n= 1,680) will not be followed up at Wave 2. Where a young person interview was achieved at Wave 1 but there was no matching parent interview an attempt will be made to recruit a parent at Wave 2.

Tables 8.1 below shows the achieved interview numbers broken down by sample source (NPD or independent schools). The NPD sample achieved a far higher number of paired interviews as a proportion of total interviews achieved compared with the independent school sample. While 80% of all interviews achieved through NPD sample were paired, only 40% of all independent school sample interviews were paired. This was almost

certainly due to the limitations of the sampling approach used for independent schools as outlined in Chapter 7.

Table 8.1: Achieved interviews by sample source

	Sample source		Total
	NPD	Independent	
Young people	13,113	674	13,787
Parents	11,368	363	11,731
Total individual interviews	***24,481***	***1,037***	***25,518***
Complete households	9,845	206	10,051
Partial households – young person only	3,268	468	3,736
Partial households – parent only	1,523	157	1,680
All households with at least one interview	***14,636***	***831***	***15,467***

Table 8.2 below shows the achieved interview numbers for the NPD sample broken down by sample batch (original or reserve) and by whether it was main or Sutton Trust boost sample. While the original sample batch yielded proportionately more young person interviews than parent interviews the reserve sample yielded roughly the same proportion of young person and parent interviews. However, this did not translate into a better outcome in terms of complete households: the number of paired interviews as a proportion of total interviews achieved was the same for both sample batches.

This difference is probably explained by a difference in mailing strategies between the two sample batches. With the original sample batch, the survey invites were sent to young people and parents in separate envelopes. However, with the reserve sample both survey invites were sent in the same envelope with the aim of boosting the proportion of complete households. The mailing was addressed to the parent which required the parent to pass on the young person's invite to their child. However, this may not have happened in some cases. This lack of direct contact with the young person therefore may have acted to depress the response among young people rather than boost the response among parents, resulting in more complete households.

Table 8.2: Achieved interviews by sample batch and whether main or boost (NPD sample)					
	Original sample batch		Reserve sample batch		Total
	Main	Boost	Main	Boost	
Young people	9,341	832	2,813	127	13,113
Parents	7,842	675	2,727	124	11,368
Total individual interviews	***17,183***	***1,507***	***5,540***	***251***	***24,481***
Complete households	6,932	620	2,192	101	9,845
Partial households – young person only	2,409	212	621	26	3,268
Partial households – parent only	910	55	535	23	1,523
All households with at least one interview	***10,251***	***887***	***3,348***	***150***	***14,636***

Table 8.3 shows the achieved number of interviews by mode for the NPD sample. This shows that face-to-face interviews made up only a small proportion of total interviews. There were 518 parent interviews conducted in-home by interviewers (456 main sample and 62 boost sample), representing around only 5% of all parent interviews achieved. For young people there were 342 young person interviews conducted in-home by interviewers (303 main sample and 39 boost sample), representing around 3% of all young person interviews achieved.

The small number of face-to-face interviews reflects the challenges of conducting face-to-face fieldwork in the pandemic period and the need to change the fieldwork protocol at short notice as outlined in sections 6.2 and 6.5. Only around three in ten households originally selected to be issued face-to-face were actually worked during the fieldwork period. Of the households worked during the fieldwork period 72% were worked during the two in-home interviewing phases and 28% were worked during the knock to nudge phase (see Table 8.9 for overall response rate among households allocated to face-to-face and issued at any point during fieldwork).

Table 8.3: Achieved interviews by mode and whether main or boost (NPD sample)					
	Main sample		Boost sample		Total
	Online	F2F	Online	F2F	
Young people	11,851	303	920	39	13,113
Parents	10,113	456	737	62	11,368
Total individual interviews	***21,964***	***759***	***1,657***	***101***	***24,481***

8.2 Response rates for the NPD sample

This section examines the response rates achieved for the NPD sample broken down by main and Sutton Trust boost sample and sample batch (original or reserve). It also examines the response rates by sample demographics, incentive value, and survey mode. The response rate estimate for the independent school sample is shown separately in section 8.3.

Response rates were calculated separately for both young people and parents and for complete households. In all cases a field response rate was calculated based on the number of achieved valid interviews as a proportion of the issued sample. Additionally, a design weighted response rate was calculated which accounted for the disproportionate sample design.

Table 8.4 shows the response rate for the NPD sample at both an individual level and a household level. The overall response rate was 37% for young people and 32% for parents. The complete household response – where both a young person and parent interview was achieved – was 28%, with at least one interview being achieved at 42% of issued households.

Table 8.4 Overall response rate for NPD sample

	Individuals		Households	
	Young people	Parents	Complete households	At least one interview
Issued sample	35,719	35,719	35,719	35,719
Achieved interviews	13,113	11,368	9,845	14,636
	%	%	%	%
Response rate	*36.7*	*31.8*	*27.6*	*41.8*
Design weighted response rate	*36.9*	*32.3*	*27.8*	*41.5*

8.2.1 Response rates for NPD sample by main or Sutton Trust boost

Table 8.5 shows the response rates for the main NPD sample broken down by sample batch. The overall response rate was 36% for young people and 31% for parents. This provides a complete household response rate of 27%. Response rates were higher for the original sample batch compared with the reserve sample: for example, the response was 41% for young people from the original sample batch compared with 26% from the reserve sample.

This difference was due to several factors including: at the online stage more reminders were sent to the original sample compared with the reserve sample; some of the original sample was issued for face-to-face follow up but this was not the case for the reserve sample; and the overall fieldwork period for the original sample was much longer compared with the reserve sample. There was a 16 percentage point difference in young people's response to the reserve sample compared with the original sample and a 10 percentage point difference in parental response. As noted in section 6.4, in the original sample separate mailings were sent to parents and young people but in the reserve sample a single mailing (but containing separate letters) was sent addressed to the parents. This probably explains why the response differential between the two batches is greater for young people than parents.

Table 8.6 shows the response rates for the Sutton Trust boost sample broken down by sample batch. This showed the same response differential between the original and reserve sample batches as seen on the main sample. However, the boost sample also had a higher response compared with the main sample. Overall, the boost response rate was 48% for young people (compared with a response rate of 36% for the main sample) and 40% for parents (compared with a main sample response rate of 31%). The complete household response rate was also much higher: 36% compared with 27%.

This higher response seen in the boost sample may have been driven by the sample characteristics of the boost sample: the boost was targeted at high performing, disadvantaged pupils who may have a higher propensity to respond to surveys compared with disadvantaged pupils in general. However, perhaps a stronger driver for the higher response is the fact that a higher proportion of the boost sample were offered a £20 incentive rather than a £10 incentive compared with the main sample. This is because the higher incentive was targeted towards disadvantaged pupils: 21% of the boost sample was offered a £20 incentive compared with 14% of the main sample.

8.2.2 Response rates for NPD sample by incentive value

Table 8.7 shows the response rates for the main and Sutton Trust boost sample combined broken down by incentive value offered. This shows that the higher incentive did boost response across all the key outcome metrics: response among young people was 40% where a £20 incentive was offered compared with 36% where a £10 incentive was offered. There was a similar difference in the parent response rate (36% and 31% respectively) and for the complete household response (32% and 27% respectively). Given that the higher incentive was targeted towards pupils in more disadvantaged areas who we hypothesised would have a lower propensity to respond the fact that response rates were higher in this group suggests targeting the higher incentive in this way was reasonably successful.

8.2.3 Response rates for NPD main sample characteristics

As noted in Chapter 2, pupils in schools sampled from NPD were implicitly stratified by eligibility for free school meals (FSM), ethnicity, gender and special education needs (SEN) prior to drawing a systematic sample. In Table 8.8 the response rate is split out by key NPD variables: FSM eligibility, ethnicity, gender, SEN, as well as EAL (speaking English as an Additional Language). This is shown for the main NPD sample only because the Sutton Trust boost was sampled in a different way (see section 2.5).

The overall response among young people was reasonably consistent across different groups. Young people with special education needs, especially those with a SEN plan, responded at slightly lower levels than those with no SEN. Boys responded at a lower level than girls although this gender difference is fairly typical of all surveys of young people. Among ethnic groups, Black Caribbean young people were the only group to have a noticeably lower response, while young people from Indian, Pakistani and Bangladeshi ethnic backgrounds had higher than average response rates.

Parental response and complete household response for different groups broadly reflected the same patterns seen for young people.

8.2.4 Response rates for sample allocated to face-to-face fieldwork

As noted in section 8.1 less sample was worked face-to-face than had been intended. While 9,854 households were selected to be worked face-to-face only 3,003 were worked: less than a third of what was originally intended. Table 8.9 shows that only 30% of households allocated for face-to-face issue were worked at some point during the fieldwork period. Of the households worked during the fieldwork period 72% were worked during the two in-home interviewing phases and 28% were worked during the knock to nudge phase.

Of all young person sample worked in field, 17% resulted in a face-to-face interview, while 19% of parent sample worked resulted in a face-to-face interview[29]. An additional 3% of young person cases and 5% of parent cases yielded an online interview after a KTN visit was made. It is noticeable that the conversion rate for partially responding households was better than completely non-responding households (especially for KTN visits). It is likely that respondents in partially responding households were more receptive to the study given that someone else in the household had already taken part.

Based on which phase of fieldwork households were issued in, the response rate to the KTN phase can be estimated at around 21% of cases worked while the response rate of the in-home phase was about 27% of cases worked. The response to the in-home phase

[29] A case was considered to have been worked if it had a field outcome code recorded whether during the KTN phase or the in-home phases. Due to the way the field management system was set up it was not possible to separate out outcomes from the KTN phase and the in-home phases.

is lower than would normally be expected because interviewers only had limited time to make calls and so did not manage to complete their normal call patterns, including making a minimum number of calls at each household.

