THE UNIVERSITY *of* EDINBURGH

Parental Socio-Economic Background and Children's School-Level GCSE Attainment

Sarah Stopforth

Thesis submitted for the degree of Doctor of Philosophy

School of Social and Political Science

University of Edinburgh

2019

Thesis Declaration

I declare that this thesis has been composed by myself, is my own work, and has not been submitted, in whole or in part, for any other degree.

Sarah Stopforth

Abstract

The principal aim of this thesis is to better understand the contemporary relationship between parental socio-economic background and children's General Certificate of Secondary Education (GCSE) attainment. Previous empirical research has demonstrated that there is a strong, persisting association between parental socio-economic background and educational outcomes, and specifically school GCSE attainment. This thesis directly contributes to the sociology of education in two main ways. First, it presents new empirical evidence about the nature of socio-economic inequalities in young people's GCSE attainment in England over the course of the 1990s and early 2010s. Second, it builds on previous empirical work and develops a more comprehensive understanding of the effects of socio-economic background on educational outcomes. Developing a better understanding of why, or how, those from more advantaged socio-economic backgrounds achieve more favourable educational outcomes by the end of compulsory schooling is important to enable young people, parents, teachers, schools, and policymakers to help to address the persisting attainment gap observed in school-level qualifications.

The thesis is organised into two parts. Part 1 examines the nature of the relationship between parental socio-economic background and children's school GCSE attainment for synthetic cohorts of English Year 12 pupils (i.e. aged 16 and 17). The analyses examine the role of parental socio-economic background in GCSE attainment using the British Household Panel Survey for young people taking their GCSE examinations in the 1990s and 2000s. A key methodological aspect of this work is sensitivity analyses of the independent variables (i.e. socio-economic background measures) and the functional form of the outcome variable (i.e. GCSE attainment). Particular attention is paid to checking the robustness of results using alternative measures and alternative statistical model specifications. The analyses are replicated using the UK Household Longitudinal Study (UKHLS, also known as Understanding Society). Analyses of the UKHLS dataset represent more contemporary cohorts of young people taking their GCSE examinations in the early 2010s. The final section of Part 1 addresses the methodological challenge of missing

data in social surveys. It takes a series of principled statistical approaches to help to address the potential distortions caused by missing data in the synthetic cohort analyses.

Part 2 of this thesis investigates the relationship between parental socio-economic background and children's school GCSE attainment in greater depth. The analyses in Part 2 empirically explore three potential explanations for the enduring socio-economic inequalities observed in educational outcomes. The first set of analyses examine the extent to which inequalities in GCSE attainment can be accounted for by prior academic attainment, for example, attainment at age 11. Cognitive and educational outcomes at earlier stages of schooling are stratified by parental socio-economic background, and therefore the inequalities observed at GCSE level may be a continuation of inequalities observed at earlier stages of a young person's schooling. Path analysis models are used to decompose the effects of parental education and parental social class on attainment at the end of compulsory secondary school.

The next set of analyses investigate the role of cultural capital in educational inequalities. The concept of cultural capital is a prominent sociological explanation for persisting educational inequalities. Developing theoretically informed measures of cultural capital using social survey data is especially challenging because there are no clear prescriptions of how to operationalise these measures. A key aspect of this work is the attention to sensitivity analyses of alternative measures. The candidate measures are compared and contrasted within a series of analyses, with particular attention paid to the effect such measures have on understanding the relationship between parental socio-economic background and GCSE attainment.

The final set of analyses explore the role of educational aspirations in educational inequalities. 'Raising aspirations' has been at the core of recent UK government rhetoric to help to address the attainment gap between the most disadvantaged and more advantaged young people. The overarching government position has been that the attainment gap has been, in part, attributed to the 'low' aspirations held by young people and their parents. The analyses explore the socio-economic gradient to young

people's aspirations over the course of their secondary school years, before examining the influence of the educational aspirations of young people and their parents on GCSE attainment.

Lay Summary

The principal aim of this thesis is to better understand the contemporary relationship between parental socio-economic background and children's General Certificate of Secondary Education (GCSE) attainment. Previous empirical research has demonstrated that there is a strong, persisting association between parental socio-economic background and educational outcomes, and specifically school GCSE attainment. Advanced quantitative methods are employed throughout this thesis to examine the relationship for young people in contemporary England between the 1990s and early 2010s.

The overwhelming substantive finding is the presence of strong, enduring parental socio-economic effects in GCSE attainment. The analyses demonstrate that socio-economic inequalities in educational attainment:

- Have persisted at GCSE level over the course of more than two decades, i.e. from the 1990s to the early 2010s;
- Persist at GCSE level regardless of the form of the GCSE measure;
- Begin early, and are clearly observable by the end of primary school;
- Continue to persist, and potentially widen, over the course of secondary school;
- Are not convincingly explained by conventional sociological explanations of cultural capital;
- Are not convincingly explained by policy explanations of educational aspirations.

Acknowledgements

This PhD research was supported by an ESRC Advanced Quantitative Methods studentship.

I gratefully acknowledge the participants in the datasets used throughout this thesis, as well as the UK Data Service, the Secure Lab team, and the IT and Facilities teams at the University of Edinburgh for enabling my use of the administrative education data.

I have immeasurable gratitude to Professor Vernon Gayle for his generosity of time, patience, support, and enthusiasm over the past three years. I cannot thank him enough for his encouragement and belief in my work and abilities from day one. Many special thanks also to Professor Ellen Boeren for her valued support and guidance throughout. I would also like to thank Dr Chris Playford for his support in the early stages of the PhD process.

I would like to thank all of my friends and family. I am especially grateful to Elisabeth and Emily for their unwavering academic and emotional support.

My greatest thanks go to my Mum, Dad, and Matthew for their constant encouragement throughout my education. Finally, I am indebted to Luke for keeping me calm, focused, and for always listening to an excess of extended monologues about this research.

Contents

List of Tables 12

List of Figures 17

List of Abbreviations 19

Introduction 20

1 Social Stratification and Socio-Economic Background 21

2 Socio-Economic Background and Educational Attainment 22
2.1 Socio-Economic Background and Educational Expansion 22
2.2 The English Education System 24
2.3 Socio-Economic Background and GCSE Attainment 25

3 Research Questions 28

4 Data and Methods 29
4.1 Household Panel Surveys 30
4.1.1 The British Household Panel Survey 30
4.1.2 The UK Household Longitudinal Study 31
4.1.3 The National Pupil Database 31
4.2 Complex Survey Design 32
4.3 Methods 33

5 Structure of the Thesis 34

Part 1: The Contemporary Relationship between Parental Socio-Economic Background and Children's School GCSE Attainment 35

Introduction to Part 1 35

Chapter 1 Synthetic Cohorts of Year 12 Pupils in the 1990s and 2000s: A Sensitivity Analysis 37

1 Introduction 37

2 Operationalising Socio-Economic Background and Educational Attainment 38
2.1 Social Class Schemas 39
2.1.1 Registrar General Class Schema 39
2.1.2 Goldthorpe Class Schema 40
2.1.3 National Statistics Socio-Economic Classification 41
2.1.4 Micro-Class Analysis 43
2.1.5 'New Model' of Social Class 44
2.2 Social Stratification Scales 46
2.2.1 Occupational Prestige Scales 46
2.2.2 Cambridge Scale and CAMSIS 47
2.3 Unit of Analysis Problem 47
2.4 Measuring Education Level 48
2.4.1 Years of Education 49
2.4.2 Highest Education Qualification 49
2.5 Measuring School GCSE Attainment 50

3 Synthetic Cohorts of English Year 12 Pupils in the BHPS 52

4 Descriptive Statistics 53

5 Modelling School GCSE Attainment 57
5.1 Sensitivity Analyses of Independent Variables 63
5.1.1 Testing Measures of Parental Social Class 63
5.1.2 Testing Measures of Parental Education 66
5.2 Functional Form of the Outcome Variable 69
5.2.1 Alternative Thresholds of GCSE Attainment 70
5.2.2 GCSE Attainment Brackets 72
5.2.3 Number of GCSEs at Grades A*-C 77

6 Discussion and Conclusions 80

Chapter 2 Synthetic Cohorts of Year 12 Pupils in the Early 2010s: A Replication Analysis 82

1 Introduction 82

2 Replicating Synthetic Cohorts of English Year 12 Pupils in the UKHLS 83

3 Descriptive Statistics 84

4 Modelling GCSE Attainment 86

4.1 Attainment of 5 or More A*-Cs 86

4.2 Number of GCSEs at Grades A*-C 91

4.3 Overall GCSE Point Score 94

5 Discussion and Conclusions 98

Chapter 3 Missing Data in the Synthetic Cohorts Analyses 100

1 Introduction 100

2 Missing Data 100

3 Missing Data in the BHPS 101

3.1 Description of the Missing Data 102

3.2 Missing Data for Parental NS-SEC 103

3.3 Multiple Imputation by Chained Equations 105

4 Missing Data in the UKHLS 107

5 Discussion and Conclusions 110

Part 2: Exploring Potential Sociological Explanations for the Persisting Socio-Economic Effect 112

Introduction to Part 2 112

Chapter 4 Socio-Economic Differences in Prior Attainment and GCSE Attainment 115

1 Introduction 115

2 School Attainment 116
2.1 Key Stages and National Curriculum Levels 116
2.2 Socio-Economic Background and School Attainment 118

3 Data and Methods 122
3.1 Sample 122
3.2 Measures 122
3.3 Structure of Analysis 123

4 Descriptive Statistics 124

5 Modelling Results 128
5.1 Modelling Key Stage 2 Attainment 128
5.2 Controlling for Prior Attainment 129
5.3 Path Analysis 135
5.4 Sensitivity Analyses 140
5.4.1 English and Maths Attainment Measures 140
5.4.2 School Type 143

6 Discussion and Conclusions 144

Chapter 5 Cultural Capital, Socio-Economic Background, and GCSE Attainment 147

1 Introduction 147

2 The Concept of Cultural Capital 148
2.1 Bourdieu and Cultural Capital 148
2.1.1 The Forms of Cultural Capital 149
2.1.2 Cultural and Social Reproduction 149
2.2 Cultural Capital and Quantitative Sociology 150
2.2.1 Operationalisation Challenges 150
2.2.2 Parental Education Level 152
2.2.3 Highbrow Cultural Participation 153

2.2.4 Reading Behaviours 155
2.2.5 Highbrow Cultural Participation and Reading Behaviours 156

3 Data and Methods 157
3.1 Sample 157
3.2 Measures 158
3.2.1 Highbrow Cultural Participation 159
3.2.2 Reading Behaviours 159
3.3 Structure of Analysis 160

4 Descriptive Statistics 161

5 Modelling Results 165
5.1 Modelling Cultural Capital 165
5.2 Modelling GCSE Attainment 169
5.3 A Cultural Capital Model 173
5.3.1 Factor Scores 174
5.3.2 Summed Scales 177
5.4 Parental Education as Cultural Capital 181

6 Discussion and Conclusions 183

Chapter 6 Educational Aspirations, Socio-Economic Background, and GCSE Attainment 186

1 Introduction 186

2 Youth Aspirations, Attainment, and Socio-Economic Differences 187
2.1 Policy Discourse 187
2.2 Aspirations and Socio-Economic Background 188
2.3 Aspirations and Attainment 191

3 Data and Methods 193
3.1 Sample 193
3.2 Measures 193
3.2.1 Longitudinal Analyses of Educational Aspirations 193

3.2.2 Cross-Sectional Analyses of GCSE Attainment 194
3.3 Structure of Analysis 195

4 Modelling Aspirations 196
4.1 Descriptive Statistics 196
4.2 Panel Models 199
4.3 Modelling Results 201

5 Modelling GCSE Attainment 208
5.1 Descriptive Statistics 208
5.2 Modelling Results 210

6 Discussion and Conclusions 214

Conclusions 216

1 Introduction 216

2 Substantive Conclusions 216
2.1 The Contemporary Relationship between Parental Socio-Economic Background and Children's School GCSE Attainment 217
2.2 The Role of Prior Attainment 218
2.3 The Role of Cultural Capital 220
2.4 The Role of Educational Aspirations 222
2.5 Substantive Reflections 223

3 Methodological Reflections 224
3.1 Large-Scale Social Science Surveys 224
3.2 Changes to Curriculum and Assessment 226

4 Final Remarks 228

Appendix 1: Additional Tables and Figures 229

Appendix 2: Missing Data in Part 2 241

Data Citation.....246

Bibliography.....247

List of Tables

Table 1. 1: Registrar General class schema 40
Table 1. 2: Goldthorpe class schema and collapses 41
Table 1. 3: NS-SEC analytic classes and collapses 42
Table 1. 4: Example occupations by category of NS-SEC 43
Table 1. 5: CASMIN framework 50
Table 1. 6: ISCED framework 50
Table 1. 7: Descriptive statistics of explanatory variables for synthetic cohorts of English Year 12 pupils in BHPS households, unweighted 57
Table 1. 8: Goodness-of-fit summaries for explanatory variables and attainment of 5+ GCSEs at grades A*-C 58
Table 1. 9: Model building goodness-of-fit summaries for logistic regression model of attainment of 5+ GCSEs at grades A*-C 59
Table 1. 10: Logistic regression model of the attainment of 5+ GCSEs at grades A*-C 60
Table 1. 11: Average marginal effects on the probability of attaining 5+ GCSEs at grades A*-C 62
Table 1. 12: Sensitivity analyses of alternative measures of parental social stratification 65
Table 1. 13: Predicted probabilities of attaining 5+ GCSEs at grades A*-C using alternative parental social stratification measures 66
Table 1. 14: Sensitivity analyses of alternative measures of parental education level 68
Table 1. 15: Predicted probabilities of attaining 5+ GCSEs at grades A*-C using alternative parental education measures. 69
Table 1. 16: Frequencies and percentages of GCSE results by varying thresholds of attainment, unweighted 70
Table 1. 17: Frequencies and percentages of GCSE results by attainment bracket, unweighted 72
Table 1. 18: Visualisation of the comparisons made in proportional odds models and continuation ratio models for GCSE attainment brackets 73

Table 1. 19: Ordered logistic regression models of GCSE attainment brackets, unweighted 74
Table 1. 20: Zero-inflated negative binomial regression model of the number of A*-Cs 78
Table 1. 21: Expected counts for the number of GCSEs attained at grades A*-C ... 80
Table 2. 1: Descriptive statistics of explanatory variables for synthetic cohorts of English Year 12 pupils in UKHLS-NPD dataset, unweighted 85
Table 2. 2: Goodness-of-fit summaries for explanatory variables and attainment of 5+ GCSEs at grades A*-C 87
Table 2. 3: Model building goodness-of-fit summaries for logistic regression model of attainment of 5+ GCSEs at grades A*-C 87
Table 2. 4: Logistic regression model of the attainment of 5+ GCSEs at grades A*-C 89
Table 2. 5: Average marginal effects on probability of attaining 5+ GCSEs at grades A*-C 90
Table 2. 6: Zero-inflated negative binomial regression model of the number of A*-Cs 92
Table 2. 7: Expected counts for the number of GCSEs attained at grades A*-C 94
Table 2. 8: Linear regression model of total GCSE point score 97
Table 3. 1: Missing data patterns for the BHPS synthetic cohorts 102
Table 3. 2: Logistic regression models of the attainment of 5+ GCSEs at grades A*-C addressing missing data for parental social class 104
Table 3. 3: Logistic regression models of the attainment of 5+ GCSEs at grades A*-C using multiple imputation by chained equations 106
Table 3. 4: Linear regression models of total GCSE point score using multiple imputation by chained equations 109
Table 4. 1: Descriptive statistics for key explanatory variables, unweighted 124
Table 4. 2: Mean Key Stage 2 and GCSE attainment scores for key explanatory variables, unstandardised 125
Table 4. 3: Linear regression model of Key Stage 2 score (standardised) 129
Table 4. 4: Linear regression models of total GCSE point score controlling for Key Stage 2 attainment (standardised) and interaction effects 131

Table 4. 5: Path analysis model of Key Stage 2 attainment and GCSE attainment with standardised coefficients 136
Table 4. 6: Total, direct, and indirect effects of socio-economic background on GCSE attainment, standardised 140
Table 4. 7: Path analysis model of Key Stage 2 English score and GCSE attainment with standardised coefficients 141
Table 4. 8: Path analysis model of Key Stage 2 Maths score and GCSE attainment with standardised coefficients 142
Table 5. 1: Candidate cultural capital measures available in Wave 2 of the UKHLS adult survey and youth questionnaire 158
Table 5. 2: Descriptive statistics for key explanatory variables, unweighted 162
Table 5. 3: Descriptive statistics for candidate cultural capital measures, unweighted 163
Table 5. 4: Chi square tests of association and Cramer's V statistics for parental socio-economic background and cultural capital measures 166
Table 5. 5: Chi square tests of association and Cramer's V statistics for parent and child cultural capital measures 167
Table 5. 6: Poisson regression model of child highbrow participation 168
Table 5. 7: Poisson regression model of child reading behaviours 169
Table 5. 8: Raw effects (bivariate) of cultural capital measures on total GCSE point score 171
Table 5. 9: Tetrachoric correlations for measures of parent and child cultural capital 175
Table 5. 10: Rotated (oblique promax) factor loadings for principal components factor analysis of cultural capital 176
Table 5. 11: Linear regression model of total GCSE point score including cultural capital factor scores 177
Table 5. 12: Linear regression model of total GCSE point score including cultural capital summed scales 179
Table 5. 13: Linear regression model of total GCSE point score with parental education level 182
Table 6. 1: Patterns of wave response in the youth questionnaire with linked NPD records 196

Table 6. 2: Overall, between, and within-person variation of key variables 198
Table 6. 3: Pooled cross-sectional logistic regression model of university aspirations 201
Table 6. 4: Random effects logistic regression model of university aspirations 203
Table 6. 5: Predicted probabilities of university aspirations by parental socio-economic background 205
Table 6. 6: Random effects logistic regression model of university aspirations with Key Stage 2 attainment 206
Table 6. 7: Descriptive statistics for cross-sectional analyses, unweighted 209
Table 6. 8: Linear regression model of total GCSE point score with educational aspirations 211
Table 6. 9: Linear regression model of total GCSE point score with Key Stage 2 attainment 213
Table A1. 1: Logistic regression models of different threshold measures of GCSE attainment 229
Table A1. 2: Multinomial logistic regression model of GCSE attainment brackets. 230
Table A1. 3: Stereotype logistic regression model of GCSE attainment brackets . 231
Table A1. 4: Comparison of count regression models of number of GCSEs attainment at grades A*-C 232
Table A1. 5: Total, direct, and indirect effects of socio-economic background on GCSE attainment, accounting for English score at Key Stage 2, standardised 234
Table A1. 6: Total, direct, and indirect effects of socio-economic background on GCSE attainment, accounting for Maths score at Key Stage 2, standardised 234
Table A1. 7: Rotated and unrotated factor loadings of cultural capital measures .. 235
Table A1. 8: Linear regression model building process of GCSE point score including cultural capital factor scores 236
Table A1. 9: Linear regression model building process of GCSE point score including cultural capital summed scales 237
Table A1. 10: Linear regression model of total GCSE point score with post-16 study aspirations 240
Table A2. 1: Linear regression models of Key Stage 2 attainment using multiple imputation by chained equations (Chapter 4) 242

Table A2. 2: Linear regression models of GCSE attainment using multiple imputation by chained equations (Chapter 4).....243
Table A2. 3: Linear regression models of GCSE attainment using multiple imputation by chained equations (Chapter 5).....244
Table A2. 4: Linear regression models of GCSE attainment using multiple imputation by chained equations (Chapter 6).....245

List of Figures

Figure 1. 1: Synthetic cohorts of Year 12 pupils in BHPS households.....52
Figure 1. 2: Number of observations of synthetic cohorts of Year 12 pupils per wave 53
Figure 1. 3: Number of GCSEs attained at grades A*-C for synthetic cohorts of Year 12 pupils.....54
Figure 1. 4: Log odds coefficients with 95% quasi-variance comparison intervals on the attainment of 5+ GCSEs at grades A*-C by parental education level.....61
Figure 1. 5: Log odds coefficients with 95% quasi-variance comparison intervals on the attainment of 5+ GCSEs at grades A*-C by parental social class.....61
Figure 1. 6: Predicted probabilities of thresholds of GCSE attainment by parental education level (separate logistic regression models).....71
Figure 1. 7: Predicted probabilities of thresholds of GCSE attainment by parental social class (separate logistic regression models) 72
Figure 1. 8: Predicted probabilities of GCSE attainment brackets by parental education level (proportional odds model) 76
Figure 1. 9: Predicted probabilities of GCSE attainment brackets by parental social class (proportional odds model) 76
Figure 2. 1: Number of GCSEs attained at grades A*-C for synthetic cohorts of Year 12 pupils in the UKHLS-NPD 84
Figure 2. 2: Total GCSE point score (grades A*-U) for synthetic cohorts of Year 12 pupils in the UKHLS-NPD 95
Figure 2. 3: Linear regression coefficients with 95% quasi-variance comparison intervals for total GCSE point score by parental education level 97
Figure 2. 4: Linear regression coefficients with 95% quasi-variance comparison intervals for total GCSE point score by parental social class.....98
Figure 4. 1: Mean Key Stage 2 test scores (standardised) by parental education level.....126
Figure 4. 2: Mean Key Stage 2 test scores (standardised) by parental NS-SEC...126
Figure 4. 3: Mean GCSE point score (standardised) by parental education level..127
Figure 4. 4: Mean GCSE point score (standardised) by parental NS-SEC.....127

Figure 4. 5: Interaction effect of parental education and Key Stage 2 attainment on GCSE score 134
Figure 4. 6: Interaction effect of parental NS-SEC and Key Stage 2 attainment on GCSE score 134
Figure 4. 7: A conceptual path model of parental socio-economic background and attainment..... 135
Figure 4. 8: A path model of parental socio-economic background, Key Stage 2 attainment and GCSE attainment..... 138
Figure 5. 1: Sensitivity analyses of the effect of cultural capital measures on parental education level regression coefficients with 95% confidence intervals 172
Figure 5. 2: Sensitivity analyses of the effect of cultural capital measures on parental NS-SEC regression coefficients with 95% confidence intervals..... 173
Figure 5. 3: Parental education level linear regression coefficients with 95% confidence intervals for models of GCSE point score with and without cultural capital measures..... 180
Figure 5. 4: Parental NS-SEC linear regression coefficients with 95% confidence intervals for models of GCSE point score with and without cultural capital measures 181
Figure 6. 1: University aspirations by age of respondent, pooled data 199
Figure 6. 2: Log odds coefficients and 95% quasi-variance comparison intervals for university aspirations by parental education level 204
Figure 6. 3: Log odds coefficients and 95% quasi-variance comparison intervals for university aspirations by parental NS-SEC 204
Figure 6. 4: Predicted probabilities of university aspirations by Key Stage 2 attainment..... 207
Figure 6. 5: Interaction effect of Key Stage 2 attainment with university aspirations on GCSE attainment 214
Figure A1. 1: Post-16 destinations by age of respondent, pooled data..... 239

List of Abbreviations

ALSPAC	Avon Longitudinal Study of Parents and Children
BHPS	British Household Panel Survey
CASMIN	Comparative Analysis of Social Mobility in Industrial Nations
FSM	Free School Meals
GCSE	General Certificate of Secondary Education
ISCED	International Standard Classification of Education
KS2	Key Stage 2
LYSPE	Longitudinal Study of Young People in England
MCS	Millennium Cohort Study
NPD	National Pupil Database
NS-SEC	National Statistics Socio-Economic Classification
ONS	Office for National Statistics
UKHLS	UK Household Longitudinal Study, also known as Understanding Society
YCS	Youth Cohort Study of England and Wales

Introduction

There is a long-standing research tradition examining social stratification and the role of socio-economic background in education (Floud et al., 1961, Blau and Duncan, 1967, Jencks, 1973, Boudon, 1974, Karabel and Halsey, 1977, Halsey et al., 1980, Shavit and Blossfeld, 1993, Crompton, 2008, Platt, 2010). Empirical evidence consistently demonstrates that socio-economic inequalities in educational outcomes continue to pervade education systems in the UK (Jackson, 1962, Lacey, 1970, Halsey et al., 1980, Heath and Clifford, 1990, Ball, 2003, Blanden and Machin, 2004, Machin and Vignoles, 2004, Reay, 2017). This thesis directly contributes to the long-standing research tradition in the sociology of education. First, it presents new empirical evidence about the nature of socio-economic inequalities in young people's General Certificate of Secondary Education (GCSE) attainment in England between the 1990s and 2010s. Second, it builds on previous empirical work and develops a more comprehensive understanding of the effects of socio-economic background on educational outcomes. Developing a better understanding of why, or how, those from more advantaged socio-economic backgrounds achieve more favourable educational outcomes by the end of compulsory schooling is important to enable young people, parents, teachers, schools, and policymakers to help to address the persisting attainment gap observed in school-level qualifications.

This thesis focuses on parental socio-economic background effects in children's school-level GCSE attainment. GCSEs are a set of examinations historically taken at the end of compulsory schooling in England and Wales. GCSEs mark a key branching point for young people, shaping their opportunities and choices for continued education, employment, or training (see, for example, analyses of the Youth Cohort Study in Payne, 1995a, Payne, 1995b, Payne, 2001a, Payne, 2001b, and see Jones et al., 2003, Babb, 2005). In this introductory chapter, the main themes of the thesis are presented. The following sections present the context of social stratification, socio-economic background, and GCSE attainment. The research questions, data, and methods are presented, and the main structure of the thesis is outlined.

1 Social Stratification and Socio-Economic Background

Social stratification can be understood as the persistence of inequalities which occur, or are reproduced, across generations (Bottero, 2005). Inequalities over time have been the subject of much empirical work, which tends to find that the structures of social stratification are largely stable (Lambert et al., 2012). Stratification in society can take a variety of forms according to socially-constructed differences of, for example, class, gender, and ethnicity (Payne, 2013b, Grusky, 2014).

There are no agreed-upon, single measures of socio-economic background in social science research (Crompton, 2008, Connelly et al., 2016c). Socio-economic stratification measures often involve occupation-based schemas and scales (Lambert et al., 2012). A focus on occupational structure or relations can provide a more stable base than the more transient, or fluctuating, nature of income in terms of economic security and future prospects (Goldthorpe and McKnight, 2004).

Social distance measures have historically taken a bottom-up approach, whereby the structure is defined by the social relations within it, rather than by a pre-determined class structure (Bottero, 2005). Stewart et al. (1973) used multidimensional scaling techniques to construct social stratification scales of social distance or social interaction between individuals, for example the Cambridge Scale. The authors used the occupations of acquaintances and friends to generate measures of social relationships.

Social class schemas use occupations as the basis of categorising people's socio-economic position. Rose and Pevalin (2003) noted that some analysts consider socio-economic background in terms of social class, using measures such as the National Statistics Socio-Economic Classification (NS-SEC). Individuals occupying a similar NS-SEC category, or social class position, are understood to have similar market and economic power (Rose and Pevalin, 2003). Crompton (2008), for example, suggested that they also share similar lifestyles and social attitudes. The UK Office for National Statistics (ONS) adopted NS-SEC for use in official statistics (Rose and Pevalin, 2003).

NS-SEC is considered as a robust measure of socio-economic position which has been found to play a central role in a range of different outcomes such as health and education (Rose and Pevalin, 2003). The NS-SEC schema will be the main social class background measure employed throughout this thesis. Parental education level will also be included in analyses as a measure of socio-economic background. This follows the suggestion of Bukodi and Goldthorpe (2013) that social class and education exert separate, and independent, influences on outcomes such as educational attainment. Alternative measures of parental background are tested as part of the routine sensitivity analyses undertaken in Chapter 1.

2 Socio-Economic Background and Educational Attainment

There is a wealth of sociological literature documenting the long-standing association between parental socio-economic background and children's educational attainment (Floud et al., 1961, Blau and Duncan, 1967, Jencks, 1973, Boudon, 1974, Karabel and Halsey, 1977, Halsey et al., 1980, Shavit and Blossfeld, 1993, Crompton, 2008, Platt, 2010). The observed association between socio-economic background and educational outcomes persists over time and across most developed countries, irrespective of differences in welfare state structures and educational systems (Shavit and Blossfeld, 1993, Erikson and Jonsson, 1996, Goldthorpe, 2003).

2.1 Socio-Economic Background and Educational Expansion

Theoretically, an increase in educational expansion could lead to greater equality in educational outcomes between more and less advantaged young people. Boudon (1974) argued that inequalities could diminish over time as educational expansion occurs. Those from less advantaged backgrounds would be able to increase their participation at a greater rate than those from more advantaged backgrounds, who were more restricted by 'ceiling effects'. Greater rates of participation in higher levels of education can be considered as inevitable, as access becomes easier to gain by children from all social backgrounds. The theoretical assumptions therefore suggest that the association between socio-economic background and educational outcomes will diminish or weaken over time. By contrast, empirical studies have shown that the

association between socio-economic background and educational attainment persists despite large-scale educational expansion (for example, see Floud et al., 1961, Jencks, 1973, Shavit and Blossfeld, 1993).

There are a number of theories which could explain the persisting inequalities in educational outcomes. Maximally maintained inequality (MMI) theory states that class inequality persists at different levels of the education system because the rate of transitions at various levels, contingent on social background, remains the same until demand outstrips supply (Raftery and Hout, 1993). Educational expansion occurs when the current level of education is saturated and opportunities open up at higher transition points. The increase in the rate of enrolment for all social classes at lower transition points is offset by the greater opportunities created at higher transition points. These are typically taken up by more advantaged young people. One example of this is the expansion of the British higher education system. There has been an overall increase in participation for students from all social backgrounds (Chowdry et al., 2013). As a result, there has been 'credential inflation', where the value of undergraduate degrees has lessened due to the higher proportion of people obtaining them. As undergraduate education becomes saturated, there has been a subsequent rise in the numbers gaining postgraduate degrees (Van de Werfhorst and Andersen, 2005).

Paterson and Iannelli (2007) suggested a refinement to the MMI theory which accounts for non-linearity in the expansion of educational opportunity. They argued that inequalities are higher at certain phases, for example, more advantaged children are better placed to take advantage of new opportunities as they arise, but this evens out over time. Another refinement of the MMI theory is Effectively Maintained Inequality (EMI) (Lucas, 2001). EMI theory states that more advantaged families maintain their advantage even when opportunities become universal. Instead of differentiating through attending higher levels of education when opportunities are not universal, more advantaged families differentiate themselves through the quality of education sought once a particular education level is saturated. For example, more advantaged young people may attend more prestigious institutions than less advantaged young people (Reay et al., 2001, Boliver, 2015). At a university level,

Reay et al. (2001) demonstrated that although overall participation rates increased, working class and middle class young people attend very different types of institutions, and the most elite universities remain the preserve of the white middle classes. Bathmaker et al. (2013) commented that as greater numbers of working class students enter university, the 'rules of the game' shift. Middle class students may mobilise their resources to advance their employability after graduation. For example, in addition to academic achievement, more advantaged young people may undertake internships and extra-curricular activities to gain an advantage in the labour market. In this way, socio-economic advantage is preserved despite widening access for all.

2.2 The English Education System

The UK consists of four territories and does not have a single school education system (Paterson and Iannelli, 2007). The focus of this thesis is on young people in the English education system (see section 4 for an explanation). Universal secondary schooling was introduced by the 1918 Fisher Act for children between the ages of 5 and 14, which was subsequently raised to age 15 with the 1944 (Butler) Education Act (McKibbin, 1998). The 1944 Act also established a tripartite system of grammar schools, secondary moderns, and technical schools (Halsey et al., 1980). Heath and Clifford (1990) argued that the tripartite school system disproportionately benefited middle class children because they were more likely to get into grammar schools. One explanation for this was that middle class children had the cultural knowledge required to pass the 11-plus examination (McKibbin, 1998). The nature of school-level streaming from the age of 11 was deemed to contribute to a 'wastage of talent' because many able working class children did not enter grammar schools (Central Advisory Council for Education, 1959). Lacey (1970) highlighted that working class boys who did enter grammar schools were disproportionately streamed into lower ability sets and sometimes became disillusioned with the school system. Widespread comprehensivisation of secondary schools began in the early 1960s, and by the 1970s English schools were mostly non-selective, although pockets of selective schools still existed in some areas (Paterson and Iannelli, 2007, Coldron et al., 2010). From the late 1980s, governments championed choice, competition, and standardisation in the secondary school system, for example, through the introduction

of school league tables and Ofsted inspections (Brown, 1990). Ball (2003) argued that middle class parents were more able to 'game' the system, for example, through greater financial resources to move house or rent a second home in the catchment areas of very good schools.

Post-school education in the UK also expanded over the 20th century (Machin and Vignoles, 2004). The expansion of higher education became UK government policy with the 1963 Robbins Report. The report suggested that expanding higher education could allow students from all classes to share in a 'common culture' which may 'compensate for any inequalities of home background' (Great Britain Committee on Higher Education, 1963: 7). Participation rates increased over the years, and increased at a much faster rate in the 1990s (Machin and Vignoles, 2004). The accelerated increase was in part a result of the 1992 Higher Education Act which granted polytechnics university status (Marcenaro-Gutierrez et al., 2007). In 1997, the Dearing Report recommended the introduction of £1000 tuition fees and income contingent loans (Blanden and Machin, 2004). Fees rose after legislation was passed in 2004 (Galindo-Rueda et al., 2004). Although higher education expansion was intended for pupils from all social backgrounds, evidence suggests that there was a disproportionate benefit of higher education expansion for those from more advantaged families compared with their less advantaged peers (Blanden et al., 2003, Machin and Vignoles, 2004, Chowdry et al., 2013).

2.3 Socio-Economic Background and GCSE Attainment

General Certificates of Secondary Education (GCSEs) are the set of examinations taken by young people historically at the end of compulsory schooling, i.e. at the age of 15 or 16. GCSE results are sociologically informative because they signal the first major point of departure in a young person's schooling. The qualifications gained at the end of compulsory schooling can be important determinants of the young person's future educational and occupational opportunities and choices (see, for example, analyses of the Youth Cohort Study in Payne, 1995a, Payne, 1995b, Payne, 2001a, Payne, 2001b, and see Jones et al., 2003, Babb, 2005). GCSE examination results are the main substantive focus of this thesis. The term 'attainment' has been

preferred, rather than ‘achievement’ or ‘outcomes’, to reflect the specific results gained in GCSE examinations, and the qualifications obtained as a result.

GCSE examinations were introduced as part of the reforms in the 1988 Educational Reform Act. GCSEs replaced General Certificate of Education Ordinary (O’) Levels and Certificates of Secondary Education (CSEs) to establish a single system of assessment, with grades ranging from A to G (Department for Education, 1985). In 1994, an additional A* grade was introduced at the highest level (Yang and Woodhouse, 2001). Assessments originally comprised of a combination of coursework and examinations. At the outset of the GCSE system, the core subjects at the heart of the National Curriculum in England and Wales were English, Maths, and Science, and further foundational subjects were Languages, Technology, History, Geography, Art, Music, and Physical Education (Department for Education and Science Welsh Office, 1987). The Government advised that seven or eight of the core and foundational subjects should be examined (Department for Education and Science Welsh Office, 1987). Young people generally studied eight or nine subjects at GCSE level (Rothon, 2007). In 2004, the National Curriculum was revised (Department for Education and Skills and Qualifications and Curriculum Authority, 2004). The core subjects of English, Maths, and Science remained, as well as ‘non-core foundational subjects’ of Information and Communication Technology (ICT), Citizenship, and Physical Education. The Arts, Design and Technology, Humanities, and Modern Foreign Languages subjects became ‘entitlement areas’, with schools encouraged to offer GCSE qualifications in a subject in each area (Department for Education and Skills and Qualifications and Curriculum Authority, 2004). The particular combination of examination subjects taken by the young person may be influenced by the individual, their parents, and the subjects offered by their school. The complexity of the assessment structure in England and Wales presents sizeable empirical challenges in the study of school attainment.

Empirical work has demonstrated that young people from more advantaged backgrounds tend to have more favourable GCSE outcomes than their less advantaged peers (Demack et al., 2000, Sullivan, 2001, Connolly, 2006, Connelly et al., 2013, Gayle et al., 2014, Strand, 2014a, Playford and Gayle, 2015). The social

stratification of GCSE attainment (often by class, gender, and ethnicity) has been explored using large-scale datasets such as the Youth Cohort Study (Demack et al., 2000, Babb, 2005, Connolly, 2006, Gayle et al., 2014). Analyses of the older Youth Cohort Study of England and Wales (YCS) have demonstrated that there are substantial differences in GCSE attainment by parental social class (as measured by NS-SEC) and parental education level (Babb, 2005, Gayle et al., 2014). Analyses of the YCS have further suggested that social class differences in GCSE attainment may be substantively greater than gender or ethnicity differences (Demack et al., 2000, Connolly, 2006, Rothon, 2007).

Analyses of GCSE attainment have also been undertaken, to a lesser extent, using the British Household Panel Survey (Murray et al., 2012, Connelly et al., 2013), the Avon Longitudinal Study of Parents and Children (Morris et al., 2016), and the Longitudinal Study of Young People in England (Chowdry et al., 2011). Murray et al. (2012) and Connelly et al. (2013) used the BHPS to examine middle attainment, and provided further evidence of the importance of gender, ethnicity, parental education level, and parental NS-SEC (using a 3-category measure) for young people's GCSE results. Morris et al. (2016) used the Avon Longitudinal Study of Parents and Children to examine the relationship between a variety of socio-economic background measures (including an occupation-based measure of social class) and school attainment. The authors found that socio-economic inequalities in GCSE attainment and General Certificate of Education Advanced (A') Level attainment persisted over and above the effects of cognitive ability measured at age 8 (Morris et al., 2016). Chowdry et al. (2011) used the Longitudinal Study of Young People in England to assess the relationship between attitudes, behaviours, and socio-economic differences in GCSE attainment. The authors demonstrated that socio-economic position, based on household income, financial difficulties, parent occupations, and housing tenure, is associated with GCSE attainment at age 16, even accounting for attainment at age 11. The persistent finding of a socio-economic gradient to GCSE attainment is consequential and worthy of further sociological investigation.

Government statistics tend to demonstrate similar findings of inequalities in GCSE attainment, using proxy measures for disadvantage instead of occupation-based

measures (see, for example, the 2009/10 results Department for Education, 2010). Census data or official statistics do not currently collect the finer-grained detail of parental socio-economic positions and therefore often rely on proxy measures of disadvantage, such as eligibility for Free School Meals, or area-level socio-economic indicators (for example, Hamnett et al., 2007, Crawford et al., 2017). This thesis uses large-scale social science surveys with detailed parental socio-economic background measures based on parental occupation and education level to examine a broader spectrum of socio-economic advantage and disadvantage.

3 Research Questions

The overarching aim of this thesis is to develop a more comprehensive understanding of the contemporary relationship between parental socio-economic background and children's school GCSE attainment. The main research methods are the application of statistical models to large-scale and complex social science survey data.

This thesis is organised into two separate, yet inter-related, parts. In Part 1, the nature of the relationship between parental socio-economic background and children's school GCSE attainment is examined. The overarching research question for Part 1 is:

1. What is the relationship between parental socio-economic background and children's school GCSE attainment?

Measures of parental socio-economic background and GCSE attainment are described and modelled using British household panel data. Part 2 builds on the work developed in Part 1 and presents exploratory analyses of three potential explanations for the socio-economic background effect in school attainment for young people in contemporary England. The potential explanations are examined based on their sociological and policy relevance, and prominence in debates concerning educational inequalities and the attainment gap at the end of compulsory schooling (see the individual chapters for more detail). There are three supplementary research questions addressed in Part 2.

2. To what extent does prior attainment explain socio-economic differences in GCSE attainment?
3. To what extent does cultural capital explain socio-economic differences in GCSE attainment?
4. What role do educational aspirations play in GCSE attainment?

4 Data and Methods

The contemporary relationship between parental socio-economic background and school GCSE attainment is examined using large-scale, nationally representative data collected from two household panel surveys. The first dataset is the British Household Panel Survey (BHPS).[1] The second dataset is the UK Household Longitudinal Study (UKHLS).[2] Official education data containing GCSE records from the National Pupil Database are linked to the UKHLS (UKHLS-NPD).[3] These data are accessible using the Secure Lab environment provided by the UK Data Service. The records are only available for young people in England. The following sections outline the suitability of these studies to examining parental socio-economic background and educational outcomes for young people, the considerations of working with complex survey designs, and an overview of the main statistical methods used throughout this thesis.

[1] University of Essex, Institute for Social and Economic Research. (2010) British Household Panel Survey: Waves 1-18, 1991-2009. [data collection]. 7th Edition. UK Data Service. SN: 5151, http://doi.org/10.5255/UKDA-SN-5151-2.
[2] University of Essex. Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2018). Understanding Society: Waves 1-7, 2009-2016 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 10th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-11.
[3] Department for Education, University of Essex. Institute for Social and Economic Research and NatCen Social Research. (2015). Understanding Society: Wave 1, 2009-2011: Linked National Pupil Database: Secure Access [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. SN: 7642, http://dx.doi.org/10.5255/UKDA-SN-7642-2.

4.1 Household Panel Surveys

The UK has an impressive set of birth cohort studies which support secondary data analysis. Britain has a history of collecting birth cohort studies in 12-yearly intervals from the post-war period. These include the National Survey of Health and Development in 1946 (Wadsworth et al., 2006), the National Child Development Study in 1958 (Power and Elliott, 2006), and the British Birth Cohort Study in 1970 (Elliott and Shepherd, 2006). There was a hiatus in the birth cohort studies in the 1980s and 1990s, as the study planned for 1982 did not take place (Gayle, 2005). The birth cohort studies resumed with the Millennium Cohort Study, with participants born between 2000 and 2002 (Connelly and Platt, 2014). There is not large-scale UK birth cohort data to cover school attainment for young people in the 1990s and early 2000s.

Previous studies examining educational attainment for young people in the 1990s have utilised the Youth Cohort Study and the British Household Panel Survey (Gayle, 2005, also see section 2.2 above). The UK Household Longitudinal Study is a relatively under-utilised dataset in the area of GCSE attainment and will contribute contemporary findings in the 21st Century. Household panel surveys are not specialist education datasets but provide a promising opportunity to study educational outcomes for young people in participating households. The BHPS and UKHLS are particularly suited to these analyses because, by design, they provide information on every member of the household, including parents and young people. The youth panels in the surveys (see below) further allow for longitudinal attitudinal and behavioural information to be collected. This is valuable because data are collected for the young people throughout secondary school and before they sit their GCSE examinations.

4.1.1 The British Household Panel Survey

The BHPS was a nationally representative survey of adults in households living in Britain in 1991. The data collection was undertaken over 18 years (or waves) from 1991 to 2008. Researchers conducted annual interviews with approximately 10,000 adults living in 5,000 households from the age of 16 (Taylor et al., 2010). From 1994,

a youth panel was introduced for young people aged between 11 and 15 who lived in the households surveyed each year (Taylor et al., 2010). The BHPS is a very useful secondary data resource for researchers exploring social change or stability over time. The BHPS collected data on young people's GCSE results when they entered the adult survey. The household design means that the data provided by the young people can be matched with their parents' data. This is useful for the purposes of this thesis because comprehensive measures of parental socio-economic background are provided by the parents themselves, rather than reported by the children.

4.1.2 The UK Household Longitudinal Study

The UKHLS is a nationally representative survey of adults in households living in Britain from 2009/10. Approximately 40,000 households are interviewed per wave (Buck and McFall, 2011). Data collection takes place over a 24-month period (per wave) due to the much larger number of households interviewed compared with the BHPS. The larger sample of households, and individuals, enables researchers to study sub-groups more adequately, for example, young people in the households. The UKHLS has four main components: the General Population Sample, the Ethnic Minority Boost Sample, the BHPS Sample (i.e. households continuing from the BHPS study), and the Innovation Panel (Buck and McFall, 2011). Unlike in the BHPS, GCSE results are not reported within the survey itself for the first six waves. This is because the ambition from the outset of the survey was to link administrative education records to the data. The National Pupil Database (NPD) was linked to the UKHLS for young people in state schools in England up to the academic year 2012/13.

4.1.3 The National Pupil Database

Access to the linked UKHLS and NPD records are available through the UK Data Service using a secure access licence. Researchers must have SURE (Safe Users of Research data Environment) researcher training. The data available are based on consents given in Wave 1 of the UKHLS. The GCSE results contained within the linked NPD data cover the academic years 2001/02 to 2012/13 (University of Essex, 2015). The GCSE information includes the number of GCSEs attained and the grades

attained in each subject. The level of detail in the linked administrative data is greater than the measures collected in household panel surveys. It is anticipated that using official education records are likely to be more reliable than self-reported measures.

In accordance with the strict procedures and regulations, all analyses using the UKHLS-NPD data have been undertaken within the Secure Lab environment (see Chapters 2, 3, 4, 5, and 6). The UK Data Service provides Accredited Researchers with remote access to the Secure Lab server. It is not possible to download any data which require secure access, and work must be undertaken from specific institutional work stations or Safe Rooms which are logged on to the remote server.[4] All of the data management, analyses, and the writing of chapters using the UKHLS-NPD data have been undertaken within the Secure Lab environment. Outputs are only released from the Secure Lab after they have been checked to satisfy statistical disclosure control regulations, which are designed to ensure anonymity and confidentiality of participants. Where outputs do not satisfy this requirement, the detailed information is not reported in this thesis. This is represented by * in tables or information is aggregated, and only non-disclosive results are reported.

4.2 Complex Survey Design

The BHPS and UKHLS are complex social surveys. It is important to appropriately represent design and selection strategies when analysing complex surveys (see Longhi and Nandi, 2015). By default, statistical software packages assume that the data it is dealing with has been collected through simple random sampling (Longhi and Nandi, 2015). Conducting analyses of the BHPS and UKHLS without adjusting for the complex survey design would be a naïve approach to inferential data analysis (Gayle and Connelly, 2017). To fail to represent the complex survey design will negatively influence results (Treiman, 2009).

The BHPS had a two-stage stratified sample design. In stage one, primary sampling units (PSUs) of postcodes were identified. In stage two, systematic sampling was

[4] See https://www.ukdataservice.ac.uk/use-data/secure-lab.aspx for more details [accessed 24.09.19].

used to select addresses for interview (Taylor et al., 2010). When the BHPS started in 1991, the sample was representative of households in Britain south of the Caledonian Canal (Longhi and Nandi, 2015). Subsequent booster samples were implemented in the BHPS, for example regional boosters, which will not be considered in these analyses.

The UKHLS has a similar (but not identical) complex survey design with stratified, clustered, and equal probability selection of addresses both north and south of the Caledonian Canal (Buck and McFall, 2011). The primary sampling units of postcodes across Great Britain formed the initial stratified sample. These postal sectors were sampled systematically, with equal probability within each strata (Buck and McFall, 2011). The analyses in this thesis use the General Population Sample and Ethnic Minority Boost Sample, which are the key analytical samples.

Within longitudinal studies, non-response and missing data can take the form of unit non-response, wave non-response, and attrition (Hawkes and Plewis, 2006). Angrist and Pischke (2009) highlighted the difficulty and complexity of using sample weights in statistical analyses for even the most advanced researchers. Different survey weights are deposited with the BHPS and UKHLS datasets with useful guidance for choosing the most appropriate weight provided in Taylor et al. (2010) for the BHPS and Knies (2018) for the UKHLS. The primary aim of the weights is to account for the complex survey design and non-response adjustment (Buck and McFall, 2011). Specialist survey commands in statistical software packages, for example, the svy suite in Stata (see svy, StataCorp., 2017a) then make suitable adjustments for complex designs and sample selection (Treiman, 2009).

4.3 Methods

This PhD research has been supported by an ESRC Advanced Quantitative Methods enhanced stipend. The analyses undertaken in this thesis represent an amalgamation of three years of intensive methods training. All research questions examined throughout this thesis are addressed using quantitative methods. The following work employs newly-acquired skills in data management, using large-scale, complex,

social science surveys, using linked administrative records, applying survey weights and missing data methods, estimating standard generalised linear models, using quasi-variance estimates, estimating structural equation models and path models, and estimating panel data models. The methods used for each set of analyses are outlined in more detail within each chapter. All analyses have been undertaken using version 15 of the statistical software Stata (StataCorp., 2017b).

5 Structure of the Thesis

This thesis is organised into two parts. Part 1 presents analyses of the relationship between parental socio-economic background and children's school GCSE attainment. The main statistical methods used are generalised linear models using household panel surveys. Chapter 1 provides comprehensive evidence of the relationship between parental socio-economic background and children's school GCSE attainment in the 1990s and 2000s. Chapter 2 replicates the analyses for young people in the early 2010s. Particular attention is paid to sensitivity analyses of alternative measures and statistical model specifications. The role of missing data is examined in Chapter 3.

Part 2 empirically investigates key sociological mechanisms to better understand the enduring relationship between parental socio-economic background and children's educational attainment. Each chapter in Part 2 is devoted to a different mechanism: prior attainment, cultural capital, and educational aspirations. Chapter 4 models the roles of parental socio-economic background and school GCSE attainment incorporating prior attainment measures. The statistical method employed is path analysis in a structural equation modelling framework. Chapter 5 assesses the utility of the concept of cultural capital in explaining socio-economic background effects in school GCSE attainment. A range of candidate measures of cultural capital are tested in a series of sensitivity analyses. Chapter 6 examines the role of educational aspirations in GCSE attainment. The socio-economic gradient in educational aspirations is explored using longitudinal (i.e. panel) models to analyse data collected over the course of the young person's secondary schooling.

Part 1:

The Contemporary Relationship between Parental Socio-Economic Background and Children's School GCSE Attainment

Introduction to Part 1

The central sociological motivation of this thesis is to better understand the contemporary relationship between parental socio-economic background and children's school GCSE attainment. The first part of this thesis aims to explore the relationship for young people growing up in the 1990s, 2000s, and early 2010s.

Young people typically sit their GCSE examinations at the age of 15 or 16 in June each academic year, when they are in school Year 11. For the cohorts in the following analyses, the end of Year 11 also marked the end of compulsory schooling. The results of all GCSE examinations are typically received a few months later in August. September marks the start of the next academic year, and therefore the young people continuing to study will continue into school Year 12. Synthetic cohorts of English school Year 12 pupils are identified in the household panel surveys, because they will have recently received, and reported, their GCSE examination results. The young people are either in school Year 12, or are of equivalent age if they have not continued with non-compulsory education (i.e. young people aged 16 or 17). For simplicity, the synthetic cohorts will be referred to as English school Year 12 pupils throughout the following chapters.

Chapter 1 examines the role of parental socio-economic background in GCSE attainment using the British Household Panel Survey. A key methodological aspect of this work is the sensitivity analyses of socio-economic background variables and the functional form of the educational outcome variable. Particular attention is paid to checking the robustness of results with alternative measures and alternative statistical

model specifications. Chapter 2 aims to replicate the findings in Chapter 1 using the UK Household Longitudinal Study. Following the same analysis plan with more contemporary data, the empirical findings in Chapter 1 are tested for young people growing up in later cohorts, in the early 2010s. The analyses are then extended due to the greater granularity of GCSE attainment measures available in the National Pupil Database. Chapter 3 investigates the role of missing data in the preceding chapters. The analyses take a series of principled statistical approaches to help to address the potential problem of missing data in the synthetic cohort analyses.

Chapter 1
Synthetic Cohorts of Year 12 Pupils in the 1990s and 2000s: A Sensitivity Analysis

1 Introduction

General Certificate of Secondary Education (GCSE) examinations were introduced in the late 1980s for pupils at the end of compulsory schooling. The results of GCSE examinations can be important determinants of a young person's future education, employment, and earnings (Babb, 2005, Leckie and Goldstein, 2009, Croll, 2009, Playford and Gayle, 2015). GCSEs are consequential because they are necessary to continue to study at General Certificate of Education Advanced (A') Level. GCSE results can also be consequential for university entrance. Earlier research has found that GCSE attainment is stratified by socio-economic background, gender, and ethnicity (Drew, 1995, Demack et al., 2000, Sullivan, 2001, Connolly, 2006, Connelly et al., 2013, Gayle et al., 2014, Strand, 2014a, Playford and Gayle, 2015).

The analyses in this study build upon and extend previous empirical work by modelling parental socio-economic background and young people's GCSE attainment throughout the 1990s and 2000s. Particular attention is paid to the measurement of GCSE attainment. The analyses in this chapter use the BHPS to construct synthetic cohorts of young people in school Year 12 (or of equivalent age, i.e. aged 16 and 17) in England who have recently sat their GCSE examinations, and reported their results in the BHPS.

Undertaking social scientific research using statistical models requires important decisions to be made about the operationalisation of key social science variables (Stacey et al., 1969, Burgess, 1986, Connelly et al., 2016d). There are no single, agreed-upon measures for many of the key concepts routinely used in the social sciences, such as social class and education level (Crompton, 2008, Connelly et al., 2016a, Connelly et al., 2016b). Therefore, it may be possible to arrive at different substantive conclusions as a consequence of using different measures. Sensitivity

analyses can be understood as post-analysis robustness checks which re-estimate analyses using alternative measures or statistical model specifications (Connelly et al., 2016c, Freese and Peterson, 2017). A unique aspect of the following work is the sensitivity analyses of alternative measures of parental socio-economic background and alternative statistical model specifications to estimate measures of GCSE attainment.

