**SPS Student Cover Sheet**

**\*Indicated fields MUST be completed by student.**

| *Exam number | B149008 |
|---|---|
| *Course Name | Multi-Level Modelling in Social Science |
| *Essay/title of work | Multi-Level Modelling in Social Science – Practical Project |
| *Word count | 2982 |

------------------------------------------------------------------------------------------------

**This Section is for office use.**

Staff should add a written bubble /comment to the relevant box to input any of the below details.

| Word count penalties | |
|---|---|
| Lateness penalties | |
| Markers name | |

# Multi-Level Modelling in Social Science - Practical Project

B149008

April 2024

# Contents

# 1 Introduction

Student performance and level of attainment are heavily influenced by numerous factors. These include factors in their personal lives, societal factors, or factors relating to their experiences in school [5]. This study aims to investigate the impact of several factors on Scottish school leavers' level of academic attainment upon leaving secondary education, using survey data about various aspects of their personal lives and experiences at secondary school.

Section 2 outlines the features of the survey data and how they were reformatted and processed before modelling. Section 3 describes the different models fitted to the data and provides a discussion about their respective results.

# 2 Data Pre-processing

The data in this study was obtained from the 1999 Scottish School Leavers Survey (SSLS) [12]. There was data for 3,689 pupils from 407 secondary schools across Scotland.

## 2.1 Demographic Variables

Table 1 outlines the demographic information recorded for each pupil, including 4 different variables and their respective levels/categories.

| Variable | Variable Description | Value | Label |
|:---:|:---|:---:|:---:|
| sexmf | Sex of participant | 0 | Female |
| | | 1 | Male |
| parsc | Parents' social status | 1 | Professional |
| | | 2 | Intermediate |
| | | 3 | Skilled non-manual |
| | | 4 | Skilled manual |
| | | 5 | Partly skilled |
| | | 6 | Unskilled |
| | | 7 | Unclassified |
| parleft | Age at which parents left full-time education | 1 | Both to 17 years or more |
| | | 2 | One to 17 years or more |
| | | 3 | One or both to 16 years |
| | | 4 | Both to 15 years or less |
| hqual (response) | Level of education attained by participant | 1 | Low |
| | | 2 | Standard grade only |
| | | 3 | 1-4 highers |
| | | 4 | 5+ highers |

Table 1: Values and their respective labels for demographic variables in the SSLS dataset.

Since `hqual` was the response variable and categorical with 4 levels, it needed to be recoded as a binary variable to implement logistic regression. Since many British universities typically require 4 or more Scottish Highers for admission, 5 Highers served as a natural divider for levels of academic attainment [20]. The levels of the recoded variable, `hqual_enc`, are outlined in Table 2. Additionally, the variable referring to the pupils' parents' social class, `parsc`, had 7 categories. To reduce computational complexity, this was also recoded to have fewer categories. The levels of the recoded variable, `parsc_enc`, are outlined in Table 3. Summary statistics for the recoded data are shown in Table 4.

Although there was an uneven split in the counts for the levels in each variable, it was not so drastic that logistic regression should fail. A minority of pupils had parents whose social class was *unskilled* or *unclassified*. Additionally, most pupils' parents had left full-time education at least after 16 years, with

a minority having 2 parents who left full-time education before the age of 16. Finally, 1,383 students of 3,689 ($\approx 37\%$) obtained less than 5 or more Scottish Highers by the time they left secondary education.

| Value | Previous Category | New Category |
|:---:|:---:|:---:|
| 1 | Low | |
| 2 | Standard grade only | Up to 4 Highers |
| 3 | 1-4 highers | |
| 4 | 5+ highers | 5+ Highers |

Table 2: The `hqual` variable recoded as a binary variable.

| Value | Previous Category | New Category |
|:---:|:---:|:---:|
| 1 | Professional | Professional |
| 2 | Intermediate | |
| 3 | Skilled non-manual | |
| 4 | Skilled manual | Skilled |
| 5 | Partly skilled | |
| 6 | Unskilled | Unskilled |
| 7 | Unclassified | Unclassified |

Table 3: The `parsc` variable recoded to have fewer categories.

| Variable | Category | Count |
|:---:|:---:|:---:|
| `sexmf` | Female | 2,171 |
| | Male | 1,518 |
| `parsc_enc` | Professional | 1,476 |
| | Skilled | 1,675 |
| | Unskilled | 92 |
| | Unclassified | 446 |
| `parleft` | Both to 17 years or more | 800 |
| | One to 17 years or more | 1,043 |
| | One or both to 16 years | 1,349 |
| | Both to 15 years or less | 497 |
| `hqual_enc` | Up to 4 highers | 1,383 |
| | 5+ Highers | 2,306 |

Table 4: Summary statistics for the variables in the SSLS dataset after recoding.

