Royal Society: Computing Education Strand 3 technical report

JN 260138933

1. Modelling the uptake of computing in secondary education

This section presents two statistical models that explore the uptake of GCSE computing at Key Stage 4 (KS4) and A level computing at Key Stage 5 (KS5) with regard to the characteristics of pupils and their schools. More specifically:

- The first model considers computing uptake at GCSE, focusing on schools where at least one pupil has completed GCSE computing (i.e. schools where it is safe to assume that the subject was offered). The model processes data about pupil-level and school-level characteristics to isolate their individual effect on GCSE computing uptake. For simplicity, we will refer to this model as the *computing entry (CE) model.*
- The second model considers computing uptake at A level for a particular subset of pupils: those who have previously taken up computing at KS4. The model examines data about pupil-level and school/college-level characteristics to understand their individual effect on whether pupils continue with computing education at KS5, once they have entered computing education at KS4. For simplicity, we will refer to this model as the *computing* c*ontinuation (CC) model.*

Combining insights from the CE and CC models, the underlying objective of the analysis presented in this section is to investigate the journey of pupils who choose computing throughout secondary education, reflecting on its possible determinants. The paragraphs that follow describe the methodological approach to constructing the CE and CC models and present their statistical outputs. We then reflect on the modelling findings and discuss the insights that emerge on the basis of the analysis.

1.1 Methodological approach to constructing the CE and CC models

The process of constructing the CE and CC models involved (a) the preparation of corresponding analysis datasets using secondary data sources; (b) the selection of pupil-level and school-level characteristics that should be accounted for when exploring computing uptake; and finally (c) fitting the models.

1.1.1 Preparing the analysis datasets

To facilitate the construction of the CE and CC models, two analysis datasets (the CE analysis dataset and the CC analysis dataset, respectively) were compiled using available data sources:

- **The CE analysis dataset** was compiled by linking data from three data sources maintained by the Department for Education: the National Pupil Database (NPD); the School Workforce Census (SWC); and Edubase. The dataset was filtered to only include pupils at year 11 in the academic year 2014-15. Further filtering excluded from the dataset pupils studying at schools where no year 11 pupils had completed computing GCSE. This exclusion aimed to narrow down the scope of the CE models to schools where it is safe to assume that computing GCSE was offered. We acknowledge that there may be some inaccuracy in the assumption that computing GCSE was not offered at schools where no pupils completed it; however, this methodological decision was deemed as more preferable than its alternative (i.e. maintaining in the dataset pupils at schools where no pupils completed GCSE computing) as it will allow us to focus more closely on the determinants of entering KS4 computing education *other than the school-level provision of KS4 computing education*.[1]

= **The CC analysis dataset** was compiled by linking four secondary data sources maintained by the Department for Education: the National Pupil Database (NPD); the Individualised Learner Record (ILR); the School Workforce Census (SWC); and Edubase. The dataset included year 12 and year 13 pupils during the academic year 2014-15. The CC analysis dataset was filtered to only include pupils who had completed GCSE computing at KS4. By including only this particular subset of KS5 pupils (as opposed to all KS5 pupils), the CC analysis dataset will help us understand which pupil-level and school/college-level characteristics are most likely to determine whether pupils *continue* their computing secondary education after KS4. [2]

It is noted that the CE and CC analysis datasets include data from a particular cross-section of pupils: the 2014-15 cross-section. However, the two analysis datasets can be seen as representative of two theoretical populations of pupils *beyond the particular 2014-15 cross-section*: (a) the (wider) population of year 11 pupils at schools where computing GCSE is offered; and (b) the (wider) population of year 12 and year 13 pupils who have taken up computing at KS4. Inferences based on the analysis presented in this section aim to use the data from the particular 2014-15 cross-section in order to draw inferences regarding the wider pupil-populations of interest.

Constructing the CE and CC analysis datasets involved an extensive phase of data pre-processing. The pre-processing was implemented using the statistical package SPSS 23 and comprised two elements:

= A **variable-inspection** element, whereby variables in the original data sources (i.e. the NPD; the SWC; the ILR, and Edubase) were examined one by one using statistics of central tendency and dispersion; frequency distributions; and appropriate visualisation tools. This process focused on ensuring that the data included in the CE and CC analysis datasets are informative and of high-quality by (a) eliminating duplicate records; (b) removing variables with high proportions of missing values which may limit the statistical power of the analysis (i.e. variables with more than 50% of their values missing); and (c) discarding variables that based on their frequency distribution are expected to perform poorly in differentiating between units of analysis (i.e. variables where the modal code has a frequency greater than 95%). At the same time, the data was inspected for outliers that could indicate potential data-quality issues.