Section 2 outlines alternative measures of socio-economic background and education attainment routinely employed in empirical research. Section 3 describes the synthetic cohorts in more detail and presents the descriptive statistics for the variables used in the subsequent analyses. Section 4 presents the modelling results and the sensitivity analyses of alternative measures. First, the independent variables of socio-economic background are compared using alternative measures of parental education and parental social class. Second, the outcome variable of GCSE attainment is explored using alternative measures, and accordingly, alternative statistical model specifications depending on the functional form of the outcome variable.

2 Operationalising Socio-Economic Background and Educational Attainment

Occupations have been central to sociological understandings of social stratification and social class (Blau and Duncan, 1967). The dominant, 'employment-aggregate' approach has been key to studying social class and stratification in empirical research in the post-war period (Crompton, 2008). An individual's position in the occupational structure strongly reflects differential material rewards, social standing (or prestige), and life chances (for example, in health and education) (see Parkin, 1971). There are a raft of existing, validated occupation-based measures for researchers to use and compare (Lambert and Bihagen, 2014). Detailed occupational information is routinely collected in large-scale social science surveys to facilitate the examination of social stratification using occupation-based measures (Connelly et al., 2016b).

The following sections provide an overview of the different occupational and educational measures often employed in social stratification and education research. These include social class schemas, stratification scales, education level, and

alternative measures of GCSE attainment. The list of measures is not intended to be exhaustive. Instead, the following summaries highlight a range of different measures commonly used and the review informs the subsequent sensitivity analyses in section 4.

2.1 Social Class Schemas

Social class has an enduring presence in British sociology (Crompton, 2008, Savage, 2016). Key sociological works have addressed issues of social class in post-war Britain, for example, in social mobility (Glass, 1954, Lockwood, 1958, Goldthorpe, 1969, Marshall, 1988, Devine, 1997, Cannadine, 1998) and educational opportunities (Jackson, 1962, Lacey, 1970, Willis, 1977, Halsey et al., 1980). In the 1980s and 1990s, the relevance of class analysis in modern society was questioned, and some heralded the 'death of class' (Pahl, 1989, Pakulski, 1996). Goldthorpe and Marshall (1992) launched a defence of class analysis which argued that class is useful as an analytical lens to view stratification, rather than as a deterministic concept in a Marxist sociological tradition. Crompton (2008) and Savage (2016) noted a revival in class analysis, with a shift in emphasis from class as an economic and deterministic concept in the structure, consciousness and agency debates to class as a cultural concept drawing on similarities in lifestyle and cultural or material consumption. Class remains a key concept in sociological studies in the 21st century (for example, Savage, 2015, Reay, 2017, Friedman and Laurison, 2019). Quantitative studies have employed varied measures of the class structure and this section provides an overview of the main social class schemas used in empirical research.

2.1.1 Registrar General Class Schema

The use of occupations in measuring socio-economic background has been a feature of categorising the British population since the 1911 Census, with information on occupation and industry routinely collected in government documentation (Rose et al., 1997). The first social class schema developed was the Registrar General class schema. This originally classified the British population into a five-fold schema based on indices of occupation and industry (Table 1.1) (Rose et al., 1997). A distinction

was made in category III between manual and non-manual work (Crompton, 2008). The schema can be collapsed to a three-category, hierarchical approach to classifying the social positions of British people. The primary purpose of the schema was to investigate why some sections of society had lower mortality rates than others (Szreter, 1984). Class schemas based on occupations have since been applied and developed for use in analyses of social stratification. The Registrar General class schema has largely been replaced by the Goldthorpe class schema and National Statistics Socio-Economic Classification schema in sociological and governmental work (Office for National Statistics, 2010).

Table 1. 1: Registrar General class schema

Class	Occupations
I	Professional occupations
II	Intermediate occupations
IIIN	Skilled non-manual occupations
IIIM	Skilled manual occupations
IV	Partly skilled occupations
V	Unskilled occupations

2.1.2 Goldthorpe Class Schema

The Goldthorpe class schema emerged from the Oxford Mobility Study in the 1970s. The schema categorised men into social classes based on their occupations (Goldthorpe et al., 1980). The theoretical foundations of the Goldthorpe class schema were comprehensively detailed in Erikson and Goldthorpe (1992). There are two elements to the Goldthorpe classification: employment status and employment contract. The employment status refers to whether the individual is an employer, self-employed, or an employee. The employment contract refers to whether the individual is under a labour contract or a service contract with their employer (Rose et al., 1997). Evans (1992) noted that the emphasis on employment position was more prominent in Erikson and Goldthorpe's 1992 study compared to the original emphasis on work and market situation in the initial Oxford Mobility Study. The resultant classification is an 11-fold schema, with several possible collapses (Table 1.2). Academic or government work tends to employ aggregated versions of the schema. In the

Goldthorpe class schema, the social classes are not considered hierarchical unless collapsed to a three-category version (i.e. salariat, intermediate, and working class positions) (Goldthorpe et al., 1980).

Table 1. 2: Goldthorpe class schema and collapses

Full version Eleven class		**Collapsed versions Seven class**	**Three class**
I	Higher-grade professionals, administrators and officials; managers in large industrial establishments, large proprietors	Service class	Non-manual workers
II	Lower-grade professionals, administrators and officials; higher-grade technicians; managers in small industrial establishments; supervisors of non-manual employees		
IIIa	Routine non-manual employees; higher-grade administration and commerce	Routine non-manual workers	
IIIb	Routine non-manual employees; lower-grade sales and services		
IVa	Small proprietors, artisans etc with employees	Petty bourgeoisie	
IVb	Small proprietors, artisans etc without employees		
IVc	Farmers and smallholders; other self-employed workers in primary production	Farmers	Farm workers
V	Lower-grade technicians, supervisors of manual workers	Skilled workers	Manual workers
VI	Skilled manual workers		
VIIa	Semi and unskilled manual workers (not in agriculture)	Non-skilled workers	
VIIc	Agricultural and other workers in primary production	Agricultural labourers	

2.1.3 National Statistics Socio-Economic Classification

The Goldthorpe class schema has been further developed by the UK Office for National Statistics (ONS) as the National Statistics Socio-Economic Classification (NS-SEC) (Office for National Statistics, 2010). The classification system of the NS-SEC is based on the Weberian concept of market situation and life chances derived from occupational position (Crompton, 2008). There are 14 functional classes for

those in employment and residual classifications for those not in employment, such as students and unclassifiable occupations (Rose and Pevalin, 2003). NS-SEC draws on similar methodology to the Goldthorpe class schema, basing categories on employment status and employment contracts (Office for National Statistics, 2010). There are several operational collapses, including eight-fold and three-fold collapses detailed in Table 1.3 (Crompton, 2008). The ONS stated that NS-SEC has clear conceptual underpinnings and is well-constructed and validated to reflect socio-economic differences in society (Office for National Statistics, 2010). NS-SEC is the official classification used in government reports and has been widely used in social scientific research.

Table 1. 3: NS-SEC analytic classes and collapses

Full Version Fourteen class		**Collapsed Versions Eight class**		**Three class**
1	Employers in large establishments	1.1	Large employers and higher managerial	Managerial and professional
2	Higher managerial and administrative			
3	Higher professional	1.2	Higher professional	
4	Lower professional and higher technical	2	Lower managerial and professional	
5	Lower managerial and administrative			
6	Higher supervisory			
7	Intermediate	3	Intermediate	Intermediate
8	Employers in small organisations	4	Small employers and own account workers	
9	Own account workers			
10	Lower supervisory	5	Lower supervisory and technical	Routine and manual
11	Lower technical			
12	Semi-routine	6	Semi-routine	
13	Routine	7	Routine	
14	Never worked/ long-term unemployed	*	Never worked/ long-term unemployed	* Never worked/ long-term unemployed

Table 1.4 presents the 8-category NS-SEC measure and example occupations in each corresponding category. NS-SEC is the preferred measure of social class for the current work, and this table can serve as a reference tool for the analyses throughout this thesis. The reference category for all statistical models in the following analyses is NS-SEC 1.2, i.e. the children of parents with higher professional occupations, such as doctors, lawyers or university lecturers. Throughout this thesis,

the GCSE attainment of young people from NS-SEC 1.2 social class backgrounds are compared with the outcomes for children of, for example, chief executives of large firms (NS-SEC 1.1), teachers or nurses (NS-SEC 2), police officers, paramedics or secretaries (NS-SEC 3), owners of small firms, childminders or taxi drivers (NS-SEC 4), mechanics, electricians or chefs (NS-SEC 5), receptionists, care workers or retail assistants (NS-SEC 6), and hairdressers, bus drivers or cleaners (NS-SEC 7).

Table 1. 4: Example occupations by category of NS-SEC

Parental NS-SEC	**Example Occupations**
1.1 Large employers and higher managerial occupations	Chief executives Managers and directors in finance, mining and energy, marketing and sales, HR Senior officers, e.g. in police, fire, ambulance, prisons, councils
1.2 Higher professional occupations	Higher education professionals, e.g. lecturers Pharmacists, dentists, doctors, veterinarians Judges, solicitors, barristers
2 Lower managerial and professional occupations	Managers in retail and wholesale, leisure and sports, health care practice, residential care Opticians, physiotherapists, nurses, midwives Primary and secondary school teachers
3 Intermediate occupations	Electronic and electrician technicians, draughtsperson Paramedics, pharmaceutical, dental and medical technicians Teaching assistants, nursing assistants, ambulance staff
4 Small employers and own account workers	Managers and proprietors in agriculture and horticulture, shopkeepers, publicans Bricklayers, masons, roofers, tilers, carpenters, joiners, plasterers, painter and decorators Taxi drivers, chauffeurs, driving instructors
5 Lower supervisory and technical occupations	Tool makers and fitters, metal working production Mechanics, electricians, plumbers Skilled constructed and building trades supervisors
6 Semi-routine occupations	Dental nurses, veterinary nurses, pharmacy dispensing assistants Sales and retail assistants, retail cashiers Process operatives in food, drink, tobacco, glass and ceramics, rubber, plastics
7 Routine occupations	Metal plate workers, welding trades, pipe fitters Hairdressers and barbers Cleaners, launderers, dry cleaners, vehicle valets

2.1.4 Micro-Class Analysis

A more recent development in social class analysis is the micro-class approach. Micro-class analysis uses occupational data to better explain differences in life

chances, patterns of behaviour or differential attitudes, than at the level of the big or agglomerate classes (Grusky and Weeden, 2001, Weeden and Grusky, 2004). Micro-class approaches have been successfully applied in social mobility research (for example, Jonsson et al., 2009). There may be broader applications, for example, in explaining educational outcomes. It is plausible that parental skills and knowledge in particular micro-classes (characterised by specific occupations and industries) may be reproduced across generations at the occupation-level, rather than at the agglomerate class level (Connelly and Gayle, 2016). Erikson et al. (2012) suggested that a micro-class approach might be useful to understand immobility (i.e. why children follow parents into specific occupations), but it is less successful in explaining social mobility out of the micro-classes. There are practical challenges associated with adopting the micro-class approach, for example, in many datasets there may be sparse information in some occupational categories to develop and apply micro-class measures, and there may be associated challenges of sample size and statistical power in analyses (Connelly et al., 2016b).

2.1.5 'New Model' of Social Class

Moving away from an employment-aggregate approach, Savage et al. (2013) constructed a 'new model' of social class based on the concept of capitals, assets, and resources (also see Crompton, 2008). Savage et al. (2013) argued that occupation-based measures of social class, such as NS-SEC, distinguish primarily between individuals in routine or semi-routine occupations with 'labour contracts' and individuals in professional and managerial occupations with 'service contracts'. The authors argued that occupation-based measures are primarily an economic definition of social class, and do not reflect broader social and cultural processes which distinguish social classes, for example lifestyles and social networks (Savage et al., 2013). The 'new model' of social class borrows Pierre Bourdieu's theoretical work and presents economic, social, and cultural capital as multi-dimensional aspects of social class.

The 'new model' of social class further distinguishes between 'highbrow taste' (i.e. a conventional Bourdieusian interpretation of cultural capital), and 'emerging' cultural

capital, based on popular consumption activities around sport, information technology, and contemporary music (Savage et al., 2013). Mills (2014) questioned the validity of 'emerging' cultural capital, which tends to capture age and cohort, rather than social class, differences. Mills (2014) further emphasised the role of age (or 'life-cycle') differences between many of the new social class categories, for example comparing the age distributions of the 'elite' compared with the 'established middle class'. Lambert and Griffiths (2013) similarly identified age and region as key criteria influencing the new class categories, which are not reflected on by Savage and colleagues.

The class schema was constructed using latent class analysis of a set of capitals, assets, and resources measures in the Great British Class Survey dataset (Savage et al., 2013). Seven new class categories were identified: the elite, established middle class, technical middle class, new affluent workers, traditional working class, emergent service workers, and the precariat (Savage et al., 2013). The new schema attracted much critical reflection, for example, on measurement, sample selection bias, model selection, and classifications (for example, see Bradley, 2014, Mills, 2014). Payne (2013a) noted the analytical and sociological usefulness of thinking about class in these terms. He argued, however, that the new class scheme does not address a recurring problem in class analysis of classifying the intermediate classes. Payne (2013a) further suggested that the new model of social class offers a detailed view of class in 21st century Britain, but that for pragmatic reasons, the conventional NS-SEC schema is more suitable in empirical analyses.

A further challenge is that large-scale, multi-purpose surveys do not routinely collect the detailed level of information required to construct a capitals, assets, and resources measure of social class. Connelly et al. (2019) undertook a principled attempt to replicate the 'new model' using the UKHLS. The authors noted methodological challenges, for example, the results are likely to be sensitive to the manifest variables available to use. This has consequences for comparability of measures across datasets and over time. Their latent class analysis model resulted in three large class categories which did not fully replicate the new class categories in Savage et al. (2013). Furthermore, there is not convincing evidence that a capitals, assets, and

resources measure improves the explanatory power or theoretical understanding of social class compared with the existing NS-SEC social class schema (Connelly et al., 2019).

2.2 Social Stratification Scales

Social stratification scales, such as social distance or interaction approaches, are alternative means of measuring social inequalities (Bottero and Prandy, 2003). Occupational prestige scales, the Cambridge scale, and CAMSIS are outlined in the sub-sections below.

2.2.1 Occupational Prestige Scales

In Britain, Hope and Goldthorpe (1974) developed an occupational prestige scale of the general desirability of occupations using data in the Oxford Mobility Study. The authors noted that there might be broader applications of the scale in other studies (Hope and Goldthorpe, 1974). International versions of occupational prestige scales have been developed in the form of reputational approaches. For example, Treiman's Standard International Occupational Prestige Scale compared occupational rankings across 60 countries (Treiman, 1977). The International Socio-Economic Index of occupational status (ISEI) was developed using measures of occupational prestige alongside education and income for men in 16 countries (Ganzeboom et al., 1992). Ganzeboom and Treiman (1996) detailed the complementarity of occupational status across country-specific scales using the 1988 International Standard Classification of Occupations. This creates a level of international standardisation, and better facilitates cross-national comparisons. Coxon and Jones (1979) critiqued the use of occupational prestige scales because of the assumptions made about the homogeneity of social desirability when respondents are invited to rank or rate occupations. Stewart et al. (1980) argued that the scales assume that all respondents are able to make an identical subjective judgement about general desirability of occupations. More recently, Ganzeboom (2019) re-investigated and reaffirmed the idea of the Treiman constant in empirical work. The Treiman constant is the

theoretical idea that occupational prestige rankings remain constant between different countries and over time (Hout and Diprete, 2006).

2.2.2 Cambridge Scale and CAMSIS

Stewart et al. (1980) constructed the Cambridge scale to measure advantage based on social associations. The original Cambridge scale was devised based on a study of male workers in the Cambridge area in 1918 (Stewart et al., 1980). The scale was constructed as a measure of associational lifestyles to highlight social and material advantage across social groups. The scale measures social distances between groups in an inductive, rather than deductive, approach to social stratification (Bottero, 2005). The original Cambridge scale was based on pairwise matching of a respondent's four closest friendships and multidimensional scaling was used to generate a social distance score (Stewart et al., 1973). Prandy (1999) asserted that the Cambridge scale is preferable to categorical class measures because of the closer affinity to measuring the underlying hierarchy in social relations. The scale has since been replicated and applied to different samples (Prandy, 1999). Prandy and Lambert (2003) noted that the Cambridge Scale, based on a specific historical sample, was at risk of being out of date. They further developed and updated the scale using 1971 census data to measure marriage as well as friendships under the generic term of CAMSIS (Prandy and Lambert, 2003). CAMSIS is based on occupational information and how this relates to social networks. It is therefore claimed that it is not constrained to a structuralist account of social stratification like class schemas, with categories chosen *a priori* (Bergman and Joye, 2005, Bottero, 2005). The scale can be extended to apply to different contexts and countries (Prandy and Lambert, 2003).

2.3 Unit of Analysis Problem

In the late 1960s and early 1970s, sociologists noted the empirical challenge of including women in class analysis (Stacey 1969, Acker 1973). Acker (1973) highlighted the intellectual sexism in social stratification research that the conventional focus on male occupations represented. This may be thought of as a

unit of analysis problem in modern British households. The problem relates to the conventional approach of using father's social class to represent the family and assuming that women's positions are equal to her partner's. Addressing the unit of analysis problem became a topic of debate in class analysis from the mid-1980s (Goldthorpe, 1983, Erikson, 1984, Goldthorpe, 1984, Heath and Britten, 1984, Stanworth, 1984, Sorensen, 1994). Goldthorpe (1983) argued that the conventional approach was appropriate. This involved using the head of the household's (i.e. father's) occupation as the unit of class analysis. Heath and Britten (1984) argued against the conventional view, particularly as female employment, and married women's employment, was increasing. Erikson (1984) argued in defence of a dominance approach, which assigned the social class of the household to the person with the highest occupational position and with the longest working hours, regardless of their gender. Erikson (1984) conceded that in practice this was most likely to be a male head of household.

Sorensen (1994) conducted a review of the empirical evidence using both the conventional view and a joint classification. Sorensen (1994) concluded that neither approach was inherently more appropriate than the other. Korupp et al. (2002) tested five approaches to the unit of class analysis on data from the Netherlands, West Germany and the USA. These models used a dominance approach, a modified dominance approach, a joint classification model, a sex-role model, and an individual model. The authors found that the best fit of the data was a modified dominance approach, which used information on the non-dominant parent as well as the dominant parent (Korupp et al., 2002). There is not a single appropriate way of measuring parental socio-economic background (Sorensen, 1994, Beller, 2009). A dominance approach to the NS-SEC schema will be taken in this thesis following the clear guidance from Rose and Pevalin (2003).

2.4 Measuring Education Level

Measuring education level can be challenging in social research (Lambert, 2012). One aspect of this thesis is to appropriately measure parental education level. The

following sub-sections detail two prominent ways of measuring highest education level: years of education (or duration) and highest qualification.

2.4.1 Years of Education

The years spent in education is a common measure of education level in economic research and it can reflect the returns to education in the form of human capital (Connelly et al., 2016a). Years of education can be easily compared between cohorts and can indicate whether or not the respondent continued in education past the compulsory school leaving age. The successive raising of the school leaving age in the UK will affect easy comparisons across generations. Duration measures can also be compared across educational contexts and are strongly correlated with other measures of education level, such as categorical measures (Schröder and Ganzeboom, 2014). Measuring the number of years can obscure more specific educational attainment, for example, levels of education completed and grades attained. This can be partially addressed by scaling methods to combine information on years spent in education and the time taken to achieve different levels within the education system (Schröder and Ganzeboom, 2014).

2.4.2 Highest Education Qualification

Qualifications signify not only an educational transition, but successful completion and certification at a certain level. Qualifications tend to be ordinal in nature. In the UK, the National Qualifications Framework separates and orders levels of qualifications into general (i.e. academic), vocationally-related, and occupational categories with internal hierarchies of achievement (Jenkins and Sabates, 2007). Qualifications are often country-specific. International measures have been developed to aid cross-country comparisons, for example, the Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) and the International Standard Classification of Education (ISCED). CASMIN is constructed based on a hierarchy of education level and a distinction between general and vocational education (Brauns et al., 2003). This has been described as providing vertical and horizontal distinctions (Schneider, 2011). There are nine categories in the CASMIN framework (Table 1.5).

Table 1. 5: CASMIN framework

Category	Description
1a	Inadequately completed general elementary education
1b	General elementary education
1c	Basic vocational qualification or general elementary education
2a	Intermediate vocational qualification or intermediate general education plus basic vocational qualification
2b	Intermediate general qualification
2c voc	Full vocational maturity certificate
2c gen	Full general maturity certificate
3a	Lower tertiary certificate
3b	Higher tertiary certificate

By contrast, ISCED recognises formal and non-formal education and measures the level of education completed, which is usually accompanied by a form of qualification (UNESCO, 2012). There are nine categories in the 2011 ISCED framework (Table 1.6).

Table 1. 6: ISCED framework

Category	Description
ISCED 0	Early childhood education
ISCED 1	Primary education
ISCED 2	Lower secondary education
ISCED 3	Upper secondary education
ISCED 4	Post-secondary non-tertiary education
ISCED 5	Short-cycle tertiary education
ISCED 6	Bachelor's or equivalent level
ISCED 7	Master's or equivalent level
ISCED 8	Doctoral or equivalent level

2.5 Measuring School GCSE Attainment

There is a substantial amount of empirical work focused on formal qualifications taken at the end of compulsory schooling, particularly on GCSEs in England and Wales (Leckie and Goldstein, 2009). GCSEs are an important branching point in a young person's education (Murray et al., 2012, Playford and Gayle, 2015). The grades achieved in GCSE examinations tend to be determinants of further education and employment opportunities (Croll, 2009).

There is no single agreed-upon way to measure GCSE attainment in social science research (Connelly et al., 2016a). This is because there are a number of different combinations of GCSE examinations that any one pupil could sit, and because the grading system is alphabetised rather than numeric (from grades A* to G). A common measure is the binary outcome variable of whether the respondent attained 5 GCSEs at grades A*-C. This was the standard attainment benchmark in UK education policy (Gayle et al., 2003, Connolly, 2006, Leckie and Goldstein, 2009). Previous empirical studies have demonstrated a clear effect of social class on the likelihood of attaining 5 GCSEs at grades A*-C, often over and above the effects of gender and ethnicity (for example, Demack et al., 2000, Connolly, 2006). Connelly et al. (2013) used categorical measures to examine those with middle attainment, i.e. the achievement of 1-4 A*-Cs. Sullivan et al. (2011) examined low, medium, and high attainment, measured as 1 or more A*-Cs, 5 or more A*-Cs, and 8 or more A*-Cs. Gorard and Taylor (2002: 7) noted the challenge of equivalence with the 5 or more A*-C benchmark measure, which treats an A* in Music, a B in Physics, and a C in Sociology as equivalent.

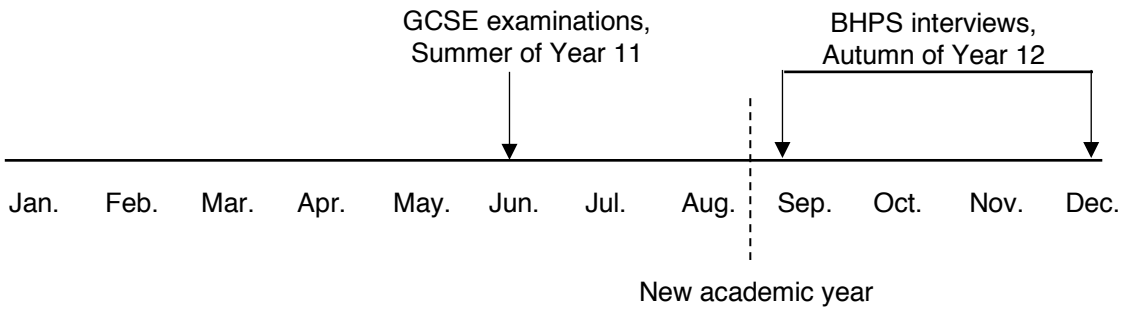
Some sociologists have constructed continuous measures of GCSE scores based on the grades attained in separate subjects. Demack et al. (2000) and Connolly (2006) constructed GCSE scores by assigning points to different grades. Demack et al. (2000) assigned 7 points to a grade A* or A, 6 points to a grade B, and so on. Connolly (2006) assigned 8 points to an A* grade, 7 points to an A grade, and so on. More innovative approaches to measuring GCSE attainment include latent class analysis of school subjects in the Youth Cohort Study of England and Wales (Playford and Gayle, 2015). These approaches rely on datasets with a greater granularity or quality of information than is routinely collected in multi-purpose, large scale datasets.

The preceding sections have detailed a variety of different measures routinely used in social scientific research to measure socio-economic background and educational attainment. There is no consensus on the most suitable and appropriate measures for the key variables which are used in the subsequent analyses. The next sections explore the variety of measures available in the BHPS and test the robustness of results when using many of the different measures reviewed in the sections above.

3 Synthetic Cohorts of English Year 12 Pupils in the BHPS

The units of analysis in these sections are synthetic cohorts of English Year 12 pupils in BHPS households. The synthetic cohorts include the 16 or 17 year olds taking part in the full adult interview in the BHPS wave following their GCSE examinations. The respondents will have sat their GCSE examinations in the summer of school Year 11 (i.e. when aged 15 or 16). They will have reported their results in the adult interview of the BHPS in the following Autumn (see Figure 1.1). The synthetic cohorts are termed 'school Year 12 pupils', but the identification of these individuals does not distinguish between the individuals who have continued with non-compulsory education and those who have left education.

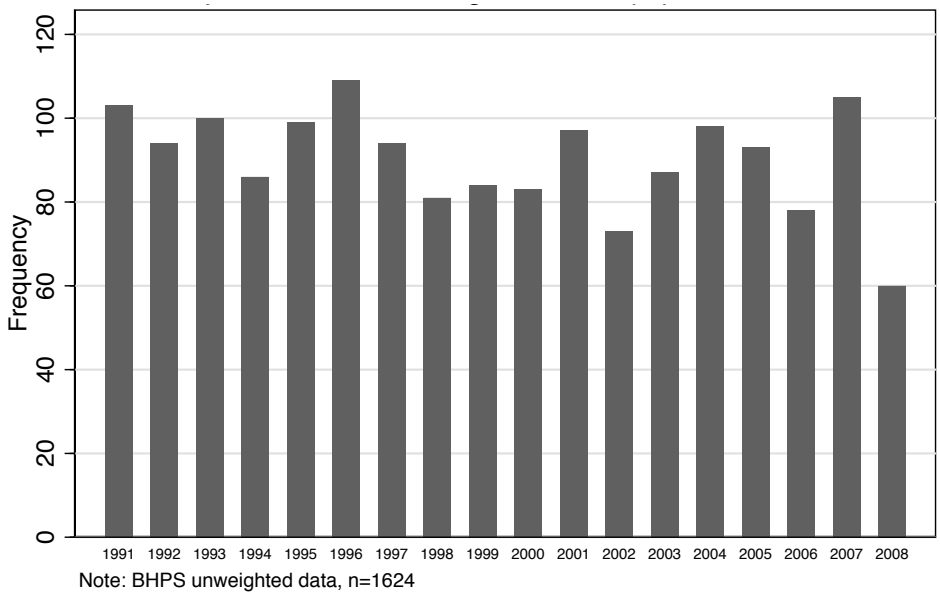
Figure 1. 1: Synthetic cohorts of Year 12 pupils in BHPS households



Synthetic school year cohorts were developed by grouping young people born in the same academic year, according to their birth months and years.[5] The young people's data were then linked to their mothers' and fathers' data. The school year cohorts were appended together across all 18 waves of the BHPS. The analytical sample of synthetic cohorts of English Year 12 pupils with complete records comprised of 1624 individuals (Figure 1.2). Carpenter and Kenward (2013) strongly advised that the first stage in statistical analyses should be to conduct a complete records analysis. Where possible, this should be followed by exploring the effects of missing data. Approaches to handling missing data are explored in Chapter 3.

[5] Pupils will be in the same academic year in England if they were aged 16 and born between January and August, or if they were aged 17 and born between September and December.

Figure 1. 2: Number of observations of synthetic cohorts of Year 12 pupils per wave



Note: BHPS unweighted data, n=1624

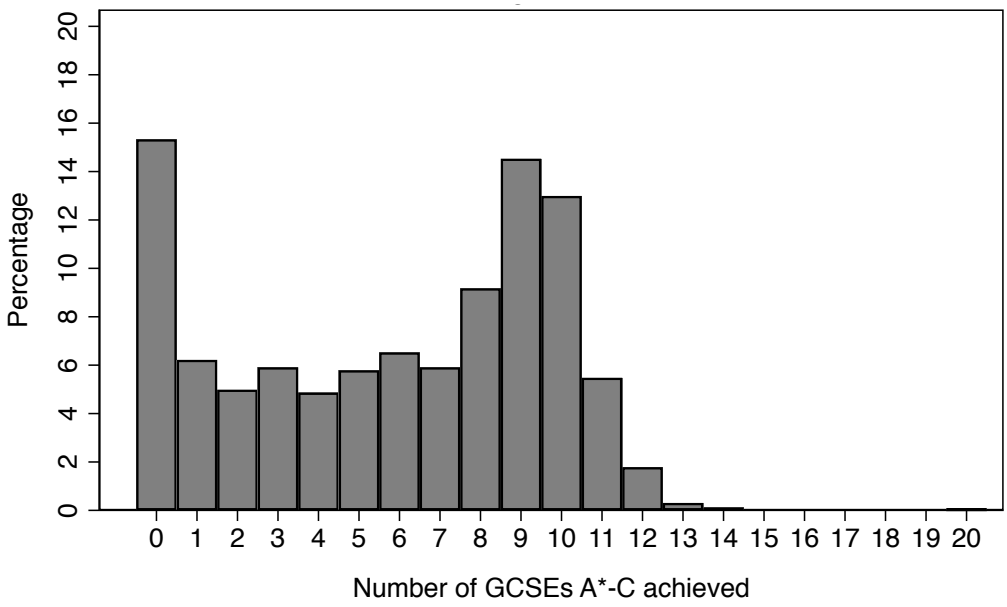
4 Descriptive Statistics

The main outcome variable of interest is school GCSE attainment. The focus of this work is on GCSEs at grades A*-C because they are conventionally considered good passes. The main measure of GCSE attainment is presented in Figure 1.3. The measure uses information on the young person's GCSE passes at grades A*-C (i.e. good passes) and at grades D-G (i.e. lower-level passes). In these analyses, individuals with zero A*-Cs have at least one GCSE at grades D-G. This is important to ensure that the measure does not conflate those who did not achieve the higher grades with those who did not sit the examinations at all. Figure 1.3 illustrates that there is a clear spike at zero, as 15.3% of respondents do not have any GCSEs at grades A*-C but at least one GCSE at grades D-G. There is also a distinctive spike around the attainment of 9 or 10 GCSEs at grades A*-C, suggesting that these are generally high achieving synthetic cohorts.[6]

[6] There is a potential outlier with one individual reporting 20 GCSEs A*-C within the space of one wave, which seems highly unlikely and may be a reporting or data entry error. In the next wave the individual has a highest education level of A' Level or equivalent, having achieved 3 A' Levels at grades C-E. In subsequent waves, the respondent dropped out of the survey and so cannot be further followed up. The A' Level examinations usually require at least 5 Cs at GCSE level to progress. Kohler and Kreuter (2012) argued strongly against removing

The national standard policy benchmark of 5 or more GCSEs at grades A*-C is used as the initial outcome variable (Gayle et al., 2003, Connolly, 2006, Leckie and Goldstein, 2009). This is also the baseline measure of GCSE attainment used when testing different measures of socio-economic background in section 4.1. The logistic regression model includes independent variables for social class background, parental education level, household tenure, gender, and ethnicity. Wave year is included as a control variable for change in GCSE attainment over time. Variations in the operationalisation of GCSE attainment are tested in section 4.2.

Figure 1. 3: Number of GCSEs attained at grades A*-C for synthetic cohorts of Year 12 pupils



Note: BHPS unweighted data, n=1624

The measures of socio-economic background in the analyses are parental social class and parental education level. Parental social class is measured using an 8-category measure of parental NS-SEC. Parental education level is measured by highest parental education qualification. The categories are higher education, further education, school-level (including O' Level and A' Level),[7] and below school-level (i.e.

observations from the dataset without strong evidence to do so and, as the logistic regression diagnostics for influential cases did not highlight this case as a potential concern, the data point remains in the analyses.

[7] General Certificate of Education (GCE) Ordinary (O') Levels were examinations taken at the end of compulsory schooling before the late 1980s when they were replaced by GCSEs. GCE

less than an O' Level pass) (Lambert, 2012). Socio-economic background measures are tested in sensitivity analyses below as robustness checks.

The analyses use a dominance approach for the measures of parental socio-economic background. For NS-SEC, the dominance approach distinguishes between two concepts: work position and class position (Rose and Pevalin, 2003). Following Erikson (1984), the class position of a family or household can be identified as a function of individuals' work positions, i.e. the work position of the more dominant individual. Dominance can be determined based on the life chances associated with employment and occupational relations (Rose and Pevalin, 2003) or, more specifically, the work position which is likely to have the greater influence on 'ideology, attitudes, behaviour and consumption patterns of the family members' (Erikson, 1984: 504). Rose and Pevalin (2003: 23) noted that 'higher qualifications dominate lower ones; non-manual work dominates manual work; self-employment dominates being employed; employers dominate own account workers; and professional work dominates all other forms of work'.

In practice, the social class of the household is represented by the parent with the higher social class position and with the longer working hours (Erikson 1984). The NS-SEC schema does not follow a strict hierarchical structure to easily identify the higher social class position. The guidance for NS-SEC in Rose and Pevalin (2003) suggests that the ordering for dominance is NS-SEC 1.2, 1.1, 4, 2, 3, 5, 6, 7. A dominance approach for parental education level is more straight-forward, i.e. using the highest education qualification of mother or father.

Other independent variables included in the models are housing tenure, gender, and ethnicity. Housing tenure has been included in previous analyses of educational attainment (Connelly et al., 2013, Playford and Gayle, 2015). In the following analyses, tenure is a measure of whether the individual lives in an owned or privately rented home, or in social housing. Demographic factors like gender and ethnicity have been found to have strong influences on educational attainment, whereby girls tend

Advanced (A') Levels are the examinations taken at the age of 17 or 18, historically after compulsory secondary school.

to outperform boys, and those from Indian or Chinese backgrounds tend to outperform children from all other backgrounds (Drew, 1995, Demack et al., 2000, Connolly, 2006, Platt, 2010, Sullivan et al., 2011, Connelly et al., 2013, Strand, 2014a). Studying ethnicity in the following analyses is challenging because of the low coverage of ethnic minority groups in the BHPS. Over 95% of the synthetic cohort sample are from white backgrounds. Only 69 individuals (4.3% of the sample) reported that they are from Black Caribbean, Black African, Indian, Pakistani, Bangladeshi, Chinese, mixed, or other ethnic backgrounds. The sample sizes for some of the disaggregated ethnicity groups are very small (for example, many have five or fewer individuals). For this reason, the resultant variable is parameterised as white and non-white. The parameterisation is highly restrictive and makes a very unrealistic assumption of within-group homogeneity. For this reason, ethnicity is included in the models as a control variable, rather than a variable which can be suitably interpreted.

Table 1.7 shows the frequencies and summary statistics of the synthetic cohorts of English Year 12 pupils in the BHPS. Overall, the mean number of GCSEs attained at grades A*-C for the synthetic cohort sample is 5.86 with a standard deviation of 3.85. In the sample, 51% are male and 49% are female with girls attaining a higher mean number of GCSEs at grades A*-C than boys. A large majority of the synthetic cohort sample live in owned or privately rented accommodation compared with social housing. Individuals living in owned or privately rented accommodation have a higher mean number of GCSEs attained at grades A*-C than individuals living in social housing. Those with parents in NS-SEC social class 1.1 and 1.2 have a higher mean number of GCSEs A*-C (7 and 8 GCSEs A*-C respectively) than those in all other social classes. There is a general negative gradient in the mean number of GCSEs attained at grades A*-C by parental social class. Those with parents with higher education qualifications have the highest mean number of GCSEs attained at grades A*-C and there is a negative gradient for those whose parents have lower education levels. The summary statistics suggest that there is an association between parental socio-economic background and the number of GCSEs that children attain at grades A*-C.

Table 1. 7: Descriptive statistics of explanatory variables for synthetic cohorts of English Year 12 pupils in BHPS households, unweighted

	Freq.	Percent	Mean A*-Cs	Std. Dev.
Parental Education Level				
Higher education	290	17.86	8.14	3.01
Further education	658	40.52	6.19	3.76
School-level education	506	31.16	5.05	3.72
Below school-level education	170	10.47	3.07	3.40
Parental NS-SEC				
1.1 Large employers and higher managerial	99	6.10	7.23	3.63
1.2 Higher professional occupations	154	9.48	8.08	2.96
2 Lower managerial and professional occupations	463	28.51	6.61	3.59
3 Intermediate occupations	176	10.84	6.03	3.75
4 Small employers and own account workers	199	12.25	5.30	3.80
5 Lower supervisory and technical occupations	172	10.59	4.75	3.86
6 Semi-routine occupations	207	12.75	4.46	3.88
7 Routine occupations	154	9.48	4.10	3.73
Housing Tenure				
Owned or privately rented	1440	88.67	6.20	3.73
Social housing	184	11.33	3.14	3.69
Gender				
Male	821	50.55	5.33	3.90
Female	803	49.45	6.40	3.73
Ethnicity				
White	1555	95.75	5.82	3.86
Non-white	69	4.25	6.59	3.66
Total *n*	**1624**	**100.00**	**5.86**	**3.85**

5 Modelling School GCSE Attainment

The main outcome variable is the attainment of 5 or more GCSEs at grades A*-C. The first set of analyses estimate a logistic regression model. Table 1.8 details the deviance, change in deviance, change in degrees of freedom, and the McFadden's Adjusted Pseudo R^2 measure to compare the null model (i.e. a model with no explanatory variables) with models with one explanatory variable. There are a number of Pseudo R^2 measures available (Smithson, 2003). In the modelling results throughout this thesis, several different Pseudo R^2 measures are presented to acknowledge that there is not a superior measure, but at this stage brevity of output has been preferred.

Table 1. 8: Goodness-of-fit summaries for explanatory variables and attainment of 5+ GCSEs at grades A*-C

Outcome Variable: 5+ GCSEs A*-C	**Deviance**	**Δ Deviance** (from Null)	**Δ d.f.** (from Null)	**McFadden's Adjusted Pseudo R^2**
Null model	2145.669	-	-	-
Null + Parental Education	1973.435	172.233	3	0.077
Null + Parental NS-SEC	2020.152	125.517	7	0.051
Null + Housing Tenure	2056.513	89.156	1	0.040
Null + Gender	2124.614	21.055	1	0.008
Null + Ethnicity	2141.584	4.085	1	0.000
Null + Wave Year	2099.750	45.919	17	0.005

Explanatory variables are entered sequentially in the logistic regression model (Gayle et al., 2009). Table 1.9 details the deviance, change in deviance, change in degrees of freedom, and the McFadden's Adjusted Pseudo R^2 measure during the model building process. Ethnicity and wave year are included as control variables in the subsequent analyses. The logistic regression diagnostic tests demonstrate that there is evidence of homoscedasticity, that the model is correctly specified, that there is no evidence of multicollinearity between the explanatory variables, and there are no influential cases worthy of further investigation (Kohler and Kreuter, 2012, Mehmetoglu and Jakobsen, 2017). There are no significant interactions between the explanatory variables.

Table 1. 9: Model building goodness-of-fit summaries for logistic regression model of attainment of 5+ GCSEs at grades A*-C

Outcome Variable: 5+ GCSEs A*-C	**Deviance**	**Δ Deviance** (from previous model)	**Δ d.f.** (from previous model)	**McFadden's Adjusted Pseudo R^2**
Null model	2145.669	-	-	-
Null + Parental Education	1973.435	172.233	3	0.077
Null + Parental Education + Parental NS-SEC	1936.643	36.792	7	0.087
Null + Parental Education + Parental NS-SEC + Housing Tenure	1905.757	30.886	1	0.101
Null + Parental Education + Parental NS-SEC + Housing Tenure + Gender	1879.527	26.230	1	0.112
Null + Parental Education + Parental NS-SEC + Housing Tenure + Gender + Ethnicity	1871.452	8.075	1	0.115
Null + Parental Education + Parental NS-SEC + Housing Tenure + Gender + Ethnicity + Wave Year	1836.439	35.013	17	0.115

The results of the logistic regression model are reported in Table 1.10. The presentation of the model results follow the useful guidance provided in Connelly et al. (2016d). In the conventional reporting of regression output, it is not possible for the reader to ascertain significance of parameters other than in relation to the reference category using the logistic regression coefficients and associated standard errors (Firth, 2003). Quasi-variances are presented to help to address the reference category problem by providing quasi-standard errors and 95% comparison intervals, which enable the reader to compare all contrasts of the categorical variables (Gayle and Lambert, 2007).

The output demonstrates that young people whose parents have higher education qualifications have significantly higher log odds of reaching the national benchmark of GCSE attainment compared with those whose parents have any lower level of education. The 95% quasi-variance comparison intervals demonstrate that there are not significant differences in the GCSE attainment of children whose parents have completed further education compared with parents who have school-level education. There is a general negative gradient in GCSE attainment across all NS-SEC categories. Young people whose parents have occupations in NS-SEC 1.2 have

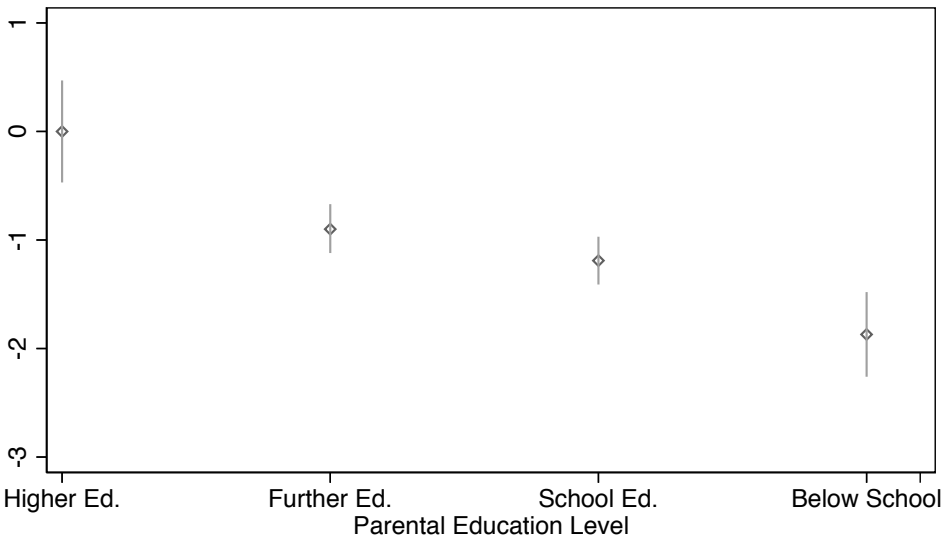
significantly higher log odds of attaining 5 or more GCSEs at grades A*-C compared with those whose parents have occupations in NS-SECs 4, 5, 6, and 7. The 95% quasi-variance comparison intervals demonstrate that neighbouring categories do not tend to be significantly different from each other. The quasi-variance comparisons for all categories of parental education and parental NS-SEC are visualised in Figures 1.4 and 1.5. Young people living in owned or privately rented homes have significantly higher log odds of attaining 5 or more GCSEs at grades A*-C than those living in social housing, and females have significantly higher log odds of attaining 5 or more GCSEs at grades A*-C than males.

Table 1. 10: Logistic regression model of the attainment of 5+ GCSEs at grades A*-C

				Quasi-Variance		
Logistic Regression: 5+ GCSEs A*-C	**Coef.**	**S.E.**		**S.E.**	**LCI**	**UCI**
Parental Education Level						
Higher education	Ref.	(.)		0.23	-0.47	0.47
Further education	-0.90	(0.24)	***	0.11	-1.12	-0.67
School-level education	-1.19	(0.26)	***	0.11	-1.41	-0.97
Below school-level education	-1.87	(0.32)	***	0.19	-2.26	-1.48
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.51	(0.36)		0.24	-1.00	-0.01
1.2 Higher professional occupations	Ref.	(.)		0.28	-0.57	0.57
2 Lower managerial and professional occupations	-0.57	(0.30)		0.13	-0.83	-0.30
3 Intermediate occupations	-0.63	(0.34)		0.19	-1.02	-0.25
4 Small employers and own account workers	-0.97	(0.33)	**	0.17	-1.33	-0.62
5 Lower supervisory and technical occupations	-1.06	(0.33)	**	0.16	-1.38	-0.74
6 Semi-routine occupations	-1.30	(0.35)	***	0.19	-1.69	-0.91
7 Routine occupations	-1.01	(0.35)	**	0.20	-1.42	-0.60
Housing Tenure						
Owned or privately rented	Ref.	(.)		(.)	(.)	(.)
Social housing	-1.13	(0.23)	***	(.)	(.)	(.)
Gender						
Male	Ref.	(.)		(.)	(.)	(.)
Female	0.58	(0.12)	***	(.)	(.)	(.)
Constant	1.45	(0.44)	**	(.)	(.)	(.)
Observations	1624					
McFadden's Adjusted Pseudo R^2	0.115					
Cox-Snell Pseudo R^2	0.173					
Nagelkerke Pseudo R^2	0.236					
BIC (d.f.)	2065.612 (31)					
BIC (based on deviance)	-9940.048					
BIC (based on likelihood ratio chi square)	-87.450					

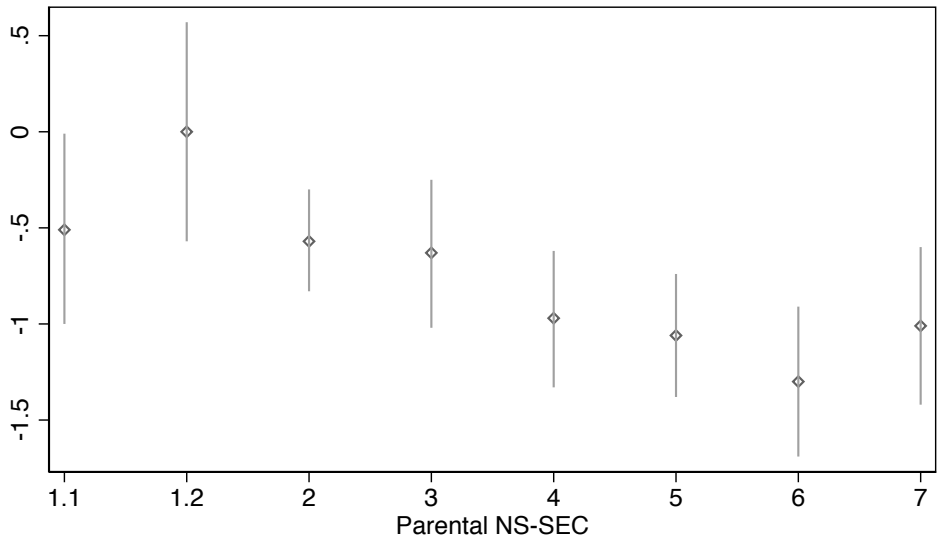
*Source: BHPS synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Model also includes ethnicity and wave year. * p<.05, ** p<.01, *** p<.001.*

Figure 1. 4: Log odds coefficients with 95% quasi-variance comparison intervals on the attainment of 5+ GCSEs at grades A*-C by parental education level



Source: BHPS data adjusted for complex survey design, n=1,624.
Parental NS-SEC, housing tenure, gender, ethnicity, and wave year also included in model.

Figure 1. 5: Log odds coefficients with 95% quasi-variance comparison intervals on the attainment of 5+ GCSEs at grades A*-C by parental social class



Source: BHPS data adjusted for complex survey design, n=1,624.
Parental education level, housing tenure, gender, ethnicity, and wave year also included in model.

The challenges of interpreting effect sizes of logistic regression models using log odd coefficients is well-documented (see, for example, Long, 1997, Treiman, 2009, Long and Freese, 2014). A suitable alternative method is to convert log odds into probabilities, for example, using marginal effects (Long and Freese, 2014). Table 1.11

details the average marginal effects, i.e. the average change in probability of attaining 5 or more A*-Cs given a change in the explanatory variable, holding all other variables at their observed values. The average marginal effects illustrate substantial differences in the probabilities of attaining the GCSE benchmark of 5 or more GCSEs at grades A*-C between more and less advantaged individuals.

Table 1. 11: Average marginal effects on the probability of attaining 5+ GCSEs at grades A*-C

	Δ Prob.	S.E.	
Parental Education Level			
Higher education	Ref.	(.)	
Further education	-0.16	(0.04)	***
School-level education	-0.22	(0.04)	***
Below school-level education	-0.37	(0.06)	***
Parental NS-SEC			
1.1 Large employers and higher managerial	-0.09	(0.06)	
1.2 Higher professional occupations	Ref.	(.)	
2 Lower managerial and professional occupations	-0.10	(0.05)	*
3 Intermediate occupations	-0.11	(0.06)	
4 Small employers and own account workers	-0.18	(0.06)	**
5 Lower supervisory and technical occupations	-0.20	(0.06)	**
6 Semi-routine occupations	-0.25	(0.06)	***
7 Routine occupations	-0.19	(0.06)	**
Housing Tenure			
Owned or privately rented	Ref.	(.)	
Social housing	-0.23	(0.05)	***
Gender			
Male	Ref.	(.)	
Female	0.11	(0.02)	***

The logistic regression model indicates that there is a statistically significant effect of parental education level, parental NS-SEC, housing tenure, and gender on the policy benchmark attainment of 5 or more GCSEs at grades A*-C. Young people from more educationally and occupationally advantaged backgrounds have significantly higher probabilities of attaining at least 5 good passes (i.e. A*-Cs) in their GCSE examinations over the course of the 1990s and 2000s. These findings support previous empirical work examining socio-economic background and the attainment of 5 or more GCSEs at grades A*-C (for example, Demack et al., 2000, Connolly, 2006, Sullivan et al., 2011). The next section tests the robustness of these findings using sensitivity analyses of parental social class and parental education.

5.1 Sensitivity Analyses of Independent Variables

It is not statistically appropriate to directly compare log odds coefficients across logistic regression models (see detailed reviews in Allison, 1999, Williams, 2009, Mood, 2010). In the following sensitivity analyses of socio-economic background measures, robustness is assessed by focusing on the substantive conclusions in the alternative logistic regression models, and comparing predicted probabilities for the explanatory variables in each model. The goodness-of-fit of each model is assessed using three BIC measures (based on degrees of freedom, model chi square, and deviance)[8], and a variety of Pseudo R^2 measures, as there is not a clear, superior Pseudo R^2 measure (Smithson, 2003, Lewis-Beck et al., 2004).

5.1.1 Testing Measures of Parental Social Class

Parental social class is measured using the NS-SEC (8-category) schema, the Goldthorpe (7-category) class schema, and Cambridge Scale score (the pre-cursor to the CAMSIS stratification measure). There are strong correlations between the three measures of parental social class. Parental NS-SEC and the Goldthorpe schema has a significant chi square statistic (3300 at 42 degrees of freedom, $p<.001$) and a Cramer's V of 0.59. Both parental NS-SEC and Cambridge Scale score, and parental Goldthorpe and Cambridge Scale score yield statistically significant, strong eta statistics (0.71 and 0.72 respectively). A dominance approach is used to construct the parent measures (Erikson 1984). The dominance approach for parental NS-SEC has been outlined above. The dominance approach using the Goldthorpe class schema follows the example of Erikson (1984): Goldthorpe class I and II, IVc, IVa, III, V, VIIa, VIIb.

Three separate logistic regression models are estimated and the results are presented in Table 1.12. The first model uses the measure of parental NS-SEC, and has been described in detail in the previous section. The next model uses the

[8] These statistics are retrieved using Long and Freese's fitstat command in the Stata SPost package. Formulas for the three BIC statistics can be found in Appendix C, at https://www3.nd.edu/~rwilliam/stats3/L05.pdf [accessed 27.07.19].

Goldthorpe class schema. Young people with parents in the service class in the Goldthorpe schema have significantly higher log odds of attaining 5 or more GCSEs at grades A*-C than those with parents in the petty bourgeoisie, skilled workers, or semi- or unskilled workers. This demonstrates a similar trend to the NS-SEC categories 4, 5, 6, and 7 (compared with 1.2). The categories for farmers and agricultural workers are not significant, but this may be due to relatively low sample sizes in these categories. The final model uses the Cambridge Scale. Parents with higher Cambridge Scale scores are significantly associated with higher log odds of attaining 5 or more GCSEs at grades A*-C.

The goodness-of-fit statistics are relatively similar for the three models. There are minor differences in Pseudo R^2 measures, suggesting that the amount of explained variance across the three models is relatively consistent. The BIC statistics demonstrate that the most parsimonious model uses the parental Cambridge Scale. This is unsurprising because the BIC statistic penalises models for estimating additional parameters. The Goldthorpe model would be considered an improvement over the NS-SEC model, i.e. BIC is lower (Raftery, 1995). However, the differences are small.