### 2.1.1  Preliminary Hypotheses

Students' level of academic attainment could be influenced and impacted by various factors. However, regarding the variables included in this study, it was expected that women would outperform men and have higher academic attainment levels in general [13]. In addition, higher levels of parental occupational prestige and higher parental academic attainment were expected to result in higher student academic attainment as well [7, 8].

## 2.2  Survey Question Responses

In addition to the demographic information, the pupils' responses to 2 questions in the survey, specifically questions 1 and 2, were recorded. Question 1 (Q1) had 12 parts, from Q1a to Q1l, and question 2 (Q2) had 9 parts, from Q2a to Q2i, and students were asked whether they agreed or disagreed with the statements in each question. The individual parts of both questions are shown in Table 5.

To model effectively, it was essential to avoid including all parts of both questions in the models as many of the questions asked similar things. Given the similarity of such questions, there was a risk of multicollinearity in the models, leading to inaccurate inference. Hence, for both Q1 and Q2, specific indicative questions were selected for modelling.

| Question no. | Question in the survey |
|---|---|
| Q1a | School has helped to give me confidence to make decisions. |
| Q1b | School has done very little to prepare me for life when I leave school. |
| Q1c | School has taught me things which would be useful in a job. |
| Q1d | My school had a wide range of after-school activities. |
| Q1e | My school dealt with any bullying that went on. |
| Q1f | There was vandalism at my school during school day. |
| Q1g | If I had a problem there was always a teacher I could talk to. |
| Q1h | Theft among pupils was common at my school. |
| Q1i | My school dealt with any harassment that went on. |
| Q1j | Pupils respected the teachers. |
| Q1k | Pupils sometimes got bullied. |
| Q1l | Pupils sometimes got harassed. |
| Q2a | School was generally worth doing. |
| Q2b | There were too many troublemakers in my S4 classes. |
| Q2c | My teachers helped me do my best. |
| Q2d | Many teachers could not keep order in class. |
| Q2e | My friends took school seriously in S4. |
| Q2f | Pupils who were punished usually deserved it. |
| Q2g | Teachers listened to my ideas and views. |
| Q2h | Teachers often gave me homework in S4. |
| Q2i | Teachers made sure I did homework they set. |

Table 5: The question variables included in the data. Students responded with either *Agree* or *Disagree*.

### 2.2.1 Parts Selected From Q1

Q1a, Q1b, and Q1c were concerned with students' self-perceived preparedness to leave school and enter the next stage of their lives, especially where they may need to live more autonomously. Self-regulated or autonomous learning often positively impacts students' academic performance [10]. Furthermore, students capable of independent decision-making often feel more confident when leaving school to begin higher education or enter the world of work [2]. Hence, these 3 questions were important to consider. However, since the responses to these responses were correlated with each other (see Figure 1), it was important to select 1 question to represent all 3. Hence, any models constructed in this study considering the responses to Q1 only included responses to Q1a while disregarding Q1b and Q1c.

Participating in extracurricular activities (EA) can often positively impact students' academic achievements [1, 9]. Since Q1d was the only question to consider EA and largely uncorrelated with other questions, it was included in subsequent models to provide insight into how the opportunity to partake in EA impacts students' academic performance.

Q1e, Q1i, Q1k, and Q1l indicated how often students encountered or witnessed bullying or harassment. Bullying and harassment in schools can have a negative impact on overall attainment levels [19]. Additionally, since the prevalence of bullying and harassment were highly correlated (see Figure 1), they

could be grouped. Since Q1e and Q1k were negatively correlated, as were Q1i and Q1l, only 1 of these 4 questions, specifically Q1k, was selected to represent the impact of bullying and harassment on academic attainment in subsequent models.

Similarly, the existence of crimes such as theft or vandalism could also negatively impact students' academic attainment overall [16]. The two questions about theft and vandalism in schools, Q1f and Q1h, were also positively correlated (see Figure 1). Hence, Q1f was selected as the representative question.