= A **variable-transformation** element, whereby new variables were derived based on variables in the original data sources (i.e. the NPD; the SWC; the ILR, and Edubase), where this was deemed as necessary. This process aimed to generate data-points that are more appropriate for the purposes of the analysis and involved (a) the computation of new categorical variables by combining codes of the original categorical variables together; and (b) the computation of new continuous variables by applying scaling transformations on original continuous variables (such as centring around their mean value).

---

[1] Information about the databases used to construct the CE analysis dataset, including their coverage, can be found at: https://www.gov.uk/government/collections/national-pupil-database (NPD); https://www.gov.uk/guidance/school-workforce-census (SWC); and http://www.education.gov.uk/edubase/about.xhtml (Edubase).

[2] Information about the databases used to construct the CC analysis dataset, including their coverage, can be found at: https://www.gov.uk/government/collections/national-pupil-database (NPD); https://www.gov.uk/government/collections/individualised-learner-record-ilr (ILR); https://www.gov.uk/guidance/school-workforce-census (SWC); and http://www.education.gov.uk/edubase/about.xhtml (Edubase).

1.1.2 Selecting CE and CC modelling variables

Constructing the CE and CC models involved a systematic variable-selection process. The objective of this process was to distinguish between two tiers of variables: (a) those that are likely to be relevant to pupils' choices regarding computing at KS4 or KS5 and (b) those that are not. The first tier of variables was subsequently included in the proposed substantive CE and CC models, while the second tier was excluded. Including relevant variables in the modelling process is important in order to analytically account for the theoretically interesting factors that potentially shape the outcomes that the analysis intends to study. At the same time, excluding non-relevant variables from the modelling process mitigates the risk of over-fitting, i.e. the risk of proposing substantive models that are specific to the particular datasets used for the analysis, rather than models that are descriptive of the mechanism that underlies the outcomes we intend to study in the wider pupil-populations of interest (see paragraph 1.1.1).

The relevance of variables in the CE and CC analysis datasets with regard to the outcomes that interest the CE and CC models (i.e. whether pupils enter computing at KS4 or whether they continue with computing at KS5, respectively), was determined by means of bivariate and multivariate techniques:

- **Bivariate techniques** examined each of the outcomes of interest against individual variables in the CE and CC analysis datasets. By assessing measures of statistical dependency and correlation, bivariate techniques highlighted variables with strong links to the outcomes of interest. Through this process, we identified *disjoint* two-dimensional data-spaces within the CE and CC analysis datasets, where it is potentially interesting to explore the distribution of the outcomes that interest the analyst.
- Selecting variables for the CE and CC models solely on the basis of bivariate analysis risks overlooking theoretically interesting variables, when their relationship to the outcomes of interest is not immediately obvious. For example, variable X may appear unrelated to the outcome variable Y; however, an interesting relationship may emerge within a sub-space of the analysis data-space defined by a third variable Z. To ensure that the variable-selection process does not exclude potentially interesting variables from the proposed substantive CE and CC models, bivariate variable-selection techniques were complemented with a **multivariate approach.** Schematically speaking, the multivariate approach controlled for an extensive mix of variables in the CE and CC analysis datasets *simultaneously* to identify variables that have a statistically significant "predictive value" in relation to the outcomes of interest. To determine this "predictive value", the analysis employed one-level logistic regression models that eliminated non-relevant variables based on a pre-defined algorithm[3].

The set of variables selected from the CE and CC datasets based on bivariate and multivariate statistical criteria was further reviewed using expert knowledge in the domain of computing in secondary education. This process aimed to ensure that key variables of theoretical interest have not been omitted from the scope of the modelling exercise. The final mix of variables used in the CE and CC models is presented in Table 1.1.2, below. Appendix 1 presents the complete list of variables that the analysis considered (some of which were eliminated by the bivariate and multivariate variable-selection techniques described just above).

[3] The algorithm used employs a statistical criterion based on the probability of the likelihood-ratio statistic and the maximum partial likelihood estimates.