Table 8.5: Final response rates for main NPD sample by sample batch

	Original NPD main sample			Reserve NPD main sample			Total NPD main sample		
	Young people	Parents	Complete households	Young people	Parents	Complete households	Young people	Parents	Complete households
Issued sample	22,719	22,719	22,719	11,000	11,000	11,000	33,719	33,719	33,719
Achieved interviews	9,341	7,842	6,932	2,813	2,727	2,192	12,154	10,569	9,124
	%	%	%	%	%	%	%	%	%
Response rate	*41.1*	*34.5*	*30.5*	*25.6*	*24.8*	19.9	*36.0*	*31.3*	*27.1*
Design weighted response rate	*42.1*	*35.6*	*31.3*	*26.3*	25.6	*20.5*	*36.9*	*32.2*	*27.7*

Table 8.6: Final response rates for Sutton Trust boost sample by sample batch

	Original NPD boost sample			Reserve NPD boost sample			Total NPD boost sample		
	Young people	Parents	Complete households	Young people	Parents	Complete households	Young people	Parents	Complete households
Issued sample	1,600	1,600	1,600	400	400	400	2,000	2,000	2,000
Achieved interviews	832	675	620	127	124	101	959	799	721
	%	%	%	%	%	%	%	%	%
Response rate	*52.0*	*42.2*	*38.8*	*31.8*	*31.0*	*25.3*	*48.0*	*40.0*	*36.1*
Design weighted response rate	*50.0*	*40.4*	*36.6*	*28.7*	*30.8*	*23.4*	*45.5*	*38.4*	*33.6*

Table 8.7: Final response rates by incentive value for main and Sutton Trust boost NPD sample combined

	£10			£20			Total NPD sample		
	Young people	Parents	Complete households	Young people	Parents	Complete households	Young people	Parents	Complete households
Issued sample	30,412	30,412	30,412	5,307	5,307	5,307	35,719	35,719	35,719
Achieved interviews	10,989	9,471	8,131	2,124	1,897	1,714	13,113	11,368	9,845
	%	%	%	%	%	%	%	%	%
Response rate	*36.1*	*31.1*	*26.7*	*40.0*	*35.7*	*32.3*	*36.7*	*31.8*	*27.6*
Design weighted response rate	*36.8*	*32.1*	*27.5*	*37.9*	*34.0*	*30.5*	*36.9*	*32.3*	*27.8*

Table 8.8: Final response rates by sample characteristics for NPD sample

	Issued sample	Young people		Parents		Complete households	
		Achieved interviews	*Response rate*	Achieved interviews	*Response rate*	Achieved interviews	*Response rate*
	n	n	*%*	n	*%*	n	*%*
Free school meals							
FSM in last 6 years	17,353	6,172	*36%*	5,323	*31%*	4,618	*27%*
Not FSM in last 6 years	18,366	6,941	*38%*	6,045	*33%*	5,227	*28%*
SEN status							
SEN plan	1,237	304	*25%*	296	*24%*	228	*18%*
SEN support	4,516	1,369	*30%*	1,237	*27%*	1,031	*23%*
No SEN	29,966	11,440	*38%*	9,835	*33%*	8,586	*29%*
English as an additional language							
Yes	7,832	3,083	*39%*	2,691	*34%*	2,380	*30%*
No	27,887	10,030	*36%*	8,677	*31%*	7,465	*27%*
Gender of pupil							
Male	17,844	6,122	*34%*	5,521	*31%*	4,691	*26%*
Female	17,875	6,991	*39%*	5,847	*33%*	5,154	*29%*
Ethnicity of pupil							
Indian	1,912	794	*42%*	675	*35%*	604	*32%*
Pakistani	2,040	811	*40%*	704	*35%*	635	*31%*
Bangladeshi	1,895	818	*43%*	723	*38%*	657	*35%*
Black Caribbean	1,637	450	*27%*	387	*24%*	339	*21%*
Black African	2,115	819	*39%*	648	*31%*	573	*27%*
Mixed	2,102	751	*36%*	659	*31%*	565	*27%*
All other ethnicities	24,018	8,670	*36%*	7,572	*32%*	6,472	*27%*

Table 8.9: Response rates for sample allocated to face-to-face fieldwork

	Young people	Parents	Complete households[30]	Partial households- young person only	Partial households – parents only
Allocated sample	6,736	8,791	5,637	1,063	3,118
Issued to field	1,989	2,696	1,682	307	1,014
% of cases allocated	*31%*	*30%*	*30%*	*33%*	*29%*
Unproductive	*1,589*	2,038	1,387	202	651
% of cases issued	*80%*	*76%*	*82%*	*66%*	*64%*
F2F interview	*342*	*518*	*273*	*69*	*245*
% of cases issued	*17%*	*19%*	*16%*	*22%*	*24%*
Online interview after KTN	*58*	*140*	*16*	*42*	*124*
% of cases issued	*3%*	*5%*	*1%*	*14%*	*12%*
Any achieved interview	*400*	*658*	*289*	*111*	*369*
% of cases issued	*20%*	*24%*	*17%*	*36%*	*36%*

[30] For complete households the numbers shown are where an interview was achieved with both a young person and a parent. Interviews achieved with only one respondent at a complete household are included under partial households.

8.3 Response rates for the independent school sample

Given the sampling and fieldwork approaches used for independent school pupils (see Chapter 7), the response rates need to be calculated separately from the NPD sample. While it is possible to do this reasonably accurately the final response rate for the independent school sample represents our best estimate as the size of the eligible issued sample cannot be known with certainty: it relies on participating schools providing us with accurate pupil counts and following the fieldwork procedures correctly, including sending invitations out to the exact number of pupils and parents that were included in the forms sampled by Kantar Public.

The response rate was calculated in three steps as outlined below:

Step 1: School-level response

A total of 240 independent schools were sampled and issued to field as described in Chapter 7. From this sample, responses were received from young people at 33 schools representing a school response rate of 13.8% and from parents at 32 schools representing a school response rate of 13.3%. The reason for this difference is that one school only issued survey invitations to pupils but not to parents. Because of this the complete household school-level response rate was also 13.3%.

Step 2: Within-school response

Based on information provided by the 33 schools that sent invitations to young people, there were 2,005 pupils in the forms that were sampled. A total of 674 young people responded from this eligible sample representing an estimated within-school response rate of 33.6% for young people. Based on the 32 schools that sent survey invitations to parents there were 1,972 pupils (and so parents) in the forms that were sampled. A total of 363 parents responded from this eligible sample representing an estimated within-school response rate of 18.4%. From these achieved pupil and parent interviews a total of 206 were paired interviews which represented an estimated within-school response rate of 10.2% for complete households.

Step 3: Final response rate

The final independent school sample response rate was calculated by multiplying the school-level response by the within-school response. Table 8.10 summarises how the independent school sample response rate was calculated.

The overall response rate for both young people and parents was extremely low. This was driven primarily by the school-level response rate which reflects the low level of school co-operation despite the efforts that were made to contact and engage with schools (see Chapter 7). While school response rates have generally fallen in recent years across all school surveys, fieldwork was conducted during an extremely challenging period for

schools who were dealing with the COVID-19 pandemic, and this is likely to have had an impact on school response.

The estimated within-school response rates for young people are not too dissimilar to those achieved on the NPD sample although the parent response rate is noticeable lower. This also affected the number of paired interviews achieved and so made the complete household response rate particularly low. This lower response rate among parents was almost certainly due to the fieldwork approach which made it more difficult to engage parents in the study and the importance of their participation.

Table 8.10: Final response rates for independent school sample

	Young people	Parents	Complete households
Number of schools issued	204	204	204
Number of participating schools	33	32	32
School response rate	*13.8%*	*13.3%*	*13.3%*
Estimated issued pupil and parent sample	2,005	1,972	1,972
Achieved interviews	674	363	206
Within-school response rate	*33.6%*	*18.4%*	*10.2%*
Final response rate	*4.6%*	*2.4%*	*1.4%*

8.4 Data linkage rates

At the end of the survey young people were asked to consent to linking data from several different administrative sources to their survey data. Separate requests were made for each data linkage as follows:

- National Pupil Database (NPD) and Individual Learner Records (ILR) held by DfE
- National Tutoring Programme database held by Educational Endowment Foundation (EEF)
- Higher Education Access Tracker (HEAT) Service database
- DWP records
- HMRC records

Table 8.11 shows the level of consent for each separate request by data collection mode. Overall, 79% of young people agreed to their data being linked to at least one source, while 57% agreed to all linkage requests. Consent rates for individual linkages ranged from 74% for linkage to DfE records to 65% for linkage to HMRC records.

The level of consent was higher among those interviewed face-to-face compared with those who completed the survey online: for example, 89% of those interviewed face-to-face agreed to at least one linkage compared with 79% who completed the survey online. This finding is entirely consistent with other mixed mode studies which have found data linkage consent rates to be much higher on interviewer-administered surveys compared with online surveys.

Table 8.11: Consent to data linkage by mode of data collection

	Online		F2F		All modes	
	n	%	N	%	n	%
DfE	9,839	*73*	299	*87*	10,138	*74*
EEF	9,398	*70*	290	*85*	9,688	*70*
HEAT	9,241	*69*	282	*82*	9,523	*69*
DWP	8,863	*66*	275	*80*	9,138	*66*
HMRC	8,678	*65*	269	*79*	8,947	*65*
At least one consent	10,612	*79*	303	*89*	10,915	*79*
All consent	7,648	*57*	258	*75*	7,906	*57*
Base:	13,445		342		13,787	

8.5 Pattern of response by contact strategy

The cumulative pattern of interviews achieved during the fieldwork period for the original and reserve sample batches is shown in Figures 8.1 and 8.2. Given the relatively small number of interviews achieved face-to-face the pattern of response shown relates primarily to online completions and reflects the online contact strategy outlined in section 6.4.

Figure 8.1 Cumulative returns during fieldwork (original sample)

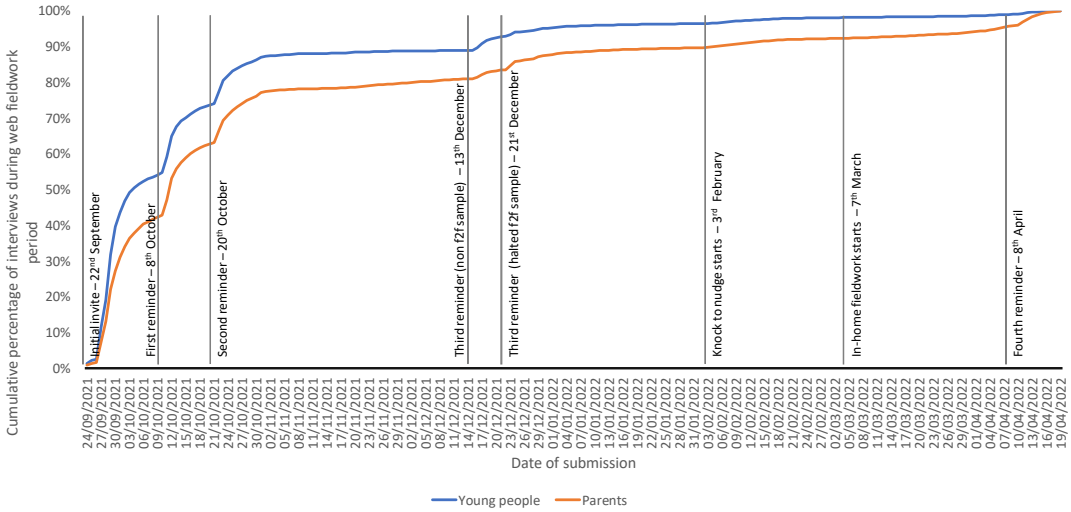
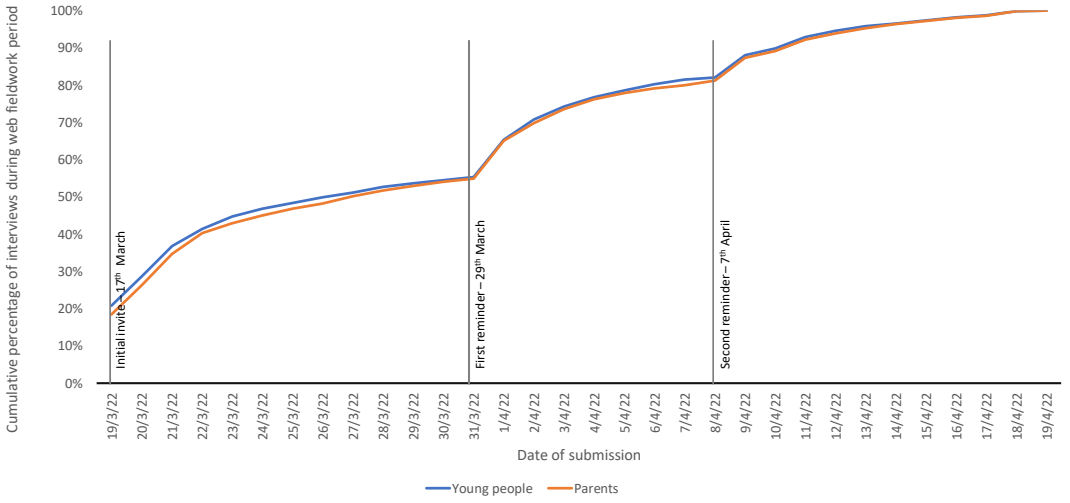