Finally, Q1g and Q1j described the student-teacher relationships at each school. Positive student-teacher relationships could result in higher attainment overall [4]. Additionally, although Q1g and Q1j are positively correlated in Figure 1, Q1g indicated the approachability and supportiveness of teachers, whereas Q1j indicated whether teachers were respected and obeyed by students. Hence, Q1g and Q1j were both included in subsequent models.
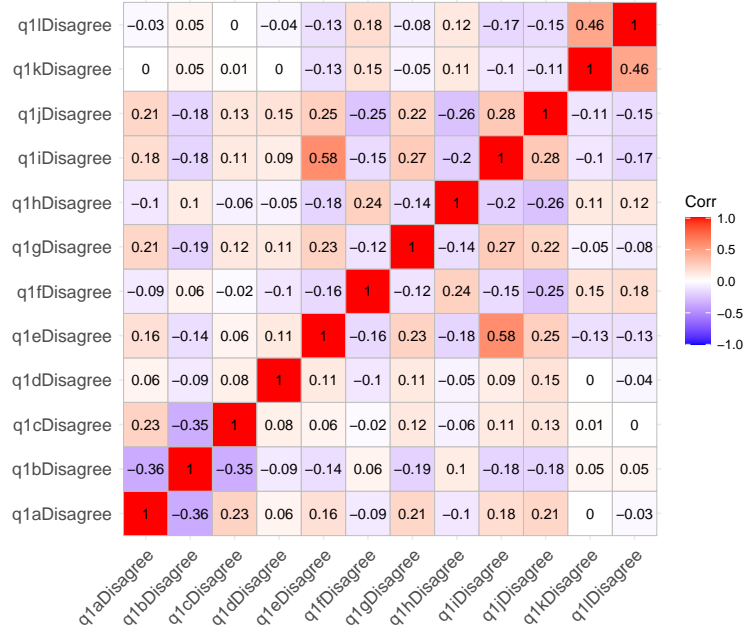


Figure 1: Correlation plot showing relationships between responses to Q1 of the survey.

### 2.2.2 Parts Selected From Q2

As with Q1, Q2 was broken down into what individual parts of it indicated. Q2a referred to students' motivation at school, which is often found to positively impact students' academic performance [18]. Hence, Q2a was included in subsequent models.

Classroom disruption often negatively impacts students' learning experience [16]. Q2b and Q2d were indicative of students encountering disruption in classrooms. As shown in Figure 2, the responses to these questions were also positively correlated. Hence, Q2d was selected as the representative question for subsequent models.

Following the positive impact of positive student-teacher relationships described in Section 2.2.1, it was necessary to consider the responses for Q2c or Q2g, as these indicated the supportiveness of teachers. Since these 2 questions were positively correlated, Q2g was selected as the representative question in this case.

Similarly, it was important to consider students' relationships with their peers. Students are often positively influenced by their peers being motivated as well [3]. Q2e indicated the level of motivation among peers. Therefore, it was included in subsequent models.

Given that the prevalence of behavioural issues in general negatively impacts academic attainment, it was important to consider the role of discipline in schools as well [17]. Q2f represented the level of discipline maintained in schools, so it was included in subsequent models.

Finally, homework is typically found to improve academic attainment [6]. Since Q2h and Q2i indicated the importance placed on homework in schools and were positively correlated (see Figure 2), Q2i was included in subsequent models.
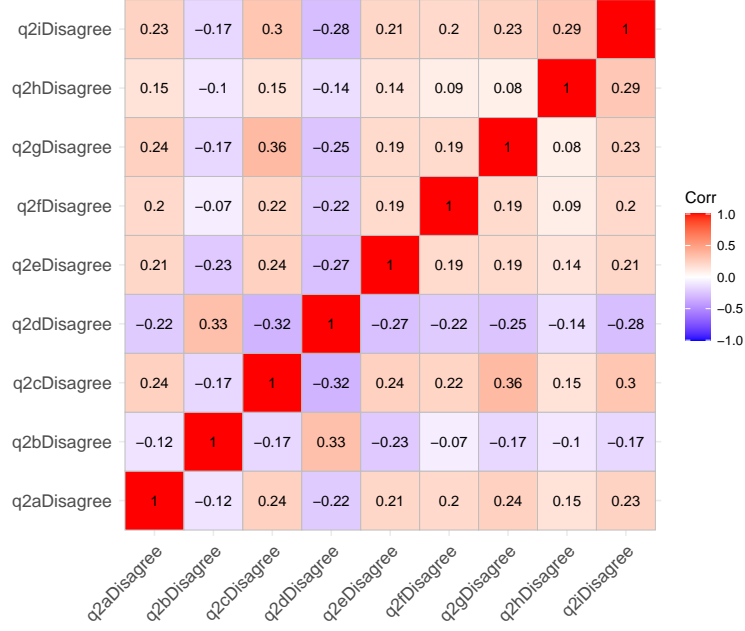


Figure 2: Correlation plot showing relationships between responses to Q2 of the survey.

# 3 Models and Results

To effectively investigate the factors impacting students' academic attainment, various multi-level models were fitted to the dataset. A variance component model (VCM) using logistic regression, VCM1, was initially fitted and compared with an equivalent variance component linear model, VCM2. The summary of fixed effects in VCM1 and VCM2 are shown in Tables 6 and 7, respectively.

| Variable | Estimate | Std. Error | $z$-value | $p$-value |
|---|---|---|---|---|
| Intercept | 0.58611 | 0.04967 | 11.81 | $< 2 \times 10^{-16}$ |

Table 6: Summary of fixed terms in VCM1.

| Variable | Estimate | Std. Error | $t$-value |
|---|---|---|---|
| Intercept | 0.63266 | 0.01071 | 59.08 |

Table 7: Summary of fixed terms in VCM2.