Table 1.1.2: Variables in the computing entry (CE) and the computing continuation (CC) models

Variables in the CE model	Variables in the CC model
Gender	Gender
Ethnic background	Ethnic background
Special education needs (SEN)	Total GCSE (and equivalents) score
First language	Attainment in KS4 Computing
Total GCSE (and equivalents) score	Attainment in KS4 Maths
Attainment in KS4 Maths	School type
Quintiles of the number of KS4 pupils in school	Deciles of the Income Deprivation Affecting Children Index (IDACI) for pupils' school postcode
Gender of school admissions	Region where pupil's school is
Quintiles of percentage of pupils who are white British in school	
Percentage of pupils whose first language is other than English	
Percentage of pupils recorded as eligible for free school meals	
Percentage of pupils achieving at least 5 GCSE's A star to C	
Percentage of pupils with special education needs (SEN) in school	
Deciles of the Income Deprivation Affecting Children Index (IDACI) for pupils' school postcode	
Region where pupil's school is	
Urbanisation level where pupil's school is	
Number of teachers at pupil's school known to have computing qualification(s)[4]	

1.1.3 Fitting the substantive CE and CC models

Having selected the sets of variables that should feature in the CE and CC models (see paragraph 1.1.2), the analysis proceeded with fitting the substantive model specifications using a *multilevel binary logistic regression* mechanism. This particular modelling mechanism accounts for the hierarchical structure of the CE and CC analysis datasets, whereby individual pupils (level 1) are nested within schools (level 2). It therefore acknowledges that (a) pupils in the same school are likely to be (collectively) more similar than pupils in different schools; and (b) the relationship between the outcomes of interest (i.e. whether pupils enter computing at KS4 or whether they continue with computing at KS5) and the variables featuring in the CE and CC models may vary between different schools.

[4] Information about teachers' qualification is provided by the School Workforce Census (SWC). The SWC describes teachers' qualifications using the Joint Academic Council's code-set of principal subjects. This can be accessed at https://www.hesa.ac.uk/support/documentation/jacs. The list of computing qualifications comprises: computational science foundations; computer architectures; computer architectures & operating systems; computer science; computer vision; computing science not elsewhere classified; human-computer interaction; mathematical and computing sciences not elsewhere classified; multi-media computing science; neural computing; other computing sciences; other mathematical and computing Sciences.

To fit the substantive CE and CC models, each of the CE and CC analysis datasets was partitioned into two *randomly selected, non-overlapping* subsets:

- The *training CE and CC datasets*, which comprised 80% of records in the complete CE and CC analysis datasets, respectively; and
- The *testing CE and CC datasets*, which comprised the remainder 20% of records in the complete CE and CC analysis datasets, respectively.

The substantive CE and CC (multilevel binary logistic) models were initially fitted upon the training datasets and the modelling outputs were inspected. Then, the CE and CC models were fitted upon the testing datasets and these modelling outputs were compared against the outputs from models fitted upon the training datasets. The motivation for this comparison was to assess if the substantive CE and CC models generate comparable, non-contradictory insights when fitted upon different datasets, which represent the same pupil populations[5].

The comparisons confirmed that the substantive CE and CC models were “stable” when fitted on different datasets representing the same pupil population, suggesting that the proposed CE and CC model specifications should be expected to generalise well to the wider pupil populations they intend to describe (see paragraph 1.1.1).

Once the stability of the substantive CE and CC models was confirmed, the models were re-fitted to the complete CE and CC analysis datasets. A final inspection of the model outputs ensured that insights derived based on the complete CE and CC analysis datasets align to the insights derived based on the training and testing subsets. The final outputs from the CE and CC models are presented in paragraph 1.2.

1.2 The outputs of the CE and CC models

This paragraph presents the statistical outputs from the proposed substantive CE and CC models, which respectively aim to help us understand the determinants of entering computing education at KS4 and the determinants of continuing computing education at KS5 (for pupils who have entered it KS4). The CE and CC models are summarised in Tables 1.2.1 and 1.2.2, respectively, which present key statistics yielded by the modelling process (while a reflection on the model outputs and their implications is then provided in paragraph 1.3)[6].

Tables 1.2.1 and 1.2.2 present the following statistics:

[5] If a substantive model produces contradictory insights when applied to different datasets that represent the same population, concerns should be raised with regard to the generalisability of the model. In such cases, it is likely that the substantive model has been *over-fitted* to the training dataset. It is therefore necessary to revisit the model specification and re-think the mix of variables to include in the model.

[6] Using the McFadden approach, the analysis calculated pseudo R-squared metrics for the substantive CE and CC models (5.4% and 3.6%, respectively). Effectively, this provides a quantification of the outcome variability that is explained by the substantive CE and CC models. However, we note that the usefulness of pseudo R-squared metrics is open to debate amongst data users, with concerns being raised regarding the extent to which these are intuitively interpretable in relation to non-linear outcomes (such as the binary outcomes modelled in this study). For a brief review of pseudo R-squared metrics, see: Tabachnick, B. G; & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston: Pearson / Allyn & Bacon.

- **Odds ratios and coefficients**. These statistics quantify the relationship between a variable and the outcome of interest. They are mathematically equivalent, as the odds ratio is equal to Euler's e (c.2.718) in the power of the coefficient. Both metrics are presented in the tables, as some readers may find one statistic more intuitive than the other.