Figure 8.2 Cumulative returns during fieldwork (reserve sample)



8.6 Survey break-offs

A break-off occurs when a participant starts the survey but does not complete it. While break-offs can happen in all data collection modes it is primarily a concern only with online surveys. The break-off rate is defined as the number of respondents who abandoned the survey before reaching the point defined as representing a valid interview

(termed an 'unusable break-off interview') divided by the number who started the survey[31].

Break-offs in the young person's survey

A total of 16,074 young people started the survey, excluding invalid cases and those removed for quality reasons (see section 10.1). Given that there were 13,787 valid cases this represents a break-off rate of 14%. While this seems relatively high compared with other surveys there are two points worth noting:

- Firstly, the break-off rate in the independent school sample was extremely high: 66% of young people in the independent school sample who started the survey did not complete a valid interview. This is almost certainly related to the fieldwork procedures. One possibility is that because young people were only contacted by email through their school, they were less informed about the survey in advance and so started the survey out of curiosity to see what it was about, but quickly decided they did not want to continue. By contrast the break-off rate for young people in the NPD sample – who were likely to have a better idea of what the survey was about before starting - was only 7% which is more in line with the experience of other similar surveys.
- Secondly, it was decided to use a more exacting definition of a valid interview than is the case with many surveys which tend to use a more relaxed definition of a 'useable break-off'. In this case for an interview to be regarded as valid the first data linkage question (ZYPCONDFE) had to be answered. If the definition of a useable break-off had been set earlier in the interview (for example, before the start of the self-completion modules) the break-off rate would have been slightly lower.

Table 8.12 shows the cumulative number of young people who did not answer certain questions at different points in the survey and so gives an indication of where break-offs occurred. This shows that almost 70% of total break-offs occurred near the start of the survey during the household grid module: this was especially true for independent school pupils with 82% of all break-offs occurring in this section. This high break-off rate near the start of the survey is a pattern seen in most online surveys and may simply be a function of the fact that respondents tend to decide very quickly whether they wish to carry on or not. However, it may also indicate that young people felt uncomfortable being asked about personal details for everyone in their household and dropped out for this reason. Some surveys put household grid information towards the end of the survey to try and avoid this.

Apart from the high rate of break-offs early in the survey the remaining drop out occurred consistently throughout the rest of the survey suggesting there were no particular questions or modules which caused respondents to stop. However, around 5% of all break-offs happened near the end of the survey in the last self-completion module

[31] This does not include those who simply clicked on the survey link and entered the survey but did not answer any questions.

of, including 10% of break-offs in the NPD sample. This may have been due to the content of this module which asked questions about smoking, drinking alcohol, and drug taking.

Table 8.12: Number of break-offs at different points in the young person's survey

	Cumulative number and percentage of break-offs at different points in the survey:					
	NPD sample		Independent school sample		All sample	
	n	%	n	%	n	%
Section B (ZNUMHH)	0	*0*	0	*0*	0	*0*
Section C (ZCURSTAT)	502	*52*	1,087	*82*	1,589	*69*
Section E (ZSCHPERSON1)	583	*60*	1,117	*84*	1,700	*74*
Section F (ZSCHPERSON2)	642	*67*	1,168	*88*	1,810	*79*
Section G (ZATSCHOOL)	657	*68*	1,182	*89*	1,839	*80*
Section J (ZUNILIKELY)	720	*75*	1,219	*92*	1,939	*85*
Section L (ZSCHOOLATT2)	781	*81*	1,253	*95*	2,034	*89*
Section N (ZPEERSUPP)	848	*88*	1,295	*98*	2,143	*94*
Section O (ZCIGFREQ)	872	*90*	1,304	*99*	2,176	*95*
Section P (ZYPCONDFE)	964	*100*	1,323	*100*	2,287	*100*

Break-offs in the parents' survey

The patterns seen in the parents' survey were largely similar to what has already been seen with young people. A total of 13,276 parents started the survey excluding invalid cases and those removed for quality reasons, which given 11,731 valid cases represents a break-off rate of 12%. As with young people the break-off rate among parents in the independent school sample was much higher compared with parents in the NPD sample (55% and 9% respectively).

Table 8.13 shows the cumulative number of parents who did not answer certain questions at different points in the survey and so gives an indication of where break-offs occurred. As with young people there is a relatively high level of dropout near the start of the survey. More than half of the total break-offs have happened by the second module (XREVIEW), including 87% of break-offs in the independent school sample. After this the rate of drop off is reasonably consistent through the rest of the questionnaire, although there is a slight increase in break-offs in the module where parents are asked detailed questions about their occupation.

Table 8.13: Number of break-offs at different points in the parents' survey

	Cumulative number and percentage of break-offs at different points in the survey:					
	NPD sample		Independent school sample		All sample	
	n	%	n	%	n	%
Section A (XGENDERTYP)	226	*20%*	306	*70%*	532	*34%*
Section B (XREVIEW)	458	*41%*	379	*87%*	837	*54%*
Section C (XCHILDOFT)	512	*46%*	386	*89%*	898	*58%*
Section D (XECONACBEFORE)	590	*53%*	390	*90%*	980	*63%*
Section E (XTENURE)	672	*61%*	397	*91%*	1,069	*69%*
Section F (XEDUC1)	871	*78%*	414	*95%*	1,285	*83%*
Section G (XINCTYP)	949	*85%*	419	*96%*	1,368	*89%*
Section H (XBEENVAC)	988	*89%*	426	*98%*	1,414	*92%*
Section I (XLIFESTYLE1)	1,014	*91%*	426	*98%*	1,440	*93%*
Section J (XGHQ1)	1,050	*95%*	429	*99%*	1,479	*96%*
Section K (XFINANCIALMAND)	1,088	*98%*	432	*99%*	1,520	*98%*
Section L (XETHNIC)	1,110	*100%*	435	*100%*	1,545	*100%*

9 Interview length and device choice

This chapter covers information on survey interview length and type of device chosen to complete the survey.

9.1 Overall interview length and by survey section

Throughout this section, we use the median rather than the mean interview length to derive an average figure, as this minimises the effect of outliers (for example people who have taken a long pause in the middle of the interview[32]).

The target interview length was 30 minutes for both audiences, a target which was broadly met (slightly under for parents). It is important to note that we did not have the scope or time within the overall timetable to verify interview length through formal piloting, and therefore we needed to rely on more informal estimates of interview length when finalising the questionnaire.

The median interview length for young people was 32:07 minutes and the median interview length for parents was 24:50 minutes. It should also be noted that these overall lengths do not consistently include the time spent claiming the incentive[33] which is estimated to add around 2.5 minutes to the medians (an estimated median of approximately 35 minutes for young people and approximately 27.5 minutes for parents if this additional time is included).

Tables 9.1 and 9.2 show the median interview lengths by survey audience both overall and by module.

[32] Interviews completed across multiple sessions have been excluded from this analysis
[33] This is always included in face-to-face interviews, but only sometimes included in online interviews. This is because if the respondent does not close their browser quickly after claiming the incentive, the extra time spent claiming their incentive is not recorded.

Table 9.1: Median survey length overall and by section: young people	
	Median (minutes:seconds)
Total interview length	32:07
By section:	
A. Introduction, verification and opening demographics	2:06
B. Household grid	1:36
C. Current status	0:22
D. Qualifications studying	1:19
E. Education during lockdown 1/Year 10 (April-July 2020)	2:22
F. Education during lockdown 3/Year 11 (January-March 2021)	1:38
G. Education during Year 11 when schools were open (September-December 2020 and March-July 2021)	1:06
H: Catch up	1:40
I: Cancelled assessments (asked to a random half sample)	1:59
J. Education and career aspirations	2:21
K. Extra-curricular activities pre and post-pandemic (asked to a random half sample)	1:56
L. Attitudes to education (including motivation)	0:33
M. Health and wellbeing (CASI)	3:26
N. Friends, peers and family support (CASI)	1:40[34]
O. Health Related Behaviours (CASI)	
P. Linkage	2:32
Q. Recontact, signposts and closing screens	1:01

[34] Due to a missing timestamp in the script modules N and O have been combined

Table 9.2: Median survey length overall and by section: parents	
	Median (minutes:seconds)
Total interview length	24:50
By section:	
A. Introduction and verification checks	2:06
B. Attitudes to education	1:17
C. Parenting, home learning, tuition & catch-up	3:24
D. Working status across the pandemic	1:51
E. Parental tenure, HRP and occupational details	2:15
F. Parental education	1:03
G. Parental income	0:59
H. COVID History and vaccination (CASI)	0:35
I. Pandemic impact on family life (CASI)	1:49
J. Parent health and wellbeing (CASI)	2:06
K. Disadvantage (CASI)	1:26
L. Closing demographics	0:29
M. Contact details, signposting and closing screens	2:10

9.2 Total interview length by different characteristics

Table 9.3 below displays the overall interview length by different characteristics of the sample and respondent. There is relatively little variation across young people and parents with different characteristics, although on average face-to-face interviews were around 10 minutes longer than online self-completion interviews for both young people and parents, which is as expected. Online interviews completed on a laptop or tablet took longer than those completed on a smartphone.