The fixed terms for both models are statistically significant and provide similar results. The intercept for VCM1 represents the log-odds (logarithm of the likelihood) of a student having less than 5 Scottish Highers [11]. Therefore, the probability of students having less than 5 Highers when no independent variables are included is

$$p = \frac{0.58661}{1 + 0.58661} = 0.6426.$$

VCM2 considers the response variable to be numeric, where 0 represents 5+ Highers and 1 represents less than 5 Highers. Hence, since the intercept 0.63266 is larger than 0.5, students are more likely to have

less than 5 Highers in the absence of independent variables. It is worth noting that these values for the intercepts in VCM1 and VCM2 are similar to the proportion of students in the dataset with less than 5 Highers ($2306/3689 = 0.63$).

The summary of the random effects in VCM1 and VCM2 are shown in Tables 8 and 9, respectively. The variance at the lowest level in multi-level logistic regression models such as VCM1 is substituted by $\pi^2/3 \approx 3.29$ [15]. Further, since the variation at the school level for VCM1 is given by 0.3722, the proportion of the variance in the response variable at the school level is

$$\% \text{ variance} = \frac{\text{variance at level 2}}{\text{variance at level 2 } + \text{variance at level 1}} \times 100 = \frac{0.3722}{0.3722 + \pi^2/3} \times 100 = \underline{10.16\%}.$$

Similarly, given the values for VCM2 in Table 9, the proportion of the variance in the response variable at the school level is

$$\% \text{ variance} = \frac{\text{variance at level 2}}{\text{variance at level 2 } + \text{variance at level 1}} \times 100 = \frac{0.0187}{0.0187 + 0.2148} \times 100 = \underline{8.08\%}.$$

Here, there is a slightly higher variance at the school level in VCM1 than in VCM2. This study implicitly aims to explore whether students' schools impact their academic attainment. Hence, since VCM1 suggests that there is a higher level of variance at the school level, it may be beneficial to use logistic regression as the differences between schools may be more pronounced.

| Groups | Name | Variance | Std. Dev. |
|--------|------|----------|-----------|
| School | (Intercept) | 0.3722 | 0.6101 |

Table 8: Summary of random terms in VCM1.

| Groups | Name | Variance | Std. Dev. |
|--------|------|----------|-----------|
| School | (Intercept) | 0.0187 | 0.1367 |
| Residuals | | 0.2148 | 0.4634 |

Table 9: Summary of random effects in VCM2.

Furthermore, it is worth comparing VCM1 with a single-level model (SLM) to investigate the impact of incorporating a level to account for variation between schools. The log-likelihoods for VCM1 and SLM are $-2393.47$ (df $= 2$) and $-2440.32$ (df $= 1$), respectively. Then, using the likelihood-ratio test, it is evident that the improvement of fit by including the variation at the school level is statistically significant, i.e., VCM1 provides a better fit than SLM [14]. Hence, using a multi-level model is most appropriate for this study. Sections 3.1 and 3.2 explore the effects of including the demographic variables and question responses as explanatory variables in multi-level models, respectively.

## 3.1 Impact of Responses to Demographic Questions

### 3.1.1 Random Intercept Model with Multiple Variables

To investigate the impact that the demographic variables, namely `sexmf`, `parsc_enc`, and `parleft` had on attainment, a multi-level model including all 3 as explanatory variables and a random intercept at the school level was fitted (hereby referred to as DVM). The summary statistics for its random and fixed terms are shown in Tables 10 and 11.

Compared to VCM1, there is less variation at the school level for DVM. This could be attributed to the addition of the demographic variables, as the added information may allow for a more accurate fit overall. Furthermore, since the estimates in Table 11 represent the log-odds of a student having less than 5 Highers, estimates less than 0 represent odds less than 0.5 and estimates greater than 0 represent odds greater than 0.5.

| Groups | Name | Variance | Std. Dev. |
|--------|------|----------|-----------|
| School | (Intercept) | 0.1739 | 0.417 |

Table 10: Summary of random terms in DVM.