Table 1. 12: Sensitivity analyses of alternative measures of parental social stratification

Logistic Regression:	**NS-SEC**			**Goldthorpe Schema**			**Cambridge Scale**		
5+ GCSEs A*-C	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-0.90	(0.24)	***	-0.91	(0.23)	***	-0.79	(0.25)	**
School-level education	-1.19	(0.26)	***	-1.19	(0.25)	***	-1.05	(0.26)	***
Below school-level education	-1.87	(0.32)	***	-1.92	(0.32)	***	-1.75	(0.34)	***
Parental NS-SEC									
1.1 Large employers/higher man.	-0.51	(0.36)							
1.2 Higher professional	Ref.	(.)							
2 Lower managerial/professional	-0.57	(0.30)							
3 Intermediate	-0.63	(0.34)							
4 Small employers & own account	-0.97	(0.33)	**						
5 Lower supervisory/technical	-1.06	(0.33)	**						
6 Semi-routine	-1.30	(0.35)	***						
7 Routine	-1.01	(0.35)	**						
Parental Goldthorpe Schema									
Service class				Ref.	(.)				
Routine non-manual workers				-0.33	(0.20)				
Petty bourgeoisie				-0.64	(0.22)	**			
Farmers				-0.19	(0.68)				
Skilled /lower grade technicians				-0.77	(0.20)	***			
Non-skilled, semi, unskilled				-0.73	(0.22)	**			
Agricultural workers				0.99	(0.99)				
Parental Cambridge Scale							0.02	(0.00)	***
Housing Tenure									
Owned or privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-1.13	(0.23)	***	-1.17	(0.23)	***	-1.11	(0.23)	***
Gender									
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	0.58	(0.12)	***	0.60	(0.12)	***	0.59	(0.12)	***
Constant	1.45	(0.44)	**	1.04	(0.34)	**	-0.19	(0.45)	
Observations	1624			1624			1624		
McFadden's Adjusted Pseudo R^2	0.115			0.116			0.119		
Cox-Snell Pseudo R^2	0.173			0.173			0.171		
Nagelkerke Pseudo R^2	0.236			0.237			0.234		
BIC (d.f.)	2065.612 (31)			2058.032 (30)			2025.127 (25)		
BIC (based on deviance)	-9940.048			-9947.628			-9980.533		
BIC (likelihood ratio chi square)	-87.450			-95.030			-127.935		

*Source: BHPS synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Models also include ethnicity and wave year. * p<.05, ** p<.01, *** p<.001.*

Table 1.13 presents the predicted probabilities of attaining 5 or more GCSEs at grades A*-C for each of the three logistic regression models. The predicted probabilities are almost identical. The minor perturbations in effect sizes suggest that the alternative measures of social stratification are similar. In practice, there is no clear statistical evidence to prefer any one measure of parental social stratification. Drawing on the evidence presented above, NS-SEC is the preferred measure in subsequent analyses. NS-SEC also has the benefit of being widely used in official

governmental and social scientific research, and is therefore more ably compared across studies.

Table 1. 13: Predicted probabilities of attaining 5+ GCSEs at grades A*-C using alternative parental social stratification measures

	NS-SEC		Goldthorpe		Cambridge Scale	
	Prob.	**95% C.I.**	**Prob.**	**95% C.I.**	**Prob.**	**95% C.I.**
Parental Education Level						
Higher education	0.81	0.74 – 0.87	0.81	0.75 – 0.87	0.79	0.72 – 0.86
Further education	0.65	0.61 – 0.69	0.65	0.61 – 0.69	0.65	0.61 – 0.69
School-level education	0.59	0.54 – 0.64	0.59	0.54 – 0.64	0.60	0.55 – 0.64
Below school-level education	0.44	0.35 – 0.52	0.43	0.34 – 0.52	0.44	0.35 – 0.53
Housing Tenure						
Owned/privately rented	0.66	0.64 – 0.69	0.66	0.64 – 0.69	0.66	0.63 – 0.69
Social housing	0.43	0.34 – 0.52	0.42	0.33 – 0.51	0.43	0.34 – 0.52
Gender						
Male	0.58	0.55 – 0.61	0.58	0.54 – 0.61	0.58	0.54 – 0.61
Female	0.69	0.66 – 0.72	0.69	0.66 – 0.72	0.69	0.66 – 0.72

5.1.2 Testing Measures of Parental Education

Three measures of parental education level are tested. The first measure is the highest UK qualification attained (a measure collected in the BHPS). Two further international measures of education level are tested, CASMIN and ISCED. The UK qualifications measure collected in the BHPS represents the highest qualification that a respondent attained without making a distinction between quantity or grade attained. The CASMIN and ISCED measures make this distinction.[9] A dominance approach is used, taking the higher education level of mother or father.

The UK-specific qualifications measure in the BHPS has 12 categories. When collapsing the UK qualifications measure, categories represent higher education qualifications, further education qualifications, school-level qualifications, and below school-level qualifications, following the example of Lambert (2012). The CASMIN measure deposited in the BHPS has 9 categories, and the ISCED measure has 8

[9] For example, if a respondent achieved at least one O' Level at any grade, they would be categorised as O' Level in the UK qualifications measure. If the O' Level was achieved between grades A and C or CSE grade 1, the respondent would be categorised as 2b in the CASMIN framework and IIIb in the ISCED framework. If the O' Level was achieved between grades D and E or a CSE grade lower than 1, the respondent would be categorised as 1b in the CASMIN framework and II in the ISCED framework.

categories. The CASMIN and ISCED measures are suitably prefixed to easily facilitate collapses. For example, CASMIN is aggregated into three categories: tertiary (3a and 3b), secondary (2a, 2b, 2c general and 2c vocational), and elementary or below (1a, 1b and 1c). ISCED is aggregated into four categories: higher degree, first degree and higher vocational (5a, 5b and 6), high and low secondary vocational (3a and 3c), lower secondary (2) and primary or below (0 and 1).

Tabulations demonstrate that there is general consistency in the assignment of individuals to categories in the CASMIN and ISCED frameworks. There is less consistency when tabulating these measures with the UK-specific qualifications. For example, it is clear that higher and first degrees (as categorised in the UK-specific qualifications measure) would be considered university-level education qualifications. CASMIN separates lower tertiary and higher tertiary qualifications, which may also include some vocational degrees and diplomas. ISCED has separate categories for higher degrees, first degrees, and higher vocational qualifications. In their aggregated forms, there are strong associations between CASMIN and ISCED (chi square 2500 at 6 degrees of freedom, $p<.001$ and Cramer's V = 0.89). Associations are weaker, but still statistically significant, for CASMIN and UK qualifications (chi square 836.95 at 6 degrees of freedom, $p<.001$ and Cramer's V = 0.51), and for ISCED and UK qualifications (chi square 1100 at 9 degrees of freedom, $p<.001$ and Cramer's V = 0.47).

Table 1.14 presents the results of the three alternative logistic regression models using the disaggregated measures for parental education level. The reference category is no qualifications, as this is a consistent category across all three measures and a substantively useful comparison point. There is a positive effect on the attainment of 5 or more GCSEs at grades A*-C for females, children whose parents have higher education qualifications, children whose parents have occupations in NS-SEC 1.2, and for those living in owned or rented accommodation. The significance of those in NS-SEC 2 and NS-SEC 3 compared with NS-SEC 1.2 is not consistent across the three measures, although the p-values are close to the conventional threshold of $p<.05$. It is anticipated that this might be due to the broad range of occupations in NS-SECs 2 and 3, and the associated intermediate levels of

Table 1. 14: Sensitivity analyses of alternative measures of parental education level

Logistic Regression:	UK Qualifications			CASMIN			ISCED		
5+ GCSEs A*-C	Coef.	S.E.		Coef.	S.E.		Coef.	S.E.	
Parental Education Level (UK)									
No qualification	Ref.	(.)							
Other e.g. youth training	0.88	(0.75)							
Apprenticeship	-0.03	(0.76)							
CSE Grade 2-5/O Level D-G	0.86	(0.37)	*						
Commercial qual., No O Level	0.76	(0.43)							
GCE O Level A-C or equiv.	0.68	(0.24)	**						
GCE A Level or equiv.	0.76	(0.30)	*						
Nursing qualification	1.58	(0.56)	**						
Other higher qualification	0.92	(0.23)	***						
Teaching qualification	1.64	(0.49)	**						
First degree certificate	2.02	(0.35)	***						
Higher degree	1.61	(0.54)	**						
Parental Education Level (CASMIN)									
1a None				Ref.	(.)				
1b Elementary				0.40	(0.26)				
1c Basic vocational				0.80	(0.20)	***			
2b Middle general				0.79	(0.21)	***			
2a Middle vocational				1.00	(0.25)	***			
2c_gen High general				1.12	(0.43)	**			
2c_voc High vocational				1.31	(0.45)	**			
3a Low tertiary				1.07	(0.24)	***			
3b High tertiary				2.30	(0.42)	***			
Parental Education Level (ISCED)									
1 Primary							Ref.	(.)	
2 Low secondary							0.36	(0.27)	
3c Low secondary (vocational)							0.82	(0.15)	***
3a High secondary/mid vocational							1.22	(0.34)	***
5b Higher vocational							1.06	(0.24)	***
5a First degree							2.37	(0.46)	***
6 Higher degree							1.73	(0.99)	
Parental NS-SEC									
1.1 Large employers/higher man.	-0.52	(0.35)		-0.44	(0.37)		-0.45	(0.37)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower managerial/professional	-0.61	(0.28)	*	-0.54	(0.32)		-0.54	(0.32)	
3 Intermediate	-0.64	(0.34)		-0.76	(0.35)	*	-0.75	(0.35)	*
4 Small employers & own account	-0.99	(0.32)	**	-0.89	(0.35)	*	-0.89	(0.35)	*
5 Lower supervisory/technical	-1.05	(0.32)	**	-1.01	(0.35)	**	-1.00	(0.35)	**
6 Semi-routine	-1.29	(0.34)	***	-1.38	(0.36)	***	-1.37	(0.36)	***
7 Routine	-1.01	(0.34)	**	-1.01	(0.37)	**	-1.01	(0.37)	**
Housing Tenure									
Owned or privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-1.11	(0.23)	***	-1.19	(0.23)	***	-1.20	(0.23)	***
Gender									
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	0.59	(0.12)	***	0.58	(0.12)	***	0.58	(0.12)	***
Constant	-0.47	(0.44)		-0.28	(0.43)		-0.27	(0.44)	
Observations	1624			1624			1624		
McFadden's Ad. Pseudo R^2	0.113			0.121			0.123		
Cox-Snell Pseudo R^2	0.179			0.185			0.184		
Nagelkerke Pseudo R^2	0.245			0.252			0.252		
BIC (d.f.)	2112.764 (39)			2079.586 (36)			2065.819 (34)		
BIC (based on deviance)	-9892.895			-9926.074			-9939.840		
BIC (likelihood ratio chi square)	-40.297			-73.475			-87.242		

*Source: BHPS synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Models also include ethnicity and wave year. * p<.05, ** p<.01, *** p<.001.*

qualifications required for different occupations. The categories are therefore more likely to be sensitive to how qualifications are measured. The goodness-of-fit measures are almost identical for the CASMIN and ISCED models, with these models

having slightly higher Pseudo R^2 and slightly lower BIC statistics than the model using UK qualifications. The aggregated educational transition variables are also tested in the sensitivity analyses and the results are very similar.

Table 1.15 presents the predicted probabilities of attaining 5 or more A*-Cs for the three logistic regression models. There are minor perturbations in the predicted probabilities across all three models using different parental education measures. This suggests that the alternative measures of parental education level are very similar. In practice, there are no clear statistical or theoretical reasons to prefer any particular measure of parental education level. The aggregated UK qualifications measure is the preferred measure in subsequent analyses.

Table 1. 15: Predicted probabilities of attaining 5+ GCSEs at grades A*-C using alternative parental education measures.

	UK Qualifications		**CASMIN**		**ISCED**	
	Prob.	**95% C.I.**	**Prob.**	**95% C.I.**	**Prob.**	**95% C.I.**
Parental NS-SEC						
1.1 Large employers/higher man.	0.69	0.59 – 0.78	0.70	0.61 – 0.78	0.70	0.61 – 0.78
1.2 Higher professional	0.77	0.69 – 0.86	0.77	0.68 – 0.86	0.77	0.68 – 0.86
2 Lower managerial/professional	0.67	0.62 – 0.72	0.68	0.63 – 0.73	0.68	0.63 – 0.73
3 Intermediate	0.66	0.59 – 0.73	0.64	0.57 – 0.71	0.64	0.57 – 0.71
4 Small employers & own account	0.60	0.53 – 0.67	0.61	0.54 – 0.68	0.61	0.54 – 0.68
5 Lower supervisory/technical	0.58	0.51 – 0.65	0.59	0.52 – 0.66	0.59	0.52 – 0.66
6 Semi-routine	0.53	0.45 – 0.61	0.51	0.43 – 0.59	0.51	0.43 – 0.59
7 Routine	0.59	0.50 – 0.68	0.59	0.50 – 0.67	0.59	0.50 – 0.67
Housing Tenure						
Owned/privately rented	0.66	0.63 – 0.69	0.66	0.64 – 0.69	0.66	0.64 – 0.69
Social housing	0.43	0.34 – 0.52	0.42	0.33 – 0.51	0.42	0.33 – 0.50
Gender						
Male	0.58	0.55 – 0.61	0.58	0.55 – 0.61	0.58	0.55 – 0.61
Female	0.69	0.66 – 0.72	0.69	0.66 – 0.72	0.69	0.66 – 0.72

5.2 Functional Form of the Outcome Variable

The results in the section above indicated that parental NS-SEC and UK-specific qualifications are suitable measures for parental socio-economic background. There is no single, agreed-upon way to measure GCSE attainment. The policy benchmark of attaining 5 or more A*-Cs has been used so far. The following section provides a series of sensitivity analyses to explore different operationalisations of GCSE attainment.

5.2.1 Alternative Thresholds of GCSE Attainment

The underlying variable of the number of GCSEs attained at grades A*-C can be cut at varying levels of attainment to signify increasing thresholds of 'success' or 'failure'. The use of the binary variable of 5 or more A*-Cs attained at GCSE level was a standard policy benchmark of attainment in England (Leckie and Goldstein, 2009). Connelly et al. (2013) used 1-4 A*-Cs at GCSE level as a measure of middle attainment. The individuals who attained 1-4 A*-Cs were considered to have fallen short of the national benchmark, but had at least one good pass. The threshold for high attainment is less clear. There is no policy standard on what constitutes high attainment. The following sections construct substantively informative cut-off points to represent varying degrees of attainment at GCSE level.

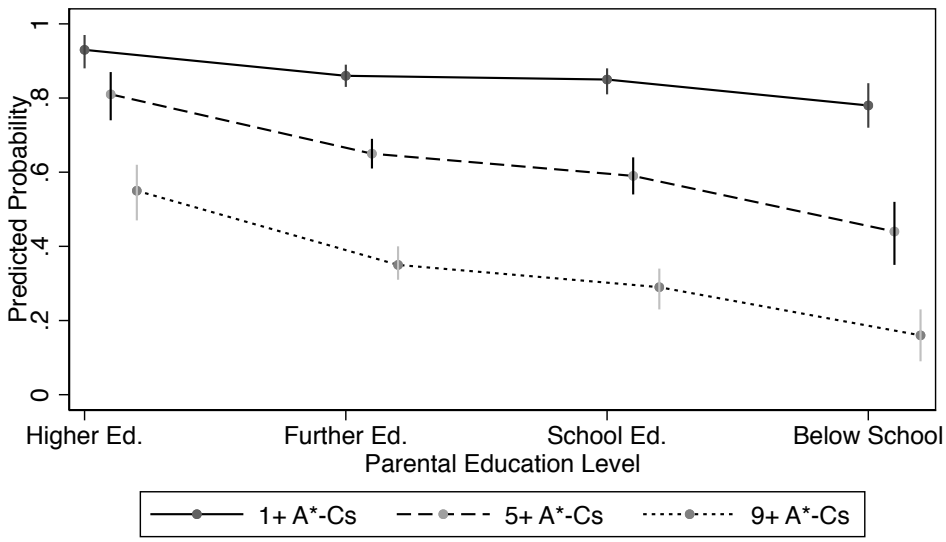
The distribution of the data for the number of GCSEs attained at grades A*-C has a lower quartile of 2 GCSEs at grades A*-C, a median of 7 GCSEs at grades A*-C, and an upper quartile of 9 GCSEs at grades A*-C. Following policy standards, 5 or more GCSEs A*-C will be taken as a benchmark of good attainment. Lower attainment will be considered as 1-4 A*-Cs for those falling short of the national benchmark. Three different categorisations of high attainment are tested: 8 or more A*-Cs, 9 or more A*-Cs, and 10 or more A*-Cs. The following analyses use 9 or more A*-Cs as the most suitable threshold because it represents the 75th percentile of the number of GCSEs attained at grades A*-C. Table 1.16 presents the frequencies of the three binary variables of GCSE attainment.

Table 1. 16: Frequencies and percentages of GCSE results by varying thresholds of attainment, unweighted

GCSE Outcome	Freq.	Percent
Attained at least 1 A-C*	1375	84.67
Did not attain at least 1 A-C*	249	15.33
Attained at least 5 A-Cs*	1018	62.68
Did not attain at least 5 A-Cs*	606	37.32
Attained at least 9 A-Cs*	573	35.28
Did not attain at least 9 A-Cs*	1051	64.72
Total *n*	**1624**	**100.00**

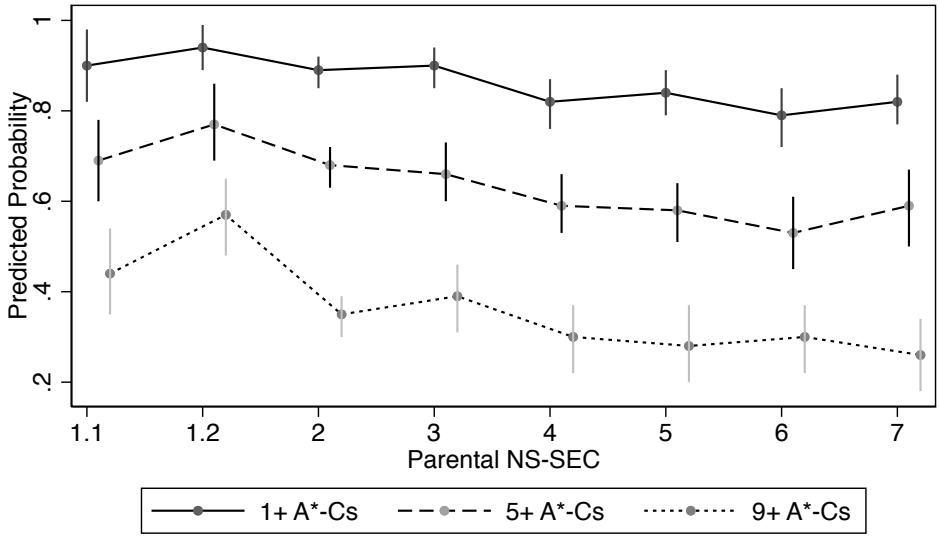
The results of the logistic regression models demonstrate the importance of parental social class, parental education level, housing tenure, and gender at all three threshold levels (see Table A1.1 in Appendix 1). Figures 1.6 and 1.7 present graphs of the predicted probabilities of attaining the different thresholds. The negative gradients are steeper for higher levels of attainment (i.e. 9 or more good passes) than lower levels of attainment (i.e. 1 or more good passes). Children with parents with higher education qualifications have higher probabilities of attaining at all three thresholds than children whose parents have any lower level of education. The effect of parental NS-SEC is particularly interesting at the highest levels of attainment, as two clusters emerge between NS-SEC 1.2 and all other social class categories. The results suggest that the gap between the most advantaged and less advantaged young people widens at higher levels of attainment.

Figure 1. 6: Predicted probabilities of thresholds of GCSE attainment by parental education level (separate logistic regression models)



Source: BHPS data adjusted for complex survey design, n=1,624.
Parental NS-SEC, housing tenure, gender, ethnicity, and wave year also included in model.

Figure 1. 7: Predicted probabilities of thresholds of GCSE attainment by parental social class (separate logistic regression models)



Source: BHPS data adjusted for complex survey design, n=1,624.
Parental education level, housing tenure, gender, ethnicity, and wave year also included in model.

5.2.2 GCSE Attainment Brackets

This section examines a categorical measure of GCSE attainment brackets (Table 1.17). There are several models suited to modelling categorical dependent variables. Multinomial logistic regression models are suitable for nominal dependent variables, and are sometimes applied to ordinal outcomes (Long, 1997, Long and Freese, 2014). Stereotype logistic regression models are parsimonious alternatives to multinomial logistic regression models (Lunt, 2001), and can be used when the assumptions of ordered logistic regression models, such as the proportional odds assumption, are violated (Liu, 2014). The estimation of these models for categorical dependent variables provide useful background information and can be found in Tables A1.2 and A1.3 in Appendix 1.

Table 1. 17: Frequencies and percentages of GCSE results by attainment bracket, unweighted

GCSE Attainment Brackets	Freq.	Percent
0 GCSEs A-C*	249	15.33
1-4 GCSEs A-C*	357	21.98
5-8 GCSEs A-C*	445	27.40
9+ GCSEs A-C*	573	35.28
Total *n*	**1624**	**100.00**

An ordered logistic regression model is more appropriate for GCSE attainment brackets, because there is a definite order and hierarchy to the categories and the distances between the categories are not assumed to be equal (Long, 1997). There are different types of ordered logistic regression models, for example, the proportional odds model (McCullagh, 1980) and the continuation ratio model (Fienberg and Mason, 1979) (also see Berridge, 1992, Gayle, 1996, O'Connell, 2006, Long and Freese, 2014).

O'Connell (2006) provides a detailed review of interpreting the proportional odds and continuation ratio models. An ascending proportional odds model estimates the probability of being in a particular category compared with being in all higher categories, estimated at cumulative cut-points. The continuation ratio model estimates the probability of being in a particular category compared with being in all higher categories, conditional on not being in any lower category. Table 1.18 illustrates the differences in the contrasts made for each category of the GCSE outcome variable in the following proportional odds and continuation ratio models.

Table 1. 18: Visualisation of the comparisons made in proportional odds models and continuation ratio models for GCSE attainment brackets

Proportional Odds Contrasts (Freq.)				
Cut 1 (0 v 1+)	249	1375		
Cut 2 (0-4 v 5+)		606	1018	
Cut 3 (1-8 v 9+)		1051		573
Continuation Ratio Contrasts (Freq.)				
Cons 1 (0 v 1+)	249	1375		
Cons 2 (1-4 v 5+)	-	357	1018	
Cons 3 (5-8 v 9+)	-	-	445	573

Table 1.19 presents the proportional odds model and the continuation ratio model to estimate GCSE attainment brackets. There is a practical analytical challenge with estimating the continuation ratio model in Stata. The commands are generally user-written and are not compatible with the svy suite in Stata. In practice, this means that appropriate adjustments for complex survey design cannot be made. The models have been estimated using unweighted data to allow for ready comparison in Table 1.19. The use of unweighted estimates is problematic for robust inferential analysis (Treiman, 2009).

An assumption of the proportional odds model is that the coefficients are the same across the separate logistic regressions, termed the parallel regression assumption (Williams, 2016). The proportional odds model in Table 1.19 does not violate the parallel regression assumption (assessed using the Brant test, see Brant, 1990). The proportional odds model results demonstrate that young people whose parents have higher education qualifications and who have occupations in NS-SEC 1.2 have higher log odds of being in any higher attainment bracket than young people whose parents have less advantaged educational and occupational positions (with the exception of NS-SEC 1.1). The log odds of being in higher attainment brackets are higher for those living in owned or privately rented homes compared with social housing, and for girls compared with boys.

Table 1. 19: Ordered logistic regression models of GCSE attainment brackets, unweighted

Ordered Logistic Regression: Attainment Brackets			**Proportional Odds Coef.**	**S.E.**		**Continuation Ratio Coef.**	**S.E.**	
Parental Education Level								
Higher education			Ref.	(.)		Ref.	(.)	
Further education			-0.87	(0.15)	***	-0.74	(0.13)	***
School-level education			-1.09	(0.16)	***	-0.94	(0.14)	***
Below school-level education			-1.81	(0.21)	***	-1.57	(0.18)	***
Parental NS-SEC								
1.1 Large employers and higher managerial			-0.45	(0.27)		-0.39	(0.24)	
1.2 Higher professional occupations			Ref.	(.)		Ref.	(.)	
2 Lower managerial and professional occupations			-0.87	(0.20)	***	-0.77	(0.18)	***
3 Intermediate occupations			-0.80	(0.23)	**	-0.70	(0.21)	**
4 Small employers and own account workers			-1.26	(0.23)	***	-1.08	(0.20)	***
5 Lower supervisory and technical occupations			-1.25	(0.24)	***	-1.07	(0.21)	***
6 Semi-routine occupations			-1.35	(0.23)	***	-1.14	(0.20)	***
7 Routine occupations			-1.47	(0.25)	***	-1.26	(0.21)	***
Housing Tenure								
Owned or privately rented			Ref.	(.)		Ref.	(.)	
Social housing			-1.11	(0.16)	***	-0.90	(0.13)	***
Gender								
Male			Ref.	(.)		Ref.	(.)	
Female			0.65	(0.10)	***	0.55	(0.08)	***
0 A*-Cs	v	1-4, 5-8, 9+	-2.77	(0.27)				
0, 1-4 A*-Cs	v	5-8, 9+ A*-Cs	-1.34	(0.27)				
0, 1-4, 5-8 A*-Cs	v	9+ A*-Cs	0.04	(0.27)				
0 A*-Cs	v	1-4, 5-8, 9+				-2.59	(0.24)	
1-4 A*-Cs	v	5-8, 9+ A*-Cs				-1.68	(0.24)	
5-8 A*-Cs	v	9+ A*-Cs				-0.61	(0.24)	
Observations			1624			1624		
BIC (d.f.)			4144.248 (33)			-		
BIC (deviance)			-7861.412			-		
BIC (likelihood ratio chi square)			-239.453			-		

Source: BHPS synthetic cohorts of English Year 12 pupils, unweighted data.
*Models also include ethnicity and wave year. * p<.05, ** p<.01, *** p<.001.*

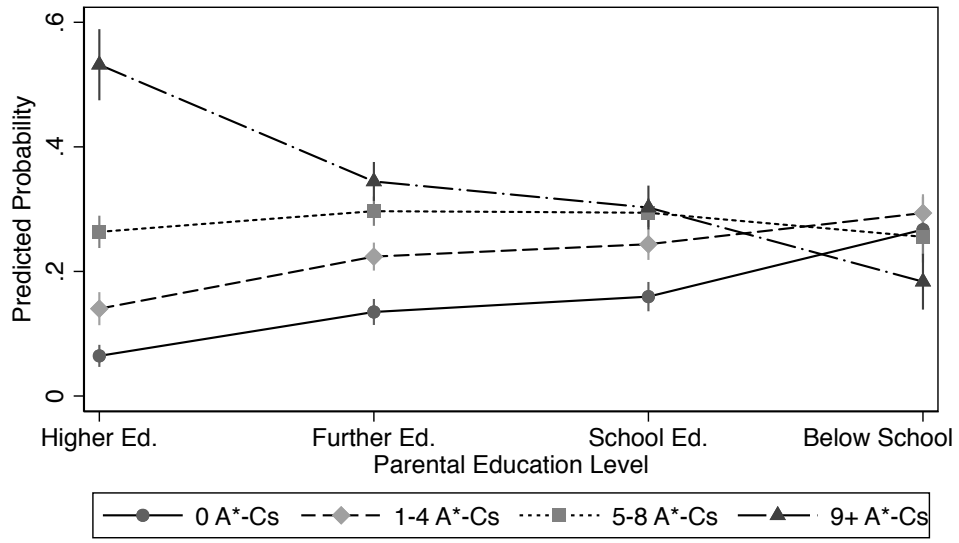
Continuation ratios can be constrained or unconstrained. The difference between the constrained and unconstrained models is in the number of parameters estimated. Constrained models make the assumption that the effects of the independent variables are constant across the categories of the dependent variable. Constrained models are therefore more parsimonious because they produce one set of coefficients (Long and Freese, 2014). A non-significant likelihood ratio test provides evidence to prefer the constrained model over the unconstrained model (ΔLR 62.03 @ 60 degrees of freedom, p=.4037). The results of the continuation ratio model suggest that young people whose parents have higher education qualifications have higher log odds of being in any higher attainment bracket, conditional on not being in lower attainment brackets, than young people whose parents have lower education qualifications. Young people whose parents have occupations in NS-SEC 1.2 have higher log odds of being in any higher attainment bracket, conditional on not being in lower attainment brackets, than young people whose parents have occupations in any other NS-SEC category (with the exception of NS-SEC 1.1). The log odds of being in higher attainment brackets, conditional on not being in lower attainment brackets, are higher for those living in owned or privately rented homes compared with social housing, and for girls compared with boys.

The substantive conclusions of the proportional odds model and the continuation ratio model in Table 1.19 are very similar. The coefficients in the two models are presented on the log odds scale. The coefficients can be converted into predicted probabilities to better understand the magnitude of effect for the independent variables. Figures 1.8 and 1.9 plot the predicted probabilities and 95% confidence intervals for the GCSE attainment brackets by parental education level and parental social class. The predicted probabilities shown are based on the proportional odds model. The predicted probabilities were re-estimated for the continuation ratio model and the results were almost identical.[10] The graphs suggest that young people with the most advantaged parents (in terms of education and occupation) have the highest probabilities of attaining the higher attainment brackets, and the lowest probabilities of attaining the lower attainment brackets. There are more distinctive gradients for the

[10] In order to do this, the continuation ratio model was re-estimated using Buis (2007) seqlogit command with constraints, which is compatible with the margins command in Stata.

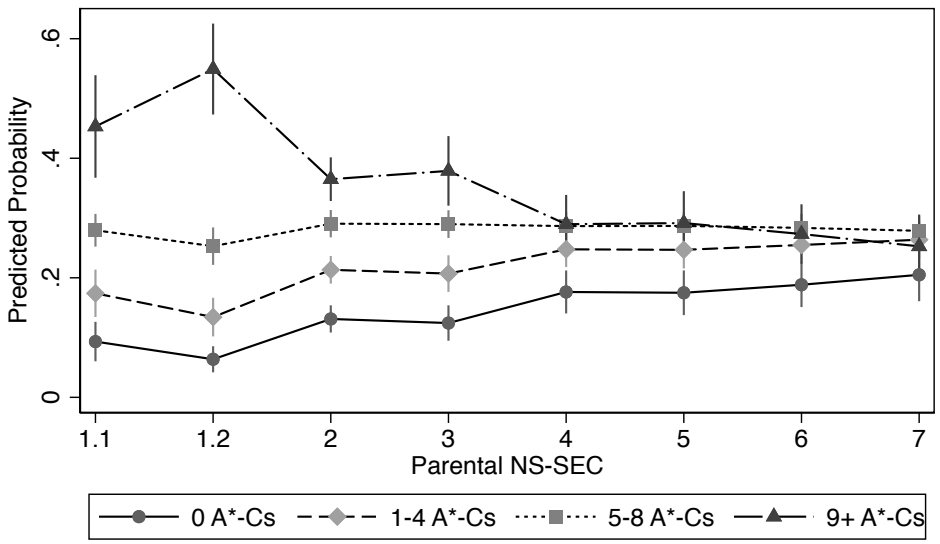
highest attainment bracket and lowest attainment bracket by parental education and parental social class. This might suggest that greater socio-economic differences are observed at the highest and lowest attainment brackets compared with differences observed in the middling attainment brackets.

Figure 1. 8: Predicted probabilities of GCSE attainment brackets by parental education level (proportional odds model)



Source: BHPS unweighted data, n=1,624.
Parental NS-SEC, housing tenure, gender, ethnicity, and wave year also included in model.

Figure 1. 9: Predicted probabilities of GCSE attainment brackets by parental social class (proportional odds model)



Source: BHPS unweighted data, n=1,624.
Parental education level, housing tenure, gender, ethnicity, and wave year also included in model.

5.2.3 Number of GCSEs at Grades A*-C

The final set of models estimate the number of GCSEs attained at grades A*-C (refer to Figure 1.3). A series of regression models suitable for count data are estimated including the Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial (Cameron and Trivedi, 1998). The estimates across all four models are relatively similar (see Table A1.4 in Appendix 1). Long and Freese (2014: 507) advise that the Poisson model seldom fits count measures in social survey data because the model does not account for over-dispersion. The negative binomial regression model is usually better suited to dealing with data with over-dispersion (Cameron and Trivedi, 1998, Long and Freese, 2014). Comparing the four models, a significant likelihood ratio test provides evidence of over-dispersion and therefore the negative binomial regression model is preferred over a Poisson model (Long, 1997, Long and Freese, 2014). The outcome variable of the number of GCSEs attained at grades A*-C also has a high proportion of zero counts (0.153). Zero-inflated models can account for high proportions of zeros in count data (see Long and Freese, 2014 for a detailed review). Comparing the negative binomial model and zero-inflated negative binomial model, a significant Vuong test (see Vuong, 1989) provides evidence that a zero-inflated model is most suitable for these data (Long, 1997, Long and Freese, 2014).

The output for the zero-inflated negative binomial model is reported in Table 1.20. The first section details the results of a logistic regression model of the zero count, i.e. the attainment of zero A*-Cs. Young people from more advantaged educational and social class backgrounds are less likely to attain zero A*-Cs than those from less advantaged backgrounds. Boys are more likely to attain zero A*-Cs than girls, and young people living in social housing are more likely to attain zero A*-Cs than young people living in owned or privately rented houses.

The second section in Table 1.20 details the results of the negative binomial regression model for the non-zero count. For young people who have attained some GCSEs at grades A*-C, the children of graduates are more likely to attain higher counts of A*-Cs than the children of parents with any lower level of education

Table 1. 20: Zero-inflated negative binomial regression model of the number of A*-Cs

				Quasi-Variance		
Logistic Regression: Zero A*-Cs	**Coef.**	**S.E.**		**S.E.**	**LCI**	**UCI**
Parental Education Level						
Higher education	Ref.	(.)		-	-	-
Further education	0.73	(0.39)		-	-	-
School-level education	0.87	(0.41)	*	-	-	-
Below school-level education	1.38	(0.45)	**	-	-	-
Parental NS-SEC						
1.1 Large employers and higher managerial	0.64	(0.67)		-	-	-
1.2 Higher professional occupations	Ref.	(.)		-	-	-
2 Lower managerial and professional occupations	0.80	(0.49)		-	-	-
3 Intermediate occupations	0.65	(0.55)		-	-	-
4 Small employers and own account workers	1.43	(0.50)	**	-	-	-
5 Lower supervisory and technical occupations	1.22	(0.55)	*	-	-	-
6 Semi-routine occupations	1.62	(0.53)	**	-	-	-
7 Routine occupations	1.36	(0.54)	*	-	-	-
Housing Tenure						
Owned or privately rented	Ref.	(.)		-	-	-
Social housing	1.23	(0.23)	***	-	-	-
Gender						
Male	Ref.	(.)		-	-	-
Female	-0.68	(0.17)	***	-	-	-
Constant	-2.67	(0.63)	***			
Negative Binomial Regression: GCSEs A*-C						
Parental Education Level						
Higher education	Ref.	(.)		0.03	-0.06	0.06
Further education	-0.14	(0.03)	***	0.02	-0.17	-0.11
School-level education	-0.23	(0.04)	***	0.03	-0.29	-0.16
Below school-level education	-0.41	(0.07)	***	0.06	-0.54	-0.29
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.07	(0.05)		0.04	-0.16	0.01
1.2 Higher professional occupations	Ref.	(.)		0.03	-0.06	0.06
2 Lower managerial and professional occupations	-0.11	(0.04)	**	0.02	-0.15	-0.06
3 Intermediate occupations	-0.12	(0.05)	*	0.04	-0.20	-0.04
4 Small employers and own account workers	-0.17	(0.06)	**	0.05	-0.26	-0.08
5 Lower supervisory and technical occupations	-0.23	(0.06)	***	0.05	-0.32	-0.13
6 Semi-routine occupations	-0.21	(0.06)	**	0.05	-0.31	-0.11
7 Routine occupations	-0.21	(0.07)	**	0.06	-0.34	-0.09
Housing Tenure						
Owned or privately rented	Ref.	(.)		-	-	-
Social housing	-0.18	(0.06)	**	-	-	-
Gender						
Male	Ref.	(.)		-	-	-
Female	0.12	(0.02)	***	-	-	-
Constant	1.95	(0.08)	***			
Log of alpha	-3.04	(0.27)	***			
Observations	1624					
McFadden's Adjusted Pseudo R2	0.036					
Cox-Snell Pseudo R2	0.237					
Nagelkerke Pseudo R2	0.238					
BIC (d.f.)	8637.312 (63)					
BIC (based on deviance)	-3368.348					
BIC (based on likelihood ratio chi square)	3.918					

*Source: BHPS synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Model also includes ethnicity and wave year. *p<.05, **p<.01, ***p<.001.*

qualification. The 95% quasi-variance comparison intervals illustrate significant differences in the number of A*-Cs attained for all categories of parental education level. Young people with parents in occupations in NS-SEC 1.2 are more likely to attain higher counts of A*-Cs than the children from all other social class backgrounds (except NS-SEC 1.1). The 95% quasi-variance comparison intervals demonstrate that there are not significant differences in the number of GCSEs attained at grades A*-C for any other contrasting pairs of NS-SEC. There is therefore a distinct advantage to having parents in higher professional occupations, such as doctors, lawyers, or university lecturers. Girls are more likely to attain higher counts of A*-Cs than boys, and young people in owned or privately rented houses are more likely to attain higher counts of A*-Cs than young people in social housing.

Post-estimation marginal effects and expected counts can be calculated to better understand effect sizes (Table 1.21). The expected counts for the key independent variables of interest demonstrate that, conditional on attaining some GCSEs at grades A*-C, young people from more advantaged educational and occupational backgrounds are expected to attain higher counts of GCSEs at grades A*-C than their less advantaged peers. Expected counts can also be calculated for a combination of characteristics, sometimes called ‘ideal types’ (Long and Freese, 2014). The characteristics of the ‘most’ advantaged can be deduced as those with graduate parents in NS-SEC 1.2 living in owned or privately rented homes. Alternatively, the characteristics of the ‘least’ advantaged can be considered those with parents with below school-level qualifications in NS-SEC 7 living in social housing. The ‘most’ advantaged males have an expected count of 8.43 A*-Cs and the ‘least’ advantaged males have an expected count of 1.69 A*-Cs. The ‘most’ advantaged females have an expected count of 9.62 A*-Cs and the ‘least’ advantaged females have an expected count of 2.58 A*-Cs. The difference between the ‘most advantaged’ and ‘least advantaged’ pupils is, on average, the attainment of between 6 and 7 GCSEs at grades A*-C.

Table 1. 21: Expected counts for the number of GCSEs attained at grades A*-C

	Expected Count	S.E.
Parental Education Level		
Higher education	7.34	0.26
Further education	6.00	0.16
School-level education	5.42	0.20
Below school-level education	4.17	0.29
Parental NS-SEC		
1.1 Large employers and higher managerial	6.51	0.36
1.2 Higher professional occupations	7.29	0.29
2 Lower managerial and professional	6.21	0.18
3 Intermediate occupations	6.21	0.29
4 Small employers and own account workers	5.41	0.29
5 Lower supervisory and technical occupations	5.26	0.27
6 Semi-routine occupations	5.05	0.32
7 Routine occupations	5.23	0.35
Housing Tenure		
Owned or privately rented	6.12	0.10
Social housing	4.22	0.36
Gender		
Male	5.36	0.13
Female	6.48	0.13

6 Discussion and Conclusions

The resounding empirical finding from the analyses is that there is an enduring effect of parental socio-economic background on school GCSE attainment over the course of almost 2 decades. More socio-economically advantaged young people in the 1990s and 2000s attain better and higher grades in their GCSE examinations, net of other important demographic factors such as gender and ethnicity. This finding is consistent with previous studies of GCSE attainment (Drew, 1995, Demack et al., 2000, Sullivan, 2001, Connolly, 2006, Connelly et al., 2013, Gayle et al., 2014, Strand, 2014a, Playford and Gayle, 2015).

A central methodological contribution of this study is the attention to sensitivity analyses of key parental socio-economic background and GCSE attainment measures. Sensitivity analyses of the independent variables demonstrate that parental NS-SEC is a robust measure of parental social class and the standard BHPS measure of UK qualifications is a suitable measure of parental education level.

Defining GCSE 'success' at different thresholds demonstrates that the effects of socio-economic background tend to increase at higher levels of attainment. This finding is consistent with previous analyses of low, middle, and high GCSE attainment using the Youth Cohort Study (Sullivan et al., 2011). Modelling the number of GCSEs attained at grades A*-C provides evidence of an attainment gap between the 'most' and 'least' advantaged young people. The 'most' advantaged young people are defined as those whose parents have higher education qualifications, are in higher professional occupations, and are living in owned or privately rented accommodation. The 'least' advantaged young people are defined as those whose parents have below school-level qualifications, are in routine occupations, and are living in social housing. Those defined as the 'most' advantaged have an expected count of between 6 and 7 more GCSEs at grades A*-C than those defined as the 'least' advantaged young people.

Government statistics and education reports tend to focus on the attainment gap between the most disadvantaged and all other pupils.[11] Often, proxy measures of disadvantage are used, such as the eligibility for Free School Meals (Department for Education, 2015c). The analyses in this chapter explore the relative gap between young people from a broader range of socio-economic backgrounds using detailed education and occupation-based parental measures. The findings provide substantial empirical evidence of a more subtle socio-economic gradient in school GCSE attainment. There are large gaps in GCSE attainment between young people from the 'most' educationally and occupationally advantaged backgrounds compared with young people from all other less socio-economically advantaged backgrounds. GCSE examination results are consequential for young people's future opportunities and choices after the compulsory school leaving age. The persistence of such socio-economic inequalities for young people throughout the 1990s and 2000s are consequential for their immediate and longer-term futures. The clear socio-economic inequalities further cast serious doubt over the success of government policies aimed at reducing inequalities in educational attainment.

[11] See the annual statistical collections from the Department for Education, available at https://www.gov.uk/government/collections/statistics-gcses-key-stage-4 [accessed 26.09.19].

Chapter 2
Synthetic Cohorts of Year 12 Pupils in the Early 2010s: A Replication Analysis

1 Introduction

The previous chapter presented convincing empirical evidence that there are strong parental socio-economic background effects in GCSE attainment for young people growing up in the 1990s and 2000s. The analyses in this chapter are designed as a replication study of these empirical findings for young people in the 2010s. Chapter 1 identified synthetic cohorts of English Year 12 pupils living in British Household Panel Survey (BHPS) households. This chapter identifies similar synthetic cohorts of English Year 12 pupils using a more contemporary dataset, the UK Household Longitudinal Study (UKHLS). The replication study involves using the same analytical procedures with a different dataset to examine the empirical regularity of the findings in Chapter 1.

Replication is the process of repeating a piece of analysis to verify the original results (Freese, 2007). King (1995) suggested that the replication standard involves verifying results using the same datasets as the original research. Herrnson (1995: 452) stated that replication should necessarily involve new data, which 'repeats an empirical study in its entirety'. Janz (2016) termed this distinction duplication and replication. Freese and Peterson (2017: 152) outlined four distinct 'forms of replication': verifiability, robustness, repeatability, and generalisability. Verification tends to focus on producing the same results when analysing the same data. Robustness checks tend to use alternative specifications using the same data. Repeatability involves using the same analytical approach as the original study with different data. Generalisability involves a combination of alternative specifications and different data to test the widespread nature of the original findings (Freese and Peterson, 2017).

The analyses in this chapter test the repeatability of the results in Chapter 1. The next section details the data management process of identifying synthetic cohorts of English Year 12 pupils in the UKHLS. Section 3 presents the descriptive statistics of the UKHLS-NPD synthetic cohort sample. Section 4 presents the modelling results of the replication study. Particular attention continues to be paid to the functional form of the educational outcome variable. Alternative specifications of GCSE attainment are tested, including the attainment of 5 or more A*-Cs and the number of GCSEs attained at grades A*-C. The analyses are then extended to use a continuous measure of GCSE point score.

2 Replicating Synthetic Cohorts of English Year 12 Pupils in the UKHLS

In Chapter 1, the synthetic cohorts of Year 12 pupils were identified on a wave-by-wave basis. School Year 12 pupils were identified because the young people sat their GCSE examinations in June of school Year 11, and reported their results in the BHPS adult survey between September and December of school Year 12. There are several key differences in the data collection and study designs of the BHPS and UKHLS. The process of identifying the synthetic cohorts of English Year 12 pupils was therefore not identical in the two studies. First, each wave of data collection in the UKHLS takes place over the course of 24 months. The data collection is overlapping, with a new wave beginning every 12 months. This is largely due to the vast increase in the number of households visited per wave compared with the BHPS. Second, GCSE results are not reported within the UKHLS main survey but are linked through official education records from the National Pupil Database (NPD). NPD data has been matched both retrospectively and prospectively based on consents given in Wave 1 of the main UKHLS survey. The GCSE results contained within the linked NPD data cover the academic years 2001/02 to 2012/13 (University of Essex, 2015).

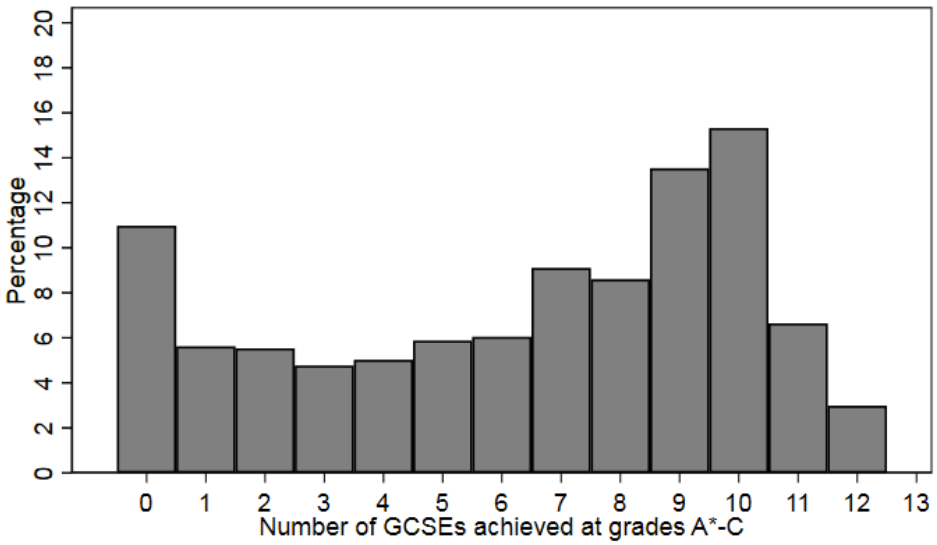
The synthetic cohorts of school Year 12 pupils in the UKHLS-NPD are identified in the adult interview of the main UKHLS survey using information on the academic year the pupil sat their GCSEs (available in the NPD data), the year they were born, and the month and year their interview took place (available in the main UKHLS survey). A complete records analysis is conducted for those of school Year 12 age. The

analytical sample size is n=1175, spanning waves 1 to 6 of the UKHLS and corresponding academic years 2009/10 to 2012/13. The potential influences of missing data on the analyses are examined in Chapter 3.

3 Descriptive Statistics

The outcome variable of interest in the subsequent analyses is GCSE attainment. Figure 2.1 graphs the number of GCSEs attained at grades A*-C for the synthetic cohorts of English Year 12 pupils in the UKHLS-NPD sample. There is a clear spike at zero as 10.98% of young people attained no A*-Cs but at least one GCSE at grades D-G. Approximately 13.5% of young people attained 9 A*-Cs, and 15.3% attained 10 A*-Cs. The maximum number of A*-Cs attained is 13. Due to statistical disclosure control, categories 12 and 13 are collapsed in Figure 2.1. The distribution of the number of GCSEs attained at grades A*-C is largely similar to the distribution for the BHPS sample.

Figure 2. 1: Number of GCSEs attained at grades A*-C for synthetic cohorts of Year 12 pupils in the UKHLS-NPD



Source: UKHLS-NPD linked unweighted data, n=1175.
Categories 12 and 13 combined due to statistical disclosure control.

The descriptive statistics for the explanatory variables are presented in Table 2.1. All explanatory variables are coded to match the previous chapter. The overall mean number of GCSEs attained at grades A*-C is 6.34, with a standard deviation of 3.71.

The mean number of GCSEs attained at grades A*-C in the UKHLS-NPD sample tend to be slightly higher than the BHPS sample for all independent variables.

Table 2. 1: Descriptive statistics of explanatory variables for synthetic cohorts of English Year 12 pupils in UKHLS-NPD dataset, unweighted

	Freq.	Percent	Mean A*-Cs	Std. Dev.
Attained 5+ GCSEs A*-C				
Yes	800	68.09	-	-
No	375	31.91	-	-
Parental Education Level				
Higher education	383	32.60	8.11	3.11
Further education	172	14.64	6.67	3.27
School-level education	504	42.89	5.47	3.66
Below school-level education	116	9.87	3.73	3.59
Parental NS-SEC				
1.1 Large employers and higher managerial	89	7.57	8.19	2.95
1.2 Higher professional occupations	113	9.62	8.67	2.40
2 Lower management and professional occupations	318	27.06	7.21	3.38
3 Intermediate occupations	121	10.30	6.57	3.45
4 Small employers and own account workers	163	13.87	5.80	3.74
5 Lower supervisory and technical occupations	71	6.04	4.70	3.35
6 Semi-routine occupations	194	16.51	4.70	3.82
7 Routine occupations	106	9.02	4.27	3.67
Housing Tenure				
Owned/privately rented	990	84.26	6.79	3.56
Social housing	185	15.74	3.88	3.50
Gender				
Male	571	48.60	5.82	3.82
Female	604	51.40	6.82	3.53
Ethnicity				
White	922	78.47	6.47	3.70
Mixed	51	4.34	6.08	4.06
Asian/Asian British	132	11.23	6.02	3.82
Black/Black British	58	4.94	5.22	3.16
Other ethnic group	12	1.02	5.83	3.41
Total *n*	**1175**	**100.00**	**6.34**	**3.71**

There are negative gradients in the mean number of good passes (i.e. A*-Cs) attained at GCSE level for parental education level and parental social class. Those living in owned or privately rented accommodation have higher mean number of GCSEs at grades A*-C than those living in social housing, and girls have a higher mean number of A*-Cs than boys. Over 78% of the UKHLS-NPD sample are from white backgrounds, but there is little difference in the mean number of GCSEs attained at

grades A*-C between those from white and those from ethnic minority backgrounds. There is a higher percentage of young people from ethnic minority backgrounds in the UKHLS-NPD sample compared with the BHPS sample. This is due to the inclusion of an Ethnic Minority Boost sample in the UKHLS survey from Wave 1. All statistical analyses in this chapter use appropriate corrections for the complex survey design and sample selection which account for the General Population Sample (GPS) and Ethnic Minority Boost (EMB) sample.

4 Modelling GCSE Attainment

The replication analyses begin by modelling the attainment of the policy benchmark of 5 or more GCSEs at grades A*-C, and the number of GCSEs attained at grades A*-C. The analyses from Chapter 1 are then extended to estimate linear regression models of GCSE point score, incorporating all grades for all GCSEs examined. This extension is possible due to the increased resolution of education data available in the linked official education records, compared with the measures routinely collected in large-scale social science surveys.

4.1 Attainment of 5 or More A*-Cs

The first measure examined is the policy benchmark measure of 5 or more GCSEs at grades A*-C. Tables 2.2 and 2.3 present the model building process for the logistic regression models. The tables outline the improvements in goodness-of-fit for models with each of the explanatory variables separately (Table 2.2), and for building the model sequentially (Table 2.3). Parental education and parental NS-SEC explain the largest portion of the variance in the outcome variable.

Despite the inclusion of an ethnic minority boost sample in the UKHLS, there continue to be low sample sizes for many ethnic groups in the UKHLS-NPD sample. Aggregation of groups is necessary and follows the Office for National Statistics guidance (Office for National Statistics, 2013). In Table 2.2, ethnicity is collapsed into a five-category variable and is not significantly associated with GCSE attainment. The within-group heterogeneity of the aggregated variable may contribute to the overall

lack of significance. For example, following the ONS collapses, Asian and Asian British comprise those from Indian, Pakistani, Bangladeshi, Chinese, and other Asian backgrounds. However, previous research has demonstrated that children from Indian or Chinese backgrounds tend to attain better results than those from white backgrounds whereas, for example, children from Pakistani or Bangladeshi backgrounds tend to attain lower results (see Demack et al., 2000, Connelly et al., 2013). In Table 2.3 and the regression models below, ethnicity is parameterised in the same way as in Chapter 1 and entered into the models as a control variable.

Table 2. 2: Goodness-of-fit summaries for explanatory variables and attainment of 5+ GCSEs at grades A*-C

Outcome variable: 5+ GCSEs A*-C	**Deviance**	**Δ Deviance** (from Null)	**Δ d.f.** (from Null)	**McFadden's Adjusted Pseudo R^2**
Null model	1471.632	-	-	-
Null + Parental Education	1359.516	112.116	3	0.071
Null + Parental NS-SEC	1332.121	139.511	7	0.084
Null + Housing Tenure	1410.729	60.903	1	0.039
Null + Gender	1453.715	17.917	1	0.009
Null + Ethnicity	1463.995	7.637	4	-0.002
Null + Academic Year	1457.906	13.725	4	0.003

Table 2. 3: Model building goodness-of-fit summaries for logistic regression model of attainment of 5+ GCSEs at grades A*-C

Outcome variable: 5+ GCSEs A*-C	**Deviance**	**Δ Deviance** (from previous model)	**Δ d.f.** (from previous model)	**McFadden's Adjusted Pseudo R^2**
Null model	1471.632	-	-	-
Null + Parental Education	1359.516	112.116	3	0.071
Null + Parental Education + Parental NS-SEC	1292.153	67.363	7	0.107
Null + Parental Education + Parental NS-SEC + Housing Tenure	1273.225	18.928	1	0.119
Null + Parental Education + Parental NS-SEC + Housing Tenure + Gender	1256.501	16.724	1	0.129
Null + Parental Education + Parental NS-SEC + Housing Tenure + Gender + Ethnicity	1255.375	1.126	1	0.128
Null + Parental Education + Parental NS-SEC + Housing Tenure + Gender + Ethnicity + Academic Year	1241.436	13.939	4	0.132

The logistic regression model is presented in Table 2.4. There are statistically significant associations between parental socio-economic background and GCSE attainment. There is not a significant interaction effect between parental education level and parental social class. Young people with at least one graduate parent have significantly higher log odds of attaining 5 or more GCSEs at grades A*-C compared with young people whose parents have lower education qualifications. The 95% quasi-variance comparison intervals demonstrate that young people whose parents have further education qualifications are not significantly different to those whose parents have school-level qualifications, and that those whose parents have school-level qualifications are not significantly different to those whose parents have below school-level qualifications.