Table 9.3: Median survey length by sample and respondent characteristics

		Young people	Parents
		Median (minutes:seconds)	Median (minutes:seconds)
Mode	Online	31:51	24:24
	F2F	41:56	33:15
Sample type (NPD)	Main	32:14	24:49
	Boost	31:16	25:22
Sample source	Original	32:24	24:56
	Reserve	31:23	24:37
School type of young person	State	32:08	24:51
	Independent	31:22	24:41
Incentive value	£10	32:06	24:50
	£20	32:22	24:59
Module[35]	Module A	32:08	-
	Module B	32:04	-
Gender of respondent	Male	31:53	24:54
	Female	32:17	24:56
Device (online only)	Laptop/PC	34:04	25:45
	Tablet	36:36	27:57
	Smartphone	30:36	23:24
Has partner	Yes	-	25:14
	No	-	24:22

[35] This indicates where content was only addressed to a random half-sample (see section 3.1.1)

9.3 Device choice

Young people and parents could complete the survey on a range of devices. The profile of completions by device is shown in Table 9.4. The device used was captured as part of survey paradata.

Overall, over six in ten (63%) young people completed the survey on a smartphone, while this figure was for slightly lower for parents (60%). A little over a third of each group completed on a laptop or PC.

Table 9.4: Devices used by respondents

	Young people	Parents
Smartphone	62.8%	59.9%
Laptop/PC	36.0%	37.6%
Tablet/other	1.2%	2.6%

10 Data preparation

This section describes all aspects of data quality assurance and data preparation.

10.1 Data quality assurance

There were two sets of interviews: one for the young person and one for the parent. The complete files including all valid and non-valid interviews initially comprised the following numbers of cases:

- 16,259 young persons
- 13,485 parents

These cases then underwent a set of data quality and validity checks:

Stage 1: Removal of non-valid cases

The files were first cleaned to remove non-valid cases as follows:

Table 10.1: Criteria for removal of non-valid cases (Stage 1)

	Description	Exclusion criteria	Number of YP flagged	Number of parents flagged
Unusable break-offs[36]	Did not reach the completion threshold (see below for definition of this)	Exclude all	2,313	1,614
Duplicates	For example, for the NPD sample, if completed both online and F2F. And, for the independent sample, if this was completed more than once by the same respondent.	In these situations, we removed the least completed interview, or if the same completion status, we removed the later interview.	88	66
Wrong year	Flag if the pupil birthday is not in period June 2004 to end of Oct 2005 (this allows some	If an interview was flagged with this status, a manual check was done to across	47	36

[36] The counts here include other invalid cases so they are higher than the count of unusable partials in section 8.8 which do not include invalid cases.

	buffer around the expected 1 September 2004 – 31 August 2005 school year) AND school year (X/ZSYCheck) is not in year 12. For independent schools a check on form name was also done to ensure we only kept cases within sampled forms.	both parent/pupil interviews (where we had both) to reach a decision about whether this looked to be a valid case. For independent schools, if the form name was not in a sampled form they were removed. For those with out of range birthdays form name was also manually checked to see if they are in the right year.		
Total to remove	Any of the above		2,419	1,680

Young person data completion threshold: Interviews are classed as complete if all sections of the questionnaire are completed (up to the end of Section P, including ZYPCONHMRC) and as usable partial interviews if the questionnaire is completed up to the beginning of linkage questions in Section P (including ZYPCONDFE).

Parent data completion threshold: Interviews are classed as complete if all sections of the questionnaire are completed (up to the end of Section L, including XBRBAND), and as usable break-off interviews if the questionnaire is completed until the end of the self-completion part, up to the beginning of Section L, including XETHNIC).

After these removals we had the following number of valid cases:

- 13,840 young persons
- 11,805 parents

Stage 2: Quality assurance of cases to identify those which indicate poor quality data

Based on the remaining valid cases we then assessed the data across a range of data quality flags including interview length, same response was selected through attitude batteries (straight-lining)[37] and repeatedly picking only one option across multi-coded questions.

Based on examining the distribution of recorded interview lengths of these remaining cases we decided to flag all cases where the interview length was < 0.25 * median

[37] Parent grids checked are XSCHOOLATT, XHOMELEARN, XLifestyle1, XLifestyle2, Young person grids checked are ZCatchup, ZCatchupConcern, ZTAGConcernG, ZTeachTime, ZJobAtt, ZSATI-GDSF, ZGAD2PHQ2, ZPeerSupp, ZSOCPROV, ZGHQ1- ZGHQ12.

interview length. The interview lengths for online and face-to-face interviews were assessed together.

For young persons, the median interview length was initially calculated as 39.3 minutes and < 0.25 of the median is < 9.82 mins.

For parents, the median interview length was initially calculated as 32.4 minutes and < 0.25 of the median is < 8.1 mins.

It should be noted that these calculations used basic script timings, which includes a 10 minute timeout period for online respondents who did not fully finish the script[38]. This means the times stated here are longer than the final times noted in chapter 9. Also, because the time cut offs are less than 10 minutes any online cases that had not fully finished the script would not have been flagged in this purely time-based check.

Table 10.2: Cases removed due to quality assurance (Stage 2)

	Description	Exclusion criteria	Number of YP flagged	Number of parents flagged
Short interview length	Flag cases where the length of interview is shorter than ¼ of the initially calculated median length	Exclude all	21	40
Other indications of speeding	Each grid question was checked to see if all answers were the same (i.e. 'straight-lining'). We also checked to see number of answers given at each multi-response question. Flag if all answers in all grid questions are straight-lined AND if only one answer at all multi-response questions.	Exclude all	15	35
Independent school pupil removal	The pupil was not at school in England during year 11 or pupil was in a state school in year 11 and surveyed as part of the NPD sample.	Exclude all	17	4
Total to remove	Any of the above		53	74

[38] The initial median interview length used at this stage was later found to be an over-estimate due to a system issue where online respondents who 'timed out' had a default extra 10 minutes added to the length. The interview lengths were later adjusted and correct timings are included in Chapter 9.

Following these Stage 1 and 2 removals the final count of interviews were as detailed in Tables 10.3 and 10.4.

Table 10.3: Breakdown of young person interviews by type of sample and completion status

	Fully completed	Break-off Complete - useable	Total
Main sample - original	9,322	19	9,341
Main sample - reserve	2,811	2	2,813
Independent school sample	669	5	674
Total: Main and independent school	12,802	26	12,828
Boost sample - original	832	0	832
Boost sample - reserve	127	0	127
Total	13,761	26	13,787

Table 10.4: Breakdown of parent interviews by type of sample and completion status

	Fully completed	Break-off Complete - useable	Total
Main sample - original	7,835	7	7,842
Main sample - reserve	2,724	3	2,727
Independent school sample	362	1	363
Total: Main and independent school	10,921	11	10,932
Boost sample - original	675	0	675
Boost sample - reserve	123	1	124
Total	11,719	12	11,731

The young person is the primary cohort member so any parent interviews with no matching young person interview are not treated as part of the analytical sample. As such in the parent data file only those with a matching young person interview are weighted.

Table 10.5: Breakdown of parent interviews by type of sample and young person/parent interview matching status		
	Parent in matched household	**Parent not in matched household**
Main sample - original	6,932	910
Main sample - reserve	2,192	535
Independent school sample	206	157
Total: Main and independent school	9,330	1,602
Boost sample - original	620	55
Boost sample - reserve	101	23
Total	10,051	1,680

10.2 Coding

The questionnaire collects some information as full verbatim answers, mainly where people select an 'Other (please specify)' response and type in a verbatim answer.

For the uncoded data files the verbatim answers are included in the data sets.

For the coded data files the verbatim answers have been removed from the data sets; and the verbatim responses were used to either back code into existing responses or some new responses were created if there were sufficient verbatim answers of the same type.

Questions where coding was applied are listed below; unless noted otherwise, only back coding was applied.

Young Person:

- zalevsub – new responses added to data
- zapny
- zaslevsub – new responses added to data
- zasux – new responses added to data
- zbtecsub – new responses added to data
- zcambsub
- zcaradv
- zcaradvinf – new responses added to data
- zcurstat
- zgcsesub
- zgradeneedbg
- zgradeneedwg

- zibsub
- zindepy11
- zneet
- zoned – data edits made based on responses
- zperm
- zschoolmiss – new responses added to data
- zsgwy – new responses added to data
- zstatey12
- zstatus2y
- zstud
- zvcqc – new responses added to data

Parents:

- xasux – new responses added to data
- xchildasp
- xeconchange – new responses added to data
- xhomqual – new responses added to data
- xrelatpar – new responses added to data
- xreligion

Responses added from coding have the note "(created from coding)" in their SPSS variable labels.

Employment details given in the parent survey are used to derive SIC 2020, SOC 2020 and NSSEC for either respondent or their partner. Detailed SIC and SOC codes are excluded from the UK Data Service (UKDS) safeguarded data deposit, but NSSEC variables were added:

- W1_XNSSEC
- W1_XPNSSEC

The NSSEC coding is based on SOC 2020 using the ONS derivation tables linked here: https://www.ons.gov.uk/file?uri=/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020/soc2020volume3thenationalstatisticssocioeconomicclassificationnssecrebasedonthesoc2020/tables912v3.xlsx

10.3 Data outputs

The survey data is available in two data files, one for young persons and one for parents.

Identifiers

Household identifiers

The parent and young person interviews are in separate data sets and a household serial is included so interviews from the same household can be matched across the 2 datasets. This is the variable "HHserial" which is a 6-digit serial.

Individual identifiers

Each interview was assigned an individual serial, this is the "HHserial" with "1" appended for young person interviews and "2" appended for parent interviews. This is the variable "INDserial" which is a 7-digit serial.

Matching young person and parent interviews into households

Because both a young person and a parent/guardian were invited to COSMO, some work has been done to ensure that we can match these interviews as young person/parent pairs (i.e. household) during data processing. Below we explain how this was done.

Data from NPD allowed the provision of unique, named invites to young people in state schools, as well as their parents (as parents of named young person). In the questionnaire, there were verification questions to make sure the invited people were filling out the survey.

For NPD sample the parent and young person were matched by sample serial. Note, the HHserial assigned in the datafiles is not the sample serial.

However, unique invites were not possible for young people in independent schools and their parents, as invitations were done at the school level and could only be unique at the school level. Therefore matching young people and parents as households required further effort.

For independent school young people, matching households were established by a process of reviewing responses to verification questions. As a first step the data was cleaned and split into separate pupil and parent datasets to simplify the household level linkage. The cleaned data contained 694 pupil and 369 parent interviews.

Afterwards, the data was grouped into candidate pairs blocked by schools. All pupils were linked to all parents within a school to create all possible candidate pairs.

To assess the probability of the candidate pair being a correct match an Expectation/Conditional Maximisation (ECM) algorithm was used.

The following features were created and fed to the algorithm to decide if candidate pairs were correct or incorrect.

Pupil reported	Parent reported	Method	Threshold
Pupil first name	Pupil first name	Jaro-Winkler[39]	0.8
Pupil surname	Pupil surname	Jaro-Winkler Damerau-Levenshtein[40]	0.8
Pupil surname	Parent surname	Jaro-Winkler Damerau-Levenshtein	0.8
Pupil date of birth	Pupil date of birth	Transposed month and day	0.5
Pupil date of birth	Pupil date of birth	Damerau-Levenshtein	0.8
Household telephone	Household telephone	Damerau-Levenshtein	0.8
Respondent email address	Respondent email address	Damerau-Levenshtein	0.8
Perks email address	Perks email address	Damerau-Levenshtein	0.8
Respondent email address	Pupil full name	Damerau-Levenshtein	0.5

All the features above are unweighted. However, as we have created two features, which measure the edit distance of the pupil's surname and date of birth, in effect, the pupil's surname and date of birth is weighted double.