| Variable (category) | Estimate | Std. Error | $z$-value | $p$-value |
|---------------------|----------|------------|-----------|-----------|
| Intercept | -0.96489 | 0.09497 | -10.160 | $< 2 \times 10^{-16}$ |
| sexmf (1) | 0.13176 | 0.07881 | 1.672 | 0.094537 |
| parleft (2) | 0.72971 | 0.10468 | 6.971 | $3.15 \times 10^{-12}$ |
| parleft (3) | 1.53316 | 0.11019 | 13.914 | $< 2 \times 10^{-16}$ |
| parleft (4) | 1.67529 | 0.14715 | 11.385 | $< 2 \times 10^{-16}$ |
| parsc_enc (Skilled) | 0.83797 | 0.08738 | 9.590 | $< 2 \times 10^{-16}$ |
| parsc_enc (Unclassified) | 1.09139 | 0.13566 | 8.045 | $8.62 \times 10^{-16}$ |
| parsc_enc (Unskilled) | 0.99400 | 0.27416 | 3.626 | 0.000288 |

Table 11: Summary of fixed terms in DVM.

Compared to VCM1, the estimate for the intercept for DVM is considerably lower at -0.96489, corresponding to odds of 0.276. Hence, DVM suggests that in the absence of all independent variables, the students are more likely to have more than 5 Highers. The remainder of the estimates for the demographic variables are positive. Hence, for each variable, they represent the increase in log-odds in comparison to the reference category for that variable. The reference category for sexmf was *male*. In contrast to the predictions in Section 2.1.1, females in this case were found to be less likely to obtain more than 5 Highers. However, it is worth noting that this was the only estimate in the model that was not statistically significant, so it may be inaccurate.

In an analogous fashion, estimates for parleft used the first level of the variable (both to 17 years or more) as the reference category. Hence, it is evident that students with parents who left full-time education before 17 were more likely to leave secondary education with less than 5 Highers. In addition to this, the higher the level of parleft, i.e., the earlier the parents left full-time education, the higher the increase in the log-odds of students obtaining less than 5 Highers. Since it is reasonable to expect that leaving full-time education earlier would be associated with lower levels of academic attainment, referring back to Section 2.1.1, these results were in line with the literature and prior predictions.

Finally, the estimates for parsc_enc used *professional* as the reference category. As expected, the lower levels of parental occupational prestige, i.e., *skilled* and *unskilled* increased the odds of students obtaining less than 5 Highers by the time they left secondary education. This was also in agreement with the relevant literature and previous predictions described in Section 2.1.1. Further, it is worth noting that the *unclassified* category of parsc_enc was associated with the largest increase in log-odds of obtaining less than 5 Highers. There was no information provided about the exact definition of this category and what it implied about the parents' social status. Hence, it was difficult to confidently infer conclusions from its estimate in the model.

### 3.1.2 Random Slope Model

To investigate the impact of students' sex on academic attainment a random slope model, RSM, was also fitted with random slopes for sexmf. Due to computational complexity and the lack of statistical significance for sexmf in DVM, RSM only included parleft and parsc_enc as fixed effects. Tables 12 and 13 summarise the random and fixed effects in RSM, respectively.

Variation at the school level in RSM is considerably larger than in DVM. Additionally, the variation between sexes at each school in RSM is very similar in value to the variation between schools. However, the estimates for the fixed effects are relatively unchanged going from DVM to RSM, with all differences being of order $10^{-2}$ or smaller. Moreover, the estimates all remained statistically significant, suggesting that while the variation in the random intercepts may increase, including the random slopes for sex has little impact on the fixed effects in the model.

| Groups | Name | Variance | Std. Dev. | Corr |
|--------|------|----------|-----------|------|
| School | (Intercept) | 0.2576 | 0.5075 | |
| | `sexmf` (1) | 0.2690 | 0.5187 | -0.62 |

Table 12: Summary of random terms in RSM.

| Variable (category) | Estimate | Std. Error | $z$-value | $p$-value |
|---------------------|----------|------------|-----------|-----------|
| Intercept | -0.91789 | 0.08874 | -10.343 | $< 2 \times 10^{-16}$ |
| `parleft` (2) | 0.74714 | 0.10682 | 6.995 | $2.66 \times 10^{-12}$ |
| `parleft` (3) | 1.55681 | 0.11290 | 13.789 | $< 2 \times 10^{-16}$ |
| `parleft` (4) | 1.69246 | 0.14963 | 11.311 | $< 2 \times 10^{-16}$ |
| `parsc_enc` (Skilled) | 0.84019 | 0.08859 | 9.484 | $< 2 \times 10^{-16}$ |
| `parsc_enc` (Unclassified) | 1.08795 | 0.13707 | 7.937 | $2.07 \times 10^{-15}$ |
| `parsc_enc` (Unskilled) | 1.01324 | 0.27655 | 3.664 | 0.000248 |

Table 13: Summary of fixed terms in RSM.