  Odds ratios greater than 1 (or coefficients greater than 0) suggest a positive relationship between the outcome and the variable. Odds ratios smaller than 1 (or coefficients smaller than 0) suggest a negative relationship between the outcome and the variable.

  Odds ratios (i.e. exponentiated coefficients) quantify the *change in the odds* of observing the outcome, given a change in the predictor variable by one unit (when we consider numeric variables, such as the total GCSE and equivalents score) or given a shift from a reference category to a different category (when we consider categorical variables, such as gender). The odds represent the ratio of the probability of the outcome occurring to the probability of the outcome not occurring.

  For example, if the odds ratio of variable X for the outcome Y is 1.5, we infer that an increase in X by one unit (if X is numeric) or a shift from the reference category to a different category (if X is categorical) means an increase in Y. We also infer than given this change in X, the odds of Y occurring are expected to increase by a factor of 1.5.

- **Standard errors of the coefficients**. These statistics help quantify the statistical uncertainty regarding the "true" magnitude of the coefficients. The uncertainty stems from the fact that the coefficients have been computed based on data from particular cross-sections of the pupil-population of interest; cross-sections, which can be deemed as representative of *wider* pupil populations of interest (see paragraph 1.1.1). The uncertainty here, therefore, reflects the fact that the analysis may have yielded different coefficients, if a different cross-section of the wider population of interest had been used.

- **The p-values**, quantify the probability of inferring that a certain relationship between a variable and the outcome occurs in the analysis datasets, if this *actually* does not occur (within the wider pupil population of interest that the analysis datasets represent). Where p-values are below the (conventionally accepted) threshold of 0.05, we infer that a certain relationship (between a variable and the outcome) has a negligibly small probability of being observed by chance if it were not real. The relationship is therefore deemed as statistically significant and can be seen as likely to generalise more widely.

- Finally, **the lower and upper bounds of a coefficient's 95% confidence interval** represent a range of plausible values that can quantify the strength of a relationship between a variable and the outcome of interest. If this range includes both positive and negative values, then there is uncertainty about the direction of the effect (i.e. whether a certain change in the predictor corresponds to an increase or a decrease in the outcome).

Table 1.2.1: CE model output (outcome: uptake of GCSE computing amongst pupils at schools where the subject is offered; specification: multilevel model; base: 248,145 pupil records from 1,296 school clusters)

Variable	Category label vs. reference category label (for categorical variables)	Odds Ratio	Coefficient	Standard error of coefficient	p-value	Coefficient lower bound of 95% Confidence Interval	Coefficient upper bound of 95% Confidence Interval
Gender	Male [vs. female]	8.847	2.180	0.019	0.000	2.143	2.218
Ethnic background	Mixed [vs. white including missing ethnic background]	0.995	-0.005	0.035	0.877	-0.074	0.063
	Black [vs. white including missing ethnic background]	0.852	-0.160	0.040	0.000	-0.238	-0.083
	Asian [vs. white including missing ethnic background]	1.375	0.319	0.031	0.000	0.257	0.380
	Chinese [vs. white including missing ethnic background]	1.547	0.437	0.084	0.000	0.272	0.601
	Other [vs. white including missing ethnic background]	1.110	0.104	0.060	0.085	-0.014	0.222
Special education needs (SEN)	SEN identified [vs. no SEN identified]	1.004	0.004	0.040	0.920	-0.075	0.083
First language	Other than English [vs. English including unspecified]	1.067	0.065	0.027	0.016	0.012	0.118
Total GCSE (and equivalents) score		1.003	0.003	0.000	0.000	0.002	0.003
Attainment in KS4 Maths		1.362	0.309	0.008	0.000	0.293	0.325
Quintiles of the number of KS4 pupils in school	Second quintile [vs. first quintile]	0.882	-0.126	0.066	0.056	-0.256	0.003
	Third quintile [vs. first quintile]	0.811	-0.210	0.071	0.003	-0.349	-0.070
	Fourth quintile [vs. first quintile]	0.674	-0.395	0.075	0.000	-0.542	-0.247
	Fifth quintile [vs. first quintile]	0.701	-0.355	0.082	0.000	-0.516	-0.195
Gender of school admissions	Single-sex school [vs. mixed school]	1.290	0.254	0.085	0.003	0.087	0.421
Quintiles of percentage of pupils who are white British in school	First quintile [vs. fifth quintile]	1.055	0.054	0.134	0.690	-0.210	0.317