Young people whose parents have occupations in NS-SEC 1.2 (i.e. higher professional occupations) have significantly higher log odds of attaining 5 or more GCSEs at grades A*-C than those whose parents have occupations in all other NS-SEC categories (with the exception of NS-SEC 1.1). The 95% quasi-variance comparison intervals demonstrate that the adjacent categories of NS-SEC are not significantly different from each other. Instead, there are distinct clustering effects between young people whose parents have managerial and professional occupations (NS-SECs 1.1, 1.2, and 2) compared with young people whose parents have routine and manual occupations (NS-SECs 5, 6, and 7).

Those living in social housing have significantly lower log odds of attaining 5 or more A*-Cs compared with those in owned or privately rented homes. Girls have significantly higher log odds of attaining 5 or more A*-Cs than boys.

Table 2. 4: Logistic regression model of the attainment of 5+ GCSEs at grades A*-C

				Quasi-Variance		
Logistic Regression: 5+ GCSEs A*-C	**Coef.**	**S.E.**		**S.E.**	**LCI**	**UCI**
Parental Education Level						
Higher education	Ref.	(.)		0.19	-0.39	0.39
Further education	-0.59	(0.28)	*	0.22	-1.05	-0.13
School-level education	-0.92	(0.22)	***	0.12	-1.18	-0.66
Below school-level education	-1.46	(0.35)	***	0.29	-2.06	-0.86
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.74	(0.60)		0.38	-1.54	0.06
1.2 Higher professional occupations	Ref.	(.)		0.47	-0.99	0.99
2 Lower management and professional occupations	-1.54	(0.50)	**	0.17	-1.91	-1.17
3 Intermediate occupations	-1.40	(0.53)	**	0.26	-1.95	-0.86
4 Small employers and own account workers	-1.98	(0.51)	***	0.20	-2.41	-1.56
5 Lower supervisory and technical occupations	-2.68	(0.57)	***	0.31	-3.33	-2.02
6 Semi-routine occupations	-2.34	(0.51)	***	0.21	-2.79	-1.89
7 Routine occupations	-2.22	(0.54)	***	0.25	-2.74	-1.70
Housing Tenure						
Owned/privately rented	Ref.	(.)		-	-	-
Social housing	-0.90	(0.23)	***	-	-	-
Gender						
Male	Ref.	(.)		-	-	-
Female	0.57	(0.17)	**	-	-	-
Constant	2.75	(0.52)	***			
Observations	1175					
McFadden's Adjusted Pseudo R^2	0.132					
Cox-Snell Pseudo R^2	0.178					
Nagelkerke Pseudo R^2	0.249					
BIC (d.f.)	1368.678 (18)					
BIC (based on deviance)	-6937.424					
BIC (based on likelihood ratio chi square)	-110.023					

Source: UKHLS-NPD synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Model also includes ethnicity and academic year.
** p<.05, ** p<.01, *** p<.001.*

As outlined in Chapter 1, the interpretation of log odds coefficients tend not to be immediately intuitive (Long and Freese, 2014). Table 2.5 presents the average marginal effects. This is the average change in probability of attaining 5 or more A*-Cs given a change in an explanatory variable, whilst holding all other variables at their observed values. The probabilities of attaining the benchmark attainment of 5 or more GCSEs at grades A*-C are significantly higher for young people from more educationally and occupationally advantaged socio-economic backgrounds.

Table 2. 5: Average marginal effects on probability of attaining 5+ GCSEs at grades A*-C

	Δ Prob.	S.E.	
Parental Education Level			
Higher education	Ref.	(.)	
Further education	-0.09	(0.05)	*
School-level education	-0.15	(0.04)	***
Below school-level education	-0.26	(0.07)	***
Parental NS-SEC			
1.1 Large employers and higher managerial	-0.07	(0.05)	
1.2 Higher professional occupations	Ref.	(.)	
2 Lower management and professional occupations	-0.18	(0.04)	***
3 Intermediate occupations	-0.16	(0.05)	**
4 Small employers and own account workers	-0.26	(0.05)	***
5 Lower supervisory and technical occupations	-0.41	(0.08)	***
6 Semi-routine occupations	-0.34	(0.06)	***
7 Routine occupations	-0.31	(0.06)	***
Housing Tenure			
Owned/privately rented	Ref.	(.)	
Social housing	-0.17	(0.05)	***
Gender			
Male	Ref.	(.)	
Female	0.09	(0.03)	**

The trends in the effect of parental socio-economic background on the attainment of 5 or more A*-Cs are consistent with the analyses of the earlier BHPS sample in Chapter 1. Young people whose parents have higher education qualifications or higher professional occupations (NS-SEC 1.2) continue to have a significant educational advantage in their GCSEs. More socio-economically advantaged young people have higher probabilities of attaining the policy standard benchmark of 5 or more GCSEs at grades A*-C than their less advantaged peers in the early 2010s. The 95% quasi-variance comparison intervals demonstrate that there might be a stronger clustering effect between the 'most' advantaged managerial and professional occupations in NS-SEC 1.1, 1.2 and 2 and the less advantaged routine and manual occupations in NS-SEC 5, 6, and 7.

The analyses in Chapter 1 and the current chapter use two datasets with different complex survey designs adjusting for survey design and selection strategies. Therefore, the analyses are undertaken in a consistent framework, rather than simultaneously in one model. Some comparisons can be made, for example, the

empirical regularity of a socio-economic background effect in the attainment of the policy standard benchmark for GCSE attainment across cohorts of young people throughout the 1990s, 2000s, and early 2010s. It is not possible to conclude that the effect of socio-economic background is stronger in the later cohorts (see Allison, 1999, Williams, 2009, Mood, 2010). There is, however, convincing evidence that parental socio-economic background continues to exert a powerful influence over the average attainment of young people into the early 2010s.

4.2 Number of GCSEs at Grades A*-C

The second measure examined in the replication analyses is the number of GCSEs attained at grades A*-C. There is a significant likelihood ratio test providing evidence that a negative binomial model is more suitable than a Poisson model (Long, 1997, Long and Freese, 2014). There is a significant Vuong test providing evidence that a zero-inflated negative binomial model is more suitable than a negative binomial model (Vuong, 1989). The output of the zero-inflated negative binomial model is reported in Table 2.6. The first section of the model presents the logistic regression model of attaining zero A*-Cs. The second section of the model presents the results of the negative binomial regression model for the count of A*-Cs attained.

Young people whose parents have school-level or below school-level qualifications have significantly higher log odds of attaining zero A*-Cs compared with young people whose parents have graduate qualifications. There is not a significant difference in attaining zero A*-Cs between young people whose parents have further education qualifications compared with higher education qualifications. Young people who have parents with occupations in NS-SEC 1.2 have significantly lower log odds of attaining zero A*-Cs compared with young people whose parents have occupations in NS-SECs 6 and 7. Boys and young people living in social housing also have higher log odds of attaining zero A*-Cs compared with girls and those living in owned or privately rented houses respectively.

Given that the young people attain some GCSEs at grades A*-C, young people whose parents have higher education qualifications have significantly higher log odds of

Table 2. 6: Zero-inflated negative binomial regression model of the number of A*-Cs

				Quasi-Variance		
Logistic Regression: Zero A*-Cs	**Coef.**	**S.E.**		**S.E.**	**LCI**	**UCI**
Parental Education Level						
Higher education	Ref.	(.)		-	-	-
Further education	0.48	(0.56)		-	-	-
School-level education	1.40	(0.41)	**	-	-	-
Below school-level education	1.80	(0.49)	***	-	-	-
Parental NS-SEC						
1.1 Large employers and higher managerial	1.49	(1.29)		-	-	-
1.2 Higher professional occupations	Ref.	(.)		-	-	-
2 Lower management and professional occupations	1.57	(1.10)		-	-	-
3 Intermediate occupations	1.15	(1.15)		-	-	-
4 Small employers and own account workers	1.68	(1.10)		-	-	-
5 Lower supervisory and technical occupations	1.73	(1.15)		-	-	-
6 Semi-routine occupations	2.53	(1.09)	*	-	-	-
7 Routine occupations	2.43	(1.11)	*	-	-	-
Housing Tenure						
Owned/privately rented	Ref.	(.)		-	-	-
Social housing	0.94	(0.29)	**	-	-	-
Gender						
Male	Ref.	(.)		-	-	-
Female	-0.64	(0.25)	*	-	-	-
Constant	-4.78	(1.14)	***			
Negative Binomial Regression: GCSEs A*-C						
Parental Education Level						
Higher education	Ref.	(.)		0.02	-0.05	0.05
Further education	-0.14	(0.04)	**	0.04	-0.22	-0.06
School-level education	-0.15	(0.04)	***	0.03	-0.21	-0.10
Below school-level education	-0.29	(0.09)	**	0.09	-0.47	-0.12
Parental NS-SEC						
1.1 Large employers and higher managerial	0.00	(0.04)		0.03	-0.06	0.07
1.2 Higher professional occupations	Ref.	(.)		0.03	-0.06	0.06
2 Lower management and professional occupations	-0.08	(0.04)	*	0.02	-0.13	-0.03
3 Intermediate occupations	-0.12	(0.06)	*	0.05	-0.22	-0.03
4 Small employers and own account workers	-0.19	(0.06)	**	0.05	-0.29	-0.09
5 Lower supervisory and technical occupations	-0.42	(0.09)	***	0.08	-0.60	-0.25
6 Semi-routine occupations	-0.21	(0.06)	**	0.06	-0.33	-0.09
7 Routine occupations	-0.28	(0.08)	***	0.07	-0.43	-0.13
Housing Tenure						
Owned/privately rented	Ref.	(.)		-	-	-
Social housing	-0.22	(0.07)	**	-	-	-
Gender						
Male	Ref.	(.)		-	-	-
Female	0.09	(0.03)	**	-	-	-
Log of alpha	-4.52	(1.11)	***			
Observations	1175					
McFadden's Adjusted Pseudo R^2	0.043					
Cox-Snell Pseudo R^2	0.256					
Nagelkerke Pseudo R^2	0.257					
BIC (d.f.)	6208.567 (37)					
BIC (based on deviance)	-2097.535					
BIC (based on likelihood ratio chi square)	-106.786					

Source: UKHLS-NPD synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Model also includes ethnicity and academic year.
** p<.05, ** p<.01, *** p<.001.*

attaining more A*-Cs compared with young people whose parents have any lower level of education qualifications. The 95% quasi-variance comparison intervals demonstrate that there are not significant differences between any other categories of parental education level. Given that young people attain some GCSEs at grades A*-C, young people whose parents have occupations in NS-SEC 1.2 have significantly higher log odds of attaining more A*-Cs compared with young people with parents in all other NS-SEC categories (except 1.1). The 95% quasi-variance comparison intervals demonstrate that there are not significant differences between adjacent categories of NS-SEC. Girls have significantly higher log odds of attaining higher counts of A*-Cs than boys, and individuals living in owned or privately rented houses have significantly higher log odds of attaining higher counts of A*-Cs than those living in social housing.

The post-estimation expected counts of the number of GCSEs attained at grades A*-C are presented in Table 2.7. Young people whose parents have higher levels of education and who have more advantaged occupations have higher expected counts of GCSEs at grades A*-C than their less advantaged peers. Expected counts can also be calculated for a combination of characteristics, sometimes called ‘ideal types’ (Long and Freese, 2014). Expected counts are calculated for the ‘most’ and ‘least’ advantaged. As in Chapter 1, the ‘most’ advantaged are defined as those with at least one graduate parent, with occupations in NS-SEC 1.2, and living in owned or privately rented houses. The ‘most’ advantaged males have an expected count of 8.84 A*-Cs and the ‘most’ advantaged females have an expected count of 9.72 A*-Cs. The ‘least’ advantaged are defined as those with parents with below school-level education, with occupations in NS-SEC 7, and living in social housing. The ‘least’ advantaged males have an expected count of 1.90 A*-Cs and the ‘least’ advantaged females have an expected count of 2.77 A*-Cs. The difference in the expected counts between the ‘most’ and ‘least’ advantaged young people is between 6.94 and 6.95 A*-Cs. For young people growing up in the early 2010s, the education attainment gap between the ‘most’ and ‘least’ advantaged continues to persist at a similar rate to the gap observed in Chapter 1 for young people growing up in the 1990s and 2000s.

Table 2. 7: Expected counts for the number of GCSEs attained at grades A*-C

	Expected Count	S.E.
Parental Education Level		
Higher education	7.63	(0.19)
Further education	6.53	(0.28)
School-level education	5.99	(0.20)
Below school-level education	4.96	(0.47)
Parental NS-SEC		
1.1 Large employers and higher managerial	7.58	(0.42)
1.2 Higher professional occupations	7.96	(0.29)
2 Lower management and professional	6.93	(0.23)
3 Intermediate occupations	6.80	(0.36)
4 Small employers and own account workers	6.17	(0.33)
5 Lower supervisory and technical occupations	4.86	(0.43)
6 Semi-routine occupations	5.53	(0.36)
7 Routine occupations	5.23	(0.42)
Housing Tenure		
Owned/privately rented	6.79	(0.12)
Social housing	4.99	(0.35)
Gender		
Male	6.09	(0.16)
Female	6.99	(0.14)

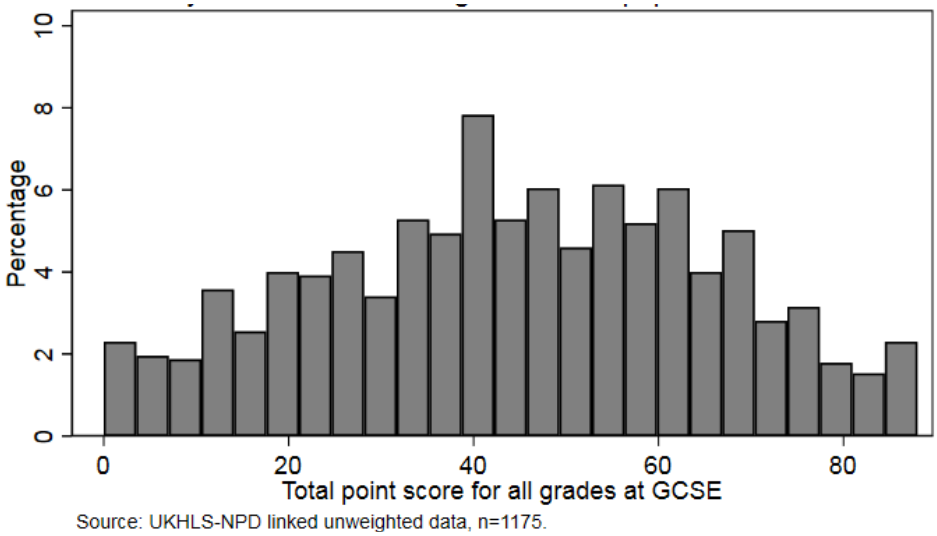
There is a remarkable empirical regularity in the results of the zero-inflated negative binomial regression models across the two synthetic cohort samples in the BHPS and the UKHLS-NPD. The datasets span over two decades and the analyses illustrate that there are clearly persisting socio-economic background effects in GCSE attainment for the young people sitting their GCSEs in the 1990s, 2000s, and early 2010s. In particular, the enduring effects of parental education and parental social class on the expected counts of GCSEs at grades A*-C is remarkably constant. This suggests that, despite overall, absolute increases in GCSE attainment across the cohorts, there is a consistent relative attainment gap between the 'most' and 'least' advantaged young people.

4.3 Overall GCSE Point Score

The NPD collects information at a finer level of detail than is routinely collected in social science surveys. The NPD data provide information on individual grades for each GCSE subject. GCSE grades are alphabetised rather than numeric. There are unlimited ways in which GCSE grades could be converted into numeric scores. This

work follows similar conventions in Demack et al. (2000) and Connolly (2006). Each grade A* is assigned 8 points, each grade A is assigned 7 points, each grade B 6 points, each C grade 5 points, each D grade 4 points, each E grade 3 points, each F grade 2 points, and each G grade 1 point. It is also possible to obtain a 'U' for unclassified, which is assigned 0 points. Some researchers have assigned 7 points to A* or As (for example, Demack et al., 2000, Croxford et al., 2007). Yang and Woodhouse (2001) commented that assigning an A* 8 points is standard practice. The A* grade was introduced in 1994 and therefore all young people in the synthetic cohort sample in the following analyses could theoretically access the A* grade. The distribution of the outcome variable is presented in Figure 2.2.[12]

Figure 2. 2: Total GCSE point score (grades A*-U) for synthetic cohorts of Year 12 pupils in the UKHLS-NPD



The results of the linear (OLS) regression model are presented in Table 2.8. Parental education, parental social class, housing tenure, and gender are statistically associated with GCSE point score. Young people with at least one graduate parent are associated with the equivalent of an extra A grade compared with those whose parents have further education qualifications; the equivalent of more than two extra C grades than those whose parents have school-level qualifications; and the equivalent

[12] Values are truncated at 88, or the equivalent of 11 A*s, in the graph due to statistical disclosure control. Regression models are estimated on the total GCSE point score, without truncation.

of more than three extra C grades than those whose parents have below school-level qualifications. The 95% quasi-variance comparison intervals (see Figure 2.3) demonstrate that there are significant differences between those whose parents have further education qualifications and all other categories. Young people whose parents have school-level qualifications are not significantly different from those whose parents have below school-level qualifications for overall GCSE point score.

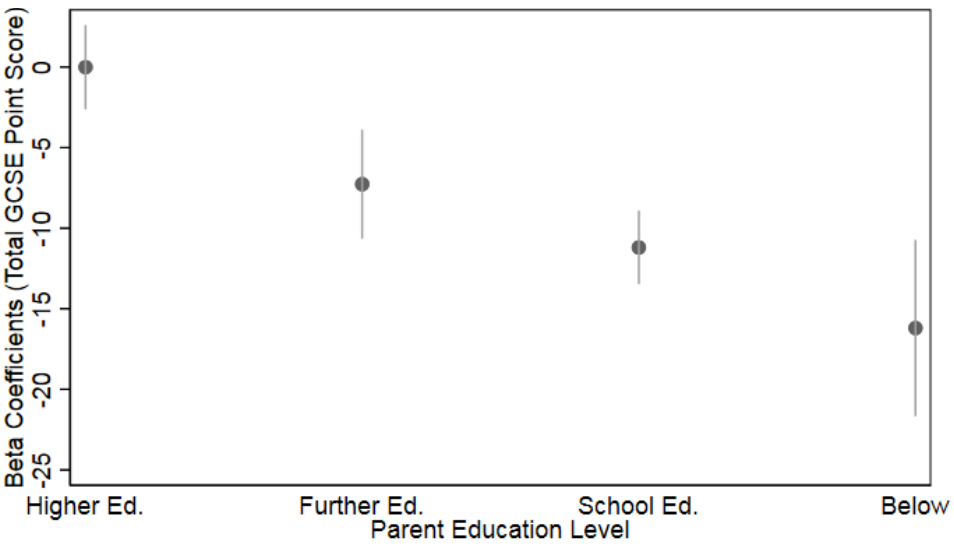
Young people with higher professional parents (NS-SEC 1.2), are associated with a higher overall GCSE point score than those with parents in less advantaged occupations, for example, in NS-SECs 2, 4, 5, 6, and 7. In particular, young people with higher professional parents are estimated to attain the equivalent of two extra A or B grades than those with parents in lower supervisory and technical, semi-routine, or routine occupations, i.e. routine and manual occupations. The quasi-variance graph of parental NS-SEC (see Figure 2.4) demonstrates a distinct clustering effect between managerial and professional occupations (NS-SECs 1.1, 1.2, and 2) and routine and manual occupations (NS-SECs 5, 6, and 7). Robustness checks for the continuous measure of GCSE point score illustrate that the trends persist when re-estimating with a truncated measure (truncated at the equivalent of 11 A*s), and with a measure of mean GCSE score (i.e. overall point score divided by total number of GCSEs taken).

Table 2. 8: Linear regression model of total GCSE point score

Linear Regression: GCSE Point Score	Coef.	S.E.		Quasi-Variance S.E.	LCI	UCI
Parental Education Level						
Higher education	Ref.	(.)		1.24	-2.62	2.62
Further education	-7.26	(2.00)	***	1.61	-10.65	-3.87
School-level education	-11.19	(1.65)	***	1.08	-13.47	-8.91
Below school-level education	-16.20	(2.93)	***	2.59	-21.66	-10.73
Parental NS-SEC						
1.1 Large employers and higher managerial	-1.89	(2.97)		2.39	-6.94	3.16
1.2 Higher professional occupations	Ref.	(.)		1.95	-4.12	4.12
2 Lower management and professional	-4.58	(2.23)	*	1.19	-7.09	-2.07
3 Intermediate occupations	-5.46	(2.93)		2.11	-9.91	-1.02
4 Small employers and own account workers	-8.69	(2.71)	**	1.89	-12.67	-4.72
5 Lower supervisory and technical occupations	-15.81	(3.16)	***	2.41	-20.90	-10.72
6 Semi-routine occupations	-13.08	(2.77)	***	1.90	-17.08	-9.07
7 Routine occupations	-14.52	(3.08)	***	2.29	-19.34	-9.70
Housing Tenure						
Owned/privately rented	Ref.	(.)		-	-	-
Social housing	-10.57	(1.99)	***	-	-	-
Gender						
Male	Ref.	(.)		-	-	-
Female	4.51	(1.20)	***	-	-	-
Constant	58.20	(2.35)	***			
Observations	1175					
Adjusted R^2	0.253					
BIC (d.f.)	10287.131 (18)					
BIC (based on deviance)	1981.029					
BIC (based on likelihood ratio chi square)	-239.967					

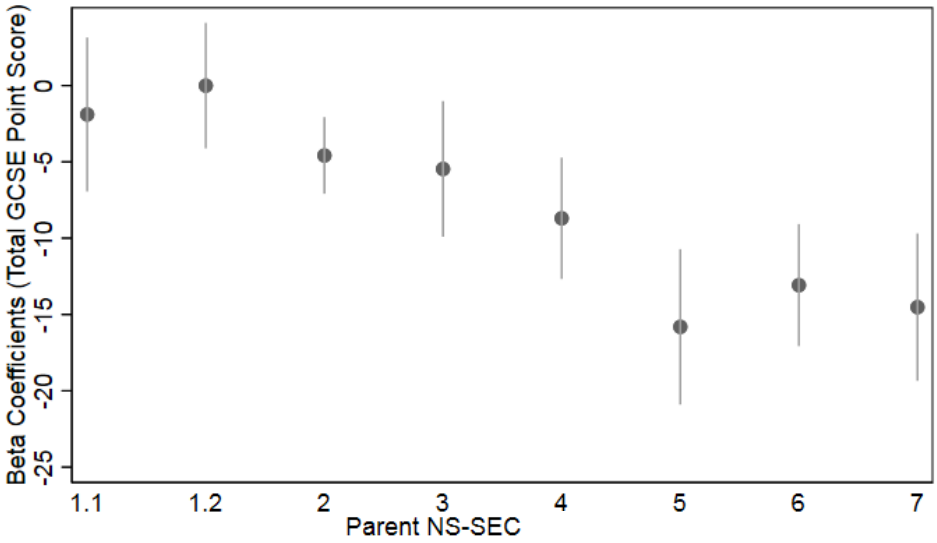
Source: UKHLS-NPD synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Model also includes ethnicity and academic year.
** p<.05, ** p<.01, *** p<.001.*

Figure 2. 3: Linear regression coefficients with 95% quasi-variance comparison intervals for total GCSE point score by parental education level



Source: UKHLS-NPD data, n=1175. Adjusted for complex survey design.
Parent NS-SEC, housing tenure, gender, ethnicity and academic year also included in model.

Figure 2. 4: Linear regression coefficients with 95% quasi-variance comparison intervals for total GCSE point score by parental social class



Source: UKHLS-NPD data, n=1175. Adjusted for complex survey design.
Parent education, housing tenure, gender, ethnicity and academic year also included in model.

5 Discussion and Conclusions

The analyses in this chapter successfully replicate the findings from Chapter 1. Chapter 1 focused on synthetic cohorts of Year 12 pupils over the course of the 1990s and 2000s. The current chapter focuses on synthetic cohorts of Year 12 pupils in the more contemporary period of the early 2010s.

The replication study employs the same analytical procedure on a new dataset in a more contemporary time period. The study has been designed to be as close as possible to the original analyses in Chapter 1. The central finding is that there is a stark empirical regularity of the role of parental socio-economic background in children's school GCSE attainment. There is an enduring advantage of having parents in higher professional occupations and with higher education qualifications. The advantage is evident when measuring GCSE attainment as the policy benchmark of 5 or more A*-Cs, or as the number of A*-Cs attained. The gap in the expected count of GCSEs attained at grades A*-C between the 'most' and 'least' advantaged young people is remarkably constant across the two sets of synthetic cohorts over the course of two decades. The 'most' advantaged (i.e. parents with higher education

qualifications, occupations in NS-SEC 1.2, and living in owned or privately rented accommodation) have an expected count of approximately 7 A*-Cs higher than the 'least' advantaged (i.e. parents with below school-level qualifications, occupations in NS-SEC 7, and living in social housing). Despite a general, absolute increase in GCSE attainment across the cohorts, a relative gap persists between the 'most' and 'least' advantaged English young people.

The analyses are extended using a continuous measure of GCSE point score. GCSE point score is a measure of attainment with much finer granularity than measures often routinely collected in social science surveys. The linear regression model results demonstrate a distinct clustering effect in overall attainment between young people from managerial and professional occupational backgrounds compared with young people from routine and manual occupational backgrounds.

The overall substantive message emerging from the empirical results is that socio-economic inequalities in GCSE attainment are not limited to the gap between the most disadvantaged and their less disadvantaged peers. The analyses suggest that there is a more nuanced, finely-grained socio-economic gradient to the educational outcomes at the end of compulsory schooling. Socio-economic inequalities are observed in relatively broad and blunt benchmark measures of attainment (for example, the attainment of 5 or more GCSEs at grades A*-C), as well as in relatively detailed, graduated measures of attainment (for example, overall GCSE point score). This has clear and substantial implications at an individual level for a young person's future opportunities and choices in their education, employment, training, and earnings after leaving school.

This has further sociological importance at a societal level. The findings in this chapter highlight that parental socio-economic inequalities in GCSE attainment are not an historic phenomenon. There is a strong, enduring socio-economic background effect in children's school GCSE attainment for some of the most contemporary cohorts of young people in the early 21st Century.

Chapter 3
Missing Data in the Synthetic Cohorts Analyses

1 Introduction

The presence of missing data is ubiquitous in social science surveys (Hawkes and Plewis, 2006, Longhi and Nandi, 2015). Missing data has the potential to produce biased estimates in statistical analyses (Treiman, 2009). Carpenter and Kenward (2013) advised that a complete records analysis should be the first step before attempting to address missing data. The authors argued that these analyses can provide valid inferences, but there is a potential for results to be inefficient (Carpenter and Kenward, 2013: 35). Mehmetoglu and Jakobsen (2017) noted that missing data are often dealt with using listwise deletion. The analyses in Chapters 1 and 2 were conducted on complete records, i.e. missing data on any of the variables in the analytical models were dealt with through listwise deletion. The analyses so far have adjusted for complex survey design and survey non-response. This chapter undertakes some principled approaches to handling missing data in the analyses of synthetic cohorts of English Year 12 pupils in the British Household Panel Survey. The missing data methods are then replicated using the UKHLS-NPD data. Due to statistical disclosure control, it is not possible to present missing data in the UKHLS-NPD sample in as much detail as the BHPS sample.

Section 2 briefly outlines the problem of missing data in social science research. Section 3 outlines the properties and patterns of missingness for the analytical variables in the BHPS sample, and then presents the results of different approaches to handling missing data for the BHPS sample. Section 4 replicates the multiple imputation methods for the UKHLS-NPD sample.

2 Missing Data

There are three general types of missingness: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). If data are MCAR,

missingness is not conditional on any other variable or on the values of the variable itself. Analyses with data MCAR are likely to provide valid estimates but potentially inefficient standard errors. If data are MAR, missingness may be conditional on other variables in the dataset, but not conditional on the values of the variable itself. Analyses with data MAR are less likely to provide valid inferences than data MCAR, but there are a number of potential statistical techniques to help to address missingness. If the data are MNAR, missingness may be conditional on other variables or on the variable itself. Analyses with data MNAR generally require the most comprehensive statistical methods to help to overcome the problems associated with missingness (see Treiman, 2009, Mehmetoglu and Jakobsen, 2017).

Listwise deletion (sometimes termed casewise deletion) has been used in the analyses so far. This removes all observations with missing values on any of the analytical variables. Following listwise deletion, the analyses have been undertaken on complete records. Treiman (2009), Carpenter and Kenward (2013), Mehmetoglu and Jakobsen (2017) advise against ad hoc methods of dealing with missing data, such as pairwise deletion, mean substitution, and creating explicit categories of missingness. Treiman (2009: 182) commented that the 'gold standard' of treating missing data in the social sciences is the use of multiple imputation methods. Multiple imputation involves the estimation of an imputation model to predict missingness in a variable, or series of variables, and the generation of several datasets imputing plausible values for the missing cases (Carpenter and Kenward, 2013).

3 Missing Data in the BHPS

The following sections examine the role of missing data in the analyses undertaken in Chapter 1 for synthetic cohorts of English school Year 12 pupils in BHPS households. The next sub-sections provide a description of the missing data in the analyses, and principled attempts to address the missing data. First, the variable with the largest proportion of missing data (parental NS-SEC) is explored. Second, all variables with missing data in the analytical model are multiply imputed by chained equations (Carpenter and Kenward, 2013). The missing data methods are undertaken

using the logistic regression models of the attainment of 5 or more GCSEs at grades A*-C at GCSE.

3.1 Description of the Missing Data

There are 1973 young people identified as English Year 12 pupils in the BHPS who indicated that they have GCSEs. The analytical sample of complete records in Chapter 1 had a sample size of n=1624. There are therefore 349 observations with missing data on at least one of the analytical variables. Of the 1973 eligible observations, 3 individuals indicated that they attained GCSEs at either grades A*-C or D-G, but they did not report how many they attained and therefore were not included in the complete records analyses. Of the 1973 eligible observations, 324 are missing data on parental NS-SEC, 99 are missing data on parental qualifications, 16 are missing data on housing tenure, and 8 are missing data on ethnicity. Gender and wave year have no missing data.

The patterns of missing data are illustrated in Table 3.1. Within the synthetic cohort sample, 82% have complete records on all analytical variables, 12% are missing values only on parental NS-SEC, and 4% have missing values on both parental measures. The remaining missing patterns constitute less than 1% of the sample (and only some of these patterns are presented).

Table 3. 1: Missing data patterns for the BHPS synthetic cohorts

Complete	Ethnicity	Housing Tenure	Parental Education	Parental NS-SEC
82%	✓	✓	✓	✓
12%	✓	✓	✓	
4%	✓	✓		
<1%	✓			
<1%	✓	✓		✓
<1%		✓	✓	✓

3.2 Missing Data for Parental NS-SEC

The variable with the largest proportion of missing data is parental NS-SEC. There are 324 young people who have missing data for parental NS-SEC. For 245 observations, there is missing data on parental NS-SEC because both parents are not in employment. This may be because the parents are, for example, unemployed, long-term sick or disabled, full-time carers, students, or retired. The NS-SEC categorisation is based on occupation type, managerial or supervisory duties, and organisation size (Rose and Pevalin, 2003). Individuals who are not active in the labour market cannot be easily assigned an NS-SEC category. Those not in employment can be considered as structurally missing. There are a further 79 missing values for parental NS-SEC. Of the 79 missing values, 54 are missing because neither mother nor father are in the household, and 25 are missing because the parents did not complete an interview.

A series of approaches to dealing with the large amount of missing values for parental NS-SEC are presented in Table 3.2. Model 1 presents an unweighted model for the complete records logistic regression model of the attainment of 5 or more GCSEs at grades A*-C. Unweighted analyses are naïve because they do not take into account complex survey design, and tend to underestimate standard errors (Treiman, 2009). Model 2 presents the model adjusting for complex survey design. This is the complete records logistic regression model presented in Chapter 1. Model 3 presents the results of a logistic regression model adjusting for complex survey design using a 9-category version of NS-SEC, i.e. including those not in employment as a separate group. This is a pragmatic solution, due to the structural missingness of this group. Further, it increases the sample size by 231. The substantive conclusions derived from Models 2 and 3 are the same. Model 4 presents the results of a logistic regression model adjusting for complex survey design and using a 9-category NS-SEC variable with the last observation carried forward (LOCF). This means that, in the event that an individual is not in employment in the year that their child reported their GCSEs, their NS-SEC from the previous wave is fed forward to fill in a missing data gap. In practice, very few cases changed as a result of carrying the last observation forward.

Table 3. 2: Logistic regression models of the attainment of 5+ GCSEs at grades A*-C addressing missing data for parental social class

Logistic Regression: 5+ GCSEs A*-C	Model 1 Unweighted: NS-SEC 8 Coef.	S.E.		Model 2 Svy-set: NS-SEC 8 Coef.	S.E.		Model 3 Svy-set: NS-SEC 9 Coef.	S.E.		Model 4 Svy-set: NS-SEC 9, LOCF Coef.	S.E.	
Parental Education Level												
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-0.93	(0.20)	***	-0.90	(0.24)	***	-1.04	(0.23)	***	-1.04	(0.23)	***
School-level education	-1.13	(0.22)	***	-1.19	(0.26)	***	-1.25	(0.25)	***	-1.24	(0.25)	***
Below school-level education	-1.89	(0.27)	***	-1.87	(0.32)	***	-2.18	(0.32)	***	-2.17	(0.32)	***
Parental NS-SEC												
1.1 Large employers and higher managerial	-0.28	(0.35)		-0.51	(0.36)		-0.49	(0.37)		-0.51	(0.36)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower managerial and professional occupations	-0.51	(0.26)	*	-0.57	(0.30)		-0.57	(0.30)		-0.54	(0.30)	
3 Intermediate occupations	-0.62	(0.30)	*	-0.63	(0.34)		-0.61	(0.34)		-0.63	(0.34)	
4 Small employers and own account	-0.96	(0.29)	**	-0.97	(0.33)	**	-0.94	(0.33)	**	-0.96	(0.33)	**
5 Lower supervisory and technical occupations	-1.06	(0.30)	***	-1.06	(0.33)	**	-1.03	(0.33)	**	-1.06	(0.33)	**
6 Semi-routine occupations	-1.15	(0.29)	***	-1.30	(0.35)	***	-1.24	(0.35)	**	-1.25	(0.35)	**
7 Routine occupations	-1.08	(0.31)	***	-1.01	(0.35)	**	-0.98	(0.34)	**	-0.99	(0.34)	**
8 Not in employment	-	-		-	-		-1.24	(0.33)	***	-1.23	(0.34)	***
Housing Tenure												
Owned or privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-1.11	(0.19)	***	-1.13	(0.23)	***	-1.09	(0.19)	***	-1.10	(0.19)	***
Gender												
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	0.58	(0.12)	***	0.58	(0.12)	***	0.60	(0.12)	***	0.60	(0.12)	***
Constant	1.38	(0.34)	***	1.45	(0.44)	**	1.47	(0.43)	**	1.44	(0.43)	**
Observations	1624			1624			1855			1855		
BIC (d.f.)	2065.612 (31)			2065.612 (31)			2335.666 (32)			2334.649 (32)		
BIC (based on deviance)	-9940.048			-9940.048			-11624.397			-11625.413		
BIC (based on likelihood ratio chi square)	-87.450			-87.450			-165.454			-166.470		

*Source: BHPS synthetic cohorts of English Year 12 pupils. Models also includes ethnicity and wave year. * p<.05, ** p<.01, *** p<.001.*

The substantive conclusions of the logistic regression model of the attainment of 5 or more GCSEs at grades A*-C remain the same across all models. The sample size has, however, increased as a consequence of adding in an extra category of NS-SEC for 'not in employment'. Including an extra category for those not in employment may be considered sub-optimum, but the results do not alter the conclusions made using the complete records analyses. Missing data for the variables of parental education, housing tenure, and ethnicity have not been addressed by these models and are addressed in the next section.

3.3 Multiple Imputation by Chained Equations

The next set of models address missing data on all analytical variables using multiple imputation by chained equations. The multiple imputation models are estimated using the mi suite in Stata (see mi, StataCorp, 2017a), which is compatible with the svy suite and can therefore adjust for complex survey design. However, there are limitations to using multiple imputation models. For example, goodness-of-fit measures, such as summary statistics of parsimony (BIC) and explanatory power (R^2), are not currently estimable using statistical software. It is not possible to compare goodness-of-fit statistics to ascertain the most appropriate model. Further, there is no clear guidance on the optimum number of imputations to estimate. The following models use 5 and 10 imputations. After 10 imputations, the imputation results are very consistent and the 10 imputation models are presented below.

Table 3.3 presents the multiple imputation analytical models. Model 1 presents an unweighted logistic regression model using a 9-category NS-SEC variable. Model 2 presents the logistic regression model results adjusting for complex survey design and using a 9-category NS-SEC variable. The results are fairly consistent between the two models, with more conservative standard error estimates for the model adjusting for complex survey design.

The next three models use variations of multiple imputation by chained equations. Model 3 is the first multiple imputation model. The imputation model and analytical model do not use survey weights or adjust for complex survey design. The mode

Table 3. 3: Logistic regression models of the attainment of 5+ GCSEs at grades A*-C using multiple imputation by chained equations

	Model 1 Unweighted			Model 2 Svy-set			Model 3 Unweighted MI			Model 4 Unweighted MI and svy model			Model 5 Svy MI and svy model		
Logistic Regression: 5+ GCSEs A*-C	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level															
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-1.06	(0.20)	***	-1.04	(0.23)	***	-1.06	(0.20)	***	-1.04	(0.23)	***	-1.04	(0.23)	***
School-level education	-1.20	(0.21)	***	-1.25	(0.25)	***	-1.19	(0.21)	***	-1.25	(0.25)	***	-1.25	(0.25)	***
Below school-level education	-2.16	(0.26)	***	-2.18	(0.32)	***	-2.10	(0.26)	***	-2.11	(0.32)	***	-2.11	(0.32)	***
Parental NS-SEC															
1.1 Large employers and higher man.	-0.27	(0.35)		-0.49	(0.37)		-0.29	(0.35)		-0.45	(0.37)		-0.45	(0.37)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower managerial and professional	-0.51	(0.26)	(*)	-0.57	(0.30)		-0.56	(0.26)	*	-0.55	(0.30)		-0.55	(0.30)	
3 Intermediate occupations	-0.60	(0.30)	*	-0.61	(0.34)		-0.58	(0.29)	*	-0.56	(0.33)		-0.56	(0.33)	
4 Small employers and own account	-0.93	(0.29)	**	-0.94	(0.33)	**	-0.94	(0.28)	**	-0.90	(0.33)	**	-0.90	(0.33)	**
5 Lower supervisory and technical	-1.03	(0.29)	***	-1.03	(0.33)	**	-1.06	(0.29)	***	-1.03	(0.33)	**	-1.03	(0.33)	**
6 Semi-routine occupations	-1.08	(0.29)	***	-1.24	(0.35)	**	-1.12	(0.29)	***	-1.24	(0.35)	**	-1.24	(0.35)	**
7 Routine occupations	-1.05	(0.31)	**	-0.98	(0.34)	**	-1.07	(0.30)	***	-1.01	(0.34)	**	-1.01	(0.34)	**
8 Not in employment	-1.09	(0.30)	***	-1.24	(0.33)	***	-1.09	(0.29)	***	-1.22	(0.34)	***	-1.22	(0.34)	***
Housing Tenure															
Owned or privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-1.11	(0.16)	***	-1.09	(0.19)	***	-1.13	(0.16)	***	-1.07	(0.19)	***	-1.07	(0.19)	***
Gender															
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	0.60	(0.11)	***	0.60	(0.12)	***	0.58	(0.10)	***	0.61	(0.12)	***	0.61	(0.12)	***
Constant	1.40	(0.34)	***	1.47	(0.43)	**	1.44	(0.33)	***	1.50	(0.43)	**	1.50	(0.43)	**
Observations	1855			1855			1973			1973			1973		

*Source: BHPS synthetic cohorts of English Year 12 pupils. Models also include ethnicity and wave year. * p<.05, ** p<.01, *** p<.001.*

does not impute the values of NS-SEC for those structurally missing, i.e. not in employment. The results are consistent with the unweighted complete records model (Model 1). Model 4 presents the results of an unweighted multiple imputation model and weighted analytical model (i.e. adjusting for complex survey design). The estimates and standard errors are consistent with the weighted model presented in Model 2. Finally, Model 5 presents the results of a weighted imputation model and a weighted analytical model. This model adjusts for complex survey design in both the imputation and analytical stages of the multiple imputation process. Models 4 and 5 are almost identical. The substantive conclusions regarding effect size, sign and significance of the explanatory variables and GCSE attainment are consistent across all models which adjust for complex survey design, selection, and item non-response. Furthermore, the results of the multiple imputation by chained equation models are consistent with the complete records analyses in Chapter 1.

4 Missing Data in the UKHLS

This section examines the role of missing data in the UKHLS-NPD sample and compares the complete records analyses presented in Chapter 2 with models using multiple imputation by chained equations.

There are a total of 1472 eligible observations of synthetic cohorts of school Year 12 pupils in the UKHLS-NPD dataset. The complete records sample in Chapter 2 had 1175 observations. Therefore, there are 297 cases with missing values on at least one of the analytical variables. The variable with the largest proportion of missing data is parental NS-SEC. Most of the missing data on parental NS-SEC is due to neither parent being in employment. The addition of an extra category for those not in employment adds an additional 237 individuals to the analysis (n = 1412). Individuals not in employment can be considered as structurally missing. After adding in a ‘not in employment’ category to the NS-SEC variable, 96% of the sample had complete records. A further 60 observations had missing data on at least one of the explanatory variables. There is a low number of missing values to be imputed. The Secure Lab procedures prevent potentially disclosive information from being released from the

secure server, including tables with very low sample sizes. A table of missingness has not been included due to statistical disclosure control.

Linear regression models of overall GCSE point score using multiple imputation by chained equations are presented in Table 3.4. Model 1 presents the complete records model used in Chapter 2 which adjusts for complex survey design. Model 2 presents the complete records model using a 9-category variable of NS-SEC, adjusting for complex survey design, and therefore increasing the sample size to 1412. The results are largely consistent across the two models. Models 3 and 4 present the 5 imputation model and 10 imputation model. The models use multiple imputation by chained equations. Both of the imputation and analytical models use survey weights and adjust for complex survey design. The results are very consistent. The substantive conclusions of the complete records linear regression models remain unchanged after attempting to handle the missing data through multiple imputation methods.

Table 3. 4: Linear regression models of total GCSE point score using multiple imputation by chained equations

Linear Regression: GCSE Point Score	Model 1 Complete records NS-SEC 8 Coef.	S.E.		Model 2 Complete records NS-SEC 9 Coef.	S.E.		Model 3 MI 5 Imputations Coef.	S.E.		Model 4 MI 10 Imputations Coef.	S.E.	
Parental Education Level												
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-7.26	(2.00)	***	-8.01	(1.93)	***	-8.09	(1.89)	***	-8.09	(1.89)	***
School-level education	-11.19	(1.65)	***	-12.05	(1.57)	***	-12.12	(1.56)	***	-12.12	(1.56)	***
Below school-level education	-16.20	(2.93)	***	-15.76	(2.38)	***	-15.64	(2.34)	***	-15.64	(2.33)	***
Parental NS-SEC												
1.1 Large employers and higher managerial	-1.89	(2.97)		-1.81	(2.98)		-1.94	(2.92)		-1.94	(2.92)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	-4.58	(2.23)	*	-4.39	(2.23)	*	-4.54	(2.20)	*	-4.54	(2.20)	*
3 Intermediate occupations	-5.46	(2.93)		-4.98	(2.89)		-5.29	(2.83)		-5.29	(2.83)	
4 Small employers and own account workers	-8.69	(2.71)	**	-8.43	(2.69)	**	-8.85	(2.64)	**	-8.85	(2.64)	**
5 Lower supervisory and technical occupations	-15.81	(3.16)	***	-15.45	(3.11)	***	-15.88	(3.06)	***	-15.88	(3.05)	***
6 Semi-routine occupations	-13.08	(2.77)	***	-12.88	(2.82)	***	-13.20	(2.75)	***	-13.20	(2.75)	***
7 Routine occupations	-14.52	(3.08)	***	-14.18	(3.02)	***	-14.50	(2.95)	***	-14.50	(2.95)	***
8 Not in employment	-	-	-	-13.81	(2.60)	***	-14.03	(2.56)	***	-14.03	(2.55)	***
Housing Tenure												
Owned/privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-10.57	(1.99)	***	-10.82	(1.65)	***	-11.13	(1.58)	***	-11.13	(1.58)	***
Gender												
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	4.51	(1.20)	***	4.04	(1.08)	***	4.00	(1.04)	***	4.00	(1.04)	***
Constant	58.20	(2.35)	***	58.83	(2.22)	***	58.73	(2.14)	***	58.73	(2.14)	***
Observations	1175			1412			1472			1472		

*Source: UKHLS-NPD synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Models also include ethnicity and academic year. * p<.05, ** p<.01, *** p<.001.*

5 Discussion and Conclusions

In Chapters 1 and 2, missing data were treated using listwise deletion, and therefore any observation with missing data on any of the analytical variables was dropped from the sample. Missing data has the potential to bias statistical estimates (Treiman, 2009). The main aim of this chapter is to check the robustness of the complete records analyses to the presence of missing data.

Parental NS-SEC is the most problematic analytical variable in terms of missing data for both samples of synthetic cohorts of English Year 12 pupils. The main explanation for missingness in both samples is that the parents are not in employment. Less prominent reasons for the missing data are that the parents are not in the household or survey, or that the parents did not respond to the interview in that wave. Those not in employment can be considered as structurally missing because they cannot have an NS-SEC if they do not have an occupation. Instead of imputing these values, a 9-category NS-SEC variable is used with an extra category for those not in employment.

There are potential limitations of using a category for those not in employment. For example, people who are not in employment are not a homogenous group. Members of this group are empirically diverse, and may be unemployed, long-term sick or disabled, retired, students, or full-time carers. Rose and Pevalin (2003) advised that researchers should make their own analytical judgements about the treatment of the non-employed. Where data allows, the non-employed can be allocated to their last paid job in order to improve sample size. The changes as a result of using the last observation carried forward in the analyses above are negligible. Rose and Pevalin (2003) further suggested that researchers may want to include a separate category of long-term unemployed or never worked, but they comment that this is not easy to define or implement. Constructing categorical variables with separate categories of missingness is not generally advised in the missing data literature (Carpenter and Kenward, 2013, Mehmetoglu and Jakobsen, 2017). The analyses in Chapters 1 and 2 of this thesis adhered to the general methodological advice. The inclusion of a separate category for those not in employment does not, however, adversely affect

the substantive conclusions of the complete records analyses and greatly improves the sample sizes. This could not have been known *a priori*.

Treiman (2009: 182) commented that the 'gold standard' of treating missing data in the social sciences is the use of multiple imputation methods. The analyses in this chapter have made a concerted attempt to move towards this 'gold standard'. Multiple imputation is not a panacea, for example, statistical software cannot currently provide goodness-of-fit measures to compare models, and it is difficult to ascertain the appropriate number of imputations. Multiple imputation models tend to be very computationally intensive and can take time to converge. General technological advances and increased computing power help to navigate this, although this still remains problematic in a Secure Lab environment.

A number of alternative methods of addressing missing data are attempted in this chapter and the results are highly consistent. The analyses improve the confidence of the robustness of the original findings in Chapters 1 and 2. The models using multiple imputation by chained equations return the same substantive conclusions as the complete records analyses for the synthetic cohorts of Year 12 pupils using the BHPS and the UKHLS. Young people from more socio-economically advantaged backgrounds tend to have better, and higher, GCSE outcomes than their less socio-economically advantaged peers and the persisting effect is observable over the course of more than two decades.

Part 2:

Exploring Potential Sociological Explanations for the Persisting Socio-Economic Effect

Introduction to Part 2

The analyses in the preceding Part have presented convincing empirical evidence that parental socio-economic background plays an enduring role in GCSE attainment. An important aspect of this work is that the results are robust to alternative measures, alternative model specifications, and have been replicated in two nationally representative British household panel surveys spanning more than two decades.

The following chapters explore the relationship between parental socio-economic background and GCSE attainment in greater depth. Part 2 of this thesis investigates the empirical basis of three potential explanations of the persisting effect: prior attainment, cultural capital, and educational aspirations. Part 2 presents three inter-related pieces of data analysis. The guiding principle underlying the three discrete pieces of work is to better understand the contemporary relationship between parental socio-economic background and children's school GCSE attainment.

Chapter 4 examines the potential implications of prior attainment. Cognitive and educational outcomes at earlier stages of schooling are stratified by parental socio-economic background (Feinstein, 2004, Gregg and Washbrook, 2011, Chowdry et al., 2011, Crawford et al., 2017, Connelly and Gayle, 2019). The inequalities observed in GCSE attainment may be the consequence of inequalities observed at earlier stages of a young person's education. Path analysis models are used to decompose the effects of parental education and parental social class on attainment at the end of compulsory secondary school using attainment at the end of primary school. Detailed measures of parental education and parental occupation are seldom available in administrative datasets. Linked official records for educational attainment are seldom

available in large-scale social science surveys. The advantage of using the UKHLS-NPD data thereby provide an original opportunity to model parental socio-economic background and school attainment using detailed individual measures for contemporary cohorts of young people.

Chapter 5 investigates the role of cultural capital in educational inequalities. A prominent sociological explanation for persisting educational inequalities is Bourdieu's concept of cultural capital (Sullivan and Brown, 2015). The use of cultural capital in quantitative sociological research provides substantial operationalisation challenges (Lamont and Lareau, 1988, Sullivan, 2002). The UKHLS contains a wealth of candidate measures for the concept of cultural capital. Measures are collected in the adult survey and in the youth questionnaires, providing an opportunity to examine the intergenerational aspect of parent and child cultural capital. There are no agreed-upon measures of cultural capital and therefore a key aspect of this work is the attention to sensitivity analyses of alternative measures. The candidate measures are compared and contrasted in the analyses and particular attention is paid to the effect such measures have on the association between parental socio-economic background and GCSE attainment.

Chapter 6 explores the role of educational aspirations in educational inequalities. 'Raising aspirations' has been at the core of recent UK government rhetoric to help to address the attainment gap between the most disadvantaged and more advantaged young people (Cabinet Office, 2008, Cabinet Office, 2009, Cabinet Office, 2011). The attainment gap has been, in part, attributed to the 'low' aspirations held by young people and their parents. Previous research has demonstrated that aspirations tend to be universally high (St Clair and Benjamin, 2011, Croll and Attwood, 2013), and there is no evidence of a 'poverty of aspirations' for young people (Treanor, 2017). Aspirations have often been modelled using cross-sectional data. The analyses use repeated contacts data and estimate panel regression models to explore socio-economic effects in young people's aspirations. A further analysis estimates linear regression models to examine the influence of parent and child educational aspirations on GCSE attainment, and the extent to which this accounts for socio-economic differences.

The analyses in Part 2 use sub-samples of the UKHLS youth questionnaire and linked NPD records. Young people aged between 11 and 15 in the UKHLS households are invited to take part in the youth panel. Each sub-sample is described in more detail within the chapters. The main outcome variable for all analyses is total GCSE point score (see Chapter 2). The key explanatory variables are parental education, parental social class, housing tenure, gender, ethnicity, and academic year. An additional control for age is used in the cultural capital and educational aspirations analyses. The analyses undertaken in Part 2 require certain questions to be answered in the main UKHLS surveys which are part of a rotating questionnaire design and are therefore not available each wave. This has consequences for sample sizes as it restricts eligible observations and waves.

There are differences in the measurement of parental social class between Part 1 and Part 2. In Part 1, NS-SEC was measured using an 8-category measure. In Part 2, NS-SEC is measured using an additional category for parents 'not in employment'. The use of an additional 'not in employment' category in the analyses in Part 2 has been largely motivated by the desire to maintain a suitably large sample size of valid observations and associated statistical power. It is appreciated that there are potential limitations of using a 'not in employment' category (see Chapter 3) and therefore less emphasis has been placed on the substantive interpretation of membership of this group. The principled and formal approaches to missing data undertaken in Chapter 3 provided encouraging evidence that the use of a 9-category NS-SEC measure does not adversely influence the substantive findings of analyses using an 8-category measure. The analyses in Part 2 were re-estimated using an 8-category NS-SEC variable, i.e. treating all those with parents 'not in employment' as missing, and the substantive conclusions presented in the following chapters remain, albeit with much reduced sample sizes and statistical power.