### 3.1.3   Comparing the Models

One method of verifying which one of DVM or RSM provides the most accurate fit for the data is through a likelihood-ratio test. The log-likelihoods for DVM and RSM are $-2111.052$ (df $= 9$) and $-2111.039$ (df $= 10$), respectively. Then, the likelihood-ratio test with these values suggests that the improvement from DVM to RSM is statistically significant, i.e., RSM provides a better fit for the data.

## 3.2   Impact of Responses to Q1 and Q2

To investigate the impact of the responses to Q1 and Q2, two final models were fitted to include responses to the questions as explanatory variables while controlling for students' sex (`sexmf`) and parental social class (`parsc_enc`).

### 3.2.1   Model for Q1

Q1M was fitted to include responses to the parts of Q1 outlined in Section 2.2.1. A summary of the question variables and what they indicated is shown in Table 14, and the summaries of the various effects in Q1M are shown in Tables 15 and 16.

| Variable label | What it indicates |
|----------------|-------------------|
| Q1a | Whether the school promoted decision-making skills and self-regulated learning. |
| Q1d | Opportunities to partake in extracurricular activities. |
| Q1f | Prevalence of crimes such as vandalism and theft. |
| Q1g | Whether teachers were approachable and supportive. |
| Q1j | Whether students respected teachers. |
| Q1k | Prevalence of bullying and harassment. |

Table 14: Parts of Q1 selected for modelling and what they indicated.

The variation at the school level is almost equal to that in DVM. Additionally, as with DVM, the estimate for `sexmf` is not statistically significant, but all levels of `parsc_enc` are statistically significant and the values of these estimates are also higher than in DVM and RSM.

The *Agree* response was used for each question variable as the reference level/category. Only the responses to Q1d, Q1f, and Q1j among these variables provided statistically significant estimates. Estimates for Q1d and Q1j were positive, suggesting that students who didn't have the opportunity to participate in

8

| Groups | Name | Variance | Std. Dev. |
|--------|------|----------|-----------|
| School | (Intercept) | 0.1759 | 0.4194 |

Table 15: Summary of random terms in Q1M.

| Variable (category) | Estimate | Std. Error | $z$-value | $p$-value |
|---------------------|----------|------------|-----------|-----------|
| Intercept | -0.550125 | 0.099134 | -5.549 | $2.87 \times 10^{-8}$ |
| sexmf (1) | 0.052615 | 0.077336 | 0.680 | 0.496288 |
| parsc_enc (Skilled) | 1.280788 | 0.081828 | 15.652 | $< 2 \times 10^{-16}$ |
| parsc_enc (Unclassified) | 1.448758 | 0.131738 | 10.997 | $< 2 \times 10^{-16}$ |
| parsc_enc (Unskilled) | 1.508450 | 0.269898 | 5.589 | $2.28 \times 10^{-8}$ |
| Q1a (Disagree) | 0.176249 | 0.094381 | 1.867 | 0.061842 |
| Q1d (Disagree) | 0.282307 | 0.078696 | 3.587 | 0.000334 |
| Q1f (Disagree) | -0.226859 | 0.079590 | -2.850 | 0.004367 |
| Q1g (Disagree) | -0.029728 | 0.089981 | -0.330 | 0.741110 |
| Q1j (Disagree) | 0.623053 | 0.082472 | 7.555 | $4.20 \times 10^{-14}$ |
| Q1k (Disagree) | -0.009652 | 0.120770 | -0.080 | 0.936301 |

Table 16: Summary of fixed terms in Q1M.

EA or those in schools that lacked respectful and amicable student-teacher relationships were more likely to leave secondary education with less than 5 Highers. The estimate for Q1f was negative, suggesting that those who encountered fewer crimes at their schools were more likely to leave secondary education with more than 5 Highers. These trends in the estimates were all in line with the literature and previous hypotheses described in Section 2.2.1.

### 3.2.2 Model for Q2

Q2M was fitted to investigate the effects of the responses to the selected parts of Q2 described in Section 2.2.2. A summary of the question variables and what they indicated is shown in Table 17, and Tables 18 and 19 show the summaries of the random and mixed effects in Q2M, respectively.

| Variable label | What the variable indicates |
|----------------|----------------------------|
| Q2a | Whether pupil was motivated to engage with school. |
| Q2d | Whether class disruption was common. |
| Q2e | Whether peers engaged with school. |
| Q2f | Whether pupils were disciplined appropriately. |
| Q2g | Whether teachers were supportive. |
| Q2i | Whether homework was treated with importance. |

Table 17: Parts of Q2 selected for modelling and what they indicated.

As with DVM and Q1M, the estimate for sexmf was statistically insignificant and the estimates for all levels of parsc_enc were significant. the values of the parsc_enc estimates were larger than DVM and RSM, and very similar to Q1M.