	Second quintile [vs. fifth quintile]	0.947	-0.055	0.087	0.531	-0.226	0.116
	Third quintile [vs. fifth quintile]	1.067	0.064	0.079	0.413	-0.090	0.219
	Fourth quintile [vs. fifth quintile]	0.992	-0.008	0.076	0.912	-0.157	0.140
Percentage of pupils whose first language is other than English		0.999	-0.001	0.002	0.725	-0.006	0.004
Percentage of pupils recorded as eligible for free school meals		1.002	0.002	0.004	0.670	-0.006	0.010
Percentage of pupils achieving at least 5 GCSE's A star to C		0.305	-1.187	0.183	0.000	-1.547	-0.828
Percentage of pupils with special education needs (SEN) in school		0.256	-1.361	1.439	0.344	-4.182	1.460
Deciles of the Income Deprivation Affecting Children Index (IDACI) for pupils' school postcode		0.986	-0.014	0.009	0.124	-0.033	0.004
Region where pupil's school is	South East [vs. London]	1.287	0.252	0.112	0.025	0.032	0.472
	South West [vs. London]	1.161	0.149	0.123	0.224	-0.092	0.390
	East of England [vs. London]	1.040	0.039	0.117	0.738	-0.191	0.269
	East Midlands [vs. London]	1.327	0.283	0.122	0.021	0.043	0.522
	West Midlands [vs. London]	1.318	0.276	0.111	0.013	0.058	0.494
	Yorkshire and the Humber [vs. London]	1.233	0.210	0.114	0.065	-0.013	0.432
	North East [vs. London]	1.147	0.137	0.151	0.364	-0.159	0.434
	North West [vs. London]	1.239	0.214	0.109	0.049	0.001	0.427
Urbanisation level where pupil's school is	Urban - city or town [vs. rural]	0.973	-0.027	0.079	0.732	-0.181	0.127
	Urban - major	0.982	-0.018	0.097	0.853	-0.209	0.173

	conurbation [vs. rural]						
Number of teachers at pupil's school known to have computing qualification(s)		1.064	0.062	0.020	0.001	0.024	0.101
Constant		0.004	-5.611	0.161	0.000	-5.926	-5.296

Table 1.2.2: CC model output (outcome: uptake of A level computing amongst pupils who completed GCSE computing; specification: multilevel model; base: 14,679 pupil records from 1,201 school clusters)

Variable	Category label vs. reference category label (for categorical variables)	Odds Ratio	Coefficient	Standard error of coefficient	p-value	Coefficient lower bound of 95% Confidence Interval	Coefficient upper bound of 95% Confidence Interval
Gender	Male [vs. female]	2.897	1.064	0.111	0.000	0.847	1.280
Ethnic background	Mixed [vs. white including missing ethnic background]	0.903	-0.102	0.141	0.472	-0.378	0.175
	Black [vs. white including missing ethnic background]	0.590	-0.527	0.185	0.004	-0.890	-0.164
	Asian [vs. white including missing ethnic background]	0.486	-0.721	0.114	0.000	-0.944	-0.499
	Chinese [vs. white including missing ethnic background]	0.619	-0.480	0.275	0.081	-1.020	0.060
	Other [vs. white including missing ethnic background]	0.818	-0.201	0.235	0.392	-0.662	0.260
Total GCSE (and equivalents) score		0.996	-0.004	0.000	0.000	-0.005	-0.004
Attainment in KS4 Computing		2.692	0.990	0.036	0.000	0.919	1.061
Attainment in KS4 Maths		1.178	0.164	0.038	0.000	0.089	0.239
School type	Selective or independent school [vs. comprehensive, modern, other maintained school]	0.738	-0.304	0.168	0.071	-0.634	0.026
	Sixth-form college [vs. comprehensive, modern, other maintained school]	1.216	0.195	0.166	0.239	-0.130	0.520

	Other Further Education (FE) college [vs. comprehensive, modern, other maintained school]	0.307	-1.181	0.199	0.000	-1.571	-0.792
Deciles of the Income Deprivation Affecting Children Index (IDACI) for pupils' school postcode		1.041	0.040	0.017	0.017	0.007	0.073
Region where pupil's school is	South East [vs. London]	1.149	0.139	0.225	0.539	-0.303	0.581
	South West [vs. London]	1.110	0.104	0.250	0.677	-0.386	0.594
	East of England [vs. London]	0.988	-0.012	0.243	0.960	-0.489	0.465
	East Midlands [vs. London]	0.947	-0.055	0.254	0.829	-0.553	0.444
	West Midlands [vs. London]	0.998	-0.002	0.229	0.995	-0.451	0.448
	Yorkshire and the Humber [vs. London]	0.485	-0.723	0.255	0.005	-1.223	-0.223
	North East [vs. London]	0.760	-0.274	0.299	0.359	-0.860	0.312
	North West [vs. London]	1.106	0.101	0.217	0.643	-0.325	0.526
Urbanisation level where pupil's school is	Urban - city or town [vs. rural]	0.979	-0.021	0.147	0.886	-0.309	0.267
	Urban - major conurbation [vs. rural]	0.951	-0.051	0.187	0.786	-0.416	0.315
Constant		0.000	-10.180	0.403	0.000	-10.970	-9.391