The ECM algorithm is an unsupervised, generalised EM classifier closely related to the Fellegi and Sunter (1969) framework. It calculates the probability of a candidate pair being a correct match. Based on all probabilities it predicts correct and incorrect pairs.

Through succeeding clerical review the following confusion matrix was constructed.

	Positive	*Negative*
True	212	8,532
False	7	0

The result then underwent a manual review to establish final parent/young person matched pairs.

[39] The Jaro-Winkler distance measures the similarity of two strings. It is normalised between 0 and 1 where 0 means complete dissimilarity and 1 an exact match.
[40] The Damerau-Levenshtein distance measures the minimum number of operations necessary to convert one string into the other (insertions, deletions, substitutions or transposition).

Variable names

Questionnaire variables in the data files were named to match the questionnaire question name whenever possible.

The standard convention used here for the naming of multi-responses and grid variables was to add a numeric suffix to the variable name in form of "VARNAME_01". For these suffixes we consistently used _96 for "Other", _97 for "None of These"/"None", _98 for "Don't Know" and _99 for "Prefer not to say".

For wave 1 a prefix of "W1_" was added to variable names.

Variable description

For questionnaire variables the variable labels used in the data files are based on the wording from the survey questionnaire, shortened and kept comprehensible.

For multi-response and grid variables the variable labels were based on the wording of the question and response text from the questionnaire. For grids the value labels used were also taken from the wording from the survey questionnaire, for multi-response variables the value labels used were No/Yes to indicate if that response was selected by the respondent.

Missing values

The missing values used in the data files are used to identify questions with no valid answer, for these there are 2 types:

1) The codes -8 and -9 are used by respondents to denote the following:

-8: Don't know

-9: Refused/Prefer not to say

These codes above, whenever they exist, were explicitly selected by respondents in the questionnaire (or communicated as such to an interviewer if face-to-face).

2) The codes -1 and -2 are used for where no respondent answer was recorded:

-1: Not applicable

-2: Question not asked due to respondent answers or script

The -1 "Not Applicable" code is used if the question was intentionally not asked due to script routing.

The -2 "Question not asked due to respondent answers or script" is used where a question should have been asked but wasn't asked/no data recorded. These would be cases where responses based on "Other" verbatim coding that were later back-coded to a different response, meant the script did not move down the right route based on the

edited response; in a small number of cases script issues may have caused an answer to not be recorded.

There is an exception in the data sets to the use of "-1" and "-2" for useable break-off interviews after the cut off points (which were XETHNIC for parents and ZYPCONDFE for young people): If the case was a usable partial interview and the codes "-1" or "-2" were required for questions after the cut off, they were set to system missing instead. As shown at the beginning of the section, this applies to a small number of useable break-off interviews and a small number of variables that exist after the cut off points.

Variable order

The order of variables in the data files follow the questionnaire order as below:

- Identifier variables
- Sample information variables
- Questionnaire variables in the same order
- SIC, SOC and NSSEC variables were added in the position of the work questions.
- Para-data variables for interview device, interview time, number of interview sessions.
- Completion flag
- Flags to denote inconsistencies in household grid data
- Geodemographic variables
- Schools level information variables
- Weighting variables

The para-data variables included are:

- W1_DeviceDetails_kantarDevice – Device used for interviews, if multiple devices used the last used is recorded. All face-to-face interviews were done on laptops.
- W1_DeviceDetails_browserName to W1_DeviceDetails_IOMUA – Detailed device details recorded by script, if multiple devices used the last used is recorded.
- W1_XTP_TIMESTAMP_SEC_A_START to W1_XTP_TIMESTAMP_SEC_M_END (Parent file), W1_ZTP_TIMESTAMP_SEC_A_START to W1_ZTP_TIMESTAMP_SEC_FINALCHECKS_END (Young person file) – survey section timestamps recorded by script.
- W1_INT_STARTTIME – Start time recorded by script
- W1_INT_FINISHTIME – End time recorded by script, if final screen is not reached this is the time 10 minutes after last action.
- W1_SURVEY_SUB – The month when the interview was completed
- W1_MULTI_SESSION – Number of different sessions the interview was completed over, recorded from the number of time the survey was opened.
- W1_COMP_FLAG – Completion status of the interview.
- "Fully completed", Parent: answered to XBRBAND, young person: answered to ZYPCONHMRC.

- "Partial Complete – useable", Parent: answered to XETHNIC, young person: answered to ZYPCONDFE.
- "Partial complete – unusable", all other started interviews that did not reach the specified questions.

Geodemographics variables included are:

- W1_msoa11cd – Middle Layer Super Output Areas (2011)
- W1_ladcd – Local Authority Districts
- W1_Polar4_quintile – POLAR4 Quintile
- W1_Region – Region
- W1_IMD_decile – English Index of Multiple Deprivation (LSOA Decile)
- W1_IDAC_decile – English Income Deprivation Affecting Children Index (LSOA Decile)

The school level information included are:

- W1_EstablishmentTypeGroupcode – School Establishment Type Group
- W1_AdmissionsPolicycode – School Admissions Policy
- W1_PercentageFSMQuintiles – Percentage FSM Pupils in School (Quintiles)
- W1_TrustSchoolFlagcode – Trust School Flag

The details of the weighting variables' are described in the weighting section.

Data edits

For the coded data files there were detailed data checks to ensure the bases for variables reflect who should have answered the questions. Most of the edits from these were due to responses for back coding. If a back coded response from a respondent suggests they should not have answered some following questions those variables were set to '-1: Not applicable'. If the back coded response suggests the respondent should have answered some following questions but no data was recorded those variables were set to '-2: Question not asked due to respondent answers or script'.

Back edits to the young person data were made to variable W1_ZCURSTAT_01 based on the respondent answer to ZNOED. If the ZNOED open text answer indicated the respondent was still in school or college the W1_ZCURSTAT_01 was set to "1". Further base edits were then done using this as a basis.

In few cases there were inconsistencies in the household member information collected in the young person data (gender, relationship to the young person, and age group). These were not edited on the data sets; however, these were denoted using flag variables (W1_HHFlag1 through W1_HHFlag4). Data users can refer to the COSMO Data User Guide for further details.

11 Weighting

Weighting was required to ensure that the sample was representative of the population and that the findings are generalisable. For this study, weights were needed for two reasons: (1) to compensate for the disproportionate sample design, and (2) to compensate for systematic non-response.

11.1 Different weights in the data files

A number of different weights have been generated for different types of analysis. In this section, we summarise the different weights and when each should be used. The full process outlining how the weights were generated can be found later in this Chapter.

11.1.1 Analysis of survey data only

Six different weights have been generated for analysis of the survey data (not subsets of sample that agreed to data linkage). These weights are summarised in the table below.

Table 11.1 – Weights to use when analysing survey data only

Weight name	N cases	This weight should be used when analysing
W1_MainFamilyFull_weight	9,330	**Main study** complete households survey data (where both the pupil and a parent in the household responded).
W1_MainYPFull_weight	12,828	**Main study** young people's survey data (i.e., this includes data from some households where just the pupil responded to the survey)
W1_BoostFamilyFull_weight	1,681	**Sutton eligible** (from both main and boost) complete households survey data
W1_BoostYPFull_weight	2,249	**Sutton eligible** (from both main and boost) young people's survey data (including partial households)
W1_AllFamilyFull_weight	10,051	**All** (main study and boost) complete households survey data
W1_AllYPFull_weight	13,787	**All** (main study and boost) young people's survey data (including partial households)

Households where only the parent was successfully interviewed are also included in published datasets. However, these have not been given a weight (the value is missing). This means that these cases will be excluded from analysis when any of the survey weights are applied.

11.1.2 Analysis of survey data linked to administrative education records

Additional weights have been produced to analyse the survey data linked to administrative education records (from the National Pupil Database (NPD)). Separate weights are required for this analysis, as not all respondents consented to having their survey responses linked to the administrative data. These weights compensate for systematic differences in agreement rates to the linkage. These weights are summarised in the table below.

Table 11.2 – Weights when analysing survey data linked to administrative educational data

Weight name	**N cases**	**This weight should be used when analysing**
W1_MainFamily_NPD_weight	6,896	**Main study** survey data for complete households linked to NPD education records.
W1_MainYP_NPD_weight	9,385	**Main study** survey data for young people only - linked to the NPD education records.
W1_BoostFamily_NPD_weight	1,293	**Sutton Trust boost eligible** (from both main and boost) complete households survey data linked to NPD education records
W1_BoostYP_NPD_weight	1,704	**Sutton Trust boost eligible** (from both main and boost) young people's survey data linked to NPD education records
W1_AllFamily_NPD_weight	7,454	**All** (main study and boost) complete households survey data linked to NPD education records
W1_AllYP_NPD_weight	10,116	**All** (main study and boost) young people's survey data linked to NPD education records

The dataset with these weights will be accessible via the ONS Secure Research Service.

11.2 Deriving weights used for analysis of survey data only

A four-stage process was used to derive the weights which should be used when analysing survey data only:

- W1_MainFamilyFull_weight
- W1_MainYPFull_weight
- W1_BoostFamilyFull_weight
- W1_BoostYPFull_weight
- W1_AllFamilyFull_weight
- W1_AllYPFull_weight

Exactly the same process was used for all these weights[41]. The weights included in the archived dataset are the final weights – once all stages of weighting outlined below have been completed (design weighting, non-response weighting, and calibration weighting with constraints).

11.2.1 Stage 1 – design weighting

All respondents were given a design weight equal to one divided by their sampling probability. The sampling sections (Chapter 2 for NPD and Chapter 7 for independent schools) outline how the sampling probabilities were calculated.

Respondents that were at a state school in Y11 and at an independent school in Y12 could potentially have been sampled from both sample sources used. The design weight calculated accounts for this joint selection probability.

11.2.2 Stage 2 – non-response modelling

All respondents were given a non-response weight equal to one divided by their estimated response probability.

For children sampled from the NPD, the estimate of response probability was a fitted value, derived from a main effects logistic regression model in which the dependent variable was a binary response indicator. The predictors included in the model were:

- Gender
- Free school meals eligibility (last 6 years)
- Ethnicity
- English as an Additional Language

[41] Although, it should be noted that the "Sutton Trust boost eligible" weights only include individuals sampled from the National Pupil Database. As such, steps related to the independent sample do not apply to these two weights.