Of the question variable estimates, only Q2g and Q2i were not statistically significant. Estimates for Q2a, Q2e and Q2f were positive, suggesting that students who weren't motivated, had unmotivated peers, or lacked discipline in their school were more likely to leave secondary education with less than 5 Highers. The negative estimate for Q2d suggests that students who encountered less disruption in their classes were less likely to leave secondary education with less than 5 Highers. As with Q1M, these trends in the estimates were in line with the literature and previous hypotheses described in Section 2.2.2.

| Groups | Name | Variance | Std. Dev. |
|--------|------|----------|-----------|
| School | (Intercept) | 0.2295 | 0.479 |

Table 18: Summary of random terms in Q2M.

| Variable (category) | Estimate | Std. Error | $z$-value | $p$-value |
|---------------------|----------|------------|-----------|-----------|
| Intercept | -0.129329 | 0.108022 | -1.197 | 0.231208 |
| sexmf (1) | -0.007120 | 0.078894 | -0.090 | 0.928092 |
| parsc_enc (Skilled) | 1.291765 | 0.083100 | 15.545 | $< 2 \times 10^{-16}$ |
| parsc_enc (Unclassified) | 1.408512 | 0.133605 | 10.542 | $< 2 \times 10^{-16}$ |
| parsc_enc (Unskilled) | 1.475296 | 0.273573 | 5.393 | $6.94 \times 10^{-8}$ |
| Q2a (Disagree) | 0.594659 | 0.147378 | 4.035 | $5.46 \times 10^{-5}$ |
| Q2d (Disagree) | -0.546027 | 0.088078 | -6.199 | $5.67 \times 10^{-10}$ |
| Q2e (Disagree) | 0.316421 | 0.084327 | 3.752 | 0.000175 |
| Q2f (Disagree) | 0.543137 | 0.116259 | 4.672 | $2.99 \times 10^{-6}$ |
| Q2g (Disagree) | 0.005529 | 0.092369 | 0.060 | 0.952271 |
| Q2i (Disagree) | 0.133829 | 0.097235 | 1.376 | 0.168716 |

Table 19: Summary of fixed terms in Q2M.

### 3.2.3 Limitations and Future Considerations

It is worth noting at this stage that the lack of statistical significance for the question variables in both Q1M and Q2M could be attributed to the poor selection of questions included in the models. The current selection was made after testing various combinations and maximizing the number of indicative questions included while ensuring the models remained mathematically sound. To improve upon these results, it may be worth considering all possible selections of indicative questions, as some may provide better results. This process requires a lot of trial and error and would be very time-consuming, but it could result in models with improved fits.

## 4 Conclusions

This study explored the impact of various personal, societal and institutional factors on students' level of academic attainment. This was achieved through the implementation of multi-level logistic regression models including random intercepts to account for variation among schools. While the variables representing parental social class and the age at which students' parents left full-time education provided statistically significant results. Students' parental social class and parents' educational attainment had positive associations with students' level of attainment. However, there was little information gained about the impact that students' sex may have on their attainment. This could be attributed to the dataset itself and may require further investigation. Despite this, it was discovered that constructing a model with random slopes for students' sex improved the fit. In addition, models were fitted to explore the impact that different school experiences had on students' overall attainment. As expected, the prevalence of behavioural issues among peers and a lack of motivation or positive student-teacher relationships resulted in lower attainment overall. However, it must be noted that many variables representing experiences failed to provide statistically significant results. Hence, it is worth exploring the data further and constructing models with different combinations of variables to obtain more meaningful results.

# Bibliography

[1]  Balaguer, Á., Benítez, E., Albertos, A. and Lara, S. 'Not everything helps the same for everyone: relevance of extracurricular activities for academic achievement'. In: *Humanities and Social Sciences Communications* 7.1 (Sept. 2020). DOI: https://doi.org/10.1057/s41599-020-00573-0.

[2]  Boekaerts, M. 'Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students'. In: *Learning and Instruction* 7.2 (June 1997), pp. 161–186. DOI: https://doi.org/10.1016/s0959-4752(96)00015-1.

[3]  Burgess, L. G., Riddell, P. M., Fancourt, A. and Murayama, K. 'The Influence of Social Contagion Within Education: A Motivational Perspective'. In: *Mind, Brain, and Education* 12.4 (Oct. 2018), pp. 164–174. DOI: https://doi.org/10.1111/mbe.12178.

[4]  Camp, M. D. 'The power of teacher-student relationships in determining student success'. English. PhD thesis. 2011, p. 284. ISBN: 978-1-124-81993-8. URL: https://www.proquest.com/dissertations-theses/power-teacher-student-relationships-determining/docview/887899361/se-2 (visited on 08/04/2024).