1.3 Reflection on the outcomes of the CE and CC models

Here, we discuss key insights derived based on the CE and CC models, using contextual information where helpful. For a description of the statistical concept of *odds* (which is frequently quoted in this paragraph), see paragraph 1.2.

1.3.1 Pupil-level characteristics

Gender

A pupil’s gender is very strongly associated with both uptake of computing at KS4 (see CE model) and continuation of the subject at KS5 (see CC model). After controlling for other factors, male pupils have almost nine times the odds of female pupils of studying GCSE computing (CE model). This is a far stronger effect than seen for any other variable in the CE model. Male pupils also have almost three times the odds of female pupils of continuing with computing at KS5 (CC model). This is a broadly similar strength effect to a pupil's GCSE computing grade; in other words, the effect of gender on continuation at KS5 appears to be roughly equivalent to achieving an extra grade in computing at KS4 after controlling for other factors.

The two models underline the very heavy influence of gender in computing education, especially concerning uptake at KS4. Taken together, the two models show that not only are male pupils much more likely than female pupils to study GCSE computing, this gender gap then worsens for A level computing, even after controlling for other factors.

Ethnicity

After controlling for other factors, Asian and Chinese pupils were significantly more likely than white pupils to study GCSE computing, while Black pupils were significantly less likely than white pupils to take the subject (CE model). Black pupils were also less likely than white pupils to continue with computing at Key Stage 5 (CC model).

Despite having higher levels of uptake at KS4, Asian pupils had lower levels of continuation to KS5 computing than white pupils, after controlling for other factors (CC model). Previous research has shown that Asian pupils are on average more likely to take A levels than white pupils[7]. In particular, Asian pupils are more likely to study subjects such as chemistry, biology and mathematics at A level (see Table 1.3.1.1). This in turn is related to the fact that Asian pupils are more likely to go on to study subjects such as medicine at university (see Table 1.3.1.2). The lower continuation levels in computing for Asian pupils may therefore be partly due to pupils positively making decisions about their future plans and career routes which they feel do not require further computing qualifications.

Nonetheless, within the broader question of higher education, there is a concern that computing is not sufficiently valued as a subject for making university applications – for example, it is not considered a facilitating subject by Russell Group universities. Persuading pupils, schools/colleges and higher education institutions of the value of computing for making higher education applications – regardless of the subject being applied for – could help encourage pupils from all backgrounds to choose to continue with the subject.

Table 1.3.1.1: A level subjects taken by major ethnic group (base: KS5 pupils in year 12 and year 13 in 2014-15)

	White	Mixed	Black	Asian	Chinese	Other
Maths	79,562 12.7%	5,077 15.1%	5,428 12.0%	18,224 21.7%	2,413 49.8%	2,153 19.6%
Biology	57,822 9.2%	3,752 11.1%	4,573 10.1%	15,049 17.9%	947 19.6%	1,648 15.0%
Chemistry	42,762 6.8%	3,112 9.2%	4,396 9.7%	14,716 17.5%	1,205 24.9%	1,608 14.6%
Physics	35,189 5.6%	2,036 6.0%	1,555 3.4%	5,138 6.1%	860 17.8%	671 6.1%
Computing	7,082	357	288	869	135	99

[7] Social and ethnic inequalities in choice available and choices made at age 16 (*Allen et al. 2016*)

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/574708/SMC_social_and_ethnic_inequalities_in_post_16_report.pdf

	1.1%	1.1%	0.6%	1.0%	2.8%	0.9%
Total	626,639	33,693	45,260	84,102	4,842	11,010

Table 1.3.1.2: Full-time HE student enrolments by ethnicity 2015/16[8] (base: full-person equivalent of UK-domiciled HE student enrolments in 2015-16)

Course involves…	White	Black	Asian	Other	Unknown ethnicity
Business and administrative studies	91,680 10.0%	15,280 17.9%	24,005 17.6%	8,310 11.5%	1,285 10.0%
Medicine / Dentistry	25,790 2.8%	1,275 1.5%	9,815 7.2%	2,675 3.7%	475 3.7%
Subjects allied to medicine	100,390 11.0%	14,855 17.4%	18,585 13.6%	5,990 8.3%	1,015 7.9%
Computer sciences	39,705 4.3%	4,195 4.9%	9,310 6.8%	2,890 4.0%	545 4.2%
Creative arts and design	104,030 11.4%	5,625 6.6%	5,335 3.9%	7,140 9.8%	940 7.3%
Total	915,030	85,275	136,585	65,410	9,755