Chapter 4
Socio-Economic Differences in Prior Attainment and GCSE Attainment

1 Introduction

Parental socio-economic background is strongly associated with attainment at the end of compulsory schooling in contemporary England (see Part 1). Socio-economic inequalities have been similarly observed at earlier stages of children's cognitive and educational development (for example, see Feinstein, 2004, Dearden et al., 2011, Gregg and Washbrook, 2011, Chowdry et al., 2011, Sullivan et al., 2013, Crawford et al., 2017, Connelly and Gayle, 2019). Cognitive ability refers to 'the ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought' (Neisser et al., 1995: 77). Educational attainment in this chapter, and throughout this thesis, refers to the test scores and grades obtained in standardised school examinations. The examinations are based on the content and curriculum studied throughout primary and secondary schooling, and the results reflect educational attainment at significant educational stages.

Early socio-economic inequalities can be consequential for inequalities in later educational stages. For example, inequalities in cognitive ability from a young age are associated with later educational attainment (see Lee and Burkam, 2002, Feinstein, 2003). Earlier education attainment can also be consequential for later education attainment. For example, socio-economic background tends to play a diminished role in explaining differences in higher education participation after controlling for prior educational attainment, due to selection effects and socio-economic inequalities influencing attainment at earlier stages (Galindo-Rueda et al., 2004, Marcenaro-Gutierrez et al., 2007, Broecke and Hamed, 2008, Chowdry et al., 2013, Smith, 2015).

This chapter focuses on the role of socio-economic inequalities in educational attainment at the end of primary school (aged 11) and secondary school (aged 16). The motivation for this chapter is to better understand the timing of socio-economic effects in school attainment. The analyses attempt to explore whether the socio-economic differences observed in GCSE attainment are the consequences of earlier stratification, or whether the effects of parental socio-economic background persist over and above the role of prior attainment. The research question for this chapter is *To what extent does prior attainment account for socio-economic differences in GCSE attainment?*

Section 2 reviews the literature on socio-economic background and school attainment over the course of primary and secondary school years. Section 3 outlines the data and methods used in the subsequent analyses. Section 4 provides descriptive statistics, and section 5 presents statistical modelling results. Sensitivity analyses of alternative measures of attainment at age 11 and school type are also detailed.

2 School Attainment

Young people in England are periodically examined throughout their primary and secondary schooling at ages 7, 11, 14, and 16 (Machin and Vignoles, 2006). The following sub-sections outline the Key Stage system and associated National Curriculum for English state schools. Next, empirical work examining the role of parental socio-economic background throughout primary and secondary schooling is summarised. Particular attention is paid to the datasets and measures of socio-economic background used in previous empirical work.

2.1 Key Stages and National Curriculum Levels

GCSE examinations were introduced by the 1988 Education Reform Act. The Act also introduced a National Curriculum for all state schools in England and Wales. The National Curriculum is divided into five Key Stages over the course of primary and secondary schooling. Government guidelines suggested that pupils should progress

through each Key Stage over the course of two years (Department for Education and Skills and Qualifications and Curriculum Authority, 2004).[13]

Standard Assessment Tests (SATs) are scheduled at the end of each Key Stage (Withey and Turner, 2015). Key Stage 1 tests are taken at the age of 7 (school Year 2). Key Stage 2 tests are taken at the end of primary school at the age of 11 (school Year 6). Key Stage 3 tests were taken at the age of 14 (school Year 9), but were abolished in 2009 (Goodman and Burton, 2012). GCSE examinations are the most common tests taken at the end of Key Stage 4 at the age of 16 (school Year 11). The examinations taken in Key Stage 5 are General Certificate of Education Advanced Subsidiary (AS) Levels and Advanced (A') Levels, taken at the ages of 17 and 18, after the end of compulsory schooling for the young people in these analyses. They are also the main entry route into university or higher education.

The National Curriculum and SATs testing was introduced, in part, to improve the accountability of schools (Whetton, 2009). The Dearing reforms from 1994 consolidated eight National Curriculum Levels as indicators of achievement as young people progress through the Key Stages (Whetton, 2009). GCSE examinations have a separate grading system to the National Curriculum Levels. English and Mathematics are assessed at each Key Stage, and Science is also assessed at Key Stages 2 and 3 (Reeves et al., 2001). Key Stage assessments were initially marked internally by teachers (Wyse and Torrance, 2009). Following the Dearing reforms, Key Stage 2 and 3 tests in English, Mathematics, and Science were to be externally examined (Whetton, 2009). 'High stakes' tests can have a restrictive effect on the curriculum and lead to 'teaching to the test' rather than taking a broader view of achievement (Wyse and Torrance, 2009). 'High stakes' tests have also been criticised for the effects on pupils, such as increased levels of stress and anxiety (Connor, 2003). The introduction of the Key Stage tests were understood to initially raise the overall achievement of pupils (Connor, 2003), but this stabilised in the early 2000s (Wyse and Torrance, 2009).

[13] Also see archived information from the Department for Education on National Curriculum Levels (October 2010) at https://webarchive.nationalarchives.gov.uk/20101021174848/https://nationalstrategies.standards.dcsf.gov.uk/print/169945 [accessed 17.05.19].

2.2 Socio-Economic Background and School Attainment

There is a large body of empirical work examining the effects of socio-economic background and formal qualifications taken at the end of compulsory schooling using detailed measures of parental social class and parental education level (see section 2.2, Introduction). Key Stage test results are not routinely reported in large-scale social science surveys and therefore tend to require the use of administrative data. Administrative datasets do not routinely collect detailed measures of socio-economic position. Empirical research examining the role of socio-economic background and educational attainment throughout primary and secondary schooling tend to use proxy indicators for low income. Eligibility for Free School Meals (FSM) is a standard measure of deprivation or disadvantage in the education system in England (Steele et al., 2007, Department for Education, 2015a, Department for Education, 2015b). FSM eligibility is an indicator of relative poverty, and is based on the receipt of other welfare benefits such as income support (Gorard and Siddiqui, 2019).

Locality-based analyses of pupil attainment in primary and secondary schools have demonstrated that eligibility for Free School Meals is strongly associated with attainment over the Key Stages (Sammons et al., 1997, Sammons and Smees, 1998, Strand, 1999, Strand, 2014b). Sammons et al. (1997) analysed Key Stage 1 assessments for pupils in 62 London primary schools and demonstrated significant effects of background characteristics, including FSM eligibility. Sammons and Smees (1998) examined Key Stage 1 assessments for pupils in 107 Surrey primary schools and found that eligibility for FSM was significantly associated with attainment, even after controlling for prior attainment. Strand (1999) examined pupils' academic progress between the ages of 4 and 7 in 55 London primary schools and found that eligibility for FSM was significantly associated with lower attainment at both ages, and that the gap widened over time. Strand (2014b) examined pupils' Key Stage attainment at ages 7 and 11 in 68 London primary schools and found evidence of significant interactions between ethnicity and socio-economic status (measured by eligibility for Free School Meals and area-level deprivation).

Longitudinal analyses of educational attainment have further demonstrated that socio-economic attainment gaps (measured by FSM eligibility) may widen over primary and secondary schooling. Sammons (1995) used the School Matters pupil cohort data over a 9-year academic period from Year 3 to Year 11 to assess relative differences in attainment. Socio-economic background was measured by father's occupation and receipt of Free School Meals. The study found that the effects of socio-economic background widened in later education stages. Crawford et al. (2017) used linked administrative data from the National Pupil Database, the Individual Learner Records, the Further Education administrative dataset, and the Higher Education Statistics Agency to analyse attainment at all Key Stages, and at university. Socio-economic position was derived from eligibility for Free School Meals and an area-level deprivation measure (Index of Multiple Deprivation). There were clear socio-economic gaps at each level of education, with evidence of a widening attainment gap between ages 7 and 19 (Crawford et al., 2017).

The FSM measure has a clear, legal definition and can be assumed to be consistently collected in education datasets for all pupils (Gorard and See, 2009). Taylor (2018) examined the reliability of FSM eligibility as a measure of socio-economic position using the Millennium Cohort Study and linked administrative data for Wales. The study found that, whilst not a perfect measure, FSM eligibility does provide a pragmatic and reliable approach to measuring disadvantage in educational research. However, using FSM eligibility at one point in time does not take into consideration longer-term effects of relative poverty (Gorard and Siddiqui, 2019). Jenkins and Jarvis (1998) commented that household incomes tend to change year-on-year, with greater income flux at the very top and bottom of the income distribution. Goldthorpe and McKnight (2004) similarly noted that measures of income are less stable indicators of socio-economic position than occupation-based measures. The NS-SEC measure of social class has been specifically developed for social and policy research to reflect socio-economic differences in society (Rose and Pevalin, 2003, Office for National Statistics, 2010). NS-SEC can therefore provide a more stable measure to examine socio-economic inequalities than a proxy indicator for low income. Further, FSM eligibility compares a minority of the most disadvantaged pupils, i.e. those living in relative poverty, to a majority of non-disadvantaged pupils (Gorard and See, 2009).

The work in Part 1 of this thesis has demonstrated a clear and substantial socio-economic gradient in GCSE attainment across a broader spectrum of educational and occupational advantage. The nuance of this is unlikely to be captured using a simple, binary measure, such as eligibility for FSM.

Large-scale social science surveys with linked administrative education records allow more detailed measures of socio-economic background to be used in educational analyses. A Special Issue of the Longitudinal and Life Course Studies journal in 2011 used a collection of large-scale social science surveys with linked administrative data to examine educational and cognitive outcomes at pre-school (Dearden et al., 2011), primary school (Gregg and Washbrook, 2011), and secondary school (Chowdry et al., 2011). All of the studies in the Special Issue measured socio-economic background based on the log of equivalised household income (i.e. household income divided by the number of household members), self-reported financial difficulties, the occupational class of mothers and fathers, and housing tenure. Principal components analysis was used to construct the measure, which was then divided into quintile ranks (Goodman et al., 2011). The Key Stage tests collected in administrative datasets were used to construct percentile scores for the young people in the respective samples. Gregg and Washbrook (2011) used the Avon Longitudinal Study of Parents and Children (ALSPAC) to explore the role of attitudes and behaviours in educational attainment at age 11 (i.e. Key Stage 2). The authors found that approximately a third of the attainment gap between the most and least advantaged observed at age 11 was evident by age 7 (i.e. Key Stage 1). Chowdry et al. (2011) reported similar findings using three waves of the Longitudinal Study of Young People in England (LSYPE). The authors found that approximately two-fifths of the attainment gap between the most and least advantaged at age 16 could be explained by prior attainment at age 11. The findings suggested that socio-economic gaps in prior attainment, i.e. at earlier stages of the education system, are important factors in a young person's later attainment (Chowdry et al., 2011).

Longitudinal social science surveys provide further opportunity to track individual attainment over time. Feinstein (2004) analysed the UK 1958 and 1970 birth cohorts and noted that there was a compounding effect of the educational attainment gap

between more and less advantaged pupils throughout primary and secondary schooling. Strand (2014a) analysed the effects of gender, ethnicity, and socio-economic status for attainment at ages 11, 14, and 16 using the LSYPE. Socio-economic status was measured using a composite measure of socio-economic classification, parent education qualifications, FSM eligibility, home ownership, and an area-level measure of deprivation. The results demonstrated that there were socio-economic differences in educational attainment and progress between the ages of 11 and 16, as well as significant interaction effects between gender, ethnicity, and socio-economic status.

Analyses of formal educational qualifications at Key Stages 4 and 5 using large-scale surveys have further demonstrated that social class effects may be more pronounced at higher levels of attainment. For example, Scott (2004) analysed the BHPS and demonstrated that social class effects were larger at A' Level than GCSE level, emphasising the importance of selection effects. Sullivan et al. (2011) analysed the Youth Cohort Study and demonstrated that social class inequalities existed in GCSE attainment and at AS and A' Level. There were larger proportionate gaps (i.e. greater distance between the most and least advantaged social class categories) at higher levels of attainment. Morris et al. (2016) analysed the ALSPAC dataset and found that socio-economic inequalities (including a measure of social class) in GCSE and A' Level attainment persisted over and above the effects of cognitive ability measured at age 8 (Morris et al., 2016).

As outlined above, previous empirical work examining socio-economic effects and attainment throughout primary and secondary schooling have generally been restricted by data and measure availability. The analyses in this chapter use the UKHLS-NPD linked data for contemporary cohorts of young people in the 2010s. The data provide an original opportunity to use detailed measures of parental education and parental social class, which is seldom available in administrative datasets. The data further provide detailed official education data throughout primary and secondary schooling, which is seldom available in large-scale social science surveys. The use of administrative records further increases the reliability of educational attainment measures compared with self-reported measures collected in social science surveys.

3 Data and Methods

3.1 Sample

The analyses in this chapter use the Wave 1 UKHLS youth questionnaires and linked NPD data. The analyses focus on attainment at the end of primary school (Key Stage 2) and the end of compulsory secondary school (Key Stage 4). There is a complete records sample size of n=1343. The presence of missing data does not adversely affect the results presented below (see Tables A2.1 and A2.2 in Appendix 2).

The socio-economic background measures are collected from parents' individual responses in Wave 1 of the UKHLS adult survey. The socio-economic measures have been collected after the young people sat Key Stage 2 tests (at age 11), and before they took their Key Stage 4, i.e. GCSE, examinations (at age 16). There is, therefore, an implicit assumption that socio-economic background is static over the secondary school years. This assumption is somewhat plausible, as preliminary analyses of the UKHLS youth questionnaires demonstrate that the majority of parents do not change education level and social class category when the young person is aged between 11 and 15 (see section 3, Chapter 6). Given the age of the young people in the samples, it is likely that their parents have reached the age of occupational maturity (see Goldthorpe et al., 1980), and therefore there is less movement between social class categories. Crawford et al. (2017) also used proxy measures for socio-economic background taken at age 16 in analyses of educational attainment from ages 7 to 19.

3.2 Measures

In the following analyses, prior attainment is measured by Key Stage 2 test scores taken at the age of 11. Key Stage 2 tests and GCSE examinations are national, externally examined tests undertaken by state school pupils. Measuring core attainment using national, standardised test scores at ages 11 and 16 is substantively informative in the context of the English education system because age 11 marks the end of primary schooling and age 16 marks the end of compulsory schooling for the young people in these analyses. The test scores attained at age 11 can provide a

useful base measure for the attainment of young people over the course of secondary schooling.

The total marks attained in English papers and Maths papers are deposited in the NPD. The measure of Key Stage 2 score in these analyses use the total marks attained in English and Maths papers, which are combined, averaged, and standardised to have a mean of 0 and standard deviation of 1. English and Maths have been combined to represent average attainment in the core subjects taken throughout primary and secondary schooling, and to reflect core literacy and numeracy in the curriculum. The reliability of a combined score of attainment is tested using robustness checks of English-only and Maths-only scores in section 5.4.

Despite the NPD data containing information from all Key Stages 1, 2, 3, 4, and 5, only Key Stages 2 and 4 data provide sufficiently high quality data to undertake analyses. For example, the data for Key Stages 1 and 3 are completely missing on test scores for the Wave 1 UKHLS youth questionnaire participants. There are more complete data for teacher-assessed National Curriculum Levels, but there is still a substantial amount of missing data. Potential reasons for the missing data could be that the Key Stage 1 examinations are internally reviewed, and the results are not necessarily submitted to the NPD. The Key Stage 3 examinations were abolished in 2009, thereby reducing the obligation for schools to administer such tests. There might be greater teacher or school variation in the reporting of results to the NPD as a result.

3.3 Structure of Analysis

The central aim of this chapter is to better understand when stratification in school-level attainment begins. The following questions are explored:

1. Is prior attainment stratified by parental socio-economic background?
2. Do the effects of parental socio-economic background on GCSE attainment persist over and above the effects of prior attainment?

4 Descriptive Statistics

Table 4.1 presents descriptive statistics of the explanatory measures used throughout the analyses. Parental education level and parental NS-SEC are included as the main socio-economic background measures of interest. Housing tenure, gender, and ethnicity are also included in the following models. Ethnicity is measured as a five-category variable following guidance from the Office for National Statistics (Office for National Statistics, 2013). Information has been omitted (represented with an *) where categories did not meet the minimum threshold for statistical disclosure control.

Table 4. 1: Descriptive statistics for key explanatory variables, unweighted

	Freq.	Percent
Parental Education Level		
Higher education	329	24.50
Further education	206	15.34
School-level education	615	45.79
Below school-level education	193	14.37
Parental NS-SEC		
1.1 Large employers and higher managerial	89	6.63
1.2 Higher professional occupations	109	8.12
2 Lower management and professional occupations	272	20.25
3 Intermediate occupations	137	10.20
4 Small employers and own account workers	147	10.95
5 Lower supervisory and technical occupations	60	4.47
6 Semi-routine occupations	178	13.25
7 Routine occupations	87	6.48
8 Not in employment	264	19.66
Housing Tenure		
Owned/privately rented	1024	76.25
Social housing	319	23.75
Gender		
Male	677	50.41
Female	666	49.59
Ethnicity		
White	1040	77.44
Mixed	*	*
Asian/Asian British	166	12.36
Black/Black British	69	5.14
Other ethnic group	*	*
Total *n*	**1343**	**100.00**

Table 4.2 reports the mean scores for non-standardised Key Stage 2 and Key Stage 4 tests for each explanatory variable. There is a strong, positive correlation between Key Stage 2 test score and GCSE point score ($r=0.71$, $p<.001$). For parental education level and parental NS-SEC, there is a negative gradient in Key Stage 2 and

Key Stage 4 attainment. More advantaged young people tend to attain better scores than their less advantaged peers at ages 11 and 16. Young people living in owned or privately rented houses have higher mean scores at ages 11 and 16 than young people living in social housing. There is a small difference in the Key Stage 2 scores for boys and girls. The gender gap in attainment seems more pronounced at the end of secondary school compared with the end of primary school. The mean scores for some ethnic groups have been omitted due threshold requirements for statistical disclosure control.

Table 4. 2: Mean Key Stage 2 and GCSE attainment scores for key explanatory variables, unstandardised

Mean	KS2 score	Std. Dev.	GCSE score	Std. Dev.
Parental Education Level				
Higher education	72.35	13.81	56.19	17.96
Further education	64.52	14.77	44.01	17.11
School-level education	61.82	16.38	37.64	19.77
Below school-level education	52.13	18.01	28.11	18.99
Parental NS-SEC				
1.1 Large employers and higher managerial	72.39	13.97	51.85	17.89
1.2 Higher professional occupations	73.23	12.13	57.39	16.52
2 Lower management and professional occupations	68.72	14.17	48.77	18.44
3 Intermediate occupations	65.54	16.04	43.84	18.38
4 Small employers and own account workers	61.73	17.09	43.65	21.22
5 Lower supervisory and technical occupations	59.93	16.17	33.38	18.52
6 Semi-routine occupations	60.01	16.74	34.98	20.00
7 Routine occupations	54.05	20.11	33.05	20.30
8 Not in employment	56.91	16.79	32.06	20.37
Housing Tenure				
Owned/privately rented	65.77	16.38	45.62	20.26
Social housing	55.87	16.71	29.50	18.43
Gender				
Male	62.87	17.29	39.28	21.02
Female	63.98	16.66	44.35	20.66
Ethnicity				
White	64.40	16.65	41.45	21.07
Mixed Race	*	*	*	*
Asian/Asian British	59.90	19.01	45.28	21.16
Black/Black British	58.49	16.06	39.54	17.43
Other ethnic group	*	*	*	*
Total	**63.42**	**16.98**	**41.79**	**20.99**

Figures 4.1 and 4.2 present the mean standardised Key Stage 2 scores by parental education level and parental social class.

Figure 4. 1: Mean Key Stage 2 test scores (standardised) by parental education level

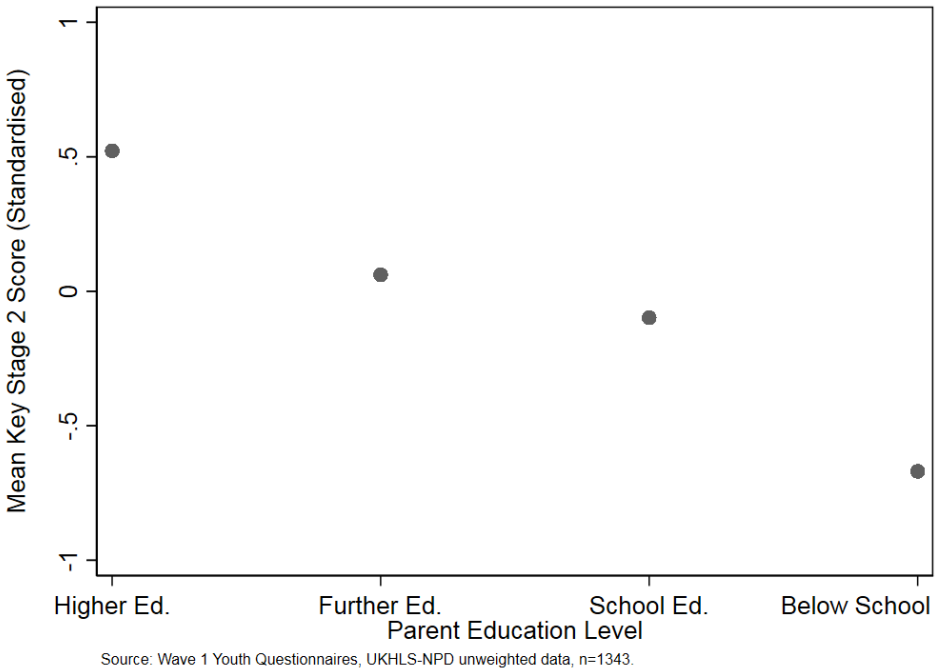
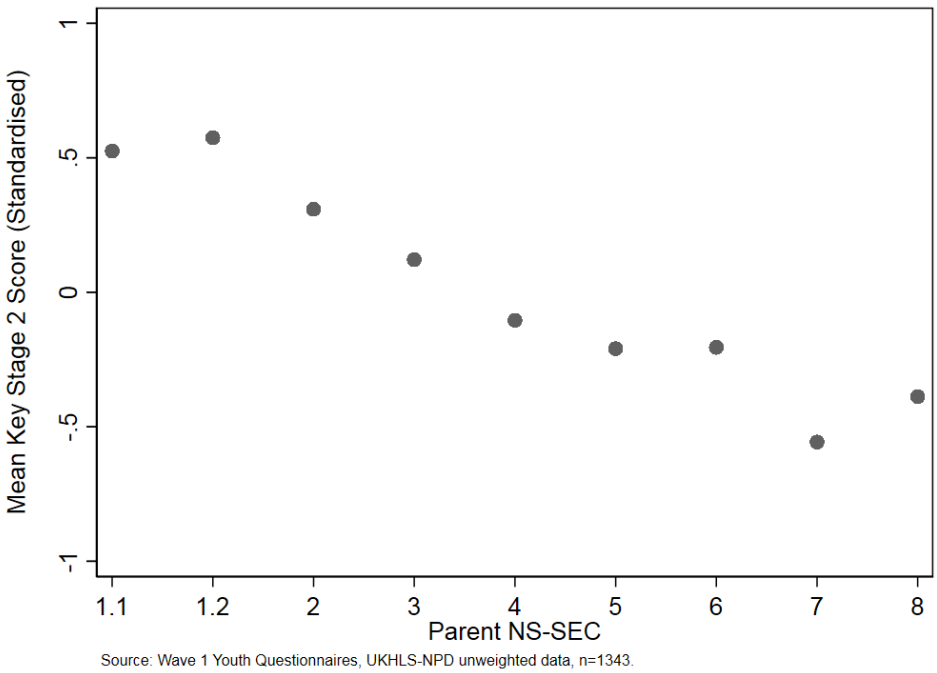


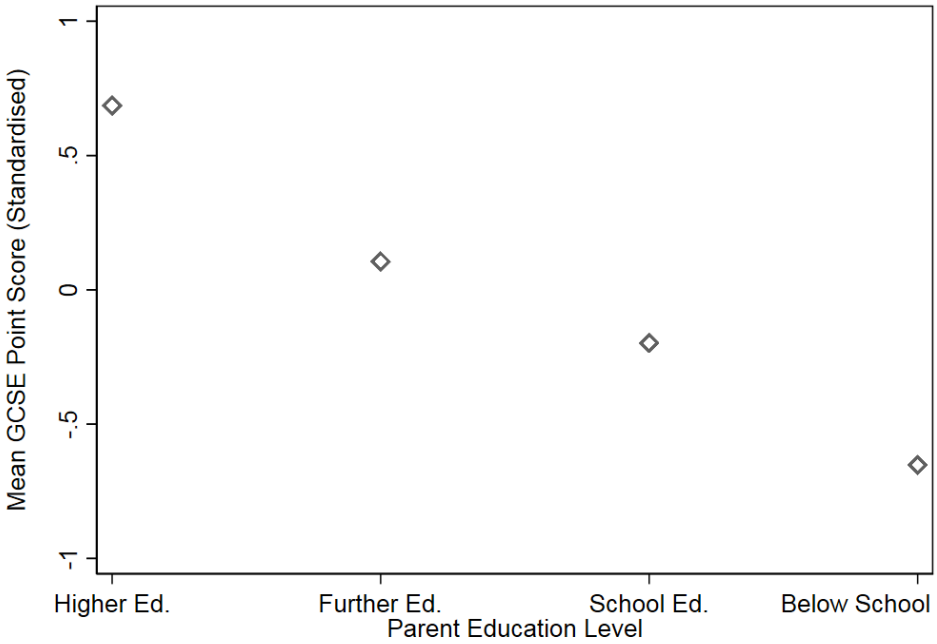
Figure 4. 2: Mean Key Stage 2 test scores (standardised) by parental NS-SEC



Figures 4.3 and 4.4 present the mean standardised Key Stage 4 (i.e. GCSE) scores by parental education and parental social class. There are clear negative gradients in the mean scores attained by young people at ages 11 and 16 by parental socio-

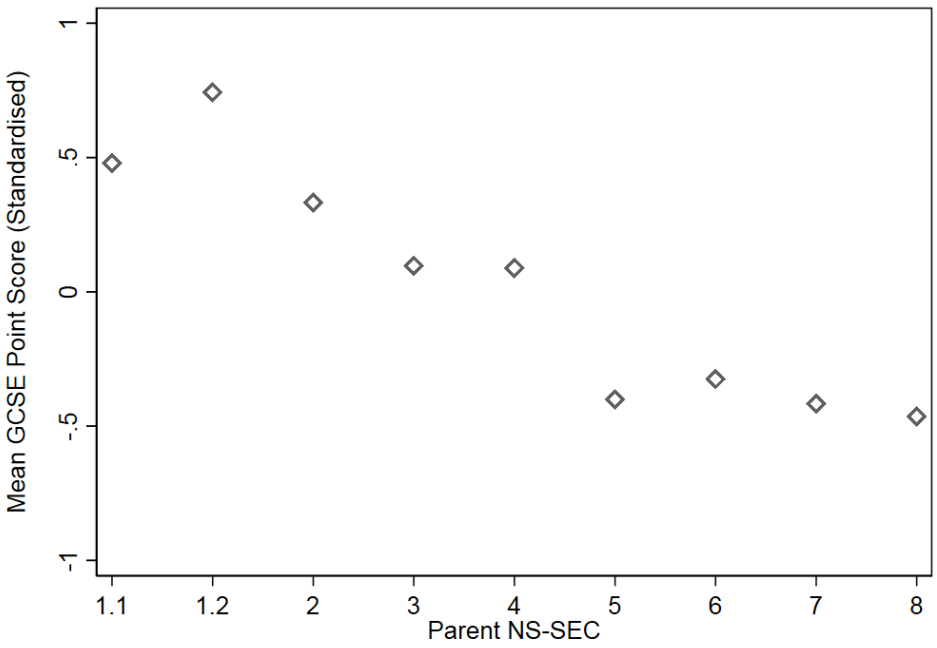
economic background. At age 16, there is a more pronounced educational advantage for young people with parents in NS-SEC 1.2 occupations compared with young people from all other backgrounds.

Figure 4. 3: Mean GCSE point score (standardised) by parental education level



Source: Wave 1 Youth Questionnaires, UKHLS-NPD unweighted data, n=1343.

Figure 4. 4: Mean GCSE point score (standardised) by parental NS-SEC



Source: Wave 1 Youth Questionnaires, UKHLS-NPD unweighted data, n=1343.

5 Modelling Results

5.1 Modelling Key Stage 2 Attainment

The first step in these analyses is to examine the extent to which socio-economic background is associated with prior attainment, i.e. Key Stage 2 scores. Table 4.3 presents the linear regression modelling output of Key Stage 2 attainment. The outcome variable is a standardised average score of English and Maths tests taken at age 11. There is a clear, negative gradient of attainment by parental education level. Young people with at least one graduate parent are significantly associated with higher test scores at age 11 compared with young people whose parents have any lower level of education. There is also evidence of a parental social class effect. Young people whose parents have higher professional occupations are significantly associated with higher test scores at age 11 than those whose parents are self-employed, who have lower supervisory and technical, semi-routine, or routine occupations, and who are not in employment. Young people living in social housing have significantly lower test scores at age 11 compared with young people living in owned or privately rented houses.

There are not significant gender or ethnicity differences in the model of age 11 attainment. The issue of within-group heterogeneity for the ethnicity measure has been discussed in Chapter 2 and may explain the non-significant effects of ethnicity on age 11 attainment. Previous research has suggested that the effects of gender and socio-economic background on attainment may increase over time (Sammons, 1995). It is possible that the gender attainment gap manifests itself over the course of secondary schooling and therefore may not be as apparent at age 11. An alternative explanation may relate to the use of a combined, average measure. Previous research has suggested that girls tend to perform better in reading than boys, and boys tend to perform better in Maths than girls (Marks, 2008). This may be influenced by gendered stereotypes of subjects, for example the characterisations of English as a ‘feminine’ subject and Maths as a ‘masculine’ subject, which can affect academic self-concept and abilities (Sullivan, 2009). More nuanced gender differences in

attainment at age 11 may be obscured by the use of a combined measure. This will be revisited in the sensitivity analyses in section 5.4.

Table 4. 3: Linear regression model of Key Stage 2 score (standardised)

Linear Regression: Key Stage 2 Score	**Coef.**	**S.E.**	
Parental Education Level			
Higher education	Ref.	(.)	
Further education	-0.37	(0.10)	***
School-level education	-0.41	(0.09)	***
Below school-level education	-0.81	(0.16)	***
Parental NS-SEC			
1.1 Large employers and higher managerial	-0.00	(0.14)	
1.2 Higher professional occupations	Ref.	(.)	
2 Lower management and professional occupations	-0.14	(0.11)	
3 Intermediate occupations	-0.18	(0.13)	
4 Small employers and own account workers	-0.38	(0.14)	**
5 Lower supervisory and technical occupations	-0.46	(0.17)	**
6 Semi-routine occupations	-0.43	(0.14)	**
7 Routine occupations	-0.69	(0.19)	***
8 Not in employment	-0.51	(0.14)	***
Housing Tenure			
Owned/privately rented	Ref.	(.)	
Social housing	-0.24	(0.10)	*
Gender			
Male	Ref.	(.)	
Female	0.07	(0.06)	
Ethnicity			
White	Ref.	(.)	
Mixed	-0.22	(0.15)	
Asian/Asian British	-0.16	(0.15)	
Black/Black British	-0.14	(0.17)	
Other ethnic group	-0.21	(0.40)	
Constant [output omitted]			
Observations	1343		
Adjusted R^2	0.180		
BIC (d.f.)	3683.380 (22)		
BIC (based on deviance)	-5989.794		
BIC (based on likelihood ratio chi square)	-136.096		

*Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Model also includes academic year. * p<.05, ** p<.01, *** p<.001.*

5.2 Controlling for Prior Attainment

The next step in the analyses is to examine models of GCSE attainment, controlling for prior attainment. This is common practice in the literature (for example, Gregg and Washbrook, 2011, Chowdry et al., 2011, Strand, 2014b, Sullivan et al., 2018). Controlling for prior attainment is a useful starting point before moving to more

comprehensive analyses. Table 4.4 presents a series of models using Key Stage 2 attainment as a covariate. Model 1 estimates GCSE point score without controlling for Key Stage 2 attainment. Model 2 adds Key Stage 2 score into the model. Parental socio-economic background continues to be significantly associated with GCSE attainment after controlling for prior attainment, although the overall effects are weaker. The model fit statistics indicate that adding in prior attainment as a covariate more than doubles the Adjusted R^2, i.e. explained variance, compared with the model without prior attainment. This should, however, be interpreted with caution due to the endogeneity effects associated with controlling for Key Stage 2 attainment (see below).

Controlling for Key Stage 2 test scores holds prior attainment constant within the model. Crawford et al. (2017) recommended that including an interaction effect between prior attainment and socio-economic background is more suitable, because it allows prior attainment to vary over the distribution of socio-economic background. Model 3 includes an interaction between prior attainment and parental education level, and Model 4 includes an interaction between prior attainment and parental NS-SEC. The significant interaction between prior attainment and parental education in Model 3 demonstrates that young people whose parents have higher education qualifications attain higher GCSE scores compared with young people whose parents have below school-level education, holding age 11 attainment constant. The significant interaction effect between prior attainment and parental social class in Model 4 demonstrates that young people whose parents have occupations in NS-SEC 1.2 attain higher GCSE scores compared with young people whose parents have occupations in NS-SECs 5, 6, 7 or who are not in employment, holding age 11 attainment constant. The interaction effects are visualised in Figures 4.5 and 4.6. For clarity, the figures only plot the significant interaction effects.

Table 4. 4: Linear regression models of total GCSE point score controlling for Key Stage 2 attainment (standardised) and interaction effect

	Model 1			Model 2			Model 3			Model 4		
Linear Regression: GCSE Point Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level												
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	- 8.18	(2.11)	***	-3.42	(1.68)	*	-2.02	(1.75)		-3.04	(1.70)	
School-level education	-13.13	(1.91)	***	-7.82	(1.50)	***	-6.50	(1.51)	***	-7.50	(1.50)	***
Below school-level education	-18.86	(2.74)	***	-8.46	(2.13)	***	-8.99	(2.45)	***	-8.57	(2.09)	***
Parental NS-SEC												
1.1 Large employers and higher managerial	- 3.52	(2.92)		-3.50	(2.23)		-3.53	(2.24)		-2.05	(2.97)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	- 4.63	(2.36)	(*)	-2.79	(1.74)		-2.55	(1.72)		-1.04	(2.19)	
3 Intermediate occupations	- 5.63	(2.55)	*	-3.27	(2.06)		-2.96	(2.08)		-1.51	(2.35)	
4 Small employers and own account workers	- 6.11	(3.03)	*	-1.29	(2.20)		-0.94	(2.22)		0.66	(2.48)	
5 Lower supervisory and technical occupations	-14.43	(3.55)	***	-8.52	(2.87)	**	-8.14	(2.87)	**	-7.13	(3.18)	*
6 Semi-routine occupations	-11.99	(2.92)	***	-6.53	(2.28)	**	-6.24	(2.27)	**	-4.93	(2.57)	
7 Routine occupations	-12.50	(3.32)	***	-3.70	(2.47)		-3.70	(2.43)		-2.71	(2.91)	
8 Not in employment	-13.33	(2.93)	***	-6.77	(2.21)	**	-6.43	(2.21)	**	-5.19	(2.63)	*
Housing Tenure												
Owned/privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	- 9.28	(1.92)	***	-6.20	(1.62)	***	-6.12	(1.61)	***	-6.33	(1.61)	***
Gender												
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	4.87	(1.30)	***	3.99	(1.02)	***	3.96	(1.01)	***	4.02	(1.02)	***
Ethnicity												
White	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Mixed	- 2.33	(3.78)		0.51	(3.06)		0.82	(3.10)		0.74	(3.08)	
Asian/Asian British	5.33	(2.91)		7.36	(1.68)	***	7.43	(1.68)	***	7.82	(1.64)	***
Black/Black British	4.13	(2.85)		5.94	(2.49)	*	6.36	(2.55)	*	6.14	(2.52)	*
Other ethnic group	1.73	(6.93)		4.43	(6.66)		5.42	(6.12)		5.56	(6.43)	

Table continued overleaf.

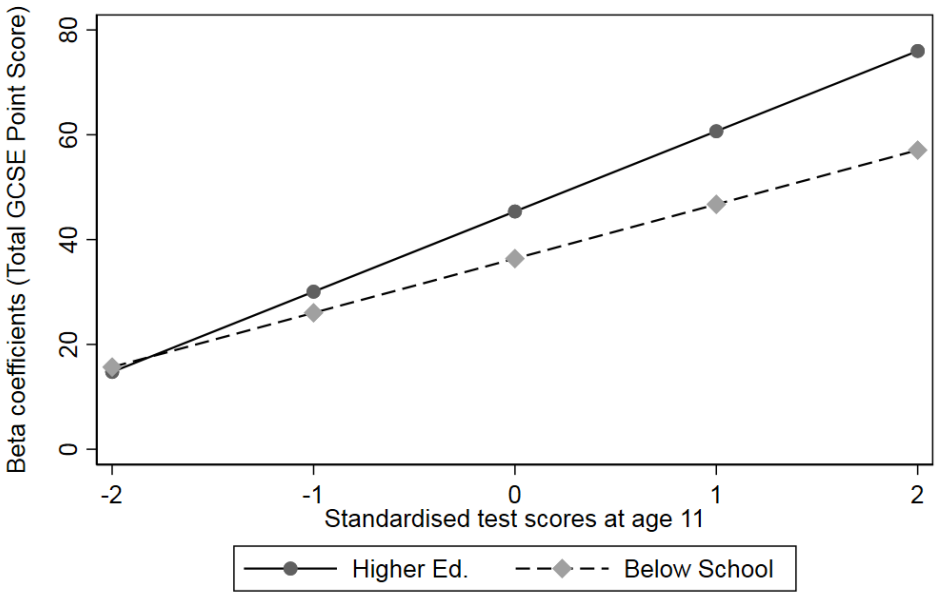
Table 4.4 continued.

	Model 1		Model 2			Model 3			Model 4		
	Coef.	**S.E.**	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Key Stage 2 Attainment			12.82	(0.56)	***	15.31	(1.08)	***	16.19	(1.68)	***
Parental Education*Key Stage 2 Attainment											
*Higher education * KS2*						Ref.	(.)				
*Further education * KS2*						-3.31	(1.72)				
*School-level education * KS2*						-2.31	(1.27)				
*Below school-level education * KS2*						-4.96	(1.72)	**			
Parental NS-SEC*Key Stage 2 Attainment											
*1.1 Large employers and higher managerial * KS2*									-2.43	(2.62)	
1.2 Higher professional occupations KS2*									Ref.	(.)	
2 Lower management /professional occupations KS2*									-2.86	(2.12)	
3 Intermediate occupations KS2*									-2.81	(2.13)	
4 Small employers and own account workers KS2*									-0.69	(2.35)	
5 Lower supervisory and technical occupations KS2*									-5.68	(2.83)	*
6 Semi-routine occupations KS2*									-5.03	(2.13)	*
7 Routine occupations KS2*									-5.07	(2.27)	*
8 Not in employment KS2*									-4.37	(2.09)	*
Constant [output omitted]											
Observations	1343		1343			1343			1343		
Adjusted R^2	0.279		0.594			0.596			0.597		
BIC (d.f.)	11683.557 (22)		10919.512 (23)			10930.701 (26)			10959.638 (31)		
BIC (based on deviance)	2010.383		1246.338			1257.527			1286.464		
BIC (based on likelihood ratio chi square)	-309.852		-1073.897			-1062.708			-1033.771		

Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Models also include academic year.
** p<.05, ** p<.01, *** p<.001.*

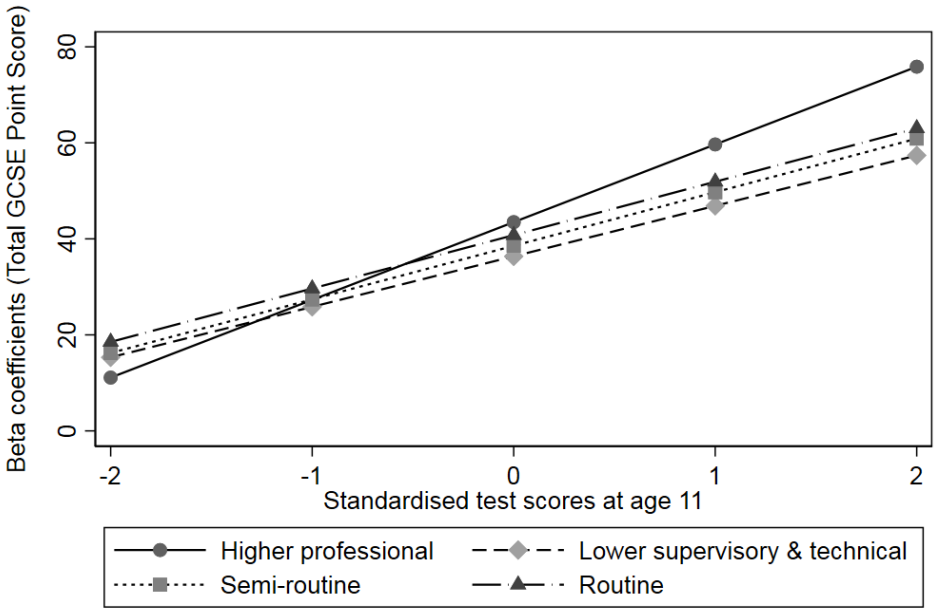
Modelling GCSE attainment and controlling for Key Stage 2 attainment, even after including an interaction effect, is not satisfactory. First, the models violate the regression assumption of multicollinearity. Key Stage 2 score has a VIF value greater than 4 and a tolerance value lower than 0.2 (Menard, 2002). Second, the determinants of the outcome measure of GCSE attainment are likely to be similar to those underlying an outcome measure of earlier attainment. For example, as observed in section 5.1 above, the socio-economic measures influencing GCSE attainment are also significantly associated with Key Stage 2 attainment. The measure of Key Stage 2 attainment in a model of GCSE attainment could therefore be understood as endogenous. One conventional approach to addressing endogeneity in statistical models is the use of instrumental variables (Angrist and Pischke, 2009). Instrumental variables are understood to be correlated with the independent variable but not with the dependent variable. Suitable instrumental variables for educational outcomes are not routinely collected in social science surveys and therefore this is not a practicable approach in these analyses. Controlling for prior attainment is an alternative, pragmatic approach in statistical models where suitable instruments do not exist. However, there are limitations to this approach. For example, the effects of residual heterogeneity for Key Stage 2 attainment are likely to also affect the estimates of a model of GCSE attainment using Key Stage 2 attainment as an independent variable. The next section undertakes path analysis as an alternative approach to modelling prior attainment.

Figure 4. 5: Interaction effect of parental education and Key Stage 2 attainment on GCSE score



Source: Wave 1 Youth Questionnaires, UKHLS-NPD data, n=1343. Adjusted for complex survey design. Parent NS-SEC, housing tenure, gender, ethnicity, and academic year also included in model.

Figure 4. 6: Interaction effect of parental NS-SEC and Key Stage 2 attainment on GCSE score



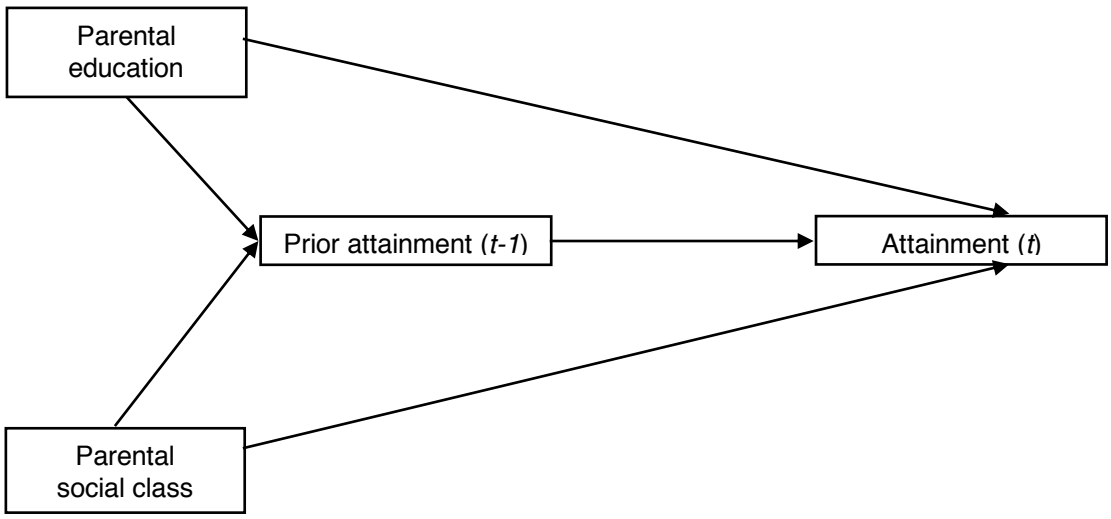
Source: Wave 1 Youth Questionnaires, UKHLS-NPD data, n=1343. Adjusted for complex survey design. Parent education, housing tenure, gender, ethnicity, and academic year also included in model.

5.3 Path Analysis

Path analysis models can estimate more than one outcome variable in the same model (Allen, 2017), including outcomes which may be temporally dependent. This is particularly useful in the present case, where parental socio-economic background measures are significantly associated with two related educational outcome variables (attainment at age 11 and attainment at age 16). Path analysis associations tend to be presented in diagrammatic form with standardised coefficients (Wright, 1960, Duncan, 1966). A statistically attractive property of path analysis is the decomposition of effects of the exogenous variables on the endogenous variables into direct, and indirect effects.

A conceptual path model of socio-economic background and educational attainment is presented in Figure 4.7. The conceptual path model suggests that parental education and parental social class exert separate influences on attainment at age 11 and attainment at age 16, and that some of the effect at age 16 works through prior attainment. This is a recursive model, because the effect is in one direction (Allen, 2017).

Figure 4. 7: A conceptual path model of parental socio-economic background and attainment



Path models are sometimes considered special types of structural equation models where all variables are manifest, i.e. observed (Kaplan, 2009, Acock, 2013). The path

analysis models presented are estimated in a structural equation modelling framework (see sem, StataCorp., 2017a). The path analysis output is presented in Table 4.5. The coefficients have been standardised to follow path model convention (Kaplan, 2009). The standardised coefficients allow for direct comparison of effect sizes across independent variables. The covariate with the strongest relative effect on GCSE attainment is Key Stage 2 attainment.

Table 4. 5: Path analysis model of Key Stage 2 attainment and GCSE attainment with standardised coefficients

	Key Stage 2 Score			GCSE Point Score		
Path Analysis Model	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	-0.14	(0.04)	***	-0.06	(0.03)	*
School-level education	-0.21	(0.04)	***	-0.19	(0.04)	***
Below school-level education	-0.27	(0.05)	***	-0.13	(0.03)	***
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.00	(0.04)		-0.05	(0.03)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	-0.06	(0.05)		-0.05	(0.03)	
3 Intermediate occupations	-0.06	(0.04)		-0.05	(0.03)	
4 Small employers and own account workers	-0.11	(0.04)	**	-0.02	(0.03)	
5 Lower supervisory and technical occupations	-0.10	(0.04)	**	-0.08	(0.03)	**
6 Semi-routine occupations	-0.14	(0.05)	**	-0.11	(0.04)	**
7 Routine occupations	-0.16	(0.05)	***	-0.04	(0.03)	
8 Not in employment	-0.19	(0.05)	***	-0.12	(0.04)	**
Housing Tenure						
Owned/privately rented	Ref.	(.)		Ref.	(.)	
Social housing	-0.09	(0.04)	*	-0.12	(0.03)	***
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	0.03	(0.03)		0.10	(0.02)	***
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	-0.03	(0.03)		0.00	(0.02)	
Asian/Asian British	-0.04	(0.03)		0.08	(0.02)	***
Black/Black British	-0.02	(0.02)		0.04	(0.02)	*
Other ethnic group	-0.01	(0.03)		0.02	(0.02)	
Key Stage 2 Attainment				0.61	(0.02)	***
Constant [output omitted]						
Error variance, Key Stage 2	0.81	(0.03)				
Error variance, GCSE	0.39	(0.02)				
Observations	1343					
Overall R^2	0.36					
Chi square (d.f.), model vs. saturated	10.621 (4) *					
Root Mean Squared Error of Approximation	0.035					
Comparative Fit Index	0.996					
Tucker-Lewis Index	0.952					
Standardised Root Mean Squared Residual	0.005					

*Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Model also includes academic year. * p<.05, ** p<.01, *** p<.001.*

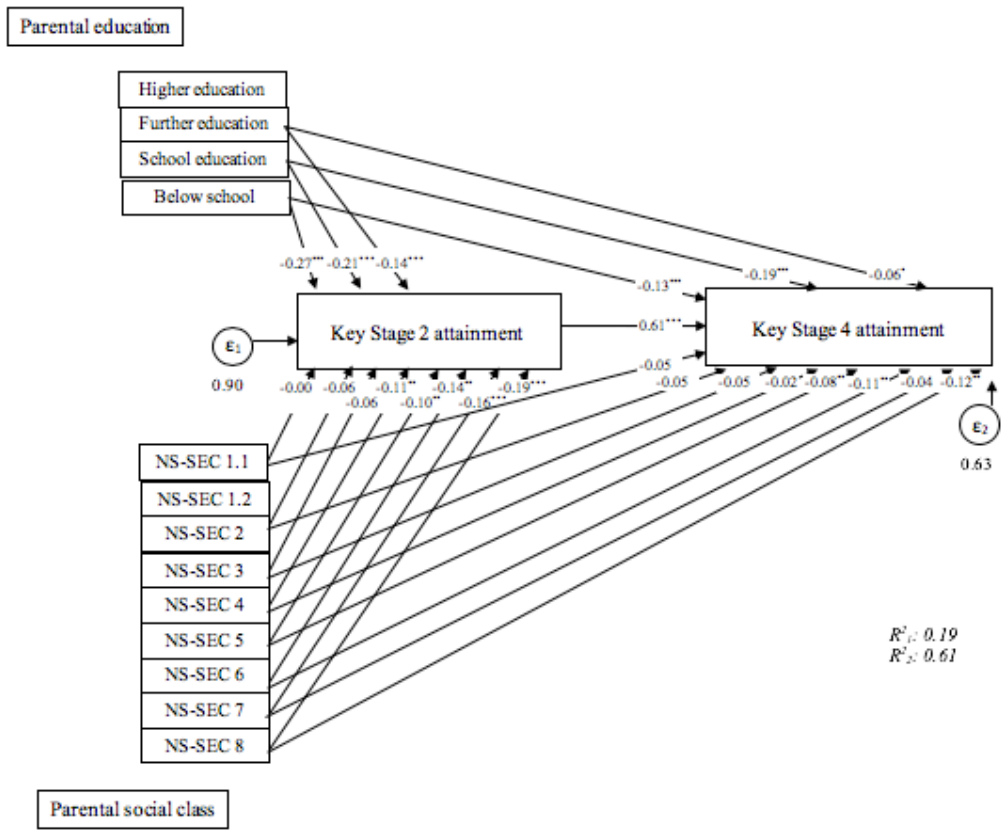
Parental education level is significant in both parts of the model. Compared with those whose parents have a degree, those whose parents have further education, school, or below school-level education are significantly associated with lower test scores at age 11, and at age 16 net of attainment at age 11. Parental social class is also significant in both parts of the model. There are significant effects of parental social class on attainment at age 11 for those whose parents are in higher professional occupations (NS-SEC 1.2), compared with those whose parents have occupations in NS-SECs 4, 5, 6, and 7, or who are not in employment. There are further significant effects of parental social class on attainment at age 16, net of attainment at age 11, for those whose parents have occupations in NS-SEC 1.2, compared with those whose parents have occupations in NS-SEC 5, 6, and who are not in employment.

Housing tenure is significant in both parts of the model. Gender is not significant at age 11 but is significant at age 16, net of attainment at age 11. Ethnicity is also not significantly associated with Key Stage 2 attainment. Ethnicity is, however, significantly associated with GCSE attainment, net of attainment at age 11. This suggests that those from Asian or Black backgrounds are not more likely to attain higher scores than those from white backgrounds at age 11. However, conditional on age 11 test score, those from Asian or Black backgrounds are significantly associated with higher GCSE point scores than those from white backgrounds.

The model fit statistics are reported at the bottom of Table 4.5. Identifying a well-specified model with suitable goodness-of-fit is an important criteria for structural equation models (Yuan, 2005), of which path models are a special case. There are a number of different model fit indicators, including absolute and relative fit indices (Hooper et al., 2008). The analyses report absolute fit indices of a chi square test, the root mean squared error of approximation (RMSEA) and the Standardised Root Mean Squared Residual (SRMSR). The analyses also report two measures of relative fit indices, the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). Schreiber et al. (2006) provide a useful guide of conventional thresholds for a variety of different model fit statistics in structural equation models. Taken together, the statistics in Table 4.5 suggest that the model is generally a good fit for the data. The overall R^2 for the model is 0.36, i.e. 36% of the variance in GCSE attainment is explained by the model.