[5]  Cao, W., Mithra, S. G. S. and R, A. B. 'Unraveling the factors shaping academic success: A structural equation modeling approach for college students'. In: *Heliyon* 10.4 (Feb. 2024). DOI: https://doi.org/10.1016/j.heliyon.2024.e25775.

[6]  Cooper, H., Robinson, J. C. and Patall, E. A. 'Does Homework Improve Academic Achievement? A Synthesis of Research, 1987–2003'. In: *Review of Educational Research* 76.1 (Mar. 2006), pp. 1–62. DOI: https://doi.org/10.3102/00346543076001001.

[7]  Dubow, E. F., Boxer, P. and Huesmann, L. R. 'Long-term Effects of Parents' Education on Children's Educational and Occupational Success: Mediation by Family Interactions, Child Aggression, and Teenage Aspirations'. In: *Merrill-Palmer Quarterly* 55.3 (July 2009), pp. 224–249. DOI: https://doi.org/10.1353/mpq.0.0030.

[8]  Eccles, J. S. 'Influences of parents' education on their children's educational attainments: the role of parent and child perceptions'. In: *London Review of Education* 3.3 (Jan. 2005), pp. 191–204. DOI: https://doi.org/10.1080/14748460500372309.

[9]  Eccles, J. S., Barber, B. L., Stone, M. and Hunt, J. 'Extracurricular Activities and Adolescent Development'. In: *Journal of Social Issues* 59.4 (Dec. 2003), pp. 865–889. DOI: https://doi.org/10.1046/j.0022-4537.2003.00095.x.

[10]  Ergen, B. and Kanadli, S. 'The Effect of Self-Regulated Learning Strategies on Academic Achievement: A Meta-Analysis Study'. In: *Eurasian Journal of Educational Research* 17.69 (May 2017), pp. 55–74. DOI: https://doi.org/10.14689/ejer.2017.69.4.

[11]  Finch, W. H., Bolin, J. E. and Kelley, K. *Multilevel Modeling Using R*. New York, NY: CRC Press (Taylor & Grancis Group), 2019, pp. 133–146. ISBN: 9781138480711.

[12]  Given, L., Nicholas, G. and Pitson, L. *Scotland's Young People in 1999: Scottish School Leavers' Survey Technical Report*. Prepared for the Scottish Executive Education Department. Aug. 2004.

[13]  Hek, M. van, Kraaykamp, G. and Wolbers, M. H. J. 'Comparing the gender gap in educational attainment: the impact of emancipatory contexts in 33 cohorts across 33 countries'. In: *Educational Research and Evaluation* 22.5-6 (Aug. 2016), pp. 260–282. DOI: https://doi.org/10.1080/13803611.2016.1256222.

[14]  King, G. *Unifying political methodology: the likelihood theory of statistical inference*. Cambridge: Cambridge University Press, 2001. ISBN: 9780521366977.

[15]  Leyland, A. H. and Groenewegen, P. P. *Multilevel modelling for public health and health services research: health in context*. Cham, Switzerland: Springer Open, 2020, pp. 89–104. ISBN: 9783030348014. DOI: https://doi.org/10.1007/978-3-030-34801-4. URL: https://doi.org/10.1007/978-3-030-34801-4.

[16]  Ma, X. and Willms, J. D. 'School Disciplinary Climate: Characteristics and Effects on Eighth Grade Achievement'. In: *Alberta Journal of Educational Research* 50.2 (July 2004), pp. 169–188. DOI: https://doi.org/10.11575/ajer.v50i2.55054.

[17]  Simba, N. O., Agak, J. O. and Kakuba, E. K. 'Impact of Discipline on Academic Performance of Pupils in Public Primary Schools in Muhoroni Sub-County, Kenya'. In: *Journal of Education and Practice* 7.6 (2016). URL: https://files.eric.ed.gov/fulltext/EJ1092484.pdf (visited on 09/04/2024).

[18]  Steinmayr, R., Weidinger, A. F., Schwinger, M. and Spinath, B. 'The Importance of Students' Motivation for Their Academic Achievement – Replicating and Extending Previous Findings'. In: *Frontiers in Psychology* 10.1730 (July 2019). DOI: https://doi.org/10.3389/fpsyg.2019.01730.

[19]  Strøm, I. F., Thoresen, S., Wentzel-Larsen, T. and Dyb, G. 'Violence, bullying and academic achievement: A study of 15-year-old adolescents and their school environment'. In: *Child Abuse & Neglect* 37.4 (Apr. 2013), pp. 243–251. DOI: https://doi.org/10.1016/j.chiabu.2012.10.010.

[20]  UCAS. *Scottish Highers*. Apr. 2019. URL: https://www.ucas.com/further-education/post-16-qualifications/qualifications-you-can-take/scottish-highers#:~:text=If%20you (visited on 10/04/2024).