Attainment

Although overall KS4 attainment[9] had a significant positive association with uptake of GCSE computing, the effect of this is very weak after controlling for other factors (CE model). Instead, pupils' GCSE mathematics grades were more strongly associated with uptake of GCSE computing: an additional grade at GCSE mathematics was associated with an increase of 1.4 times the odds of studying GCSE computing.

A pupil's GCSE mathematics grade was also positively associated with continuation of computing study at KS5 (CC model). In this case, an additional grade was associated with an increase of 1.2 times the odds of continuing with computing.

There thus appears to be a clear relationship with more mathematically able pupils being more likely to study GCSE computing and more likely to continue with the subject at KS5. In the case of continuation to KS5, the relationship

[8] HESA statistical first release SFR242, Jan. 2017
https://www.hesa.ac.uk/news/12-01-2017/sfr242-student-enrolments-and-qualifications

[9] We conceptualise overall KS4 attainment and GCSE mathematics grade as proxies for pupils' general academic ability and mathematical ability at age 14 when pupils choose their GCSE courses.

between studying computing and mathematics attainment holds even after controlling for a pupil's computing attainment at GCSE[10].

After controlling for other factors, there was a small significant negative association between overall attainment at KS4 and continuation at KS5; that is, pupils with higher attainment in their GCSEs were less likely to continue with computing at KS5 (CC model). It should be stressed that, although statistically significant, the effect of this is very small. Given that pupils' attainment in computing and maths are controlled for separately, this may simply be reflective of more able pupils prioritising subjects other than computing which they consider more relevant for their own future study and career plans.

1.3.2 School/college level characteristics

Given the differences in school/college characteristics between the two models, we discuss first, school level characteristics associated with uptake of GCSE computing; second, school/college level characteristics associated with continuation of computing study at KS5. Finally, we discuss regional variation in in both uptake and continuation.

Uptake of GCSE computing: Attainment

After controlling for other factors, school level attainment was negatively associated with uptake of computing at KS4. It is important to remember that individual attainment is also controlled for within the model. In other words, if there were two equally able pupils, identical in every regard apart from the school they attended, the pupil at a school with lower general levels of attainment would be more likely to study computing at KS4.

One possible explanation is that higher performing schools may be encouraging their pupils to prioritise other subjects that may be considered more useful for continuing on to higher education. If this is the case, there is an important challenge to persuade pupils, schools and higher education institutions of the value of the computing GCSE.

Uptake of GCSE computing: Size of school

Size of school was negatively associated with uptake of computing; after controlling for other factors, pupils in smaller schools were more likely to study GCSE computing than pupils in larger schools.

Again, it is important to remember that the model is concerned with uptake within schools where at least one pupil took GCSE computing. Considering all schools, we note that the smallest schools were less likely to enter any pupils for computing GCSE: 1.7% of schools with a GCSE cohort size of 1-11 pupils and 10.5% of schools with a cohort size of 12-89 pupils offered GCSE computing, compared with 51.9% of schools with a cohort size of more than 200 pupils[11]. There thus remains a challenge to support the smallest schools in offering GCSE computing.

Uptake of GCSE computing: Gender mix

There was a positive association between uptake of GCSE computing and attending a single sex school; pupils at single sex schools had 1.3 times the odds of pupils at mixed schools of studying GCSE computing.

---

[10] A pupil's GCSE computing grade was strongly associated with likelihood to continue with the subject at A level, each grade at GCSE being associated with an increase of 2.7 times the odds of studying computing A level. This relationship is to be expected as pupils will generally prioritise continuing with their strongest subjects at A level.

[11] The Roehampton annual computing education report: 2015 data from England (Kemp, Wong and Berry 2016) https://www.researchgate.net/publication/311595274_The_Roehampton_Annual_Computing_Education_Report_2015_data_from_England

Looking at pupils' gender within mixed and single sex schools (Table 1.3.2.1), there appears to be a particularly strong influence of a single sex environment on female pupils: uptake of GCSE computing among female pupils was 12% at single sex schools where at least one pupil took the subject, compared with only 3% at mixed schools. Girls at single sex schools will have quite a different experience of computing to girls at mixed schools, for whom the vast majority of their classmates are likely to be male. This kind of difference in learning environment could be helping to reduce the substantial gender-related barriers to uptake for female pupils.