Figure 4.8 is a diagrammatic representation of the path analysis model in Table 4.5. To ascertain the full effects of parental socio-economic background, the coefficients are traced back from the main outcome of interest to the exogenous variables (Wright, 1960), i.e. from GCSE attainment to parental education and parental social class. The direct effects of parental education and parental social class are represented by the connecting paths between the explanatory variables and GCSE attainment. The indirect effects are calculated as the product of all paths between parental education or parental social class and GCSE attainment. The total effect is the sum of the direct and indirect effects. R^2_1 represents the R^2 for the first outcome variable (i.e. attainment at age 11) and R^2_2 represents the R^2 for the second outcome variable (i.e. attainment at age 16).

Figure 4. 8: A path model of parental socio-economic background, Key Stage 2 attainment and GCSE attainment[14]



[14] Model also includes housing tenure, gender, ethnicity, and academic year.

Table 4.6 presents the total, direct, and indirect effects of parental socio-economic background on GCSE attainment. All effects use standardised coefficients and standard errors, with corresponding significance levels for the standardised solution. There are significant total, direct, and indirect effects of parental education level on GCSE attainment. Young people whose parents have higher education qualifications have significantly higher GCSE point scores net of attainment at age 11 compared with young people whose parents have any lower level of education. There is therefore a parental education level effect beyond the effects of prior attainment. There are also significant indirect effects for young people whose parents have higher education qualifications compared with young people whose parents have any lower level of education, suggesting that some of the socio-economic differences are due to inequalities occurring at earlier stages of education.

Parental social class also has significant direct, indirect, and total effects on GCSE attainment. The direct effects of parental social class on GCSE attainment are significant for those whose parents are in lower supervisory and technical occupations, semi-routine occupations, or who are not in employment compared with those in higher professional occupations. The significant direct effects demonstrate that parental social class effects in GCSE attainment exist beyond the role of prior attainment. The indirect effects of parental social class on GCSE attainment are significant for those whose parents are small employers or own account workers, or who are in lower supervisory and technical, semi-routine or routine occupations. NS-SECs 5, 6, and 7 together comprise routine and manual occupations (Rose and Pevalin, 2003). The significant indirect effects suggest that some of the socio-economic differences in GCSE attainment between young people with higher professional parents compared with young people with parents in routine or manual occupations are due to stratification occurring at earlier stages of the educational process.

Table 4. 6: Total, direct, and indirect effects of socio-economic background on GCSE attainment, standardised[15]

	Direct Effects			Indirect Effects			Total Effects		
	Coef.	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-0.06	(0.03)	*	-0.08	(0.02)	***	-0.14	(0.04)	***
School-level education	-0.19	(0.04)	***	-0.13	(0.03)	***	-0.31	(0.04)	***
Below school-level education	-0.13	(0.03)	***	-0.16	(0.03)	***	-0.29	(0.04)	***
Parental NS-SEC									
1.1 Large employer/higher man.	-0.05	(0.03)		-0.00	(0.02)		-0.05	(0.04)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower man./professional	-0.05	(0.03)		-0.04	(0.03)		-0.09	(0.05)	
3 Intermediate	-0.05	(0.03)		-0.03	(0.03)		-0.08	(0.04)	*
4 Small employer/own account	-0.02	(0.03)		-0.07	(0.03)	**	-0.09	(0.05)	
5 Lower supervisory/technical	-0.08	(0.03)	**	-0.06	(0.02)	**	-0.14	(0.04)	***
6 Semi-routine	-0.11	(0.04)	**	-0.09	(0.03)	**	-0.19	(0.05)	***
7 Routine	-0.04	(0.03)		-0.10	(0.03)	***	-0.14	(0.04)	***
8 Not in employment	-0.12	(0.04)	**	-0.11	(0.03)	***	-0.24	(0.05)	***

5.4 Sensitivity Analyses

5.4.1 English and Maths Attainment Measures

The following section re-estimates the path analysis models using separate measures of English and Maths test scores. The results are very similar to the model presented above using a combined measure. Table 4.7 presents the path analysis model using a measure of standardised English score at age 11. There are significant effects of parental education and parental social class on attainment in English at age 11 and overall GCSE attainment at age 16. The estimated decomposition of direct and indirect effects can be found in Table A1.5 in Appendix 1. Table 4.8 presents the path analysis model using a measure of standardised Maths score at age 11. There are similarly significant effects of parental education and parental social class on GCSE attainment, accounting for Maths score at age 11. The analyses demonstrate a distinct educational advantage for young people whose parents are graduates, or have occupations in NS-SEC 1.2. The decomposition of the direct and indirect effects can be found in Table A1.6 in Appendix 1.

[15] The effects might not add exactly due to rounding to 2 decimal places.

Table 4. 7: Path analysis model of Key Stage 2 English score and GCSE attainment with standardised coefficients

Path Analysis Model	Key Stage 2 English Coef.	S.E.		GCSE Point Score Coef.	S.E.	
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	-0.11	(0.04)	**	-0.08	(0.03)	**
School-level education	-0.18	(0.05)	***	-0.21	(0.04)	***
Below school-level education	-0.25	(0.05)	***	-0.15	(0.03)	***
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.02	(0.04)		-0.04	(0.03)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	-0.07	(0.05)		-0.05	(0.04)	
3 Intermediate occupations	-0.07	(0.05)		-0.04	(0.03)	
4 Small employers and own account workers	-0.12	(0.05)	**	-0.02	(0.04)	
5 Lower supervisory and technical occupations	-0.10	(0.04)	*	-0.09	(0.03)	**
6 Semi-routine occupations	-0.17	(0.05)	**	-0.10	(0.04)	*
7 Routine occupations	-0.16	(0.05)	**	-0.05	(0.03)	
8 Not in employment	-0.21	(0.06)	***	-0.12	(0.04)	**
Housing Tenure						
Owned/privately rented	Ref.	(.)		Ref.	(.)	
Social housing	-0.08	(0.04)		-0.13	(0.03)	***
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	0.16	(0.03)	***	0.03	(0.03)	
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	-0.03	(0.02)		0.00	(0.03)	
Asian/Asian British	-0.03	(0.04)		0.07	(0.02)	***
Black/Black British	-0.00	(0.02)		0.03	(0.02)	
Other ethnic group	-0.03	(0.03)		0.03	(0.02)	
Key Stage 2 English Attainment				0.56	(0.02)	***
Constant [output omitted]						
Error variance, Key Stage 2	0.81	(0.03)				
Error variance, GCSE	0.43	(0.02)				
Observations	1343					
Overall R^2	0.35					
Chi square (d.f.), model vs. saturated	18.781 (4) **					
Root Mean Squared Error of Approximation	0.052					
Comparative Fit Index	0.989					
Tucker-Lewis Index	0.881					
Standardised Root Mean Squared Residual	0.006					

*Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Model also includes academic year. * p<.05, ** p<.01, *** p<.001.*

Table 4. 8: Path analysis model of Key Stage 2 Maths score and GCSE attainment with standardised coefficients

	Key Stage 2 Maths			GCSE Point Score		
Path Analysis Model	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	-0.14	(0.04)	***	-0.07	(0.03)	*
School-level education	-0.21	(0.04)	***	-0.20	(0.04)	***
Below school-level education	-0.24	(0.05)	***	-0.16	(0.03)	***
Parental NS-SEC						
1.1 Large employers and higher managerial	0.01	(0.04)		-0.05	(0.03)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	-0.04	(0.05)		-0.07	(0.04)	
3 Intermediate occupations	-0.04	(0.04)		-0.06	(0.03)	
4 Small employers and own account workers	-0.09	(0.04)	*	-0.04	(0.04)	
5 Lower supervisory and technical occupations	-0.08	(0.04)	*	-0.10	(0.03)	**
6 Semi-routine occupations	-0.10	(0.05)	*	-0.14	(0.04)	**
7 Routine occupations	-0.14	(0.04)	**	-0.06	(0.03)	*
8 Not in employment	-0.15	(0.05)	**	-0.16	(0.04)	***
Housing Tenure						
Owned/privately rented	Ref.	(.)		Ref.	(.)	
Social housing	-0.09	(0.05)	*	-0.13	(0.03)	***
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	-0.07	(0.03)	*	0.15	(0.03)	***
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	-0.03	(0.03)		0.00	(0.02)	
Asian/Asian British	-0.04	(0.03)		0.08	(0.02)	***
Black/Black British	-0.03	(0.02)		0.04	(0.02)	*
Other ethnic group	0.00	(0.03)		0.01	(0.02)	
Key Stage 2 Maths Attainment				0.54	(0.03)	***
Constant [output omitted]						
Error variance, Key Stage 2	0.84	(0.03)				
Error variance, GCSE	0.45	(0.02)				
Observations	1343					
Overall R^2	0.37					
Chi square (d.f.), model vs. saturated	6.774 (4)					
Root Mean Squared Error of Approximation	0.023					
Comparative Fit Index	0.998					
Tucker-Lewis Index	0.976					
Standardised Root Mean Squared Residual	0.005					

*Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Model also includes academic year. * p<.05, ** p<.01, *** p<.001.*

The substantive conclusions about the effects of parental socio-economic background on GCSE attainment accounting for prior attainment are very similar when using alternative measures of age 11 attainment. There are more nuanced gender differentials in GCSE attainment, however, which are masked by the use of a combined measure of prior attainment. In the model using an average score of English and Maths, gender is not significantly associated with attainment at age 11, but there is a significant gender effect on attainment at age 16. This suggests that

girls perform better than boys in their GCSE examinations conditional on primary school attainment. Gender is significantly associated with attainment at age 11 when using measures of English and Maths separately. Girls are significantly associated with higher English scores at age 11 than boys, and boys are significantly associated with higher Maths scores at age 11 than girls. This supports previous research which finds gender differences in attainment by reading and Maths (Marks, 2008). In the path analysis models, there is not a significant difference between the attainment of boys and girls at age 16 after accounting for English scores at age 11. This might suggest that gender differentials in English scores occur more strongly during primary schooling. There are, however, significant differences between the attainment of boys and girls at age 16 after accounting for Maths scores at age 11. A particularly striking finding is the change in the direction of effect. At age 11, boys are more likely to perform better than girls in Maths. After controlling for Maths attainment at age 11, girls are more likely to perform better at GCSE level than boys. This suggests that gender differentials in attainment at age 16 continue to persist over and above Maths attainment at age 11. This finding has consequences for the understanding of gender stratification by subjects and age in school-level education.

5.4.2 School Type

There is a large literature on school effectiveness which emphasises the role that schools play in young people's educational outcomes (see Goldstein and Woodhouse, 2000 for a comprehensive review). The analyses undertaken in this chapter have used test scores at the end of primary schooling and the end of compulsory secondary schooling. It is plausible that some of the effects observed are due to school factors. The UKHLS is a nationally representative study of households in the UK, not of young people of secondary school age in the UK. There are likely to be few pupils in the same schools and strong school clustering effects are less likely to be evident. The NPD does not deposit school identifiers and therefore it is not possible ot undertake formal analyses about the role of schools. Nonetheless, it is prudent to examine the potential role that schools play in the attainment of young people over their secondary school careers.

The NPD collects a measure of the type of school attended. The school type in the Key Stage 2 database represents the primary school attended, and the school type in the Key Stage 4 database represents the secondary school attended. This measure does not take into account any school changes during this time. The categories of schools are community schools, voluntary or foundation schools, independent schools (with very few cases), academies, and very few cases of other school types such as technology colleges or pupil referral units. The path models are re-estimated to control for the type of primary and secondary school attended. The inclusion of these variables does not change the results of the models presented and does not alter the conclusions regarding the effects of parental socio-economic background. If the type of school measure were collected and categorised according to selectiveness, such as comparisons of comprehensive schools to grammar schools or private schools, then the inclusion of school type may have greater substantive significance in relation to socio-economic background.

6 Discussion and Conclusions

The analyses in this chapter provide convincing evidence that parental socio-economic background is strongly associated with educational attainment over the course of a young person's secondary school education. Differences in educational attainment by parental education level and parental social class are clearly observed at the end of primary schooling. Parental education level and parental social class are also significantly associated with GCSE attainment at the end of secondary schooling beyond the effects of prior attainment. The analyses further demonstrate that there are nuanced socio-economic differences between more and less occupationally and educationally advantaged young people at ages 11 and 16 than is routinely observed when using a binary summary measure of socio-economic disadvantage, such as eligibility for Free School Meals.

The path analysis models indicate that attainment at the end of primary school has the largest substantive and significant impact on attainment at the end of secondary school. The decomposition of effects indicate that there are significant direct and indirect effects of parental socio-economic background on GCSE attainment. This

provides contemporary evidence to support previous empirical work investigating the effects of earlier educational inequalities in later attainment (Gregg and Washbrook, 2011, Chowdry et al., 2011). There is a clear difference in GCSE attainment between those whose parents have higher professional occupations compared with those whose parents have routine and manual occupations, net of parental education, gender, ethnicity, housing tenure, and academic year.

The results indicate that there is a clear doubled-edged benefit for young people from more occupationally and educationally advantaged backgrounds. In the first instance, more advantaged young people are more likely to do better in their tests at the end of primary schools than their less advantaged peers. If they do, this is likely to have positive effects on their attainment at the end of compulsory school. If they do not, more advantaged young people are more likely to get ahead of their less advantaged peers over the course of their secondary schooling and still attain better outcomes in their GCSE examinations than their less advantaged peers. The concept of a compounding advantage and disadvantage over primary and secondary schooling has been demonstrated in analyses of earlier cohorts of young people (Feinstein, 2004).

Knowing when socio-economic inequalities open up is important because inequalities often reproduce and accumulate over time (for example, Feinstein, 2003, Feinstein, 2004, Chowdry et al., 2011, Chowdry et al., 2013). Previous empirical research has demonstrated that socio-economic inequalities in cognitive outcomes begin at a very early age and continue throughout childhood (Feinstein, 2003, Feinstein, 2004, Dearden et al., 2011, Sullivan et al., 2013, Connelly and Gayle, 2019). Empirical research has further demonstrated that earlier socio-economic inequalities are consequential at later stages, for example, much of the socio-economic differences in higher education can be explained by prior attainment in school (Galindo-Rueda et al., 2004, Marcenaro-Gutierrez et al., 2007, Broecke and Hamed, 2008, Chowdry et al., 2013, Smith, 2015). Narrowing the attainment gap between more and less advantaged young people earlier in the education system should be a priority. The analyses in this chapter demonstrate that socio-economic inequalities are clearly observed by the end of primary schooling, and that inequalities at the end of

compulsory schooling are in part a continuation of the inequalities observed at the end of primary schooling. A key implication of these findings is that narrowing the attainment gap at earlier stages is important to avoid introducing additional disadvantage for young people from less socio-economically advantaged backgrounds by the time they approach non-compulsory education, employment, or training.

Chapter 5
Cultural Capital, Socio-Economic Background, and GCSE Attainment

1 Introduction

Cultural capital was originally theorised to explain social class differences in educational outcomes (see Bourdieu, 1986). The theory suggests that the unequal distribution of cultural capital by socio-economic background explains why more advantaged young people tend to have more favourable educational outcomes. The main motivation for this chapter is to assess the role of cultural capital in explaining the socio-economic background effect in GCSE attainment. There are, however, no agreed-upon measures of cultural capital and a broad range of operationalisations have been used in previous empirical work (Sullivan, 2002). This chapter presents a series of sensitivity analyses of a variety of alternative candidate measures to directly engage with the operationalisation challenge. The overall research question for the analyses in this chapter is *To what extent does cultural capital account for socio-economic differences in GCSE attainment?*

Section 2 outlines the concept of cultural capital. First, the theoretical foundations of Pierre Bourdieu's concept of cultural capital are summarised. Second, a review of the quantitative sociological literature is presented, and particular attention is paid to the operationalisation challenge of cultural capital. Section 3 presents the data and methods used in the original analyses in this chapter. Section 4 provides descriptive statistics of the key explanatory variables and candidate cultural capital measures. Section 5 presents the modelling results which are presented as a series of sensitivity analyses of the candidate measures. First, cultural capital is modelled to better understand the inter-relationship between cultural capital and socio-economic background, and the intergenerational associations between parents and children. Second, GCSE attainment is modelled using the candidate cultural capital measures.

Alternative combinations of cultural capital measures are then compared and contrasted.

2 The Concept of Cultural Capital

Cultural capital can be understood as the accumulation of a set of skills, knowledge, attitudes, and behaviours which are sanctioned by the 'dominant' class in society (see Lamont and Lareau, 1988). Access to, or possession of, cultural resources can help individuals access 'scarce rewards' (Lareau and Weininger, 2003). In an education setting, rewards might be more favourable examination results, participation in higher education, or access to more prestigious universities. Lareau and Weiniger (2003: 587) stressed the exclusionary aspect of cultural capital and its potential to be monopolised, i.e. it is not universally available to all and is transmitted from parents to children so that advantage is passed down between generations.

2.1 Bourdieu and Cultural Capital

The concept of cultural capital is most closely associated with Pierre Bourdieu (Davies and Rizk, 2017). Lareau and Weininger (2003: 567) asserted that cultural capital is a 'signature concept' of Pierre Bourdieu. Bourdieu's original theorisation of cultural capital was intended to explain social class inequalities in educational attainment:

> *'The notion of cultural capital initially presented itself to me, in the course of research, as a theoretical hypothesis which made it possible to explain the unequal scholastic achievement of children originating from the different social classes by relating academic success, i.e. the specific profits which children from the different classes and class fractions can obtain in the academic market, to the distribution of cultural capital between the classes and class fractions'* (Bourdieu, 1986: 47).

Goldthorpe (2007) suggested that the role of cultural differences in educational outcomes was not new, and was reflective of a longer tradition in the sociology of education. For example, Bernstein noted that children from different social backgrounds were taught different (i.e. elaborated compared with restricted) linguistic codes at home (Bernstein, 1964). Jackson and Marsden's (1962) ethnographic study

in 1950s Huddersfield found that northern working class children struggled to adjust to the cultural environment of grammar schools. The cultural dissonance of the 'scholarship boy' in grammar schools was similarly presented in contemporaneous works, such as Hoggart (1957) and Lacey (1970).

2.1.1 The Forms of Cultural Capital

Cultural capital has three states: embodied, objectified, and institutionalised (Bourdieu, 1986). Embodied cultural capital is a set of predispositions associated with the culture of the 'dominant' class. The preference of certain tastes and understanding of the cultural codes of the dominant class symbolise familiarity and ease with 'high status' cultural practices (see Lamont and Lareau, 1988). Objectified cultural capital is the possession of cultural goods, which may be symbolic of status, such as paintings, antiques, or musical instruments. Institutionalised cultural capital is an accreditation or qualification of a certain level of education. Institutionalised cultural capital is a resource which can be converted into monetary rewards, for example, through increased salary in the labour market. It is not the certificate itself which infers cultural capital, but the amount of time and labour invested, and assumed skills and knowledge gained. It is theorised that the qualification therefore has symbolic value (Bourdieu, 1986).

The capital metaphor implies an element of accumulation of time and labour investment for each form of cultural capital before rewards are accessible (see Bourdieu, 1986, Field, 2003). Savage et al. (2005) used the terminology of 'capitals, assets, and resources', suggesting that the distinction between resources and capital was accumulation over time.

2.1.2 Cultural and Social Reproduction

Cultural capital plays a central role in Bourdieu's theory of cultural and social reproduction (Bourdieu, 1973). The theory of cultural and social reproduction is concerned with the 'hereditary transmission of power and privileges' (Bourdieu, 1973: 258). Bourdieu and Passeron (1990) argued that only those who already possess the

dominant cultural capital will gain from the education system, notably those with the linguistic and cultural competence of the dominant culture. Those who possess such cultural capital, i.e. those in more advantaged social positions to begin with, perform better educationally (Bourdieu and Passeron, 1990). Young people from less advantaged social class backgrounds are more likely to self-select out of particular educational routes 'based upon the unconscious estimation of the objective probabilities of success possessed by the whole category.... of a class deprived of cultural capital' (Bourdieu, 1973: 269). Bourdieu and Passeron (1990) further asserted that the education system has an appearance of rewarding based on ability, thereby legitimating the reproduction of the existing social hierarchies. The theory of cultural and social reproduction is therefore an explanation of how societies remain socially stratified.

2.2 Cultural Capital and Quantitative Sociology

Bourdieu's concept of cultural capital was theorised for the French educational context (see Bourdieu, 1984). Subsequent operationalisations of cultural capital have been applied to different educational contexts, for example, in Australia (Crook, 1997), Denmark (Jaeger, 2009), Germany (De Graaf, 1988), Greece (Katsillis and Rubinson, 1990), the Netherlands (De Graaf et al., 2000), Sweden (Jonsson, 1987), the UK (Sullivan, 2001), and in the USA (DiMaggio, 1982). The following sections present a review of previous quantitative applications of cultural capital and pay particular attention to operationalisation differences of measures that may be routinely collected within a social survey context. The review is organised around three of the most conventional operationalisations of cultural capital: parental education level, highbrow cultural participation, and reading behaviours.

2.2.1 Operationalisation Challenges

Sullivan (2002) argued that there is a lack of conceptual clarity in the original theorisation of cultural capital. Lamont and Lareau (1988) identified a variety of definitions of cultural capital used by Bourdieu. Cultural capital referred to informal academic standards such as linguistic ability in *The Inheritors* (Bourdieu and

Passeron, 1979) and *Reproduction in Education, Society and Culture* (Bourdieu and Passeron, 1990). Cultural capital referred to markers of taste and stratification between the middle and working classes, for example highbrow cultural participation like going to the theatre or the opera, in *Distinction* (Bourdieu, 1984). Lamont and Lareau (1988) further argued that the various forms and functions of the concept of cultural capital may not necessarily be compatible. This presents substantial empirical problems for researchers attempting to undertake analyses with suitable measures of the concept.

One of the key consequences of the conceptual vagueness has been varied subsequent operationalisations of cultural capital (DiMaggio, 1979, Sullivan, 2002). There are no agreed-upon measures of cultural capital, and a broad range of measures have been used in statistical analyses (Sullivan, 2002). Cultural capital has been tested using the following measures separately or in a number of different combinations:

- Highbrow arts and cultural participation, for example attendance at museums, theatres, or the opera (DiMaggio, 1982, De Graaf, 1986, Katsillis and Rubinson, 1990, Kalmijn and Kraaykamp, 1996, Aschaffenburg and Maas, 1997, Roscigno and Ainsworth-Darnell, 1999, De Graaf et al., 2000, Sullivan, 2001, Eitle and Eitle, 2002, Jaeger, 2009, Kraaykamp and Eijck, 2010, Jaeger and Breen, 2016);
- Cultural knowledge or interests relating to highbrow culture, music, and literature (DiMaggio, 1982, De Graaf, 1988, Sullivan, 2001);
- Number of books in the childhood home (Graetz, 1988, De Graaf, 1988, Sullivan, 2001, Jaeger and Breen, 2016);
- Reading or language behaviours (De Graaf, 1986, Kalmijn and Kraaykamp, 1996, De Graaf et al., 2000, Sullivan, 2001, Dumais, 2002, Jaeger and Holm, 2007, Jaeger and Breen, 2016);
- Attendance in classes outside of school (Aschaffenburg and Maas, 1997, Roscigno and Ainsworth-Darnell, 1999, Dumais, 2002, Eitle and Eitle, 2002, Jaeger and Breen, 2016);

- Educational resources in the home (Roscigno and Ainsworth-Darnell, 1999, Eitle and Eitle, 2002, Jaeger, 2009);
- Discussions between parents and children about books or school (Jaeger, 2009);
- Cultural possessions (Jaeger, 2009, Kraaykamp and Eijck, 2010);
- Parental education level (Jonsson, 1987, Egerton, 1997, Holm, 2003, Jaeger and Holm, 2007, Kraaykamp and Eijck, 2010).

Some studies have used parental cultural capital in statistical models of educational outcomes (De Graaf, 1986, De Graaf, 1988, Jonsson, 1987, Kalmijn and Kraaykamp, 1996, Egerton, 1997, De Graaf et al., 2000, Holm, 2003), others have used children's cultural capital (DiMaggio, 1982, Graetz, 1988, Farkas et al., 1990, Katsillis and Rubinson, 1990, Roscigno and Ainsworth-Darnell, 1999, Dumais, 2002, Eitle and Eitle, 2002, Jaeger and Holm, 2007), and some have used both parents' and children's cultural capital (Aschaffenburg and Maas, 1997, Sullivan, 2001, Jaeger, 2009, Kraaykamp and Eijck, 2010, Jaeger and Breen, 2016). Dumais (2002) remarked that there is a lack of easy comparability of measures across studies.

2.2.2 Parental Education Level

Bourdieu suggested that parental education reflects parental cultural capital (Bourdieu, 1973). Highest education level has been used as a proxy indicator for parent cultural capital in subsequent studies (Jonsson, 1987, Egerton, 1997, Holm, 2003, Jaeger and Holm, 2007, Kraaykamp and Eijck, 2010). Jonsson (1987) found that cultural origin (highest parental educational level) had a distinct effect from class origin on educational attainment in Sweden over the first half of the 20th century. Egerton (1997) analysed the National Child Development Study in the UK and demonstrated the importance of cultural capital (highest education level) for respondents entering the same occupational destinations as their parents. Using the Danish Youth Longitudinal Panel Survey, Jaeger and Holm (2003) found significant effects of cultural capital (father's highest level of schooling) on the highest level of education attained by their children by age 38. The authors commented that this effect was not as strong as that of father's social class background.

Parental education level has also been used in combination with other candidate measures of cultural capital. Jaeger and Holm (2007) analysed the choice of secondary school in Denmark by economic, cultural, and social capital and found that cultural capital explained much of the social class effect. The authors measured cultural capital using level of education as well as indicators of foreign languages, newspaper subscriptions, reading fiction, and interest in visual arts. Kraaykamp and Eijck (2010) found evidence of intergenerational transmission of cultural capital between parents and children in the Netherlands. The authors used highest education level as a measure of institutionalised cultural capital, alongside measures of embodied cultural capital (highbrow cultural participation) and objectified cultural capital (cultural possessions).

Sullivan (2002: 154) argued that Bourdieu was 'not entitled to assume that a high parental level of education reveals a high level of parent cultural capital'. The use of parental education level as a measure of cultural capital assumes that those attaining higher levels of education automatically possess greater levels of cultural capital or engage in classically highbrow cultural practices. The cultural omnivore thesis suggests that higher status individuals consume both highbrow and lowbrow cultural activities and practices (Peterson and Kern, 1996, Chan and Turner, 2017). In this way, the cultural consumption of higher status individuals is better characterised by 'omnivorousness' rather than 'snobbishness' (Peterson and Kern, 1996).

2.2.3 Highbrow Cultural Participation

An orthodox Bourdieusian operationalisation of cultural capital may use measures of highbrow cultural participation, such as attendance at theatres, museums, art galleries, the opera, or classical music performances (De Graaf et al., 2000). This operationalisation is commonplace in empirical work (Aschaffenburg and Maas, 1997, Reay, 2004). The findings are not always consistent. For example, DiMaggio (1982) used data on high school seniors in the USA and found that children's cultural capital (factor score of children's cultural arts participation) had a strong, positive impact on their high school grades. Cultural knowledge, cultural interests, and middlebrow activities were also included in the analyses. DiMaggio (1982) found that the cultural

capital effect was independent of the class background effect. This provided support for the cultural mobility thesis rather than Bourdieu's cultural reproduction thesis. In a replication study in Greece, Katsillis and Rubinson (1990) found that cultural capital (factor score of child cultural participation) did not have direct or indirect effects on academic attainment. Instead, the authors found that ability and effort were more important predictors of academic success. Ability was measured by an average of previous attainment in elementary and junior school exams and effort was measured by the average time spent on homework per day. Jaeger (2009) used measures of child cultural participation (museums and galleries, opera or classical music concerts, and live theatre) alongside parental socialisation (cultural possessions of art, literature and poetry), educational resources in the home, and parental investment (discussions with children about important issues, books, and school). The different aspects of cultural capital had independent, and perhaps complementary, effects on choice of secondary school in Denmark after controlling for socio-economic background.

A distinction between 'active' investment in cultural activities and, perhaps more, passive highbrow cultural participation has been made in the measures of cultural capital (Aschaffenburg and Maas, 1997, Roscigno and Ainsworth-Darnell, 1999). For example, Aschaffenburg and Maas (1997) measured active investment in cultural capital based on children's attendance at extra-curricular classes alongside parental highbrow cultural participation. Classes included music, visual arts, performance, art appreciation and music appreciation, and measures were collected at three timepoints. The results suggested that there is not necessarily a cumulative effect of investment in cultural capital, and that parental influences are distinct from the child's cultural capital (Aschaffenburg and Maas, 1997). Roscigno and Ainsworth-Darnell (1999) similarly measured investment in classes relating to art, music and dance as well as highbrow cultural participation. The results suggested that cultural capital had positive effects on Grade Point Averages in the USA. The effects were lower for Black students compared with white students and those from less advantaged socio-economic backgrounds compared with their more advantaged peers (Roscigno and Ainsworth-Darnell, 1999).

Highbrow cultural activities have been influenced by Bourdieu's descriptions of highbrow tastes and judgements in *Distinction* (Bourdieu, 1984). Prieur and Savage (2013) remarked that conventional operationalisation of highbrow cultural participation may be considered outdated as indicators of 'dominant' middle class leisure activities. Peterson and Kern (1996) similarly suggested that omnivorousness may better characterise cultural tastes and consumption of more advantaged individuals (also see Bennett, 2009). Savage et al. (2013) distinguished highbrow cultural capital from 'emerging' cultural capital, for example, belonging to social network sites, playing sport, or going to rock concerts.

2.2.4 Reading Behaviours

Reading was originally included as a form of highbrow culture in *Distinction* (Bourdieu, 1984). De Graaf (1988) examined the role of parental reading behaviours (the number of books in the home, the number of books read in the previous year, and interest in books) on educational transitions to Gymnasiums (i.e. the most academic secondary schools) in Germany. The results demonstrated a positive effect of favourable parental reading climate on their child attending Gymnasiums. Graetz (1988) modelled educational transitions in Australia and used the number of books in the childhood home as the sole indicator of cultural capital. The number of books in the home had a positive association with educational transitions net of social background, although the size of effect was relatively weak (Graetz, 1988). Reading behaviours have also been encompassed in broader definitions of educational resources. For example, Roscigno and Ainsworth-Darnell (1999) and Eitle and Eitle (2002) used the National Educational Longitudinal Study in the USA and constructed scales of educational resources. The items included having a daily newspaper, regular magazine, encyclopaedia, atlas, dictionary, computer, over fifty books, and a calculator in the home. Jaeger (2009) similarly measured educational resources using a four-item scale of having a dictionary, quiet place to study, a desk, and textbooks in the family home.

2.2.5 Highbrow Cultural Participation and Reading Behaviours

Empirical work using both operationalisations of cultural capital has tended to find that reading behaviours have a greater influence on educational outcomes than conventional highbrow cultural participation (De Graaf, 1986, Crook, 1997, De Graaf et al., 2000, Sullivan, 2001). For example, De Graaf (1986) examined the educational attainment of the two oldest siblings in the family in the Netherlands and measured parental cultural capital in terms of reading behaviours (number of hours spent reading per week and number of visits to the library per month) and highbrow cultural participation (number of visits to museums, theatres or concerts, and historical buildings per month). The effects of reading behaviours were greater than highbrow cultural participation, but there remained a large socio-economic effect which was not explained by the measures of cultural capital. Kalmijn and Kraaykamp (1996) used measures of parental cultural participation (going to classical music concerts, the theatre, museums or art galleries) and the extent to which the parents encouraged their children's reading habits. Access to cultural capital was associated with more years of schooling for students in the USA. In an Australian study, Crook (1997) found that cultural capital, operationalised in terms of reading behaviours, was associated with higher educational outcomes. This was not the case when cultural capital was operationalised as highbrow cultural participation, such as museum and gallery visits, or attendance at theatres or the opera. A similar finding emerged from a study in the Netherlands which found that cultural capital operationalised as reading behaviours (parents reading to their children) was associated with higher levels of education completed whereas cultural capital operationalised as cultural participation was not (De Graaf et al., 2000). This study further found that parents reading to their children had a more substantial effect for children from less advantaged socio-economic backgrounds, although those from more advantaged socio-economic backgrounds continued to have more favourable educational outcomes overall (De Graaf et al., 2000). Sullivan (2001) undertook a survey of secondary school pupils in the UK and found that reading behaviours (type and number of books read, use of the library, and reading newspapers) had positive effects on GCSE attainment whereas highbrow cultural participation (going to galleries, the theatre, or concerts) did not.

Previous empirical work has demonstrated that activities which develop linguistic knowledge (i.e. reading behaviours) are more strongly associated with educational outcomes than activities which may signal high familial status (i.e. highbrow cultural participation). The distinction between cultural capital as measured by reading behaviours and highbrow cultural participation has been described as status-seeking and information-processing patterns of cultural consumption (Ganzeboom, 1982), communicating status and educative resources (Crook, 1997), or the difference between verbal and non-verbal cultural capital (Sullivan, 2001).

The previous sections have illustrated the sizeable empirical challenge of defining and operationalising cultural capital for educational outcomes. The following analyses in this chapter directly address the operationalisation challenge by undertaking a series of sensitivity analyses of alternative candidate measures of cultural capital, informed by the literature reviewed above.

3 Data and Methods

3.1 Sample

The role of cultural capital is analysed using the Wave 2 UKHLS youth questionnaire and adult survey responses with linked NPD records. The potential measures of cultural capital are included as part of the rotating questionnaire content and are therefore not available in every wave of the UKHLS. The measures are reported every other wave in the youth questionnaire (Waves 2, 4, 6) and every three waves in the adult survey (Waves 2, 5). The total analytical sample is n=736 for Wave 2 participants with linked NPD records, and complete records on all potential measures included in the sensitivity analyses. Missing data are examined using multiple imputation methods (see Table A2.3 in Appendix 2). The missing data models do not alter the overall substantive conclusions drawn from the complete records results presented below.

3.2 Measures

The UKHLS collects a wealth of potential measures of cultural capital. The literature reviewed in section 2 has informed the selection of potential measures of cultural capital in the following analyses. The measures broadly represent highbrow cultural participation and reading behaviours for both parents and children. There are also measures which might represent 'emerging' cultural capital identified by Savage et al. (2013), such as going to rock concerts or watching live sport. Measures of 'emerging' cultural capital are not used in the following analyses, to focus on the role of more conventional cultural capital. Table 5.1 presents a summary of the candidate measures reported in the adult survey (termed parent measures) and in the youth questionnaire (termed child measures).

Table 5. 1: Candidate cultural capital measures available in Wave 2 of the UKHLS adult survey and youth questionnaire

	Parent	Child
Highbrow Cultural Participation		
Goes to the theatre	✓	✓
Goes to museums or galleries	✓	✓
Goes to visit an historic place or stately home	✓	✓
Goes to the opera or operetta	✓	
Goes to classical music performances	✓	
Adults take young person to the theatre		✓
Adults take young person to museums or galleries		✓
Extra Classes		
Music	✓	✓
Dance	✓	✓
Art	✓	
Sport	✓	
Religion	✓	
Tutorials in school subjects	✓	
Reading Behaviours		
Reads for pleasure	✓	✓
Goes to the library	✓	✓
Discusses books at home		✓
Is a member of a book club	✓	
Adults give young person books as gifts		✓

Throughout this thesis, parental education level has been employed as a measure of socio-economic background and is used throughout the following analyses as such. Parental education level is also often used as a measure of cultural capital. To reflect this, parental education is reconsidered as a measure of cultural capital instead of socio-economic background in the final section of the analyses.

3.2.1 Highbrow Cultural Participation

There are measures of cultural activities collected in the UKHLS which encompass Bourdieu's highbrow participation (Bourdieu, 1984). The measures most appropriately corresponding to highbrow cultural participation for parents include going to museums or galleries, going to the theatre, visiting historical places, going to the opera, and seeing classical music performances. The measures in the adult survey were dichotomous variables for whether the respondent 'mentioned' or 'did not mention' the stated activity. The youth questionnaire also asked respondents how often they went to museums or galleries, the theatre, and visited historical places. In the following analyses, summed scales are created for parent and child highbrow participation as quantity measures of engagement. The child and parent highbrow participation scales comprise whether or not the parents and young people go to the theatre, museums or galleries, or visit historical places.

Young people are also asked how often adults take them to the theatre and how often adults take them to museums or art galleries. A further set of seven questions asked youth respondents about extra-curricular classes: music, art, dance, sport, religion, and whether the respondent is tutored in any school subjects. These variables can represent 'active' investment in capital (Aschaffenburg and Maas, 1997). Music, art, and dance in particular can be understood as cultural activities (Roscigno and Ainsworth-Darnell, 1999, Dumais, 2002, Eitle and Eitle, 2002).

3.2.2 Reading Behaviours

The youth questionnaire asked respondents to report the number of books they had read in the past month. This variable has also been dichotomised to represent reading

for pleasure for these analyses. Youth respondents were also asked how often they discuss books with adults in the household; how often adults give them books as gifts; and how often they use the library. The adult survey contained similar measures of parents' reading habits, including whether the parent reads for pleasure or not, whether they are a member of a book club, and whether they use the library or not. The measures are tested individually and as summed scales of quantity of engagement. The child and parent reading scales comprise whether the respondent reads for pleasure, discusses books, and goes to the library.

3.3 Structure of Analysis

Sullivan (2001) and Jaeger (2009) both noted that researchers tend to test partial components of a cultural capital theory rather than the broader theory of cultural reproduction. The broader theory of cultural reproduction emphasised that there must be evidence of parents owning or accessing cultural capital, that they transmit this to their children, and that it is the unequal distribution of cultural capital which aids children's educational performances (Sullivan, 2001, Jaeger, 2009). De Graaf et al. (2000) provided refinements to testing the role of cultural capital and cultural reproduction. First, evidence in favour of the cultural reproduction thesis would demonstrate that the impact of cultural capital is greater for children from more advantaged than less advantaged backgrounds. Second, evidence for the cultural reproduction thesis would demonstrate that cultural capital mediates the role of social background in educational outcomes (De Graaf et al., 2000). Building on this work, the following questions are explored in the subsequent analyses.

1. Is there a socio-economic gradient to cultural capital?
2. Is cultural capital transmitted from parents to children?
3. Does cultural capital have a positive effect on GCSE attainment?
    a. Is there a 'raw' effect of cultural capital on GCSE attainment?
    b. Does cultural capital mediate the role of socio-economic background in GCSE attainment?

The analyses in this chapter are presented as a series of sensitivity analyses of alternative candidate measures of cultural capital to directly address the operationalisation challenge outlined above.

4 Descriptive Statistics

The descriptive statistics and mean GCSE scores for the key explanatory variables are presented in Table 5.2. The descriptive statistics demonstrate that young people from more advantaged socio-economic backgrounds tend to attain more favourable GCSE outcomes. Young people living in owned or privately rented homes tend to attain higher GCSE scores than young people living in social housing, and girls tend to attain higher GCSE scores than boys. Information has been omitted (represented by *) where the sample sizes do not meet the minimum threshold for statistical disclosure control.

Table 5.3 presents the descriptive statistics and mean GCSE scores for the candidate cultural capital measures. Categories have been aggregated for many of the frequency measures where sample sizes are too low to satisfy statistical disclosure control requirements. Frequency measures of child cultural capital are also re-coded as binary measures for comparability with the parent cultural capital measures. The mean GCSE scores are detailed in the last column. Mean GCSE scores tend to be higher if young people or their parents participate in highbrow cultural activities, have more positive reading behaviours, and attend extra-curricular classes. The trends are relatively consistent across all measures of child and parent cultural capital measures. Generally, young people who engage in these activities ‘several times a year’ tend to attain higher overall GCSE scores than those who engage less frequently. Interestingly, those who engage in an activity ‘several times a year’ tend to also attain higher overall GCSE scores than those who engage in the activity more frequently. This might suggest that very regular active engagement in cultural activities is not as consequential as more casual engagement. For example, going to museums or art galleries every weekend or once a month may not provide additional educational benefits compared with going a few times a year.

Table 5. 2: Descriptive statistics for key explanatory variables, unweighted

	Freq.	Percent	GCSE score	Std. Dev.
Parental Education Level				
Higher education	216	29.35	56.98	19.43
Further education	112	15.22	43.05	18.84
School-level education	323	43.89	38.06	20.90
Below school-level education	85	11.55	28.33	20.41
Parental NS-SEC				
1.1 Large employers and higher managerial	42	5.71	53.31	18.35
1.2 Higher professional occupations	64	8.70	62.08	16.11
2 Lower management and professional occupations	151	20.52	49.40	19.30
3 Intermediate occupations	76	10.33	45.25	20.41
4 Small employers and own account workers	73	9.92	44.23	21.43
5 Lower supervisory and technical occupations	43	5.84	39.26	19.89
6 Semi-routine occupations	88	11.96	38.65	23.60
7 Routine occupations	56	7.61	33.75	21.84
8 Not in employment	143	19.43	31.55	21.01
Housing Tenure				
Owned/privately rented	551	74.86	47.71	21.23
Social housing	185	25.14	29.95	19.90
Gender				
Male	356	48.37	39.89	22.62
Female	380	51.63	46.40	21.50
Ethnicity				
White	573	77.85	43.15	22.16
Mixed	*	*	*	*
Asian/Asian British	90	12.23	45.80	23.86
Black/Black British	32	4.35	37.03	17.98
Other ethnic group	*	*	*	*
Age				
12	29	3.94	40.38	23.32
13	166	22.55	43.49	22.34
14	278	37.77	43.73	22.89
15	263	35.73	42.90	21.54
Total *n*	**736**	**100**	**43.25**	**22.27**

Table 5. 3: Descriptive statistics for candidate cultural capital measures, unweighted

	Freq.	Percent	GCSE score	Std. Dev.
Highbrow Cultural Participation				
Child goes to the theatre				
Most days/once a week/once a month	24	3.26	40.54	24.60
Several times a year	113	15.35	54.30	19.33
Once a year or less	243	33.02	46.53	21.36
Never/almost never	356	48.37	37.68	21.93
Child goes to the theatre				
Yes	380	51.63	48.46	21.33
No	356	48.37	37.68	21.93
Child goes to museums/art galleries				
Most days/once a week/once a month	25	3.40	40.92	30.07
Several times a year	156	21.20	52.26	22.78
Once a year or less	205	27.85	47.96	21.40
Never/almost never	350	47.55	36.64	19.76
Child goes to museums/art galleries				
Yes	386	52.45	49.24	22.74
No	350	47.55	36.64	19.76
Child visits historical places				
Most days/once a week/once a month	22	2.99	50.14	29.39
Several times a year	154	20.92	51.73	22.44
Once a year or less	190	25.82	46.13	21.48
Never/almost never	370	50.27	37.82	20.67
Child visits historical places				
Yes	366	49.73	48.73	22.52
No	370	50.27	37.82	20.67
Child highbrow scale				
0	203	27.58	34.46	19.46
1	158	21.47	38.86	21.32
2	151	20.52	46.13	22.41
3	224	30.43	52.36	21.47
Parents take to the theatre				
Often	34	4.62	47.12	25.65
Sometimes	107	14.54	48.58	22.75
Rarely	216	29.35	46.65	20.16
Never	379	51.49	39.45	22.37
Parents take to museums/ galleries				
Often	34	4.62	51.32	23.62
Sometimes	147	19.97	49.33	23.08
Rarely	260	35.33	47.18	22.05
Never	295	40.08	35.82	19.73
Parent goes to the theatre				
Yes	352	47.83	49.36	21.16
No	384	52.17	37.64	21.81
Parent goes to museums/ galleries				
Yes	336	45.65	49.38	21.07
No	400	54.35	38.09	21.97
Parent visits historical places				
Yes	485	65.90	47.93	21.35
No	251	34.10	34.20	21.23

Table 5.3 continued

Parent goes to the opera					
	Yes	27	3.67	53.82	17.44
	No	709	96.33	42.85	22.35
Parent goes to watch classical music					
	Yes	79	10.73	59.76	16.88
	No	657	89.27	41.26	22.03
Parent highbrow scale					
	0	177	24.05	32.93	20.95
	1	162	22.01	38.11	21.78
	2	180	24.46	46.90	20.94
	3	217	29.48	52.47	20.23
Extra Classes					
Child has music classes					
	Yes	241	32.74	53.30	20.68
	No	495	67.26	38.36	21.38
Child has art classes					
	Yes	35	4.76	44.26	24.74
	No	701	95.24	43.20	22.16
Child has dance classes					
	Yes	95	12.91	51.62	20.45
	No	641	87.09	42.01	22.28
Child has sports practice					
	Yes	324	44.02	45.81	22.34
	No	412	55.98	41.24	22.03
Child has religious classes					
	Yes	46	6.25	53.17	18.10
	No	690	93.75	42.59	22.38
Child is tutored in school subjects					
	Yes	46	6.25	51.85	18.97
	No	690	93.75	42.67	22.37
Child classes scale					
	0	210	28.53	33.39	20.30
	1	329	44.70	43.36	21.79
	2	143	19.43	53.14	20.25
	3+	54	7.34	54.70	20.76
Reading Behaviours					
Child reads for pleasure					
	Yes	528	71.74	46.53	22.14
	No	208	28.26	34.91	20.40
Child discusses books with adults					
	Often	38	5.16	50.90	30.29
	Sometimes	167	22.69	46.76	23.06
	Rarely	270	36.68	44.72	21.17
	Never	261	35.46	38.36	20.63
Child receives books from adults					
	Often	89	12.09	47.36	23.97
	Sometimes	244	33.15	49.00	22.27
	Rarely	235	31.93	41.72	20.61
	Never	168	22.83	34.86	20.78

Table 5.3 continued

Child goes to the library				
Most days/once a week	36	4.89	38.50	26.86
Once a month	60	8.15	49.72	21.96
Several times a year	127	17.26	53.78	21.20
Once a year or less	112	15.22	44.55	19.88
Never/almost never	401	54.48	39.01	21.54
Child goes to the library				
Yes	335	45.52	48.33	22.10
No	401	54.48	39.01	21.54
Child reading scale				
0	150	20.38	34.50	20.42
1	241	32.74	40.31	20.45
2	208	28.26	48.25	22.04
3	137	18.61	50.39	23.60
Parent reads for pleasure				
Yes	579	78.67	45.98	21.41
No	157	21.33	33.16	22.55
Parent member of book club				
Yes	33	4.48	55.67	21.95
No	703	95.52	42.66	22.13
Parent goes to the library				
Yes	371	50.41	48.39	21.49
No	365	49.59	38.02	21.86
Parent reading scale				
0	98	13.32	29.84	21.57
1	318	43.21	40.33	21.54
2	295	40.08	49.68	20.48
3	25	3.40	57.08	22.15
Total *n*	**736**	**100**	**43.25**	**22.27**

5 Modelling Results

5.1 Modelling Cultural Capital

The first set of analyses examine the extent to which cultural capital is unequally distributed by socio-economic background and is transmitted from parents to children. Table 5.4 presents the results of the chi square tests of associations between parental education, parental social class, and the candidate measures of cultural capital. All candidate measures of parental highbrow participation, child highbrow participation, and parental reading behaviours are significantly associated with parental education and parental social class. The strength of association is weak to moderate for all measures. The candidate measures of child reading behaviour are significantly

associated with parental education level, but not with parental social class. This suggests that there is evidence of significant, but modest, bivariate associations between the measures of parental socio-economic background and cultural capital.

The statistics in Table 5.5 demonstrate that there are significant chi square associations between all corresponding parent and child cultural capital measures. The associations are relatively stronger for highbrow participation than reading behaviours. The significant associations of the summed scale measures further suggest that the children of more engaged parents engage in more activities themselves.

Table 5. 4: Chi square tests of association and Cramer's V statistics for parental socio-economic background and cultural capital measures

	Parental Education Level			**Parental NS-SEC**		
	Chi Square (d.f.)		**V**	**Chi Square (d.f.)**		**V**
Child Highbrow						
Goes to the theatre	28.66 (3)	***	0.20	35.38 (8)	***	0.22
Goes to museums or galleries	30.89 (3)	***	0.21	37.57 (8)	***	0.23
Visits historical places	35.60 (3)	***	0.22	42.42 (8)	***	0.24
Adults take to theatre	24.41 (3)	***	0.18	27.79 (8)	**	0.19
Adults take to museums/galleries	26.32 (3)	***	0.19	22.68 (8)	**	0.18
Highbrow scale (0-3)	55.23 (9)	***	0.16	66.16 (24)	***	0.17
Parent Highbrow						
Goes to the theatre	89.76 (3)	***	0.35	98.46 (8)	***	0.37
Goes to museums or galleries	105.12 (3)	***	0.38	131.92 (8)	***	0.42
Visits historical places	87.31 (3)	***	0.34	128.11 (8)	***	0.42
Goes to the opera/operetta	23.43 (3)	***	0.18	21.94 (8)	**	0.17
Goes to see classical music	68.17 (3)	***	0.30	52.11 (8)	***	0.27
Highbrow scale (0-3)	191.05 (9)	***	0.29	217.66 (24)	***	0.31
Child Reading						
Reads for pleasure	14.09 (3)	**	0.14	8.84 (8)		0.11
Goes to the library	8.89 (3)	*	0.11	11.43 (8)		0.13
Adults give books as gifts	18.58 (3)	***	0.16	14.65 (8)		0.14
Discuss books at home	4.03 (3)		0.07	5.97 (8)		0.09
Reading scale (0-3)	13.39 (9)		0.07	18.90 (24)		0.09
Parent Reading						
Reads for pleasure	58.99 (3)	***	0.28	66.93 (8)	***	0.30
Goes to the library	42.46 (3)	***	0.24	20.44 (8)	**	0.17
Is a member of a book club	28.10 (3)	***	0.20	18.88 (8)	*	0.16
Reading scale (0-3)	121.44 (9)	***	0.24	79.24 (24)	***	0.19
Child Extra Classes						
Music classes	59.89 (3)	***	0.29	30.12 (8)	***	0.20
Art classes	0.50 (3)		0.03	3.95 (8)		0.07
Dance classes	9.66 (3)	*	0.12	10.06 (8)		0.12
Sports practice	7.02 (3)		0.10	17.18 (8)	*	0.15
Religious classes	5.08 (3)		0.08	4.76 (8)		0.08
Tutored for school subjects	1.93 (3)		0.05	7.85 (8)		0.10

Table 5. 5: Chi square tests of association and Cramer's V statistics for parent and child cultural capital measures

	Parent and Child Chi Square (d.f.)		V
Highbrow Measures			
Goes to the theatre	69.01 (1)	***	0.31
Goes to museums or galleries	56.63 (1)	***	0.28
Visits historical places	67.47 (1)	***	0.30
Highbrow scale (0-3)	130.45 (9)	***	0.24
Reading Measures			
Reads for pleasure	5.40 (1)	*	0.09
Goes to the library	46.65 (1)	***	0.25
Discuss books (at home/book club)	12.25 (1)	***	0.13
Reading scale (0-3)	51.35 (9)	***	0.15
Extra Classes			
Music	44.25 (1)	***	0.25
Dance	8.29 (1)	**	0.11

Tables 5.6 and 5.7 present Poisson regression models of the young person's cultural capital engagement. The outcome variables are the child's highbrow participation scale (Table 5.6) and child's reading scale (Table 5.7), measured as count variables (which range from 0 to 3).[16] Higher values of the outcome variables represent greater engagement in cultural capital activities. In both tables, Model 1 includes parental education and parental social class. Parental education is significantly associated with both child highbrow participation and reading behaviours. Parental NS-SEC is significantly associated with child highbrow participation, but not associated with child reading behaviours. This is consistent with the chi square tests outlined above. Model 2 includes a measure of parental cultural capital. The results illustrate that the greater parental engagement in highbrow participation and reading behaviours is associated with greater child engagement. This lends support to the idea that cultural capital may be transmitted from parent to child. Comparing Models 1 and 2 in both tables, parent cultural capital seems to mediate the role of parental socio-economic background.

[16] There was no evidence of overdispersion or a high proportion of zero counts, and therefore a Poisson model was preferred over a negative binomial model and a zero-inflated Poisson model.