- SEN provision type
- KS2 reading score (banded)
- KS2 maths score (banded)
- KS2 GPS (Grammar, Punctuation and Spelling) score (banded into terciles)
- Establishment Type (GIAS)
- Number of pupils at the school banded (GIAS)
- Percentage of population with Level 4+ qualification (Census 2011 quintiles) - based on the MSOA the school is located in
- Percentage of homes that are owned (Census 2011 quintiles) - based on the Middle Layer Super Output Area (MSOA) the school is located in
- Region (former government office region) – based on school location
- Census 2011 Output Area Classification group - based on pupil address
- ONS rural/urban classification - based on pupil address
- Income Deprivation Affecting Children Index (IDACI) quintile - based on pupil address

For the non-response weighting, missing data points were included as valid categories for variables with high levels of missing data (in particular, the KS2 variables that each had 8-9% of data missing). For other variables that had a low proportion of missing data (e.g., ethnicity) the missing data points were imputed.

For children sampled from an independent school, the estimate of response probability was a compound value based on (i) the probability that the sampled school co-operated, and (ii) the probability that the young person participated given that the school co-operated. A pair of main effect logistic regression models were used to estimate these probabilities. The predictors included in each model were the same:

- Mixed or single sex (GIAS)
- Whether school has boarders (GIAS)
- Number of pupils at the school banded (GIAS)
- Census 2011 Output Area Classification supergroup - based on school location
- Region (former government office region) - based on school location
- ONS rural/urban classification - based on school location
- Percentage of population with Level 4+ qualification (Census 2011 quintiles) - based on the MSOA the school is located in
- Percentage of homes that are owned (Census 2011 quintiles) - based on the MSOA the school is located in

For respondents who could have been sampled from both the NPD and the independent school sample frame[42], the non-response weight was derived as follows:

*(p(sampled, NPD) * p(response | sampled from NPD))*

+

[42] Those that attended a state school in Y11 and an independent school in Y12

*(p(sampled, independent school) * p(response | sampled from independent school))*

-

*(p(sampled, NPD) * p(response | sampled from NPD) * p(sampled, independent school) * p(response | sampled from independent school))*

This was divided by the already-calculated sampling probability to yield an estimate of response probability for these respondents.

11.2.3 Stage 3 – calibration weighting

Every respondent was given a 'base' weight equal to one divided by the product of the sampling and estimated response probabilities.

The base-weighted respondent sample was then calibrated so that its distribution with respect to some critical variables was an exact match for the estimation population, so far as this is known.

In practice, we must use a proxy for the true estimation population, with two divisions:

1. Those who were studying at a state school in Year 11 (regardless of whether sampled from NPD or from an independent school)
2. Those who were studying at an independent school in both Years 11 and 12 (these individuals could only be sampled via an independent school)

The size of the first division of the population was equal to the number of valid records in the NPD extract of Year 11 students in Spring 2021 (= 580,278).

The calibration weight for a respondent from the first population division was equal to their base weight multiplied by a calibration factor. The iterative proportional fitting algorithm (also known as 'raking' or 'rim weighting') was then used to generate these calibration factors.

The following subclasses were included in the calibration matrix. The benchmarks used as targets for the weighting were based on the Y11 NPD Spring 2021 extract used to draw the sample:

- Size of school's Year 11:
  - Under 150 pupils
  - 150-249
  - 250+ pupils
- Type of school provision:
  - Special
  - Alternative
  - Selective Other

  - Other
- Region: 9 English regions
- FSM eligibility * SEN status:
  - FSM last 6 years & Education Health and Care (EHC) plan
  - FSM last 6 years & other SEN status
  - FSM last 6 years & no SEN status
  - No FSM last 6 years & EHC plan
  - No FSM last 6 years & other SEN status
  - No FSM last 6 years & no SEN status
- Language
  - English is primary language / not recorded
  - English is an additional language
- Sex:
  - Male
  - Female
- Ethnic group:
  - Indian
  - Bangladeshi
  - Pakistani
  - Black African
  - Black Caribbean
  - White British / no data
  - White non-British
  - Mixed / Other
- Sex * broad ethnic group:
  - Male White British
  - Male Other
  - Female White British
  - Female Other
- KS2 scores (maths / reading / GPS)
  - Upper tertile in all three
  - Upper tertile in two, middle tertile in one
  - Upper tertile in one, middle tertile in two
  - Others with at least one in upper tertile or at least two in middle tertile
  - Lower tertile in two, middle tertile in one
  - Lower tertile in all three
  - Missing data

The size of the second division of the population was estimated – as there are no published official statistics for this group. There are different ways of estimating this population size, all of which are likely to be somewhat inaccurate:

- **Approach 1** – using GIAS data to estimate the population size (this is consistent with how the sample was drawn). The GIAS database provides the number of pupils across all year groups in eligible independent schools. To estimate the number of students in Y12 at each school - we divided the total number of pupils attending the school by the number of year groups. By adding this up for the 1,112 eligible independent schools, we estimate a total population of c.33,422.
- **Approach 2** – using published DfE Key Stage 4 (KS4) data and information from the Independent Schools Council census. DfE data[43] indicates 49,597 independent school pupils took part in KS4 in 2020/2021 (which we can use as an estimate for the Y11 population size). However, available data suggests that there are fewer pupils attending Y12 of independent schools than Y11. For instance, the ISC census[44] suggests that there is a drop of c.8%pts from Y11 to Y12 (for their member schools across the UK). On this basis we might estimate that there are c.45k pupils in independent schools in England in Y12.

These two approaches lead to different population size estimates. Reflecting this uncertainty, for the purpose of weighting we have used an estimated total population size of 40,000 Y12 independent school pupils. However, it is important to note that this total population includes young people that studied at state school in Y11 and that are included in the first division for the calibration stage of weighting.

The weighted first division data was used to estimate the number of pupils that attended state school in Y11 but then moved to independent for Y12 (questions were included in the survey to capture this). The size of the second division could then be estimated by subtracting this figure from 40,000.

Finally, the calibration weight for a respondent from the second population division was calculated: their base weight, divided by the sum of all base weights for this division, and multiplied by the estimated population size of children studying at an independent school in both Year 11 and Year 12.

11.2.4 Stage 4 – constrained calibration weighting

The calibration weight divided by the design weight may be thought of as a combination of non-response weight and non-coverage weight but is mainly a non-response weight because the non-coverage level for this study was very small. We refer to this as the non-inclusion weight.

Constraining the variance of the non-inclusion weight should improve the precision of survey estimates. This can be done by trimming the non-inclusion weights (also

[43] https://explore-education-statistics.service.gov.uk/data-tables/permalink/e8942369-b2a3-406c-80b0-06a94a7881d6
[44] https://www.isc.co.uk/media/7496/isc_census_2021_final.pdf

sometimes referred as truncating). The process of trimming ensures that the minimum and maximum non-inclusion weights do not exceed (different) set values.

A respondent's non-inclusion weight should have a theoretical lower bound equal to the response rate multiplied by the mean non-inclusion weight. There are no theoretical upper bounds for non-inclusion weights but very large, outlier values are likely to inflate the mean squared error of weighted descriptive statistics, compared to a trimmed version. It was decided that non-inclusion weights should be trimmed to be no larger than c.4 times the median value.

After trimming, the respondent sample was re-calibrated using the trimmed weights as base weights rather than the original base weights. This process was repeated until no non-inclusion weight exceeded c.4 times the median value.

With the final state school weight applied (with the stage 4 constraints), we obtained the following estimates for the number of pupils that attended a state school in Y11 and an independent school in Y12:

- W1_MainFamilyFull_weight – 4,764
- W1_MainYPFull_weight – 4,784
- W1_AllFamilyFull_weight – 4,780
- W1_AllYPFull_weight – 5,293

This left us with the following population estimates for the population that attended independent school in both Years 11 and 12:

- W1_MainFamilyFull_weight – 35,236 pupils (40,000 – 4,764)
- W1_MainYPFull_weight – 35,216 pupils (40,000 – 4,784)
- W1_AllFamilyFull_weight – 35,220 pupils (40,000 – 4,780)
- W1_AllYPFull_weight – 34,707 pupils (40,000 – 5,293)

These figures were used in the final constrained calibration weight as the estimated population size of children studying at an independent school in both Year 11 and Year 12.

11.3 Approach to derive weights for analysis of survey data linked to administrative education records

The following weights were created to use when analysing the sub-set of respondents that have agreed to NPD data linkage.[45]:

- W1_MainFamily_NPD_weight
- W1_MainYP_NPD_weight
- W1_BoostFamily_NPD_weight
- W1_BoostYP_NPD_weight
- W1_AllFamily_NPD_weight
- W1_AllYP_NPD_weight

In generating these weights, we used the weights previously generated for analysis of all survey data (including those that did not consent to linkage). These 'base weights' were adjusted to compensate for systematic differences in consent rates for the linkage. The approach we used was as follows:

$$\textbf{NPD weight}_i = \textbf{Base weight}_i * [1/ \textbf{Pr(NPD)}_i]$$

Where:

- Base weighti is the final weight generated for analysis of survey data assigned to respondent i; and
- Pr(NPD)i is the estimated probability that respondent i has provided consent to NPD linkage

The following table shows which base weight was used to generate each NPD weight.

[45] For independent sampled pupils – these are the young people that both consented to linkage and also provided the personal information required for the linkage (full name, date of birth, school they attended in Y11).

Table 11.3 – The base weight used to generate each NPD weight	
Base weight	**NPD weight generated from the base weight**
W1_MainFamilyFull_weight	W1_MainFamily_NPD_weight
W1_MainYPFull_weight	W1_MainYP_NPD_weight
W1_BoostFamilyFull_weight	W1_BoostFamily_NPD_weight
W1_BoostYPFull_weight	W1_BoostYP_NPD_weight
W1_AllFamilyFull_weight	W1_AllFamily_NPD_weight
W1_AllYPFull_weight	W1_AllYP_NPD_weight

A logistic regression was used to estimate Pr(NPD) – the probability that a survey respondent also gave NPD linkage consent. The predictors used for this modelling are listed below.

For pupils at state school in Y11, the variables used as predictors were:

- Size of school's Year 11
- Type of school provision
- Region
- FSM eligibility * SEN status
- Language
- Ethnic group
- Sex * broad ethnic group
- KS2 scores (maths / reading / GPS)

For pupils at independent school in Y11 and Y12[46], the variables used as predictors were:

- Whether school is mixed or single sex
- Whether school has boarders
- Number of pupils at the school (banded)
- Census 2011 Output Area Classification supergroup - based on school location
- Region - based on school location
- ONS rural/urban classification - based on school location
- Percentage of population with Level 4+ qualification (Census 2011 quintiles) - based on the MSOA the school is located in
- Percentage of homes that are owned (Census 2011 quintiles) - based on the MSOA the school is located in

[46] The weights generated for Sutton Eligible cases do not include this stage, as Sutton eligible cases could only be sampled from the NPD.

11.4 Effectiveness of weighting

As shown in the following tables, the profile of the design weighted sample was a reasonably close match to the population profile. The additional stages of weighting were then used to compensate for remaining small imbalances. To examine the effectiveness of the final weights in restoring sample representativity we have compared the final weighted survey sample profiles to the benchmark population statistics (which were used when calibrating the data).