Table 1.3.2.1: Uptake of GCSE computing within schools where at least one pupil completed GCSE computing (base: KS4 pupils in year 11 at schools where at least one pupil completed GCSE computing in 2014-15)

	Male pupils		Female pupils	
	Boys schools	Mixed schools	Girls schools	Mixed schools
Uptake of GCSE computing	2,735 21.5%	23,307 19.8%	1,259 12.3%	3,774 3.4%
Total	12,740	117,898	10,274	109,626

Uptake of GCSE computing: Teachers with a computing qualification

The number of teachers with a computing qualification was positively associated with uptake of GCSE computing. This finding is particularly important in light of the difficulties in recruiting computing teachers. Each additional teacher with a computing qualification was associated with an increase of 1.1 times the odds of studying GCSE computing.

Uptake of GCSE computing: Deprivation and SEN

It is notable that measures of deprivation – IDACI rank, eligibility for free school meals, percentage of pupils in the school eligible for free school meals – do not appear to have a significant association with uptake of GCSE computing, after controlling for other factors. In addition, a statistically significant association between uptake and SEN status was not detected either at the pupil level (whether the individual has an identified SEN or learning disability) or at the school level (the percentage of pupils in the school with an identified SEN or learning disability).

It appears then the main barrier to access for these groups may be schools failing to offer computing in the first place. Within schools where at least one pupil studies GCSE computing these factors do not appear to have a significant impact on uptake, after controlling for other pupil-level and school-level characteristics.

Continuation of pupils to A level computing: Institution type and deprivation

Aside from region (discussed below), two school/college level variables had a significant association with continuation of computing at KS5. First, institution type was significantly associated, with pupils who completed GCSE computing and then studying at Further Education colleges much less likely to continue to KS5. This reflects the fact that pupils at FE colleges are less likely to study A levels and more likely to take other kinds of courses. This variable is therefore controlling for the different educational paths pupils take after KS4.

Second, deprivation as measured by IDACI was also significantly associated with continuation; pupils in less deprived areas were more likely to continue with computing at KS5.

Regional variation

There is some regional disparity in uptake of computing with pupils in the South East, East Midlands, West Midlands and North West more likely to study GCSE computing than pupils in London, after controlling for other factors.

In terms of continuation of computing study at KS5, geographic region generally has little impact after controlling for other factors, although continuation is notably lower in Yorkshire and the Humber.

Appendix 1 – The complete list of variables considered by the analysis conducted at Strand 3

Table A1 presents the complete list of variables that the CE and CC modelling process considered. Some of the variables presented in Table A1 were not used in the final substantive CE and CC models; they were eliminated based on the bivariate and multivariate variable-selection processes described in paragraph 1.1.2.

Table A1: Complete list of variables considered for the computing entry (CE) and the computing continuation (CC) models

Complete list of variables considered for the CE model	Complete list of variables considered for the CC model
Attainment at KS4 Maths	Attainment at KS4 Computing
Eligibility for free school meals	Attainment at KS4 Maths
Ethnic background	Eligibility for free school meals
Language	Ethnic background
Number of GCSE entries	Language
Special education needs (SEN)	Special education needs (SEN)
Gender	Gender
Total GCSE (and equivalents) point score	Total GCSE (and equivalents) point score
Gender of school admissions	Deciles of the Income Deprivation Affecting Children Index (IDACI) for pupils' school postcode
Deciles of the Income Deprivation Affecting Children Index (IDACI) for pupils' school postcode	Gender of school admissions
Number of KS4 pupils in school	Number of KS5 pupils in school
Number of teachers in school known to have computing qualification	Percentage of KS4 pupils with Special Educational Needs (SEN) with a statement or Education, health and care (EHC) plan (at pupil's school)
Percentage of KS4 pupils with Special Educational Needs (SEN) with a statement or Education, health and care (EHC) plan (at pupil's school)	Percentage of pupils achieving at least 5 GCSEs at A*-C including English and Maths (at pupil's school)
Percentage of pupils achieving at least 5 GCSEs at A*-C including English and Maths (at pupil's school)	Percentage of pupils recorded as eligible for free school meals (at pupil's school)
Percentage of pupils recorded as eligible for free school meals (at pupil's school)	Percentage of pupils who are White British (at pupil's school)
Percentage of pupils who are White British (at pupil's school)	Percentage of pupils whose language group is 'other than English' (at pupil's school)
Percentage of pupils whose language group is 'other than English' (at pupil's school)	Region where pupil's school is
Region where pupil's school is	Urbanisation level where pupil's school is

Urbanisation level where pupil's school is	**School Type**
School type	
Whether pupil's school is a selective school	