Table 5. 6: Poisson regression model of child highbrow participation

	Model 1			Model 2		
	Coef.	**S.E.**		**Coef.**	**S.E.**	
Poisson Regression: Child Highbrow Scale						
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	-0.05	(0.11)		0.01	(0.10)	
School-level education	-0.23	(0.10)	*	-0.07	(0.09)	
Below school-level education	-0.64	(0.20)	**	-0.40	(0.18)	*
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.12	(0.16)		-0.16	(0.15)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	-0.10	(0.10)		-0.10	(0.10)	
3 Intermediate occupations	-0.17	(0.14)		-0.08	(0.13)	
4 Small employers and own account workers	-0.24	(0.15)		-0.16	(0.14)	
5 Lower supervisory and technical occupations	-0.58	(0.23)	*	-0.45	(0.22)	*
6 Semi-routine occupations	-0.10	(0.14)		-0.02	(0.12)	
7 Routine occupations	-0.26	(0.19)		0.02	(0.21)	
8 Not in employment	-0.41	(0.14)	**	-0.17	(0.15)	
Parent Highbrow Scale (0-3)						
0				Ref.	(.)	
1				0.23	(0.17)	
2				0.59	(0.15)	***
3				0.80	(0.14)	***
Constant [output omitted]						
Observations	736			736		
McFadden's Adjusted Pseudo R^2	0.019			0.043		
Cox-Snell Pseudo R^2	0.088			0.161		
Nagelkerke Pseudo R^2	0.092			0.168		
BIC (d.f.)	2312.991 (12)			2271.102 (15)		
BIC (based on deviance)	-2545.514			-2587.404		
BIC (based on likelihood ratio chi square)	5.029			-36.861		

*Source: Wave 2 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. * p<.05, ** p<.01, *** p<.001.*

Table 5. 7: Poisson regression model of child reading behaviours

	Model 1			Model 2		
	Coef.	**S.E.**		**Coef.**	**S.E.**	
Poisson Regression: Child Reading Scale						
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	-0.24	(0.12)	*	-0.18	(0.11)	
School-level education	-0.19	(0.09)	*	-0.08	(0.09)	
Below school-level education	-0.20	(0.14)		-0.04	(0.14)	
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.02	(0.20)		0.02	(0.22)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	0.05	(0.12)		0.06	(0.12)	
3 Intermediate occupations	0.01	(0.15)		0.01	(0.15)	
4 Small employers and own account workers	-0.01	(0.15)		-0.02	(0.15)	
5 Lower supervisory and technical occupations	-0.01	(0.20)		0.06	(0.19)	
6 Semi-routine occupations	0.06	(0.15)		0.05	(0.14)	
7 Routine occupations	0.17	(0.17)		0.15	(0.17)	
8 Not in employment	-0.02	(0.15)		0.03	(0.15)	
Parent Reading Scale (0-3)						
0				Ref.	(.)	
1				0.15	(0.17)	
2				0.46	(0.17)	**
3				0.56	(0.23)	*
Constant [output omitted]						
Observations	736			736		
McFadden's Adjusted Pseudo R^2	-0.006			0.003		
Cox-Snell Pseudo R^2	0.015			0.047		
Nagelkerke Pseudo R^2	0.016			0.050		
BIC (d.f.)	2188.374 (12)			2183.256 (15)		
BIC (based on deviance)	-2670.132			-2675.249		
BIC (based on likelihood ratio chi square)	61.741			56.624		

*Source: Wave 2 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. * p<.05, ** p<.01, *** p<.001.*

5.2 Modelling GCSE Attainment

The concept of cultural capital was originally theorised to explain why those from more advantaged backgrounds tend to have more favourable educational outcomes than their less advantaged peers. Those from more advantaged backgrounds are understood to have greater cultural resources which are rewarded in educational settings. The analyses of bivariate associations above provide tentative evidence that cultural capital is stratified by socio-economic background and transmitted from parents to children. De Graaf et al. (2000) and Sullivan (2001) noted that the inclusion of cultural capital measures in models of educational attainment should weaken, i.e. mediate, the effects of parental socio-economic background.

As a first step in understanding the role of cultural capital in GCSE attainment, Table 5.8 presents the 'raw' effects of the candidate cultural capital measures on GCSE point score. The 'raw' effects are identified by estimating linear regression models of GCSE point score with one explanatory variable. The table presents the model fit statistics (Adjusted R^2 and BIC), and significance of the raw effects of parental education level, parental social class, and each candidate measure of cultural capital. Parental socio-economic background has the largest effect on GCSE point score. There are also positive, significant raw effects for most measures of cultural capital. Young people or parents who engage in highbrow cultural activities, reading for pleasure, and extra-curricular activities tend to have higher overall GCSE scores compared with those who engage less frequently or not at all. The number of books a young person reads does not have a significant raw effect on GCSE score, but there is a significant effect of reading for pleasure (as a binary measure). This suggests that quantity is not necessarily important, but taking an interest in reading, is. Attending art classes and having sports practice are also not significantly associated with GCSE score.

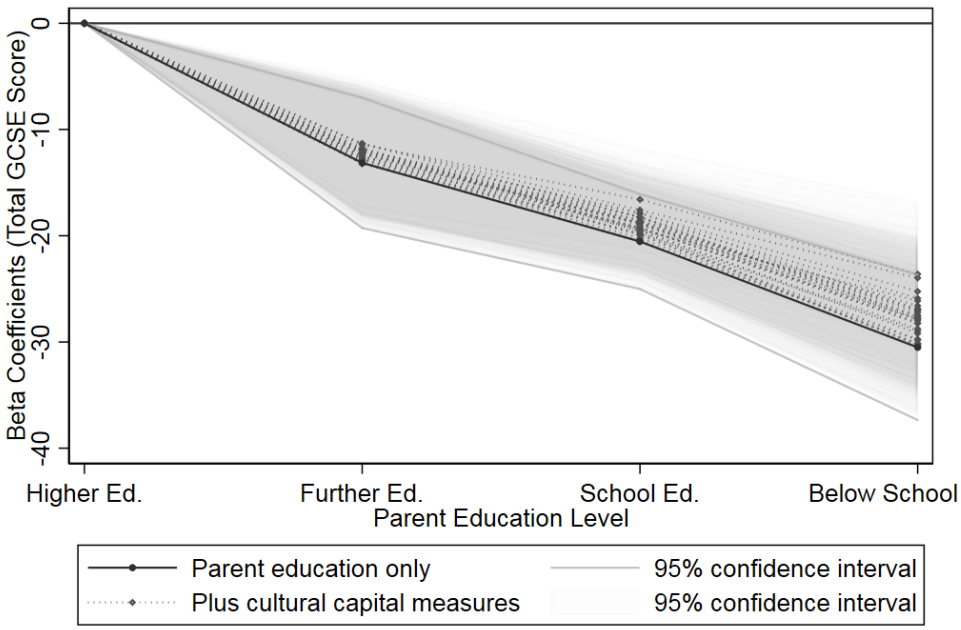
In the next stage of the analyses, cultural capital measures are entered separately into linear regression models of GCSE attainment along with parental education or parental social class. Particular attention is paid to how the effects of parental education or parental social class change as the cultural capital measures are included. Partial mediation occurs where the effect of the independent variable is reduced after controlling for the mediating variable. Perfect mediation occurs where the independent variable is not significantly associated with the dependent variable after controlling for the mediating variable (Baron and Kenny, 1986). The models are re-estimated for each separate candidate measure as a series of sensitivity analyses. Given the large number of models estimated, a visual representation of the results has been preferred to presenting tables of output. The results are visualised in Figures 5.1 and 5.2.

Table 5. 8: Raw effects (bivariate) of cultural capital measures on total GCSE point score

	Type of Variable	Adj. R^2	BIC (d.f.)	p<.05
Parental Socio-Economic				
Parental Education Level	Categorical	0.184	6529.364 (4)	Yes
Parental NS-SEC	Categorical	0.156	6582.302 (9)	Yes
Child Highbrow Participation				
Goes to the theatre	Categorical	0.072	6624.082 (4)	Yes
	Binary	0.057	6624.438 (2)	Yes
Goes to museums/art galleries	Categorical	0.086	6613.111 (4)	Yes
	Binary	0.079	6607.602 (2)	Yes
Visits historical places	Categorical	0.064	6630.694 (4)	Yes
	Binary	0.059	6623.348 (2)	Yes
Parents take to museums/ galleries	Categorical	0.073	6623.399 (4)	Yes
	Binary	0.026	6648.400 (2)	Yes
Parents take to the theatre	Categorical	0.028	6658.494 (4)	Yes
	Binary	0.011	6660.121 (2)	Yes
Child highbrow scale (0-3)	Continuous	0.104	6587.234 (2)	Yes
	Count	0.102	6599.818 (4)	Yes
Parent Highbrow Participation				
Goes to the theatre	Binary	0.068	6616.100 (2)	Yes
Goes to museums/art galleries	Binary	0.063	6620.333 (2)	Yes
Visits historical places	Binary	0.084	6603.031 (2)	Yes
Goes to see opera	Binary	0.007	6662.559 (2)	Yes
Goes to see classical music	Binary	0.065	6618.502 (2)	Yes
Parent highbrow scale (0-5)	Continuous	0.134	6562.071 (2)	Yes
	Count	0.138	6581.188 (6)	Yes
Parent highbrow scale (0-3)	Continuous	0.118	6575.195 (2)	Yes
	Count	0.117	6587.471 (4)	Yes
Child Reading Behaviours				
Reads for pleasure	Binary	0.054	6627.084 (2)	Yes
Number of books read	Continuous	0.003	6665.602 (2)	No
Discusses books with adults	Categorical	0.026	6659.375 (4)	Yes
	Binary	0.013	6658.347 (2)	Yes
Goes to the library	Categorical	0.063	6636.816 (5)	Yes
	Binary	0.042	6636.177 (2)	Yes
Receives books as presents	Categorical	0.056	6636.416 (4)	Yes
	Binary	0.046	6633.427 (2)	Yes
Child reading scale (0-3)	Continuous	0.066	6617.541 (2)	Yes
	Count	0.067	6628.176 (4)	Yes
Parent Reading Behaviours				
Reads for pleasure	Binary	0.054	6626.714 (2)	Yes
Discusses books (book club)	Binary	0.013	6658.067 (2)	Yes
Goes to the library	Binary	0.053	6627.777 (2)	Yes
Parent reading scale (0-3)	Continuous	0.101	6589.814 (2)	Yes
	Count	0.099	6602.621 (4)	Yes
Active Investment				
Has music classes	Binary	0.098	6591.992 (2)	Yes
Has art classes	Binary	-0.001	6668.829 (2)	No
Has dance classes	Binary	0.020	6653.299 (2)	Yes
Has sports practice	Binary	0.009	6661.217 (2)	No
Has religious classes	Binary	0.012	6659.078 (2)	Yes
Is tutored in school subjects	Binary	0.009	6661.541 (2)	Yes
Classes summed scale (0-3)	Continuous	0.107	6584.817 (2)	Yes
	Count	0.110	6593.138 (4)	Yes

Figure 5.1 presents the parental education coefficients in models of GCSE point score with and without the alternative measures of cultural capital. The bold, dark grey line represents the linear regression coefficients for GCSE point score in a model with parental education as the only explanatory variable. There is a clear downward gradient in the overall GCSE score by parental education level. The dotted lines represent the linear regression coefficients for GCSE point score by parental education after including the cultural capital measures separately. Each line represents a separate model. The coefficients for parental education change minimally after adding in each cultural capital measure. Young people with at least one graduate parent continue to attain higher overall GCSE scores than young people whose parents have any lower level of education, even after accounting for the cultural capital measures.

Figure 5. 1: Sensitivity analyses of the effect of cultural capital measures on parental education level regression coefficients with 95% confidence intervals

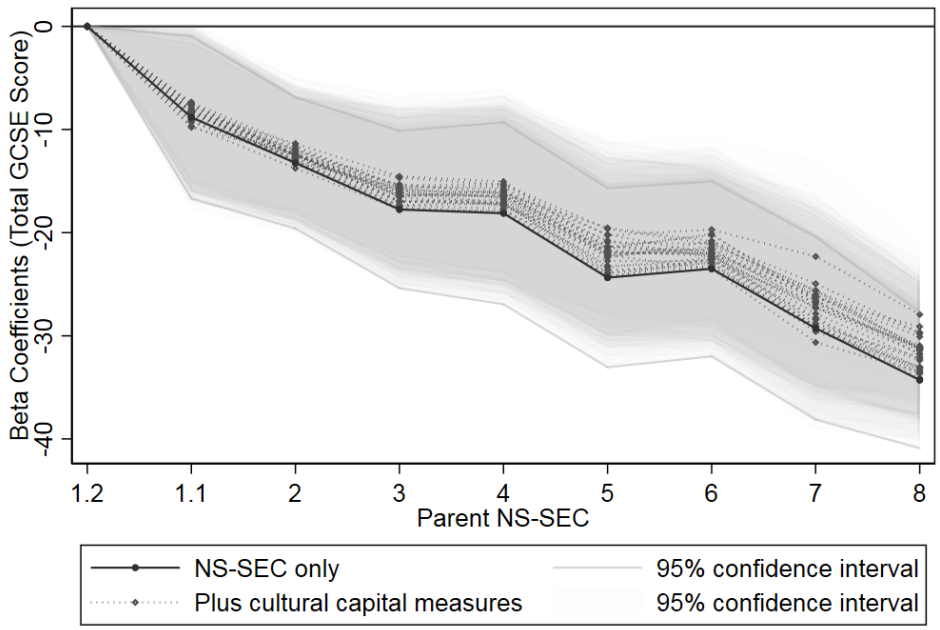


Source: Wave 2 Youth Questionnaires, UKHLS-NPD data, n=736. Adjusted for complex survey design.
Models include parent education level and cultural capital measures only.

Figure 5.2 presents the parental NS-SEC coefficients in models of GCSE score with and without the alternative measures of cultural capital. The coefficients do not change substantially after adding cultural capital measures into the model. The change in coefficients is very minimal for the professional occupations (NS-SECs 1.1

and 2), modest for the intermediate occupations of NS-SEC 3 and 4, and relatively largest for routine and manual occupations (NS-SECs 5, 6, and 7). The cultural capital measures do not, however, substantially reduce the socio-economic gradient or remove the persisting association between NS-SEC and overall GCSE score. Young people whose parents have occupations in NS-SEC 1.2 continue to attain higher overall GCSE scores than young people whose parents have occupations in NS-SECs 2, 3, 4, 5, 6, 7, or who are not in employment. There are still clear parental socio-economic background effects in GCSE attainment which are much stronger than, and largely independent of, the role of cultural capital.

Figure 5. 2: Sensitivity analyses of the effect of cultural capital measures on parental NS-SEC regression coefficients with 95% confidence intervals



Source: Wave 2 Youth Questionnaires, UKHLS-NPD data, n=736. Adjusted for complex survey design.
Models include parent NS-SEC and cultural capital measures only.

5.3 A Cultural Capital Model

The previous section illustrated that the separate cultural capital measures have significant raw effects on GCSE attainment, but do not convincingly mediate the role of parental socio-economic background. This section builds a model of cultural

capital, combining different candidate measures to further test the effects of cultural capital in GCSE attainment.

5.3.1 Factor Scores

Factor analysis is a robust method of data reduction routinely used in social science research (Bartholomew, 2008). This method has been used in previous empirical work to operationalise cultural capital (for example, DiMaggio, 1982, De Graaf, 1986, De Graaf, 1988, Katsillis and Rubinson, 1990, Hartas, 2016). The aggregation of separate measures into a single factor score is attractive because individual measures tend to be highly correlated. The resultant factor score can be interpreted as a quantity measure, whereby the individual has 'more' or 'less' cultural capital.

Factor analysis is estimated using the correlation matrix of variables (Pett et al., 2003). In this case, the tetrachoric correlation matrix is used for dichotomous variables (Uebersax, 2015). The tetrachoric correlations are presented in Table 5.9. The results presented use principal components analysis, and the results were also re-estimated using the default (principal factor) method in Stata (see factor, StataCorp., 2017a). The 14 parent and child variables have internal reliability, with a Cronbach's alpha value of 0.76. This is over the conventional 0.70 threshold (Acock, 2013). Mehmetoglu and Jakobsen (2017) advise that the number of retained factors should be based on eigenvalues, scree test, and theoretical sense. The first three factors had eigenvalues over 1. The first factor has an eigenvalue of 5.55, the second factor has an eigenvalue of 2.42, and the third factor has an eigenvalue of 1.46. The scree plot further confirms that three factors should be retained. The interpretation of the factors, using theoretical reasoning, is explored below.

Table 5. 9: Tetrachoric correlations for measures of parent and child cultural capital[17]

	Child Cultural Capital						Parent Cultural Capital							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-													
2	0.51	-												
3	0.50	0.85	-											
4	0.28	0.49	0.41	-										
5	0.32	0.60	0.46	0.50	-									
6	0.24	0.51	0.43	0.45	0.52	-								
7	0.46	0.27	0.31	0.09	-0.01	0.11	-							
8	0.38	0.42	0.48	0.20	0.07	0.18	0.59	-						
9	0.35	0.32	0.48	0.25	0.09	0.20	0.54	0.73	-					
10	0.27	0.21	0.29	-0.02	-0.06	-0.08	0.66	0.68	0.42	-				
11	0.40	0.37	0.45	0.22	0.17	0.17	0.59	0.59	0.57	0.59	-			
12	0.21	0.17	0.28	0.15	0.00	0.14	0.50	0.56	0.49	0.18	0.37	-		
13	0.25	0.17	0.22	0.23	0.39	0.26	0.34	0.38	0.34	0.23	0.25	0.24	-	
14	0.22	0.28	0.32	0.11	0.11	0.32	0.50	0.44	0.33	0.12	0.35	0.94	0.30	-

Table 5.10 presents the factor loadings for the three factor solution. The rotated factor loadings are based on oblique promax rotations to allow for correlations between factors (Kim, 1978). The rotated and unrotated solutions are presented in full in Table A1.7 in Appendix 1. There are clear theoretical distinctions between parent and child cultural capital in the three factor solution. Parent highbrow cultural participation measures load most strongly onto the first factor, child's cultural capital (both highbrow participation and reading behaviours) load most strongly onto the second factor, and parent reading behaviour measures load most strongly onto the third factor. The measure of parent library use loads weakly on all factors, and the measure of young person theatre attendance loads strongly (just over the 0.4 threshold) on two factors. The results are re-estimated separately without the parent library and child theatre variables, but the substantive results remain the same as in the model presented below.

[17] Key: 1 Theatre, 2 Museum, 3 Historic place, 4 Read for pleasure, 5 Library, 6 Discuss books, 7 Theatre, 8 Museum, 9 Historic place, 10 Opera, 11 Classical music, 12 Read for pleasure, 13 Library, 14 Discuss books.

Table 5. 10: Rotated (oblique promax) factor loadings for principal components factor analysis of cultural capital

	Variable	Factor 1	Factor 2	Factor 3	Uniqueness
Parent Cultural Capital	Theatre	**0.74**			0.29
	Museum	**0.75**			0.23
	Historic place	**0.63**			0.40
	Opera	**1.00**			0.20
	Classical music	**0.74**			0.36
	Reads			**0.97**	0.06
	Library				0.74
	Discuss books			**0.99**	0.06
Child Cultural Capital	Theatre	**0.44**	**0.45**		0.54
	Museum		**0.83**		0.23
	Historic place		**0.70**		0.30
	Reads		**0.74**		0.49
	Library		**0.90**		0.30
	Discuss books		**0.74**		0.40

Factor scores are generated and used as covariates in the linear regression models of GCSE attainment. The factor scores are standardised with mean of 0 and standard deviation of 1. The cultural capital factor scores are entered sequentially into the model. Parent highbrow participation is not significant after adding in child highbrow participation, and is removed from the model. Table 5.11 presents the output of the final regression model (see Table A1.8 in Appendix 1 for the sequential model building process). Model 1 includes explanatory variables of parental education, parental NS-SEC, housing tenure, gender, ethnicity, age, and academic year. Model 2 presents the model including cultural capital factor scores. Young people with higher levels of cultural capital are significantly associated with higher GCSE point scores. Young people whose parents have more positive reading behaviours are significantly associated with higher GCSE point scores. There are no significant interaction effects between child cultural capital and parent reading behaviours. There are minor changes in the socio-economic coefficients after the inclusion of parent and child cultural capital factor scores. Parental education and parental social class continue to exert significant influences on GCSE attainment over and above the role of cultural capital when measured using factor scores.

Table 5. 11: Linear regression model of total GCSE point score including cultural capital factor scores

	Model 1			Model 2		
Linear Regression: GCSE Point Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	- 7.86	(2.94)	**	- 6.82	(2.80)	*
School-level education	-12.88	(2.42)	***	-10.87	(2.32)	***
Below school-level education	-16.21	(3.51)	***	-12.17	(3.59)	**
Parental NS-SEC						
1.1 Large employers and higher managerial	- 9.01	(4.70)		- 8.50	(4.78)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional	- 9.55	(3.24)	**	- 9.41	(3.19)	**
3 Intermediate occupations	- 9.51	(3.56)	**	- 8.80	(3.73)	*
4 Small employers and own account workers	-10.20	(4.12)	*	- 9.28	(4.16)	*
5 Lower supervisory and technical occupations	-14.01	(4.32)	**	-11.38	(4.13)	**
6 Semi-routine occupations	-12.94	(4.53)	**	-12.74	(4.41)	**
7 Routine occupations	-14.74	(4.40)	**	-14.04	(4.35)	**
8 Not in employment	-16.87	(4.08)	***	-15.27	(4.00)	***
Housing Tenure						
Owned/privately rented	Ref.	(.)		Ref.	(.)	
Social housing	-11.70	(2.54)	***	-10.07	(2.36)	***
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	7.05	(1.82)	***	5.03	(1.84)	**
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	- 0.93	(4.27)		0.63	(4.51)	
Asian/Asian British	0.80	(4.71)		1.29	(4.18)	
Black/Black British	2.77	(4.45)		2.71	(4.16)	
Other	9.80	(7.25)		11.88	(7.17)	
Child Cultural Capital Factor Score				3.92	(1.05)	***
Parent Reading Factor Score				2.70	(1.16)	*
Constant [output omitted]						
Observations	736			736		
Adjusted R^2	0.267			0.315		
BIC (d.f.)	6550.585 (22)			6512.802 (24)		
BIC (based on deviance)	1692.080			1654.297		
BIC (based on likelihood ratio chi square)	-111.719			-149.501		

Source: Wave 2 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Models also include age and academic year.
** p<.05, ** p<.01, *** p<.001.*

5.3.2 Summed Scales

Summed scales are often used in empirical work using cultural capital indicators (for example, Roscigno and Ainsworth-Darnell, 1999, Dumais, 2002, Eitle and Eitle, 2002, Kraaykamp and Eijck, 2010). The following analyses combine the summed scales of parent highbrow participation, parent reading behaviours, child highbrow participation, and child reading behaviours into a model of GCSE point score.

The summed scale measures were added sequentially into the model and removed if not significant (see Table A1.9 in Appendix 1). Parent highbrow participation is not significant after including child highbrow participation. The addition of parent reading behaviours rendered parent highbrow participation non-significant. Child's highbrow participation was also rendered non-significant after including child's reading behaviours in the model. The mediation of highbrow cultural participation after the inclusion of reading behaviours has been demonstrated in previous empirical work (see Crook, 1997, De Graaf et al., 2000, Sullivan, 2001). Parent and child reading behaviours remain significant in the same model and there is no significant interaction effect.

Table 5.12 presents the comparison model (Model 1) and the final model (Model 2) using summed scales of cultural capital. Model 2 demonstrates that positive reading behaviours within the household are significantly associated with higher overall GCSE scores. The effect of parent reading behaviours is stronger than child reading behaviours. The addition of the cultural capital measures improves the explanatory power (Adjusted R^2) and parsimony (BIC) of the model. There are minimal differences in the coefficients for parental education and parental social class after the inclusion of cultural capital measures. Therefore, there is a large parental socio-economic background effect in GCSE attainment which persists even after including measures of cultural capital, i.e. parent and child reading behaviours.

Table 5. 12: Linear regression model of total GCSE point score including cultural capital summed scales

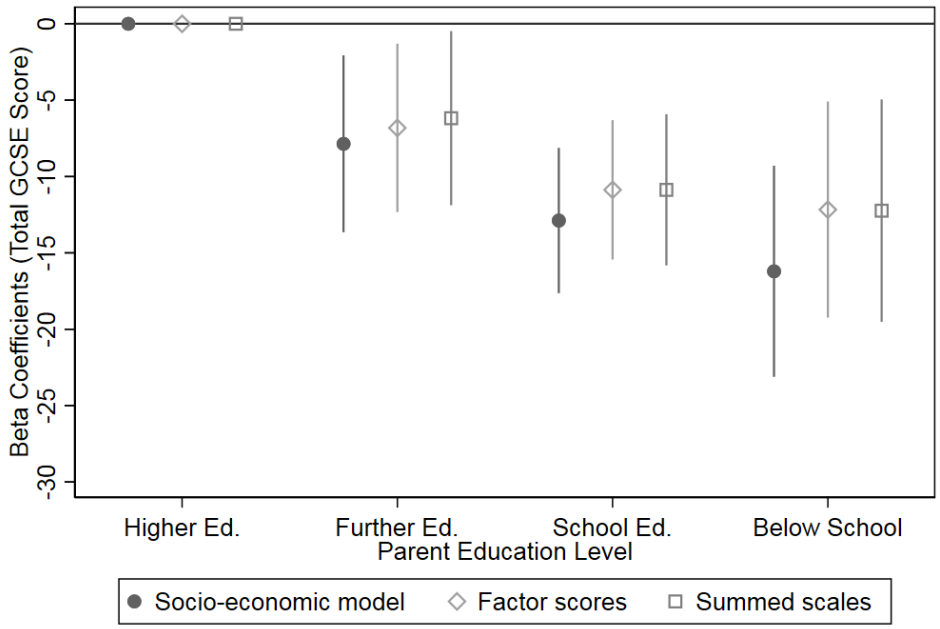
	Model 1			Model 2		
Linear Regression: GCSE Point Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	- 7.86	(2.94)	**	- 6.18	(2.90)	*
School-level education	-12.88	(2.42)	***	-10.88	(2.52)	***
Below school-level education	-16.21	(3.51)	***	-12.24	(3.70)	**
Parental NS-SEC						
1.1 Large employers and higher managerial	- 9.01	(4.70)		- 8.09	(4.66)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional	- 9.55	(3.24)	**	- 9.61	(3.19)	**
3 Intermediate occupations	- 9.51	(3.56)	**	- 9.42	(3.66)	*
4 Small employers and own account workers	-10.20	(4.12)	*	-10.05	(4.26)	*
5 Lower supervisory and technical occupations	-14.01	(4.32)	**	-12.02	(4.14)	**
6 Semi-routine occupations	-12.94	(4.53)	**	-12.76	(4.46)	**
7 Routine occupations	-14.74	(4.40)	**	-15.39	(4.37)	***
8 Not in employment	-16.87	(4.08)	***	-15.75	(4.03)	***
Housing Tenure						
Owned/privately rented	Ref.	(.)		Ref.	(.)	
Social housing	-11.70	(2.54)	***	-10.44	(2.36)	***
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	7.05	(1.82)	***	4.75	(1.85)	*
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	- 0.93	(4.27)		- 0.66	(4.26)	
Asian/Asian British	0.80	(4.71)		0.72	(4.21)	
Black/Black British	2.77	(4.45)		1.72	(4.05)	
Other	9.80	(7.25)		13.29	(7.94)	
Parent Reading Scale						
0				Ref.	(.)	
1				7.31	(3.53)	*
2				10.71	(3.48)	**
3				12.67	(5.17)	*
Child Reading Scale						
0				Ref.	(.)	
1				3.92	(2.22)	
2				7.68	(2.56)	**
3				9.00	(2.93)	**
Constant [output omitted]						
Observations	736			736		
Adjusted R^2	0.267			0.308		
BIC (d.f.)	6550.585 (22)			6542.462 (28)		
BIC (based on deviance)	1692.080			1683.957		
BIC (based on likelihood ratio chi square)	-111.719			-119.842		

Source: Wave 2 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Models also include age and academic year.
** p<.05, ** p<.01, *** p<.001.*

Figures 5.3 and 5.4 present the linear regression coefficients of parental education (Figure 5.3) and parental NS-SEC (Figure 5.4) on overall GCSE scores, comparing the models with and without cultural capital factor scores and summed scales. The

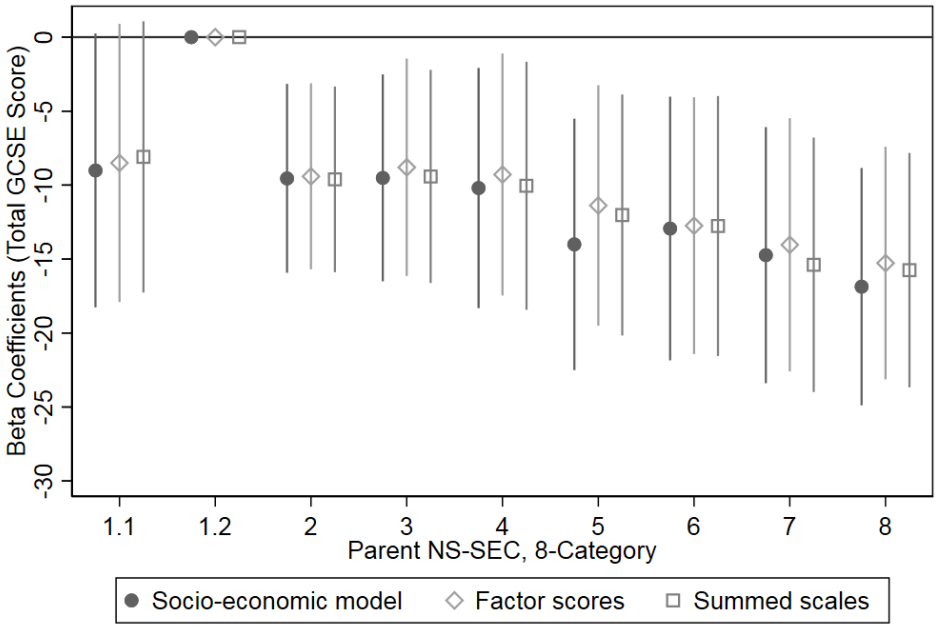
difference in regression coefficients and 95% confidence intervals between the models is minimal. The overall effect is modest and there is not convincing evidence that cultural capital mediates the role of parental socio-economic background in GCSE attainment. Cultural capital does not explain the socio-economic background effect. These findings do not support a Bourdieusian interpretation of the role of cultural capital in educational outcomes as outlined in section 2.

Figure 5. 3: Parental education level linear regression coefficients with 95% confidence intervals for models of GCSE point score with and without cultural capital measures



Source: Wave 2 Youth Questionnaires, UKHLS-NPD data, n=736. Adjusted for complex survey design. Parent NS-SEC, housing tenure, gender, ethnicity, age, and academic year also included in model.

Figure 5. 4: Parental NS-SEC linear regression coefficients with 95% confidence intervals for models of GCSE point score with and without cultural capital measures



Source: Wave 2 Youth Questionnaires, UKHLS-NPD data, n=736. Adjusted for complex survey design. Parent education, housing tenure, gender, ethnicity, age, and academic year also included in model.

5.4 Parental Education as Cultural Capital

Section 2 illustrated that parental education level has often been used as a proxy for cultural capital (Jonsson 1987, Jaeger and Holm 2007). Educational qualifications were originally cited by Bourdieu as an example of institutionalised cultural capital (see Bourdieu, 1984, Bourdieu, 1986). Sullivan (2002: 154) argued that Bourdieu 'is not entitled to assume' that higher education levels are the same as higher cultural capital. In the analyses throughout this thesis, parental education level is used as a measure of socio-economic background alongside parental social class. This follows the observation of Bukodi and Goldthorpe (2013) that parental education and parental social class exert separate influences. In this section, parental education level is theorised as a measure of cultural capital. Table 5.13 presents the results of a model of GCSE attainment with and without parental education level as a covariate.

Table 5. 13: Linear regression model of total GCSE point score with parental education level

	Model 1			Model 2		
Linear Regression: GCSE Point Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level						
Higher education				Ref.	(.)	
Further education				- 7.86	(2.94)	**
School-level education				-12.88	(2.42)	***
Below school-level education				-16.21	(3.51)	***
Parental NS-SEC						
1.1 Large employers and higher managerial	- 9.74	(4.71)	*	- 9.01	(4.70)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional	-13.10	(3.24)	***	- 9.55	(3.24)	**
3 Intermediate occupations	-16.53	(3.42)	***	- 9.51	(3.56)	**
4 Small employers and own account workers	-16.72	(4.10)	***	-10.20	(4.12)	*
5 Lower supervisory and technical occupations	-21.61	(4.37)	***	-14.01	(4.32)	**
6 Semi-routine occupations	-21.71	(4.23)	***	-12.94	(4.53)	**
7 Routine occupations	-24.41	(4.19)	***	-14.74	(4.40)	**
8 Not in employment	-25.22	(3.80)	***	-16.87	(4.08)	***
Housing Tenure						
Owned/privately rented	Ref.	(.)		Ref.	(.)	
Social housing	-13.57	(2.52)	***	-11.70	(2.54)	***
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	7.40	(1.90)	***	7.05	(1.82)	***
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	1.82	(3.83)		- 0.93	(4.27)	
Asian/Asian British	2.78	(4.45)		0.80	(4.71)	
Black/Black British	3.85	(4.49)		2.77	(4.45)	
Other	12.83	(6.81)		9.80	(7.25)	
Constant [output omitted]						
Observations	736			736		
Adjusted R^2	0.221			0.267		
BIC (d.f.)	6579.377 (19)			6550.585 (22)		
BIC (based on deviance)	1720.872			1692.080		
BIC (based on likelihood ratio chi square)	-82.926			-111.719		

Source: Wave 2 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Models also include age and academic year.
** p<.05, ** p<.01, *** p<.001.*

The magnitude of the effect of parental NS-SEC in Model 1 is substantially larger than in Model 2. This suggests that parental education partially mediates the effects of parental NS-SEC in GCSE attainment. The results serve as an interesting comparison to the candidate measures of cultural capital used throughout this chapter. Parental education level, if theorised as an indicator of cultural capital, does much more to mediate the effect of parental social class than any other measure. Parental NS-SEC continues, however, to be significant even after including the measure of parental education. This supports Bukodi and Goldthorpe's (2013) argument that parental

education and parental social class exert separate influences on educational attainment.

6 Discussion and Conclusions

There is clear evidence that cultural capital is significantly associated with GCSE attainment. Young people who actively engage in typically highbrow cultural activities may perform better in their GCSE examinations due to the direct skills, knowledge, and behaviours learned from undertaking highbrow cultural activities, such as going to the theatre, museums or art galleries, or visiting historical places. In linear regression models of GCSE score, parent highbrow cultural participation measures are not significant after the inclusion of child highbrow cultural participation measures.

The role of highbrow cultural participation is, however, substantively and significantly less important than the role of reading behaviours in GCSE attainment. Reading behaviours mediate the role played by highbrow cultural participation in models of GCSE attainment. The results provide contemporary empirical support to a distinction prevalent in older cultural capital literature between highbrow cultural participation and reading behaviours (Crook, 1997, De Graaf et al., 2000, Sullivan, 2001). The main effects of both parent and child reading behaviours are significantly associated with GCSE attainment, without evidence of an interaction effect. The increased vocabulary and linguistic ability gained from reading for pleasure can have direct applications to scholastic work (Sullivan, 2001, Sullivan and Brown, 2015). The added importance of parent reading behaviours is an interesting nuance to the role of reading in educational attainment. Parents' reading behaviours have greater effect sizes (i.e. larger coefficients in the linear regression models) and are stratified by socio-economic background to a greater degree than children's own reading behaviours. Children who see their parents reading may be more likely to form the habit themselves (De Graaf et al., 2000). Encouraging parents as well as children to read may provide a suitable policy intervention for increasing children's educational attainment due to its practicability, scalability, and relative low cost.

There is, however, convincing evidence that the cultural capital effect in GCSE attainment is largely independent of the effects of parental socio-economic background. Including cultural capital measures in models of GCSE attainment does not substantially alter the conclusions about the roles of parental education and parental social class. There is no evidence of cultural capital having a mediation effect, and therefore helping to explain the socio-economic background effect. This finding is robust across a range of potential operationalisations, including testing cultural capital measures individually, and building models of cultural capital based on factor scores and combinations of summed scales. Parental education level has relatively stronger (partial) mediation effects on the role of parental social class in GCSE attainment than any other potential cultural capital measure, although parental social class continues to be strongly and significantly associated with GCSE attainment. The presumption that higher education levels automatically relates to higher cultural capital is not convincing, particularly as the omnivore thesis suggests that higher status individuals do not only consume highbrow culture (Peterson and Kern, 1996). Overall, the inclusion of cultural capital does not offer a credible explanation for the persisting socio-economic effect. The findings in this chapter reflect similar conclusions from older empirical research (for example, Sullivan, 2001).

The empirical findings therefore do not support Bourdieu's original theorisation. Bourdieu's theory of cultural capital, and the broader cultural and social reproduction, has been criticised as being overly deterministic (see Jenkins, 1992, Sullivan, 2002). The concept has been widely criticised for being vague and ill-defined (Lamont and Lareau, 1988, Sullivan, 2002). Goldthorpe (2007) suggested that cultural 'capital' is associated with much 'theoretical baggage', and that cultural 'resources' should be preferred terminology. Despite this, the schools' inspectorate (Ofsted) introduced the concept of cultural capital into the schools assessment framework for the first time from September 2019. Ofsted define cultural capital as:

> *'The essential knowledge that pupils need to be educated citizens, introducing them to the best that has been thought and said, and helping to engender an appreciation of human creativity and achievement' (Office for Standards in Education, 2019: 10).*

This definition is not necessarily compatible with the sociological tradition, i.e. cultural capital as an accumulation of skills, knowledge, attitudes, and behaviours of the 'dominant' class which signal familiarity and ease with 'high status' practices (Lamont and Lareau, 1988). It is not currently clear how Ofsted expect schools to improve young people's cultural capital, or how they will evaluate schools' effectiveness based on their definition. The work in this chapter suggests that interventions focused on improving reading behaviours and positive attitudes towards reading for parents and children would be more fruitful than a, potentially outdated, focus on highbrow cultural participation. The interventions may further benefit from avoiding the confusing, vague language of cultural capital and using the more straight-forward, useful terminology of reading.

Chapter 6
Educational Aspirations, Socio-Economic Background, and GCSE Attainment

1 Introduction

'Raising aspirations' has been central to political rhetoric for recent UK Governments. Raising aspirations has been seen as a potential strategy to close the attainment gap between the most disadvantaged and less disadvantaged young people, and to aid social mobility in wider society (Baker et al., 2014, Berrington et al., 2016). This has involved the language of 'high' and 'low' aspirations. 'High' aspirations tend to mean continuing in education, going to university, or entering professional occupations (see Cabinet Office, 2008, Cabinet Office, 2009, Cabinet Office, 2011). Contemporaneous empirical studies in the UK have demonstrated that young people tend to have generally high aspirations, and that the narrative of a 'poverty of aspirations' is unfounded (St Clair and Benjamin, 2011, Croll and Attwood, 2013, Treanor, 2017). The overall research question explored in this chapter is *What role do educational aspirations play in GCSE attainment?*

Recent empirical work has examined explanatory factors associated with young people's educational aspirations using cross-sectional approaches (for example, Croll, 2009, Goodman et al., 2011, Chowdry et al., 2011, Gregg and Washbrook, 2011, Croll and Attwood, 2013, Berrington et al., 2016, Hartas, 2016). Cross-sectional approaches are constrained by a lack of a temporal element. The first part of the analyses in this chapter therefore extend previous work by using a longitudinal approach. A core assumption in this work is that aspirations are not *de facto* static. The panel design of the UKHLS allows for suitable examination of change or stability over time within and between individuals. The second part of the analyses then look at the role of the educational aspirations of parents and young people in GCSE attainment.

Section 2 outlines previous literature examining the association between socio-economic background, aspirations, and attainment. Section 3 outlines the data and methods used in the following analyses. Section 4 presents the descriptive statistics and modelling results of educational aspirations. Panel models are used to estimate university aspirations. Section 5 presents the descriptive statistics and modelling results of GCSE attainment to examine the roles of the aspirations of young people and their parents.

2 Youth Aspirations, Attainment, and Socio-Economic Differences

Youth educational aspirations have often been explored in the form of a desire to continue to study after compulsory schooling (Croll, 2009, Gorard et al., 2012) or to apply to university (Croll and Attwood, 2013, Khattab, 2015, Anders, 2017, McCulloch, 2017). Youth occupational aspirations have often been explored with focus on gender, ethnicity, and social class dimensions (Dumais, 2002, Archer et al., 2014, Platt and Parsons, 2017, Platt and Parson, 2018). The following sections outline the policy discourse on 'raising aspirations' and presents empirical literature examining the relationship between educational aspirations, educational attainment, and socio-economic background in the UK.

2.1 Policy Discourse

> *'Young people's aspirations - the goals they set for the future, their inspiration and their motivation to work towards these goals - have a significant influence both on their educational attainment and their future life outcomes' (Cabinet Office, 2008).*

> *'The education system should challenge low aspirations and expectations, dispelling the myth that those from poorer backgrounds cannot aim for top universities and professional careers' (Cabinet Office, 2011).*

Young people's aspirations have been a political focal point for explaining the educational attainment gap (Cabinet Office, 2008), and as an obstacle to social mobility (Cabinet Office, 2009). Aspirations are understood to differ by gender, ethnicity, and social class, and are most influenced by parents (Cabinet Office, 2008).

Girls, young people from ethnic minority backgrounds, and those from higher socio-economic backgrounds are more likely to have 'higher' aspirations than their peers (Gutman and Akerman, 2008). 'Higher' aspirations tend to refer to going to university or entering the professions (Cabinet Office, 2009, Cabinet Office, 2011). The underachievement, and 'lower' aspirations, of white, working class boys has received particular policy attention (House of Commons Education Committee, 2014). More recent policy papers have noted that less advantaged young people do not have 'low' aspirations altogether, but their aspirations are less likely to be educationally orientated (House of Commons Education Committee, 2014, Department for Education, 2014).

Baker et al. (2014) argued that the policymaker framing of 'high' and 'low' aspirations is problematic. Treanor (2017) challenged the language of 'high' aspirations, arguing that the alternative, i.e. 'low' aspirations, infers a deficit view for those from less advantaged backgrounds. St Clair and Benjamin (2011) remarked that perceived 'low' aspirations have been framed as something of a 'personal shortcoming' for individuals and their parents. In a speech to the Labour party conference in 2007, Prime Minister Gordon Brown used the phrase 'poverty of aspiration'. He asked 'how much talent that could flourish is lost through a poverty of aspiration: wasted not because young talents fail to reach the stars but because they grow up with no stars to reach for?'.[18] The 'poverty of aspiration' narrative suggests that 'low' aspirations are derived from parents under-valuing education and thereby contributing to the underachievement of less advantaged young people (Baker et al., 2014). Instead, academics have argued that aspirations are not deficient, but that they are cultivated within specific social contexts (St Clair and Benjamin, 2011, Treanor, 2017).

2.2 Aspirations and Socio-Economic Background

The status attainment literature in the USA in the late 1960s suggested that aspirations were central to understanding socio-economic inequalities in young people's education (Sewell and Shah, 1968a, Sewell et al., 1969). Sewell and Shah

[18] See a full transcript of the speech at http://news.bbc.co.uk/1/hi/uk_politics/7010664.stm [accessed 26.06.19].

(1968b: 559) noted ‘it is a sociological truism [...] that children of higher social class origins are more likely to aspire to high educational and occupational goals than are children of lower social class origins’.

The role of socio-economic background is often investigated in more contemporary studies of youth aspirations (see, for example, St Clair and Benjamin, 2011, Croll and Attwood, 2013, Baker et al., 2014, Khattab, 2015, Berrington et al., 2016, Anders, 2017). Khattab (2015) disentangled expectations and aspirations, arguing that although linked, the two represent differences between what one hopes to achieve, and what one can realistically achieve. Conceptually, aspirations may operate outside of structural constraints, or ‘socio-economic realities’ (Gorard et al., 2012).

It is also conceptually plausible that educational aspirations may differ according to social class or education level (Anders, 2017). For example, parents who have been to university may be more likely to see the transition to higher education as a normalised route for their children (Anders, 2017). The role of parental social class in aspiration formation may be understood in ‘relative risk aversion’ terms (see Breen and Goldthorpe, 1997). Following this theoretical position, undertaking further educational opportunities may reduce the risk of downward social mobility, and ensure that the young person reaches at least the class position of their parents (Breen and Goldthorpe, 1997, Breen and Yaish, 2006, Holm and Jaeger, 2008).

Previous empirical work has found that young people have generally high aspirations (Archer et al., 2014, Hartas, 2016), even where significant associations between socio-economic background and aspirations exist (St Clair and Benjamin, 2011, Croll and Attwood, 2013, Baker et al., 2014, Treanor, 2017). Baker et al. (2014) used the Effective Provision of Pre-School, Primary and Secondary Education (EPPSE) to examine the determinants of ‘high’ aspirations. High aspirations were measured by whether the respondent considered it ‘very important’ to go to university or not. Attainment was strongly linked to young people having ‘high’ aspirations at the age of 14. Socio-economic background factors, such as mothers’ education and income bracket, continued to be significant in predicting ‘high’ aspirations for university, even after controlling for prior attainment. The results demonstrated generally ‘high’

aspirations for all young people, and therefore did not support a 'poverty of aspiration' narrative for those from less advantaged backgrounds (Baker et al., 2014). Similarly, Treanor (2017) argued that the socio-economic differences in aspirations does not constitute a 'poverty of aspiration', but high aspirations are context-specific.

Longitudinal approaches to modelling aspirations over time have demonstrated socio-economic differences in young people's trajectories. For example, Anders (2017) used the Longitudinal Study of Young People in England (LSYPE) to examine young people's expectations of going to university between the ages of 14 and 17. Duration models were estimated to examine how socio-economic background and other factors may affect the transitions between states. The findings demonstrated that young people had similar aspirations of going to university at the age of 14, but those from less advantaged backgrounds were more likely to revise their aspirations downwards (from 'likely' to 'unlikely'), and those from more advantaged backgrounds were more likely to revise their aspirations upwards (Anders, 2017). McCulloch (2017) used latent class analysis to track young people's educational trajectories using the LSYPE data, and found evidence of cumulative advantage and disadvantage with regards to socio-economic background, aspiration formation, and attainment. Young people with higher educational aspirations had higher likelihoods of attaining more favourable educational outcomes, but high aspirations tended to be held by young people from more advantaged social backgrounds, for girls more than boys, and for young people from ethnic minority backgrounds compared with white young people (McCulloch, 2017).

The role of socio-economic background may more accurately explain the gap between aspirations and eventual educational outcomes. For example, Croll and Attwood (2013) suggested that the 'aspiration gap' was much smaller than the actual 'participation gap' in university attendance. Croll (2009) analysed intentions to stay on beyond the compulsory school leaving age of 16 using the BHPS and found that lower attaining individuals from more advantaged backgrounds were more likely to stay on in education than lower attaining individuals from less advantaged backgrounds. The larger participation gap may be a result of socio-economic differences in access to

resources or capitals (Hartas, 2016), or the material circumstances of the family (Hoskins and Barker, 2016).

2.3 Aspirations and Attainment

A number of contemporary studies have reported an association between aspirations and attainment (Croll, 2009, Goodman et al., 2011, Chowdry et al., 2011, Gregg and Washbrook, 2011, Croll and Attwood, 2013). Descriptive analyses have suggested that there are associations between aspirations to stay on in education and educational attainment. For example, Croll (2009) analysed data from the British Household Panel Survey youth panel using a series of tabulations and found that intentions to stay in education after the compulsory leaving age were more stable over time than intentions to leave education at the age of 16. The descriptive statistics presented suggested that aspirations were associated with later GCSE attainment. Croll and Attwood (2013) used the LSYPE data and presented tabulations to suggest that higher expectations of attending university are largely due to prior attainment at earlier stages of schooling.

More comprehensive statistical analyses have further illustrated the importance of aspirations in attainment. Gregg and Washbrook (2011) analysed the Avon Longitudinal Study of Parents and Children focusing on educational attainment at age 11 using a variety of potential explanatory measures, including mothers' aspirations for children to go to university (asked when the children were aged 9). There were socio-economic differences in university aspirations between the richest and poorest mothers when their child was in primary school. Parents from more advantaged socio-economic backgrounds were much more likely to state that they wanted their primary school child to go to university than parents from less advantaged socio-economic backgrounds. The models were estimated with and without controlling for outcomes at age 7. Mothers' university aspirations continued to be significant after controlling for prior attainment. Chowdry et al. (2011) analysed LSYPE data for educational attainment at age 16. The authors found that attitudes and behaviours had a socio-economic gradient, and that children and parents with aspirations to apply to university were likely to attain better grades in their GCSE examinations. The effect

of children's and parents' aspirations decreased, but persisted, after controlling for prior attainment at age 11. The findings suggested that aspirations were high across all socio-economic groups, and therefore there was no evidence of an aspirations 'deficit' (Chowdry et al., 2011).

Gorard et al. (2012: 41) stressed the non-recursive (i.e. not in one direction) nature of the association between aspirations and attainment, as 'aspirations can be both a predictor ... and an outcome' of educational achievement. Anders (2017) commented that when measuring university aspirations, attainment at later levels of schooling, for example GCSE examinations at age 16, are potentially endogenous. This is largely because high attainment at this level is a requirement for A' Level study, and subsequent university application. Those who have performed better in prior examinations and schooling contexts are more likely to consider further study as a credible option for them, which will in turn provide motivation to do well in school. By contrast, those receiving poorer marks than anticipated earlier in school may be less motivated to prepare for, and subsequently do well in, school-level examinations; especially if they do not expect to continue studying (Jackson, 2013).

The process of aspiration formation has been similarly described as 'dynamic', involving internal feedback between attainment and aspirations (St Clair and Benjamin, 2011). A continual feedback loop between attainment and aspirations would adjust a young person's concept of what is feasible for them to achieve. For example, Sullivan et al. (2013) used the Millennium Cohort Study to model cognitive outcomes and noted that mothers' aspirations for their children at age 7 were 'universally high' but that such aspirations were likely to vary more as their children progress through education. Aspirations may be replaced by more realistic expectations as a result of, for example, known academic attainment. Research has demonstrated that attainment at earlier stages is strongly associated with socio-economic differences in parental social class and education level (see Chapter 4, and Feinstein, 2003, Goodman et al., 2011, Chowdry et al., 2011, Strand, 2014b). Academic self-concept can be informed by prior attainment, and can also play an important role in educational and occupational aspirations (Winterton and Irwin, 2012). A young person's perception of where they fit in the academic hierarchy is

important in developing their future educational and occupational aspirations, and are likely to be solidified during secondary school years (Furlong and Biggart, 1999).

3 Data and Methods

3.1 Sample

In this chapter, educational aspirations are analysed using two approaches. First, the socio-economic determinants of university aspirations are modelled using panel regression models. A long format dataset of the UKHLS youth questionnaires and linked National Pupil Database records is constructed. A total of 1252 young people contribute 2178 observations over five waves in the analytical sample.

Second, the role of parents' and young people's aspirations and GCSE attainment is examined using a cross-sectional dataset of Wave 1 youth questionnaires and linked NPD records. Data collected on parents' aspirations are asked in alternate wave (1, 3, 5), but parents are asked whether they would like their children to go to university in Wave 1 only. The complete records analytical sample has a total n=1062. Missing data are examined using multiple imputation by chained equations (see Table A2.4 in Appendix 2). The substantive conclusions from the complete records analyses presented below remain the same after considering the role of missing data.

3.2 Measures

3.2.1 Longitudinal Analyses of Educational Aspirations

The main outcome measure in the longitudinal analyses is university aspirations, which are considered 'high' aspirations in political rhetoric (see Cabinet Office, 2008, Cabinet Office, 2009, Cabinet Office, 2011, Croll and Attwood, 2013, House of Commons Education Committee, 2014, Baker et al., 2014). The youth respondents

are asked whether they would like to go to university or college.[19] University aspirations are modelled with the following independent variables: parental education level, parental NS-SEC, housing tenure, gender, ethnicity, age, and Key Stage 2 attainment. The Key Stage 2 attainment measure is constructed in the same way as in Chapter 4, i.e. it is a standardised measure of average attainment in English and Maths tests at age 11.

3.2.2 Cross-Sectional Analyses of GCSE Attainment

The cross-sectional analyses of the Wave 1 UKHLS and linked NPD records explore a broader range of educational aspirations from both parents and children. University aspirations are included, as detailed above. A measure of post-16 destinations (i.e. what the young person wants to do after completing compulsory education) is collapsed into a binary variable of 'full-time education' compared with any other answer for the purposes of comparing results to the university aspirations reported throughout this chapter. This is also a pragmatic approach because the exact wording of the answer options changed between waves 2 and 3. The broad answer options are full-time job, full-time education, study and job e.g. apprenticeship, and something else or don't know. 'Study full-time' remained a constant option across all waves.

A further measure of the young person's educational attitudes is used. The young people are asked how important it is to do well in their GCSEs. The answer options are 'very important', 'important', 'not very important', and 'not at all important'. Not very important and not at all important are aggregated due to low sample sizes.

Two measures of the parent's educational aspirations are used from the Wave 1 adult survey. First, parents are asked how important they think A' Levels are for their child. A' Levels are post-compulsory school examinations. It could be argued that this measure therefore conflates how important it is to do well in school with how important

[19] This is a filter question to a question about post-16 destinations in the youth self-completion questionnaire. Those who do not answer 'full-time job' are asked to answer the university question. Therefore, those who respond with 'full-time job' in the previous question are coded as 'no ' for these analyses. A few respondents answer both 'full-time job' and the university question, in which case the university response is used.

it is for their child to continue in education beyond the compulsory age. The answer options are aggregated into three categories due to low sample sizes in the negative responses: ‘very important’, ‘important’, and ‘not very or not at all important’.

The second measure is whether parents would like their child to go to university. In the Wave 1 adult survey, parents are asked the question separately for each of their children. The yes or no answer is coded and deposited in the dataset along with corresponding person numbers to enable researchers to ascertain which child the parent is referring to. For example, a mother may report that they think their eldest child will go to university but not their youngest child. The ‘yes’ answer for their eldest child is deposited with the corresponding person number for that child, and their ‘no’ answer for their youngest child is deposited with a different person number for that child. The dominance approach, i.e. the highest expectation of mother or father, is taken for both these measures. This may introduce an upward bias in estimates. For those living with two parents, the mother and father responses tend to be, but are not always, the same. A second, categorical variable is also tested with answer categories of ‘both said yes’, ‘both said no’, and ‘mixed response’, and the results are consistent with the binary measure.

3.3 Structure of Analysis

As outlined above, the analyses in this chapter are undertaken on two samples of the dataset. First, longitudinal panel models are estimated to examine university aspirations over the secondary school years. Second, cross-sectional regression models are estimated to examine GCSE attainment and the roles of socio-economic background and parent and child aspirations. The following research questions are explored.