Table 11.4 – Main study full households (9,330)

	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)[47]	Final weight (linked to NPD 6,896)[48]
FSM eligibility * SEN status	Percent	Percent	Percent	Percent	Percent
FSM last 6 years & EHC plan	1.9	1.4	1.1	1.9	1.9
FSM last 6 years & other SEN status	4.3	6.4	3.6	4.3	4.4
FSM last 6 years & no SEN status	18.3	34.0	18.8	18.3	18.3
No FSM last 6 years & EHC plan	2.1	1.0	1.4	2.1	2.1
No FSM last 6 years & other SEN status	6.6	4.1	5.9	6.6	6.4
No FSM last 6 years & no SEN status	61.0	51.0	68.9	61.0	61.2
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7
Ethnicity					
Indian	2.7	6.3	3.3	2.7	2.7
Bangladeshi	1.7	6.5	2.2	1.7	1.7
Pakistani	4.2	5.9	4.8	4.2	4.2
Black African	3.8	5.2	3.7	3.8	3.8
Black Caribbean	1.2	3.6	0.9	1.2	1.2
White British / no data	64.9	55.6	69.1	64.9	65
White non-British	5.8	4.4	5.3	5.8	5.6
Mixed / Other	9.9	10.5	10.3	9.9	10
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7
Gender					
Male	48.2	46.8	47.4	48.2	48.1

[47] With **W1_MainFamilyFull_weight** applied
[48] With **W1_MainFamily_NPD_weight** applied

Table 11.4 – Main study full households (9,330)					
	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)[47]	Final weight (linked to NPD 6,896)[48]
Female	46	51.1	52.3	46	46.2
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7
Ethnicity * Gender					
Male White British	33.3	26.3	32.5	33.3	33.2
Male Other	15.0	20.5	14.9	15.0	14.9
Female White British	31.7	29.2	36.6	31.7	31.7
Female Other	14.4	21.9	15.6	14.4	14.4
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7
KS2 - maths, reading, GPS					
Upper tertile in all three	14.3	18.3	21.7	14.3	14.3
Upper tertile in two, middle tertile in one	11.8	13.9	15.4	11.8	11.7
Upper tertile in one, middle tertile in two	10.7	11.7	12.3	10.7	10.6
Others with at least one in upper tertile or at least two in middle tertile	24.5	24.7	24.4	24.5	24.2
Lower tertile in two, middle tertile in one	10.2	9.2	8.5	10.2	10.3
Lower tertile in all three	14.1	12.9	10.7	14.1	14.4
Missing data	8.7	7.2	6.7	8.7	8.7
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7
English as an Additional Language					
English is primary language / not recorded	78.4	74.9	83.0	78.4	78.4
English is an additional language	15.9	23.1	16.7	15.9	15.9
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7
School size					
Under 150	22.8	24.8	24.1	22.8	22.6
150-249	53.7	55.8	57.3	53.7	53.9
Over 249	17.8	17.3	18.3	17.8	17.8
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7

Table 11.4 – Main study full households (9,330)					
	Population	**Unwtd (all cases)**	**Design weighted (all cases)**	**Final weight (all cases)[47]**	**Final weight (linked to NPD 6,896)[48]**
School provision					
Special	1.2	0.4	0.7	1.2	1.1
Alternative	0.8	0.7	0.5	0.8	0.8
Selective Other	4.2	5.3	6.1	4.2	4.3
Other	88	91.6	92.4	88	88.1
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7
School region					
East Midlands	8.2	7.9	8.6	8.2	8
East of England	10.6	9.8	11.7	10.6	10.7
London	14.1	19.6	14.4	14.1	13.9
North East	4.4	4.8	4.9	4.4	4.4
North West	13.1	12.8	13.0	13.1	13.1
South East	14.8	13.5	16.5	14.8	14.9
South West	8.8	7.5	9.3	8.8	8.7
West Midlands	10.7	12.5	11.6	10.7	10.8
Yorkshire and the Humber	9.6	9.4	9.7	9.6	9.7
Independent in Y11 and Y12	5.7	2.0	0.3	5.7	5.7

Table 11.5 – Main study Young People (12,828)					
	Population	**Unwtd (all cases)**	**Design weighted (all cases)**	**Final weight (all cases)[49]**	**Final weight (linked to NPD 9,385)[50]**
FSM eligibility * SEN status	Percent	Percent	Percent	Percent	Percent
FSM last 6 years & EHC plan	1.9	1.3	1.1	1.9	1.9
FSM last 6 years & other SEN status	4.3	6.3	3.6	4.3	4.3
FSM last 6 years & no SEN status	18.3	33.0	18.8	18.3	18.3
No FSM last 6 years & EHC plan	2.1	1.0	1.3	2.1	2.1
No FSM last 6 years & other SEN status	6.6	4.0	5.8	6.6	6.4
No FSM last 6 years & no SEN status	61.0	49.5	68.6	61.0	61.1
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7
Ethnicity					
Indian	2.7	6.0	3.3	2.7	2.7
Bangladeshi	1.7	5.9	2.0	1.7	1.7
Pakistani	4.2	5.6	4.6	4.2	4.2
Black African	3.8	5.5	4.0	3.8	3.8
Black Caribbean	1.2	3.5	0.9	1.2	1.2
White British / no data	64.9	54.2	68.9	64.9	65.0
White non-British	5.8	4.2	5.3	5.8	5.7
Mixed / Other	9.9	10.2	10.2	9.9	9.9
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7
Gender					
Male	48.2	44.5	46.3	48.2	48.2
Female	46.0	50.6	52.9	46.0	46.1
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7
Ethnicity * Gender					
Male White British	33.3	25.3	31.9	33.3	33.2
Male Other	15.0	19.2	14.4	15.0	14.9

[49] With **W1_MainYPFull_weight** applied
[50] With **W1_MainYP_NPD_weight** applied

Table 11.5 – Main study Young People (12,828)					
	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)[49]	Final weight (linked to NPD 9,385)[50]
Female White British	31.7	28.9	36.9	31.7	31.7
Female Other	14.4	21.7	16.0	14.4	14.4
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7
KS2 – maths, reading, GPS					
Upper tertile in all three	14.3	17.6	21.2	14.3	14.3
Upper tertile in two, middle tertile in one	11.8	13.4	15.3	11.8	11.7
Upper tertile in one, middle tertile in two	10.7	11.3	12.2	10.7	10.7
Others with at least one in upper tertile or at least two in middle tertile	24.5	24.2	24.3	24.5	24.5
Lower tertile in two, middle tertile in one	10.2	9.1	8.8	10.2	10.2
Lower tertile in all three	14.1	12.5	10.8	14.1	14.2
Missing data	8.7	7.1	6.7	8.7	8.7
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7
English as an Additional Language					
English is primary language / not recorded	78.4	73.3	82.9	78.4	78.4
English is an additional language	15.9	21.8	16.4	15.9	15.9
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7
School size					
Under 150	22.8	23.8	23.7	22.8	22.7
150-249	53.7	54.4	57.2	53.7	53.8
Over 249	17.8	17.0	18.4	17.8	17.8
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7
School provision					
Special	1.2	0.4	0.7	1.2	1.1
Alternative	0.8	0.7	0.5	0.8	0.8

Table 11.5 – Main study Young People (12,828)					
	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)[49]	Final weight (linked to NPD 9,385)[50]
Selective Other	4.2	5.0	6.0	4.2	4.3
Other	88.0	89.1	92.1	88.0	88.0
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7
School region					
East Midlands	8.2	7.8	8.7	8.2	8.0
East of England	10.6	9.9	11.6	10.6	10.7
London	14.1	19.0	14.5	14.1	13.9
North East	4.4	4.4	4.6	4.4	4.5
North West	13.1	12.3	13.0	13.1	13.1
South East	14.8	13.3	16.5	14.8	14.8
South West	8.8	7.4	9.3	8.8	8.8
West Midlands	10.7	12.2	11.6	10.7	10.7
Yorkshire and the Humber	9.6	8.9	9.5	9.6	9.7
Independent in Y11 and Y12	5.7	4.9	0.8	5.7	5.7

Table 11.6 – All full households – main and Sutton Trust boost (10,051)					
	Population	**Unwtd (all cases)**	**Design weighted (all cases)**	**Final weight (all cases)[51]**	**Final weight (linked to NPD (7,454))[52]**
FSM eligibility * SEN status	Percent	Percent	Percent	Percent	Percent
FSM last 6 years & EHC plan	1.9	1.3	1.1	1.9	1.9
FSM last 6 years & other SEN status	4.3	6.4	3.6	4.3	4.4
FSM last 6 years & no SEN status	18.3	38.2	19.0	18.3	18.2
No FSM last 6 years & EHC plan	2.1	0.9	1.4	2.1	2.1
No FSM last 6 years & other SEN status	6.6	3.9	5.9	6.6	6.4
No FSM last 6 years & no SEN status	61.0	47.3	68.8	61.0	61.2
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7
Ethnicity					
Indian	2.7	6.0	3.3	2.7	2.7
Bangladeshi	1.7	6.5	2.2	1.7	1.7
Pakistani	4.2	6.3	4.8	4.2	4.2
Black African	3.8	5.7	3.7	3.8	3.9
Black Caribbean	1.2	3.4	0.9	1.2	1.2
White British / no data	64.9	54.5	68.9	64.9	65.0
White non-British	5.8	4.4	5.4	5.8	5.6
Mixed / Other	9.9	11.2	10.4	9.9	10.0
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7
Gender					
Male	48.2	46.7	47.4	48.2	48.1
Female	46.0	51.4	52.3	46.0	46.2
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7
Ethnicity * Gender					
Male White British	33.3	25.7	32.4	33.3	33.2
Male Other	15.0	21.0	15.0	15.0	14.9

[51] With **W1_AllFamilyFull_weight** applied
[52] With **W1_AllFamily_NPD_weight** applied

Table 11.6 – All full households – main and Sutton Trust boost (10,051)					
	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)[51]	Final weight (linked to NPD (7,454))[52]
Female White British	31.7	28.8	36.5	31.7	31.8
Female Other	14.4	22.7	15.8	14.4	14.4
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7
KS2 - maths, reading, GPS					
Upper tertile in all three	14.3	20.2	21.7	14.3	14.3
Upper tertile in two, middle tertile in one	11.8	15.8	15.5	11.8	11.7
Upper tertile in one, middle tertile in two	10.7	11.8	12.3	10.7	10.6
Others with at least one in upper tertile or at least two in middle tertile	24.5	23.1	24.3	24.5	24.3
Lower tertile in two, middle tertile in one	10.2	8.5	8.5	10.2	10.3
Lower tertile in all three	14.1	12.0	10.7	14.1	14.4
Missing data	8.7	6.7	6.6	8.7	8.7
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7
English as an Additional Language					
English is primary language / not recorded	78.4	74.4	83.0	78.4	78.4
English is an additional language	15.9	23.7	16.7	15.9	15.9
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7
School size					
Under 150	22.8	24.0	24.1	22.8	22.6
150-249	53.7	56.4	57.4	53.7	53.8
Over 249	17.8	17.7	18.3	17.8	17.8
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7
School provision					
Special	1.2	0.4	0.7	1.2	1.1
Alternative	0.8	0.7	0.5	0.8	0.8
Selective Other	4.2	5.2	6.1	4.2	4.3

Table 11.6 – All full households – main and Sutton Trust boost (10,051)					
	Population	**Unwtd (all cases)**	**Design weighted (all cases)**	**Final weight (all cases)[51]**	**Final weight (linked to NPD (7,454))[52]**
Other	88.0	91.9	92.4	88.0	88.1
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7
School region					
East Midlands	8.2	7.8	8.6	8.2	8.0
East of England	10.6	9.6	11.7	10.6	10.7
London	14.1	20.7	14.4	14.1	14.0
North East	4.4	5.0	4.9	4.4	4.4
North West	13.1	12.7	12.9	13.1	13.1
South East	14.8	13.1	16.5	14.8	14.9
South West	8.8	7.4	9.3	8.8	8.7
West Midlands	10.7	12.6	11.6	10.7	10.8
Yorkshire and the Humber	9.6	9.4	9.6	9.6	9.6
Independent in Y11 and Y12	5.7	1.9	0.3	5.7	5.7

Table 11.7 – All Young People – main and Sutton Trust boost (13,787)					
	Population	**Unwtd (all cases)**	**Design weighted (all cases)**	**Final weight (all cases)[53]**	**Final weight (linked to NPD 10,116)[54]**
FSM eligibility * SEN status	Percent	Percent	Percent	Percent	Percent
FSM last 6 years & EHC plan	1.9	1.3	1.1	1.9	1.9
FSM last 6 years & other SEN status	4.3	6.3	3.6	4.3	4.3
FSM last 6 years & no SEN status	18.4	37.2	19.0	18.4	18.3
No FSM last 6 years & EHC plan	2.1	0.9	1.3	2.1	2.1
No FSM last 6 years & other SEN status	6.7	3.7	5.8	6.7	6.5
No FSM last 6 years & no SEN status	61.1	46.1	68.5	61.1	61.2
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7
Ethnicity					
Indian	2.8	5.8	3.3	2.8	2.7
Bangladeshi	1.7	5.9	2.1	1.7	1.7
Pakistani	4.2	5.9	4.6	4.2	4.2
Black African	3.8	6.0	4.0	3.8	3.8
Black Caribbean	1.2	3.3	0.9	1.2	1.2
White British / no data	65.0	53.4	68.8	65.0	65.0
White non-British	5.8	4.3	5.4	5.8	5.7
Mixed / Other	9.9	10.9	10.2	9.9	10.0
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7
Gender					
Male	48.3	44.5	46.3	48.3	48.3
Female	46.1	51.0	53.0	46.1	46.1
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7
Ethnicity * Gender					
Male White British	33.3	24.9	31.9	33.3	33.3
Male Other	15.0	19.6	14.4	15.0	15.0
Female White British	31.7	28.6	36.9	31.7	31.7

[53] With **W1_AllYPFull_weight** applied
[54] With **W1_AllYP_NPD_weight** applied

Table 11.7 – All Young People – main and Sutton Trust boost (13,787)					
	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)[53]	Final weight (linked to NPD 10,116)[54]
Female Other	14.4	22.4	16.1	14.4	14.4
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7
KS2 – maths, reading, GPS					
Upper tertile in all three	14.3	19.4	21.2	14.3	14.3
Upper tertile in two, middle tertile in one	11.8	15.4	15.4	11.8	11.7
Upper tertile in one, middle tertile in two	10.7	11.4	12.2	10.7	10.7
Others with at least one in upper tertile or at least two in middle tertile	24.5	22.7	24.3	24.5	24.5
Lower tertile in two, middle tertile in one	10.3	8.5	8.8	10.3	10.2
Lower tertile in all three	14.1	11.7	10.8	14.1	14.2
Missing data	8.7	6.6	6.7	8.7	8.7
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7
English as an Additional Language					
English is primary language / not recorded	78.4	73.0	82.8	78.4	78.4
English is an additional language	15.9	22.5	16.4	15.9	15.9
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7
School size					
Under 150	22.8	23.0	23.6	22.8	22.7
150-249	53.8	55.1	57.3	53.8	53.8
Over 249	17.8	17.4	18.4	17.8	17.9
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7
School provision					
Special	1.2	0.3	0.7	1.2	1.1
Alternative	0.8	0.6	0.5	0.8	0.8
Selective Other	4.2	5.0	5.9	4.2	4.3
Other	88.1	89.5	92.1	88.1	88.1

Table 11.7 – All Young People – main and Sutton Trust boost (13,787)					
	Population	**Unwtd (all cases)**	**Design weighted (all cases)**	**Final weight (all cases)[53]**	**Final weight (linked to NPD 10,116)[54]**
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7
School region					
East Midlands	8.2	7.7	8.7	8.2	8.1
East of England	10.6	9.6	11.6	10.6	10.6
London	14.1	20.0	14.5	14.1	14.0
North East	4.4	4.7	4.6	4.4	4.5
North West	13.2	12.3	13.0	13.2	13.1
South East	14.8	12.9	16.5	14.8	14.8
South West	8.8	7.2	9.3	8.8	8.8
West Midlands	10.7	12.2	11.5	10.7	10.7
Yorkshire and the Humber	9.6	8.9	9.5	9.6	9.7
Independent in Y11 and Y12	5.6	4.5	0.8	5.6	5.7

11.5 Design effects

To ensure that standard errors are estimated correctly it is important to take into account the impact of the weighting, clustering and pre-stratification. If this is not done, the confidence intervals estimated are likely to be too narrow and there is an increased risk of Type I errors (false positives).

The variables that need to be used:

- Weight variable – as outlined earlier in this section, the correct weight needs to be selected for each analysis. The weights that can be used are:
  - W1_MainFamilyFull_weight
  - W1_MainYPFull_weight
  - W1_BoostFamilyFull_weight
  - W1_BoostYPFull_weight
  - W1_AllFamilyFull_weight
  - W1_AllYPFull_weight
  - W1_MainFamily_NPD_weight
  - W1_MainYP_NPD_weight
  - W1_BoostFamily_NPD_weight
  - W1_BoostYP_NPD_weight
  - W1_AllFamily_NPD_weight
  - W1_AllYP_NPD_weight
- Cluster variable: W1_PSU_all
- Stratification variable*: W1_AnalysisStratum_v2

*If users run into issues when conducting sub-group analysis because of there not being two clusters in each stratum, we would suggest conducting the analysis with W1_SchoolStratum_v2. If there are further singleton stratum problems when using W1_SchoolStratum_v2, we would recommend omitting the stratification variable entirely from the survey design. While these adjustments may be necessary for standard errors to be estimated, it should be noted that they are likely to lead to slightly inflated.

11.5.1 Example design effects

In this section we provide some example design effects for two of the survey weights and for a range of different variables (estimated using the complex samples module of SPSS). This analysis is intended to illustrate how design effects vary depending on the weight used and the outcome of interest (and that design effects vary between categories for factors).

For each measure we present the estimated design effect for each category, and we have also produced a variable-level design effect estimate (the mean of the category level design effects weighted by the size of each category).

W1_AllYP_NPD_weight

Design effects have been estimated for four measures from the young person survey.

Table 11.8 – Example design effects when using the *W1_AllYP_NPD_weight* weight		
Whether school provided real-time online learning during first lockdown Apr to Jul 2020?	**%**	**Design effect**
Yes, for subject lessons	65.7%	2.34
Yes, for other reasons	14.6%	2.24
No	27.6%	2.38
Don't know	2.6%	1.46
Weighted mean design effect for variables		***2.32***
Likelihood that youth will ever apply to go to university to do a degree?		
Very likely	43.1%	2.27
Fairly likely	28.5%	1.64
Not very likely	16.0%	1.94
Not at all likely	12.4%	2.21
Weighted mean design effect for variable		***2.03***
Catch up tutoring? (derived)		
Offered neither individual nor small group tutoring	58.1%	1.77
Offered either individual or small group tutoring, but took up neither	15.3%	1.76
Received either individual or small group tutoring	26.5%	1.79
Weighted mean design effect for variable		***1.77***
GHQ12 - poor mental health indicator (derived - binary)		
No	55.4%	1.56
Yes	44.6%	1.56
Weighted mean design effect for variable		***1.56***
GHQ12 - poor mental health indicator (derived – 0-12 score)	*Mean*	
Mean	*3.9*	1.60
Design effect for variable		***1.60***

W1_AllFamilyFull_weight

Design effects have been estimated for four measures from the young person survey, and one measure from the parent survey.

Table 11.9 – Example design effects when using the *W1_AllFamilyFull_weight* weight

Whether school provided real-time online learning during first lockdown Apr to Jul 2020?	**%**	**Design effect**
Yes for subject lessons	65.7%	2.65
Yes for other reasons	14.9%	2.81
No	26.1%	2.44
Don't know	4.0%	1.51
Weighted mean design effect for variables		***2.58***
Likelihood that youth will ever apply to go to university to do a degree?		
Very likely	42.4%	3.12
Fairly likely	28.6%	2.00
Not very likely	16.0%	1.92
Not at all likely	13.0%	2.17
Weighted mean design effect for variable		***2.48***
Catch up tutoring? (derived)		
Offered neither individual nor small group tutoring	59.4%	2.12
Offered either individual or small group tutoring, but took up neither	14.7%	2.12
Received either individual or small group tutoring	25.8%	2.31
Weighted mean design effect for variable		***2.17***
GHQ12 - poor mental health indicator (derived - binary)		
No	57.4%	1.69
Yes	42.6%	1.69
Weighted mean design effect for variable		***1.69***
GHQ12 - poor mental health indicator (derived – 0-12 score)	*Mean*	
Mean	*3.8*	*1.72*
Design effect for variable		***1.72***

Table 11.9 (continued) – Example design effects when using the *W1_AllFamilyFull_weight* weight		
NSSEC for HRP (derived from parent survey)		
Cannot derive who HRP is	4.9%	1.65
Higher managerial and administrative occupations	6.4%	3.27
Higher professional occupations	11.7%	3.79
Lower professional and higher technical occupations	13.7%	1.79
Lower managerial and administrative occupations	4.7%	1.81
Higher supervisory occupations	0.8%	1.83
Intermediate occupations	9.0%	1.61
Employers in small organisations	0.2%	1.25
Own account workers	7.2%	1.77
Lower supervisory occupations	0.9%	1.46
Lower technical occupations	3.7%	1.78
Semi-routine occupations	8.8%	1.57
Routine occupations	10.4%	1.73
Never worked and long-term unemployed	3.8%	1.08
Occupations not stated or inadequately described	13.9%	1.71
Weighted mean design effect for variable		***2.03***