1. How stable are the aspirations of young people during secondary school?
2. To what extent are there socio-economic differences between young people’s aspirations?
3. What is the role of parent and child educational aspirations in GCSE attainment?

4 Modelling Aspirations

4.1 Descriptive Statistics

A total of 1252 young people contribute 2178 observations in the panel dataset. The maximum number of waves is 5. The dataset is an unbalanced panel because not all young people contribute to all waves available. The youth panel is also a rotating panel, and therefore young people drop in and out of the panel between ages 11 and 15. Within the youth dataset there is both unit non-response (i.e. drop out) and item non-response. Gayle and Lambert (2018) commented that balanced panels are uncommon in social survey datasets. Table 6.1 presents the patterns of wave response in the longitudinal dataset of the youth questionnaires. The largest percentage of respondents (41%) contributed to Wave 1 only. Almost 22% of young people contributed to the first two waves, 11% of young people contributed to the first three waves, and 4% of young people contributed to the first four waves. At least half of young people contributed to more than one wave, and up to 5% contributed to 3 waves.

Table 6. 1: Patterns of wave response in the youth questionnaire with linked NPD records

Pattern of Wave Response					**Freq.**	**Percent**
1	**2**	**3**	**4**	**5**		
✓	-	-	-	-	512	40.89
✓	✓	-	-	-	269	21.49
✓	✓	✓	-	-	139	11.10
-	✓	-	-	-	70	5.59
✓	✓	✓	✓	-	49	3.91
✓	-	✓	-	-	45	3.59
-	✓	✓	-	-	41	3.27
-	-	✓	-	-	37	2.96
✓	✓	-	✓	-	16	1.28
			Other pattern		74	5.91
Total *n*					**1252**	**100.00**

Table 6.2 presents the overall variation, between variation, and within variation for university aspirations, parental education, parental social class, gender, ethnicity, and age. The overall variation is based on pooled observations across persons and waves

(i.e. treats the observations as a simple cross-sectional dataset). The between variation represents the variation between persons at any time point (i.e. the percentage of young people who have ever reported being in each category in any wave of the survey). Within variation represents change or stability within-person across all waves (i.e. the percentage of time spent in a category for an individual who has ever reported being in that category). Higher percentages represent greater within-person stability, i.e. that individuals remain in the same state for a longer length of time. Lower percentages represent greater within-person change.

There is greater variation between-persons than within-persons for all socio-economic and demographic factors. There is remarkable within-person stability for many of the socio-economic and demographic variables. There is a notable exception of age which must necessarily increase over time. In this sample, gender and ethnicity do not vary over time, as represented by the within-person variation of 100%, i.e. those who report being in a certain category do so 100% of the time they are observed in the panel. Parental education and parental social class are conceivably time-varying variables, but the rate of change is very low. The greatest within-person stability for parental social class is in NS-SEC 1.2, i.e. higher professional occupations. The greatest within-person changes are in NS-SECs 5, 6, and 7, i.e. in routine and manual occupations.

Aspirations to go to university are generally high, with 89.9% of young people ever stating that they have university aspirations, and only 17.6% ever stating they would not like to go to university. Of those who ever stated they wanted to go to university, they did so in 96.4% of the waves contributed. The within-person stability was much lower for those who ever stated they did not want to go to university. Overall, this suggests that young people's aspirations tend to be largely stable over time.

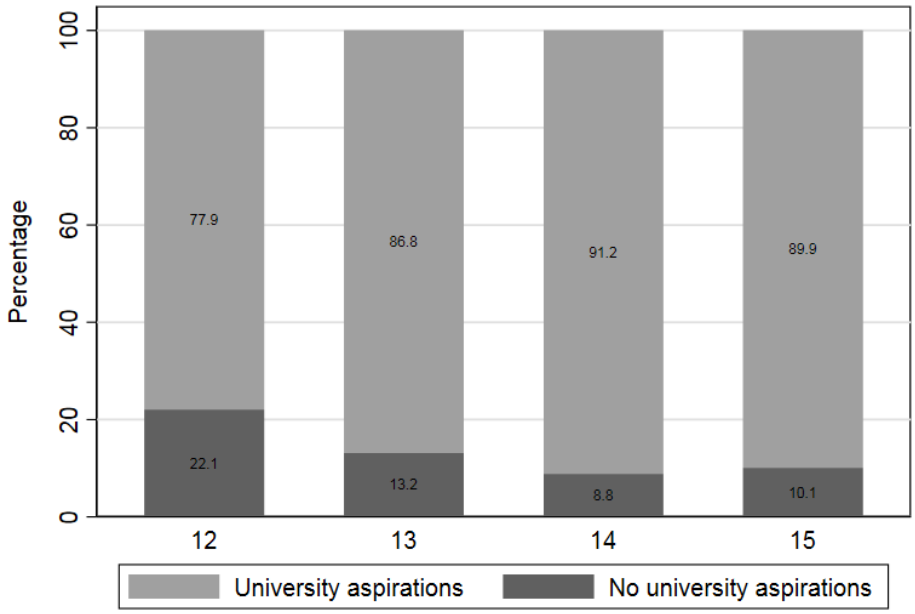
Table 6. 2: Overall, between, and within-person variation of key variables

	Overall Variation		Between Variation		Within Variation
	Freq.	Percent	Freq.	Percent	Percent
Wants to go to University					
Yes	1925	88.38	1125	89.86	96.40
No	253	11.62	220	17.57	76.14
Total	**2178**	**100.00**	**1345**	**107.43**	**93.09**
Parental Education Level					
Higher education	624	28.65	334	26.68	97.95
Further education	333	15.29	198	15.81	96.72
School-level education	948	43.53	569	45.45	98.68
Below school-level education	273	12.53	176	14.06	97.63
Total	**2178**	**100.00**	**1277**	**102.00**	**98.04**
Parental NS-SEC					
1.1 Large employers/higher managerial	132	6.06	86	6.87	91.76
1.2 Higher professional occupations	190	8.72	108	8.63	93.06
2 Lower management/professional	475	21.81	299	23.88	88.23
3 Intermediate occupations	222	10.19	146	11.66	87.49
4 Small employers/own account workers	224	10.28	145	11.58	87.18
5 Lower supervisory/technical occupations	117	5.37	75	5.99	85.22
6 Semi-routine occupations	275	12.63	190	15.18	85.84
7 Routine occupations	148	6.80	96	7.67	84.72
8 Not in employment	395	18.14	267	21.33	92.24
Total	**2178**	**100.00**	**1412**	**112.78**	**88.67**
Housing Tenure					
Owned/privately rented	1681	77.18	960	76.68	99.35
Social housing	497	22.82	305	24.36	97.79
Total	**2178**	**100.00**	**1265**	**101.04**	**98.97**
Gender					
Male	1052	48.30	611	48.80	100.00
Female	1126	51.70	641	51.20	100.00
Total	**2178**	**100.00**	**1252**	**100.00**	**100.00**
Ethnicity					
White	1665	76.45	957	76.44	100.00
Mixed	113	5.19	59	4.71	100.00
Asian/Asian British	279	12.81	160	12.78	100.00
Black/Black British	100	4.59	63	5.03	100.00
Other ethnic group	21	0.96	13	1.04	100.00
Total	**2178**	**100.00**	**1252**	**100.00**	**100.00**
Age					
11	30	1.38	30	2.40	46.72
12	201	9.23	200	15.97	50.13
13	440	20.20	438	34.98	49.49
14	656	30.12	652	52.08	55.01
15/16	851	39.07	842	67.25	66.78
Total	**2178**	**100.00**	**2162**	**172.68**	**57.91**

The following analyses focus on the socio-economic determinants of university aspirations. Figure 6.1 presents university aspirations by age group, using the data as a pooled cross-sectional dataset. As the young people age, the percentage of respondents wanting to go to university increases, although the changes are modest.

The measures reflect similar educational aspirations to continue studying after the end of compulsory schooling (see Figure A1.1 in Appendix 1).

Figure 6. 1: University aspirations by age of respondent, pooled data



Source: UKHLS Youth Questionnaire with linked NPD. Pooled cross-sectional data, n=2178.
Note: Ages 11/12 and 15/16 combined due to statistical disclosure control.

4.2 Panel Models

Repeated contacts data are sub-optimally modelled using standard regression models, because they violate the assumption of independence of observations (Gayle and Lambert, 2018). Panel regression models, for example fixed or random effects models, account for repeated contacts within the dataset (Bell et al., 2018). There are advantages and limitations with both fixed effects and random effects models (Clarke et al., 2010, Clark and Linzer, 2015, Bell et al., 2018, Hill et al., 2019). Fixed effects panel models account for within-person change and tend to theoretically produce consistent estimates. However, fixed effects panel models estimate potentially inefficient standard errors and cannot estimate models with time-constant explanatory variables. Random effects panel models better account for both within-person and between-person change, and can estimate both time-constant and time-varying explanatory variables. Theoretically, random effects panel models tend to estimate efficient standard errors. These models, however, impose a strong assumption that there is no correlation between the observed effects and the unobserved, or

unobservable, effects (Clarke et al., 2010, Gayle and Lambert, 2018). Gayle and Lambert (2018) provide a comprehensive review of panel models and their applications using Stata. Longhi and Nandi (2015) provide a useful guide to the practical applications of panel models using longitudinal social science surveys including the BHPS and UKHLS.

There is an enduring tension for researchers when choosing the most appropriate panel model (Clarke et al., 2010, Clark and Linzer, 2015, Bell et al., 2018). Fixed effects models drop observations with unchanging outcomes. This is clearly problematic for the current sample which exhibits high levels of within-person stability. In linear panel models, the Mundlak correction can be applied to random effects models to retrieve the coefficients and standard errors of a fixed effects model (Mundlak, 1978). The Mundlak correction has the attractive property of estimating time-constant variables and helps to overcome the strong assumption of error terms being uncorrelated with unobserved variables. The Mundlak correction involves adding the means of time-varying variables into the random effects model. Allison (2009) provides an alternative, the hybrid model, which can be used in a logistic regression framework. The hybrid method involves adding the means and the deviations from the mean of time-varying variables into the random effects model (Allison, 2009).

In the following analyses, fixed and random effects models are estimated to compare results. Although tracking individual trajectories is conceptually attractive, the fixed effects model did not converge. This is likely due to the large proportion of observations with unchanging outcomes. Next, the hybrid method is used to compare the fixed effects and random effects estimates. Allison (2009: 3) noted that where time-varying variables do not vary very much, the estimates of the fixed effects model are very imprecise. The descriptive statistics in Table 6.2 clearly demonstrate a high within-person stability of the key socio-economic background measures of parental education and parental social class. The measures are theoretically time-varying, but the variation over time is low and the between-effects of parental education and parental social class are much larger than the within-effects. Therefore, the random effects models are preferred in the analyses below.

4.3 Modelling Results

A useful first step in modelling panel data is to estimate a pooled cross-sectional model to provide an initial insight into relationships within the data. The pooled cross-sectional logistic regression model of university aspirations (with robust standard errors) is presented in Table 6.3.

Table 6. 3: Pooled cross-sectional logistic regression model of university aspirations

Pooled Cross-Sectional Logistic Regression: University Aspirations	**Coef.**	**Robust S.E.**		**Quasi-Variance S.E.**	**LCI**	**UCI**
Parental Education Level						
Higher education	Ref.	(.)		0.28	-0.58	0.58
Further education	-1.51	(0.33)	***	0.20	-1.93	-1.09
School-level education	-1.60	(0.29)	***	0.10	-1.81	-1.40
Below school-level education	-1.74	(0.35)	***	0.21	-2.17	-1.31
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.58	(0.66)		0.42	-1.46	0.29
1.2 Higher professional occupations	Ref.	(.)		0.53	-1.12	1.12
2 Lower management / professional occupations	-0.90	(0.56)		0.21	-1.34	-0.46
3 Intermediate occupations	-0.51	(0.59)		0.25	-1.05	0.02
4 Small employers and own account workers	-1.13	(0.58)		0.23	-1.62	-0.64
5 Lower supervisory and technical occupations	-1.41	(0.63)	*	0.33	-2.11	-0.72
6 Semi-routine occupations	-1.33	(0.57)	*	0.19	-1.72	-0.93
7 Routine occupations	-1.93	(0.59)	**	0.24	-2.43	-1.43
8 Not in employment	-1.65	(0.57)	**	0.18	-2.02	-1.27
Gender						
Male	Ref.	(.)		-	-	-
Female	1.16	(0.17)	***	-	-	-
Ethnicity						
White	Ref.	(.)		0.10	-0.22	0.22
Mixed	1.05	(0.46)	*	0.45	0.11	2.00
Asian/Asian British	1.28	(0.29)	***	0.27	0.71	1.85
Black/Black British	0.80	(0.40)	*	0.38	-0.00	1.61
Other ethnic group	1.50	(1.19)		1.19	-1.00	4.00
Age	0.27	(0.07)	***	-	-	-
Constant [output omitted]						
Observations	2178					
McFadden's Adjusted Pseudo R^2	0.128					
Cox-Snell Pseudo R^2	0.103					
Nagelkerke Pseudo R^2	0.200					
BIC (d.f.)	1467.151 (18)					
BIC (based on deviance)	-15273.310					
BIC (based on likelihood ratio chi square)	-105.240					

Source: Waves 1-5 UKHLS youth questionnaires with linked NPD records, unweighted.
** p<.05, ** p<.01, *** p<.001.*

The results demonstrate that parental education and parental social class are significantly associated with university aspirations. Young people whose parents have higher education qualifications or who have higher professional occupations have significantly higher log odds of having university aspirations compared with young

people with less educationally and occupationally advantaged parents. Girls and young people from ethnic minority backgrounds also have significantly higher log odds of having university aspirations compared with boys and young people from white backgrounds respectively.

The next stage of the analysis examines university aspirations in a panel model framework. As a binary variable, university aspirations are estimated using a panel logistic regression model (see xt, StataCorp., 2017a). The xt suite is, however, not compatible with the svy suite and therefore the models are unweighted and do not adjust for complex survey design.

An unweighted random effects logistic regression model of university aspirations is presented in Table 6.4. Parental education level, parental social class, gender, ethnicity, and age are all significantly associated with young people having university aspirations. Young people with graduate parents have significantly higher log odds of aspiring to go to university than those whose parents have any other (lower) level of education. Those whose parents have occupations in NS-SEC 1.2 have significantly higher log odds of aspiring to go to university than those whose parents have occupations in NS-SECs 5, 6, 7, and who are not in employment. Girls have significantly higher log odds of having university aspirations than boys. Ethnicity is significant for young people from white backgrounds compared with young people from mixed, Asian, and Black backgrounds. These estimates should be analysed with caution because the sample includes an ethnic minority boost and the data are unadjusted for complex survey design. Age is also significantly associated with university aspirations. This may suggest that university becomes a more realistic ambition as young people progress through secondary school.

Rho is the proportion of the total variance attributed to the panel aspect of the data. The rho statistic for the model is 0.39, i.e. 39% of the variance is at the panel level. If rho were zero, there would be no difference between the estimates from the random effects panel model and a pooled cross-sectional model (i.e. a model which does not account for repeated contacts). The likelihood ratio chi square test of rho is significant, meaning that the panel model is preferred to a pooled cross-sectional model.

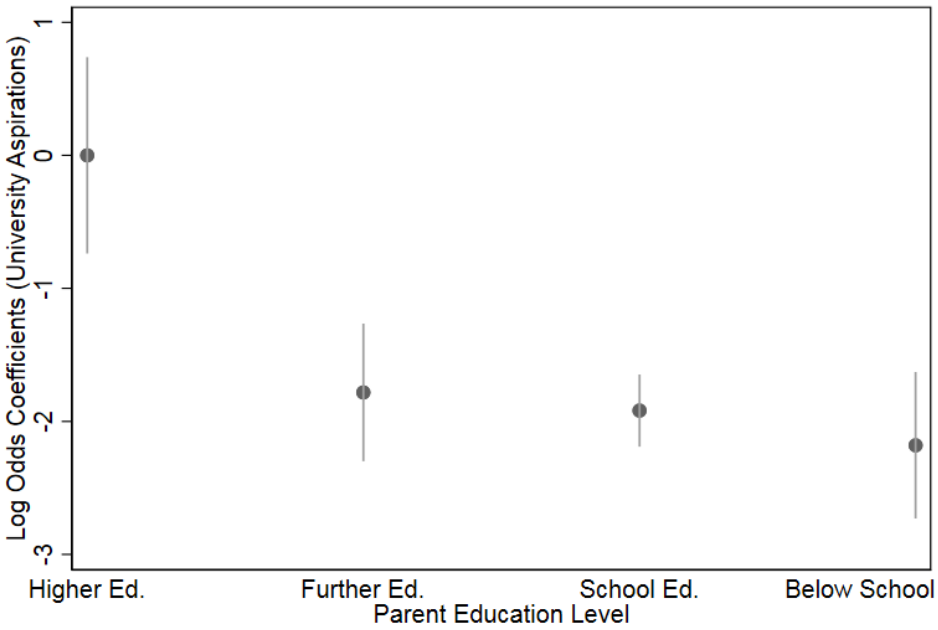
Table 6. 4: Random effects logistic regression model of university aspirations

Random Effects Logistic Regression: University Aspirations	**Coef.**	**S.E.**		**Quasi-Variance S.E.**	**LCI**	**UCI**
Parental Education Level						
Higher education	Ref.	(.)		0.35	-0.74	0.74
Further education	-1.78	(0.41)	***	0.25	-2.30	-1.27
School-level education	-1.92	(0.38)	***	0.13	-2.19	-1.65
Below school-level education	-2.18	(0.45)	***	0.26	-2.73	-1.63
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.52	(0.79)		0.53	-1.63	0.60
1.2 Higher professional occupations	Ref.	(.)		0.62	-1.30	1.30
2 Lower management / professional occupations	-1.10	(0.65)		0.24	-1.60	-0.59
3 Intermediate occupations	-0.55	(0.70)		0.34	-1.25	0.16
4 Small employers and own account workers	-1.30	(0.68)		0.29	-1.91	-0.68
5 Lower supervisory and technical occupations	-1.52	(0.73)	*	0.39	-2.33	-0.71
6 Semi-routine occupations	-1.45	(0.67)	*	0.24	-1.95	-0.95
7 Routine occupations	-2.13	(0.70)	**	0.31	-2.78	-1.47
8 Not in employment	-1.85	(0.66)	**	0.21	-2.29	-1.41
Gender						
Male	Ref.	(.)		-	-	-
Female	1.45	(0.22)	***	-	-	-
Ethnicity						
White	Ref.	(.)		0.15	-0.31	0.31
Mixed	1.31	(0.61)	*	0.59	0.07	2.55
Asian/Asian British	1.59	(0.39)	***	0.36	0.84	2.35
Black/Black British	1.05	(0.50)	*	0.48	0.03	2.07
Other ethnic group	1.73	(1.32)		1.30	-1.02	4.48
Age	0.34	(0.08)	***	-	-	-
Constant [output omitted]						
Standard deviation of panel-level variance	1.46	(0.25)				
Log of panel-level variance	0.76	(0.35)				
Observations (Total)	2178					
Observations (Individuals)	1252					
Likelihood ratio chi square (d.f.)	193.77 (17) ***					
Rho	0.39	(0.08)				
Likelihood ratio test of Rho (d.f.)	22.12 (1)	***				

Source: Waves 1-5 UKHLS youth questionnaires with linked NPD records, unweighted.
** p<.05, ** p<.01, *** p<.001.*

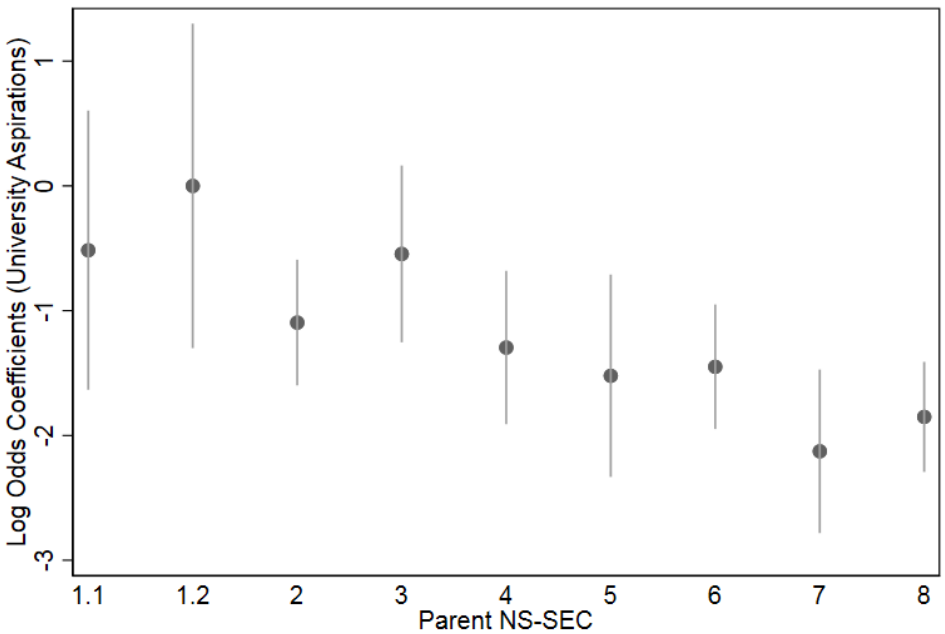
Figures 6.2 and 6.3 plot the log odds coefficients and quasi-variances (with 95% comparison intervals) for parental education and parental social class in the random effects logistic regression model of university aspirations. It is clear that there is a distinction between those whose parents have higher education qualifications compared with parents who have any (lower) education qualification. There are no significant differences in having university aspirations for those whose parents have further education, school-level, or below school-level education. The polarisation makes substantive sense because parents who have been through the higher education system themselves are probably more likely to consider it a natural progression for their children (Anders, 2017).

Figure 6. 2: Log odds coefficients and 95% quasi-variance comparison intervals for university aspirations by parental education level



Source: UKHLS-NPD unweighted data, n=2178.
Parent NS-SEC, gender, ethnicity, and age also included in model.

Figure 6. 3: Log odds coefficients and 95% quasi-variance comparison intervals for university aspirations by parental NS-SEC



Source: UKHLS-NPD unweighted data, n=2178.
Parent education, gender, ethnicity, and age also included in model.

There is a general negative gradient in the coefficients for parental NS-SEC. However, many adjacent categories of parental NS-SEC are not significantly different

to each other. The clearest distinction in university aspirations is between those whose parents have higher managerial and professional occupations (NS-SECs 1.1 and 1.2) compared with those whose parents have routine occupations (NS-SEC 7) and who are not in employment.

Table 6.5 presents the predicted probabilities of aspiring to go to university by parental socio-economic background. There are very high predicted probabilities of aspiring to go to university across all categories of parental education level and parental social class. Examining the 95% confidence intervals, the predicted probabilities of aspiring to go to university are over 75% across all socio-economic backgrounds. Whilst there are some statistical differences according to socio-economic background, there is no evidence of a 'poverty of aspiration' amongst the young people in this survey. These results provide new, contemporary evidence to support previous empirical research (for example, Treanor, 2017).

Table 6. 5: Predicted probabilities of university aspirations by parental socio-economic background

Longitudinal Predicted Probabilities	**Prob. (%)**	**S.E.**	**95% C.I.**
Parental Education			
Higher education	0.96	0.01	[0.95, 0.98]
Further education	0.87	0.02	[0.83, 0.91]
School-level education	0.86	0.01	[0.83, 0.88]
Below school-level education	0.83	0.03	[0.78, 0.88]
Parental NS-SEC			
1.1 Large employers and higher managerial	0.93	0.03	[0.88, 0.98]
1.2 Higher professional occupations	0.95	0.02	[0.91, 1.00]
2 Lower management and professional occupations	0.90	0.02	[0.87, 0.93]
3 Intermediate occupations	0.93	0.02	[0.90, 0.96]
4 Small employers and own account workers	0.88	0.02	[0.84, 0.93]
5 Lower supervisory and technical occupations	0.87	0.03	[0.80, 0.93]
6 Semi-routine occupations	0.87	0.02	[0.83, 0.91]
7 Routine occupations	0.81	0.03	[0.75, 0.87]
8 Not in employment	0.84	0.02	[0.80, 0.87]

The next stage of the analysis introduces a measure of prior attainment into the model of university aspirations. The measure is the average Key Stage 2 test scores in English and Maths taken at age 11 and is standardised (as constructed in Chapter 4).

Gorard et al. (2012) argued that caution should be applied to interpreting a causal link between attainment and aspirations. A key methodological benefit of using the UKHLS-NPD, repeated contacts, data is the temporal ordering of the measures in these data is very useful to understanding the direction of effect. The attainment measure is observed at age 11 and therefore has generally taken place prior to reporting university aspirations. The following results can be understood as the effects of attainment at the end of primary school on the aspirations formed throughout secondary school. This is a neat attraction of using the UKHLS youth panel. Table 6.6 presents the results of the panel logistic regression model.

Table 6. 6: Random effects logistic regression model of university aspirations with Key Stage 2 attainment

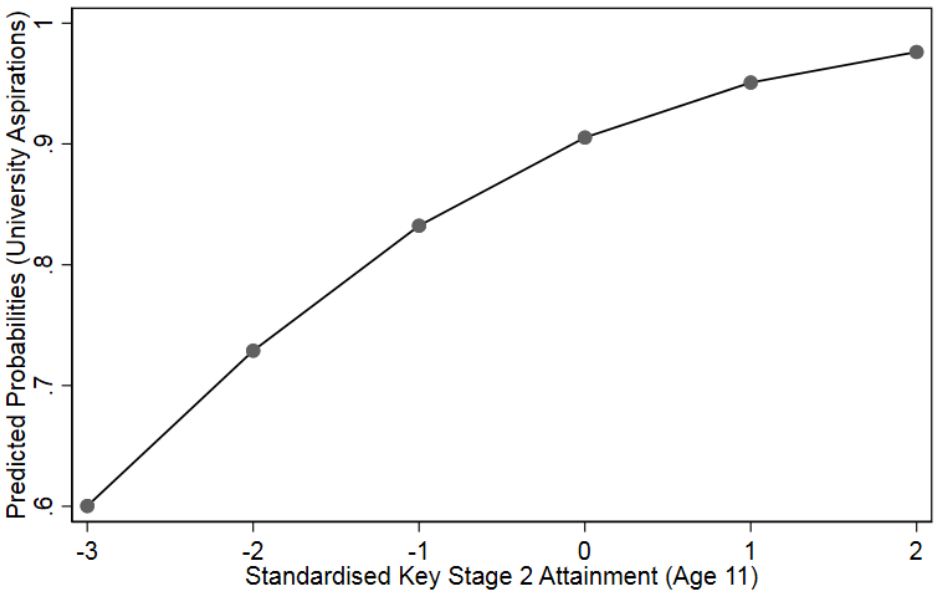
Random Effects Logistic Regression: University Aspirations	**Model 1 Coef.**	**S.E.**		**Model 2 Coef.**	**S.E.**	
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	-1.78	(0.41)	***	-1.43	(0.39)	***
School-level education	-1.92	(0.38)	***	-1.45	(0.36)	***
Below school-level education	-2.18	(0.45)	***	-1.22	(0.43)	**
Parental NS-SEC						
1.1 Large employers and higher managerial	-0.52	(0.79)		-0.31	(0.77)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	-1.10	(0.65)		-0.91	(0.62)	
3 Intermediate occupations	-0.55	(0.70)		-0.28	(0.67)	
4 Small employers and own account workers	-1.30	(0.68)		-0.79	(0.66)	
5 Lower supervisory and technical occupations	-1.52	(0.73)	*	-1.01	(0.70)	
6 Semi-routine occupations	-1.45	(0.67)	*	-0.95	(0.64)	
7 Routine occupations	-2.13	(0.70)	**	-1.57	(0.67)	*
8 Not in employment	-1.85	(0.66)	**	-1.33	(0.63)	*
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	1.45	(0.22)	***	1.07	(0.21)	***
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	1.31	(0.61)	*	1.30	(0.57)	*
Asian/Asian British	1.59	(0.39)	***	1.50	(0.37)	***
Black/Black British	1.05	(0.50)	*	1.16	(0.49)	*
Other ethnic group	1.73	(1.32)		1.51	(1.26)	
Age	0.34	(0.08)	***	0.35	(0.08)	***
Key Stage 2 Attainment (standardised)				0.86	(0.11)	***
Constant [output omitted]						
Standard deviation of panel-level variance	1.46	(0.25)		1.19	(0.25)	
Log of panel-level variance	0.76	(0.35)		0.35	(0.42)	
Observations (Total)	2178			2178		
Observations (Individuals)	1252			1252		
Likelihood ratio chi square (d.f.)	193.77 (17) ***			269.64 (18) ***		
Rho	0.39	(0.08)		0.30	(0.09)	
Likelihood ratio test of Rho (d.f.)	22.12 (1) ***			11.86 (1) ***		

Source: Waves 1-5 UKHLS youth questionnaires with linked NPD records, unweighted.
** p<.05, ** p<.01, *** p<.001.*

Attainment at age 11 is significantly associated with having university aspirations during adolescence. There is some evidence that attainment at age 11 partially mediates parental social class. This suggests that some, if not most, of the social class effect reported in Table 6.5 might be working through prior attainment. There are no significant interactions effects between prior attainment and parental education or parental social class. The effects of gender and ethnicity remain even after controlling for prior attainment.

There are clear, strong associations between attainment at the end of primary school and the aspirations formed during secondary school. Figure 6.4 presents the predicted probability of university aspirations by attainment at age 11. There is a clear, positive gradient. Young people with higher attainment at the end of primary school have higher predicted probabilities of having university aspirations during secondary school.

Figure 6. 4: Predicted probabilities of university aspirations by Key Stage 2 attainment



Source: UKHLS-NPD unweighted data, n=2178.
Parent education, parent NS-SEC, gender, ethnicity, and age also included in model.

5 Modelling GCSE Attainment

5.1 Descriptive Statistics

The previous analyses have demonstrated that university aspirations tend to be stable over secondary school years, and that attainment at age 11 is strongly associated with educational aspirations formed in adolescence. The next set of analyses use cross-sectional data from Wave 1 of the UKHLS to examine the effects of aspirations on GCSE attainment. This analysis includes measures of educational aspirations from both parents and young people.

The descriptive statistics for the sub-sample are presented in Table 6.7. The third and fourth columns present the mean GCSE scores and standard deviations for each category of the independent variables. Young people whose parents have higher education levels, and who are in professional occupations tend to have higher mean GCSE scores than their less advantaged peers. The young people and their parents who have more positive educational aspirations, i.e. placing greater importance on doing well in school examinations and aspiring to continue in education beyond the compulsory age, tend to also have higher mean GCSE scores.

There are significant chi square associations between parent and child measures of educational aspirations. There is evidence of significant, but relatively weak, associations between the educational aspirations of young people and their parents. The strength of the association between parents' and children's university aspirations is moderate (chi square 61.03 at 1 d.f., $p<.001$, Cramer's $V=0.24$). The association between the importance placed on GCSE and A' Level examinations is weaker (chi square 38.64 at 4 d.f., $p<.001$, Cramer's $V = 0.14$).

Table 6. 7: Descriptive statistics for cross-sectional analyses, unweighted

	Freq.	Percent	Mean GCSE	Std. Dev
Parental Education Level				
Higher education	271	25.52	57.61	17.58
Further education	162	15.25	45.96	16.50
School-level education	482	45.39	38.72	20.04
Below school-level education	147	13.84	27.30	18.80
Parental NS-SEC				
1.1 Large employers and higher managerial	67	6.31	53.79	16.98
1.2 Higher professional occupations	93	8.76	58.20	16.05
2 Lower management and professional occupations	227	21.37	49.73	18.31
3 Intermediate occupations	108	10.17	45.26	18.92
4 Small employers and own account workers	112	10.55	45.85	21.84
5 Lower supervisory and technical occupations	48	4.52	35.38	19.07
6 Semi-routine occupations	134	12.62	35.41	20.46
7 Routine occupations	67	6.31	33.60	19.05
8 Not in employment	206	19.40	32.57	20.95
Housing Tenure				
Owned/privately rented	815	76.74	47.24	20.22
Social housing	247	23.26	29.28	18.29
Gender				
Male	527	49.62	40.82	21.34
Female	535	50.38	45.27	20.82
Ethnicity				
White	814	76.65	42.71	21.50
Mixed	*	*	*	*
Asian/Asian British	138	12.99	46.89	20.25
Black/Black British	56	5.27	39.96	18.13
Other ethnic group	*	*	*	*
Age				
11	30	2.82	44.97	20.66
12	180	16.95	44.01	21.51
13	286	26.93	42.51	22.16
14	287	27.02	42.25	21.03
15/16	279	26.27	43.64	20.25
Wants to go to University (Child)				
Yes	899	84.65	46.13	20.49
No	163	15.35	26.14	16.50
Post-16 Destination (Child)				
Full-time employment	134	12.62	24.78	16.28
Full-time study	425	40.02	50.67	20.42
Study and employment, e.g. an apprenticeship	475	44.73	42.07	19.57
Something else/don't know	28	2.64	32.00	18.80
Importance of GCSEs (Child)				
Very important	848	79.85	44.22	20.77
Important	198	18.64	39.64	22.16
Not very/at all important	16	1.51	23.94	18.29
Wants Child to go to University (Parent)				
Yes	1034	97.36	43.43	21.16
No	28	2.64	29.36	17.89
Importance of A' Levels for Child (Parent)				
Very important	835	78.63	45.27	21.10
Important	176	16.57	36.98	19.75
Not very/at all important	51	4.80	27.96	17.07
Total *n*	**1062**	**100.00**	**43.06**	**21.19**

5.2 Modelling Results

The outcome variable is GCSE point score. The regression model including aspiration measures is built sequentially. The young person's educational aspirations has greater explanatory power than educational attitudes (i.e. the importance of doing well in GCSEs). For the parent educational aspiration measures, the importance of doing well in A' Levels has a greater effect than wanting their child to go to university. Overall, the young person's own aspirations have greater explanatory power than the parents' aspirations for their children, and parent university aspirations are not significant in a model with the young person's own university aspirations.

The final models are presented in Table 6.8. Young people from more advantaged backgrounds (i.e. parents have higher education qualifications and are in higher professional occupations) have significantly higher GCSE point scores compared with less advantaged young people. Having positive educational aspirations has significant effects over and above the effects of socio-economic background. The inclusion of educational aspiration measures in the model results in a modest decrease in the effect sizes of parental education and parental social class coefficients. The substantive conclusions about the relative importance of parental education level and parental social class in GCSE attainment remain unchanged after including measures of educational aspirations. The results are re-estimated using the post 16 destinations (continuing in education) as a substitute for university aspirations and the results were very similar (see Table A1.10 in Appendix 1).

Table 6. 8: Linear regression model of total GCSE point score with educational aspirations

	Model 1			Model 2		
Linear Regression: GCSE Point Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	- 7.95	(2.16)	***	- 6.23	(2.02)	**
School-level education	-14.00	(2.05)	***	-11.83	(1.99)	***
Below school-level education	-21.40	(2.94)	***	-19.72	(2.94)	***
Parental NS-SEC						
1.1 Large employers and higher managerial	- 1.48	(3.15)		- 1.82	(3.22)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	- 4.12	(2.46)		- 3.85	(2.36)	
3 Intermediate occupations	- 4.40	(2.78)		- 4.37	(2.67)	
4 Small employers and own account workers	- 4.40	(3.41)		- 3.88	(3.24)	
5 Lower supervisory and technical occupations	-12.07	(3.48)	**	-10.60	(3.29)	**
6 Semi-routine occupations	-11.38	(3.38)	**	- 9.40	(3.20)	**
7 Routine occupations	-10.83	(3.35)	**	- 8.32	(3.33)	*
8 Not in employment	-11.99	(3.54)	**	- 9.44	(3.31)	**
Housing Tenure						
Owned/privately rented	Ref.	(.)		Ref.	(.)	
Social housing	- 9.53	(2.33)	***	- 9.36	(2.20)	***
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	4.35	(1.41)	**	2.70	(1.38)	
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	- 1.22	(3.29)		- 2.14	(3.20)	
Asian/Asian British	6.53	(2.96)	*	4.14	(2.79)	
Black/Black British	2.39	(3.34)		0.24	(3.12)	
Other	- 4.86	(7.39)		- 6.30	(5.83)	
University Aspirations						
Yes				10.41	(1.95)	***
No				Ref.	(.)	
Importance of A' Levels (Parent)						
Very important				Ref.	(.)	
Important				- 3.81	(1.79)	*
Not very/at all important				- 6.70	(3.24)	*
Importance of GCSEs (Child)						
Very important				Ref.	(.)	
Important				- 1.94	(1.64)	
Not very/at all important				-12.60	(6.00)	*
Constant [output omitted]						
Observations	1062			1062		
Adjusted R^2	0.302			0.355		
BIC (d.f.)	9254.038 (23)			9201.178 (28)		
BIC (based on deviance)	1854.118			1801.259		
BIC (based on likelihood ratio chi square)	-251.271			-304.130		

Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Models also include age and academic year.
** p<.05, ** p<.01, *** p<.001.*

The significant effects of gender and ethnicity on GCSE attainment are mediated by the inclusion of aspirations measures in the model. Conditional on having the same educational aspirations, there are no significant differences in the GCSE attainment

of boys and girls, or between young people from white and Asian backgrounds. The inclusion of educational aspirations for both the young person and their parents improves the explanatory power of the model (i.e. increased Adjusted R^2) and provides a more parsimonious model (i.e. decreased BIC statistics).

The panel models in section 3 demonstrated that the role of socio-economic background in explaining university aspirations is partially mediated by the role of prior attainment. This section includes the measure of Key Stage 2 attainment as an additional covariate in the regression model of overall GCSE point score. This follows similar procedures undertaken by Gregg and Washbrook (2011) and Chowdry et al. (2011), who used measures of prior attainment in models of later attainment alongside other covariates of interest, including educational aspirations.

Table 6.9 presents the linear regression model output of GCSE attainment including additional measures of prior attainment. Model 1 includes Key Stage 2 attainment as a covariate. The educational aspiration measures are no longer significant after the inclusion of prior attainment at age 11. The inclusion of prior attainment further weakens the effects of parental education and parental social class in GCSE attainment. Gender and ethnicity are significant in the model with aspirations and prior attainment. In the previous model, gender and ethnicity were not significant after including educational aspirations. This suggests that both aspirations and prior attainment are processes with gendered and ethnic dimensions, which are interlinked with later educational attainment, i.e. GCSE attainment. These results indicate that there is a more subtle gender and ethnicity effect in school-level education.

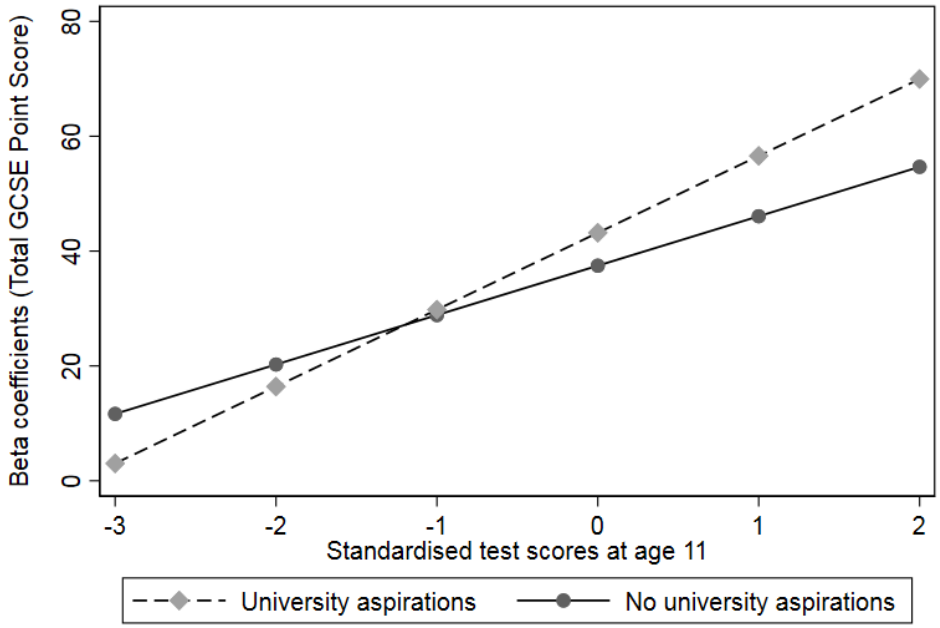
The panel analyses in section 3 demonstrated that attainment at age 11 is strongly associated with university aspirations developed over secondary school. To further develop this analytical theme, Model 2 introduces an interaction effect between Key Stage 2 attainment and university aspirations. There is a significant interaction effect between attainment at age 11 and university aspirations. The interaction effect is plotted in Figure 6.5. For the young people attaining higher than average test scores at age 11, those who have university aspirations have higher GCSE point scores than those who do not have university aspirations.

Table 6. 9: Linear regression model of total GCSE point score with Key Stage 2 attainment

	Model 1			Model 2		
Linear Regression: GCSE Point Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level						
Higher education	Ref.	(.)		Ref.	(.)	
Further education	- 2.96	(1.67)		- 2.44	(1.67)	
School-level education	- 7.71	(1.50)	***	- 7.30	(1.51)	***
Below school-level education	- 9.77	(2.36)	***	- 9.68	(2.35)	***
Parental NS-SEC						
1.1 Large employers and higher managerial	- 1.93	(2.38)		- 2.09	(2.39)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)	
2 Lower management and professional occupations	- 2.64	(1.71)		- 2.46	(1.68)	
3 Intermediate occupations	- 2.35	(2.18)		- 2.33	(2.18)	
4 Small employers and own account workers	- 0.06	(2.39)		- 0.20	(2.38)	
5 Lower supervisory and technical occupations	- 5.36	(2.82)		- 5.48	(2.80)	
6 Semi-routine occupations	- 5.33	(2.59)	*	- 5.24	(2.58)	*
7 Routine occupations	- 4.36	(2.69)		- 4.24	(2.66)	
8 Not in employment	- 6.28	(2.44)	*	- 6.28	(2.44)	*
Housing Tenure						
Owned/privately rented	Ref.	(.)		Ref.	(.)	
Social housing	- 5.59	(1.81)	**	- 5.39	(1.79)	**
Gender						
Male	Ref.	(.)		Ref.	(.)	
Female	3.03	(1.08)	**	3.06	(1.07)	**
Ethnicity						
White	Ref.	(.)		Ref.	(.)	
Mixed	1.12	(2.12)		1.65	(2.13)	
Asian/Asian British	6.45	(1.68)	***	6.41	(1.67)	***
Black/Black British	4.33	(2.72)		4.50	(2.78)	
Other	0.60	(7.76)		0.74	(8.03)	
University Aspirations						
Yes	3.10	(1.79)		5.73	(2.16)	**
No	Ref.	(.)		Ref.	(.)	
Importance of A' Levels (Parent)						
Very important	Ref.	(.)		Ref.	(.)	
Important	- 0.13	(1.47)		- 0.02	(1.47)	
Not very/at all important	- 4.03	(2.90)		- 4.00	(2.90)	
Importance of GCSEs (Child)						
Very important	Ref.	(.)		Ref.	(.)	
Important	- 1.31	(1.19)		- 1.46	(1.18)	
Not very/at all important	- 8.25	(5.63)		- 8.40	(5.07)	
Key Stage 2 Score (Standardised)	12.52	(0.59)	***	8.61	(1.49)	***
Key Stage 2 Score X University (Interaction)				4.79	(1.54)	**
Constant [output omitted]						
Observations	1062			1062		
Adjusted R^2	0.623			0.629		
BIC (d.f.)	8636.951 (29)			8625.164 (30)		
BIC (based on deviance)	1237.031			1225.244		
BIC (based on likelihood ratio chi-square)	-868.358			-880.145		

Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Models also include age and academic year.
** p<.05, ** p<.01, *** p<.001.*

Figure 6. 5: Interaction effect of Key Stage 2 attainment with university aspirations on GCSE attainment



6 Discussion and Conclusions

'Raising aspirations' has been central to recent political rhetoric and considered a strategy to close the attainment gap. The language of 'high' aspirations has focused on having aspirations to go to university and to enter professional occupations. A 'poverty of aspiration' narrative was developed to explain why those from less advantaged backgrounds tend to have poorer educational outcomes than their more advantaged peers. The analyses presented in this chapter do not provide any empirical support for the narrative of a 'poverty of aspirations'. Instead, there are universally 'high' educational aspirations reported by the young people, regardless of socio-economic background. This chimes with previous empirical research (St Clair and Benjamin, 2011, Baker et al., 2014, Treanor, 2017). The analyses of the panel data indicate that aspirations tend to be very stable over time, particularly for young people expressing interest in continuing their education beyond the compulsory age.

As stated by Gorard et al. (2012), establishing a causal link between aspirations and attainment is problematic. The structural ordering of the measures used in these

analyses provide useful insights into the temporal ordering of these events. Measures of attainment at ages 11 and 16 are retrieved from the National Pupil Database. The UKHLS youth questionnaire is completed by young people aged 11 to 15 in UKHLS households. There is a clear temporal ordering of test scores at age 11, reporting aspirations during the youth panel, and then GCSE examinations at age 16. The analyses demonstrate that socio-economic background is associated with university aspirations, but this relationship is much weakened by the inclusion of attainment at age 11. Prior attainment is a strong predictor of both university aspirations and GCSE attainment. The effect of prior attainment may contribute directly to later educational attainment, or may work indirectly through other factors such as educational aspirations. It could therefore be theorised that aspirations adjust for the young person's lived school experience, for example, aspirations are informed by known attainment and considerations of academic self-concept.

University aspirations, the importance of GCSEs from the young person's perspective, and the importance of A' Levels from the parent perspective are all significantly associated with GCSE attainment. Parents' university aspirations for their child are not significant after including the child's own aspirations. However, the role of socio-economic background in GCSE attainment continues to persist over and above the effects of both parents' and children's educational aspirations. The inclusion of educational aspiration measures in the model of GCSE attainment do not convincingly explain the role of socio-economic background for the Wave 1 respondents. The clear findings emerging from these analyses is that the Government emphasis of raising aspirations to close the attainment gap (see Cabinet Office, 2008, Cabinet Office, 2009, Cabinet Office, 2011) are not supported by empirical data. These findings therefore strongly challenge the validity of contemporary political rhetoric around 'raising aspirations'.

Conclusions

1 Introduction

This thesis was an empirical investigation into the relationship between parental socio-economic background and children's school GCSE attainment. The analyses were organised into two parts. Part 1 examined the role of parental education and parental social class in GCSE attainment constructing synthetic cohorts of English school Year 12 pupils in two British household panel surveys. The nature of the association between parental socio-economic background and children's school GCSE attainment was estimated using a variety of statistical models from the generalised linear framework. Particular attention was paid to providing sensitivity analyses of alternative measures and model specifications. The empirical findings using the British Household Panel Survey were replicated and extended using more recent data from the UK Household Longitudinal Study.

In Part 2, three potential explanations for the persisting socio-economic background effect in children's school GCSE attainment were explored. The role of prior attainment was examined using path analysis models. The particular analytical attraction of the path analysis approach was that it enabled the decomposition of direct and indirect effects. This facilitated an alternative method to examining parental socio-economic background and prior attainment. The role of cultural capital was examined using a series of sensitivity analyses of alternative candidate measures for parents and children. Finally, the role of educational aspirations was examined through the application of panel models.

2 Substantive Conclusions

The overwhelming substantive finding is the presence of strong, enduring parental socio-economic effects in GCSE attainment. Socio-economic inequalities in educational attainment:

- Have persisted at GCSE level over the course of more than two decades, i.e. from the 1990s to the early 2010s;
- Persist at GCSE level regardless of the form of the GCSE measure;
- Begin early, and are clearly observable by the end of primary school;
- Continue to persist, and potentially widen, over the course of secondary school;
- Are not convincingly explained by conventional sociological explanations of cultural capital;
- Are not convincingly explained by policy explanations of educational aspirations.

2.1 The Contemporary Relationship between Parental Socio-Economic Background and Children's School GCSE Attainment

The central empirical finding in Part 1 was a remarkable enduring effect of parental socio-economic background on children's school GCSE attainment. Using contemporary household panel survey data, the analyses demonstrated that relative socio-economic inequalities in GCSE attainment persist for young people growing up between the 1990s and the early 2010s. This finding is consistent with previous studies of young people's GCSE outcomes throughout the 1990s and 2000s (Drew, 1995, Demack et al., 2000, Sullivan, 2001, Connolly, 2006, Connelly et al., 2013, Gayle et al., 2014, Strand, 2014a, Playford and Gayle, 2015), and provides new empirical evidence for young people in the early 2010s. These findings are consequential at an individual level because of the potential restrictions on the opportunities and choices for future education, employment, and training. These findings are unsettling at a societal level, because they clearly demonstrate that socio-economic inequalities in school attainment are not an historic phenomenon, but strongly persist in the 21st Century.

In education policy, the attainment gap is often reported as the gap between the most disadvantaged young people and their non-disadvantaged peers. The Government's definition of the most disadvantaged tends to be indicated by binary measures, such as the eligibility for Free School Meals (see Department for Education, 2015c). The

analyses in Part 1 demonstrated that there is a much more nuanced socio-economic background effect. In particular, the analyses provided substantial empirical evidence of a finely-graded socio-economic effect on GCSE attainment for young people across a broad spectrum of parental educational and occupational advantage. The Government threshold measure for the attainment gap tends to be the attainment of 5 or more GCSEs at grades A*-C (Leckie and Goldstein, 2009). The results in Part 1 further demonstrated that there are socio-economic inequalities for a wider range of GCSE attainment measures than the benchmark measure of 5 or more GCSEs at grades A*-C. The implication of the current findings is that the definition of the attainment gap could be more broadly termed to address the clear inequalities across a more finely-graded spectrum of educational and occupational advantage.

A key methodological contribution of the work in Part 1 was the attention to the replicability of results. The robustness of the original findings was tested using a series of sensitivity analyses. These included alternative measures of parental socio-economic background, and alternative model specifications regarding the functional form of the GCSE outcome variable. This thesis further undertook a replication analysis using more contemporary household panel data to test the 'repeatability' of the original findings. The results were robust to repeating the same analyses using a different dataset. Socio-economic inequalities were evident when measuring GCSE 'success' as 5 or more GCSEs at grades A*-C, a set of attainment brackets, the number of good passes attained, or overall point score at the end of compulsory schooling. This could not have been known *a priori*.

2.2 The Role of Prior Attainment

The analyses in Chapter 4 demonstrated that socio-economic inequalities begin to emerge before young people finish primary school and continue to persist throughout secondary school. This finding is consistent with previous studies of educational attainment over the course of primary and secondary schooling for young people in the UK (for example, Feinstein, 2004, Gregg and Washbrook, 2011, Chowdry et al., 2011, Strand, 2014b). Parental socio-economic background plays an enduring effect in attainment at age 11, and stratification at earlier ages is consequential for

stratification at later ages. Policy interventions aimed at closing attainment gaps between individuals from different socio-economic backgrounds must begin earlier in the educational system.

In addition, substantial effects of parental education and parental social class persist over and above the effect of prior attainment. More socio-economically advantaged young people, therefore, have a double advantage. First, they are more likely to attain better test results at age 11 and, as a consequence, also attain better results at age 16. Second, even if they do not attain good test scores at age 11, they are still likely to get ahead of their less advantaged peers and attain better GCSE examinations results at the age of 16. This supports previous research which has demonstrated that socio-economic gaps widen between Key Stage 2 (age 11) and Key Stage 4 (age 16) (Crawford et al., 2017). If governments are serious about narrowing the attainment gap, then these analyses suggest that focusing on earlier educational stages are important to prevent the widening of inequalities at later stages over the educational and occupational life course.

It is anticipated that this is one of the first studies to model prior attainment and GCSE attainment using the UKHLS-NPD data. The use of these data is important, and builds upon previous research, due to the detailed measures available. In particular, the use of household panel data provides a detailed measure of occupational social class, reported by the parents, which is usually absent in administrative datasets. Studies relying solely on administrative data tend to use sub-optimum, proxy variables for socio-economic background such as eligibility for Free School Meals or area-level deprivation measures. The use of detailed social class measures allows for a more finely-grained analysis of socio-economic advantage at the individual level. The use of the NPD, as administrative education records, allows for the study of detailed attainment data, which is seldom available in omnibus, large-scale social science surveys.

The data further facilitated a detailed consideration of the role of prior attainment at age 11. It is highly plausible that social stratification occurs at even earlier stages in a young person's schooling. The NPD also contains information on young people's

attainment over the course of their primary and secondary schooling. The availability and quality of attainment measures at Key Stage 1 (age 7) and Key Stage 3 (age 14) was problematic due to large proportions of cases with missing data. Sample sizes were already small, which complicates statistical analyses. In future research, an interesting extension would be to map the trajectories of young people's school attainment for a longer period, beginning at earlier stages, to better understand when stratification starts to influence educational outcomes.

2.3 The Role of Cultural Capital

The analyses in Chapter 5 examined the role of cultural capital in GCSE attainment. The analyses disentangled potential operationalisations of cultural capital into highbrow cultural participation and reading behaviours. The analyses indicated that reading behaviours have stronger effects than conventional indicators of highbrow cultural participation, such as going to the theatre, going to museums and art galleries, or visiting historical places. These findings provide new empirical support to the distinction between reading behaviours and highbrow cultural participation demonstrated in older empirical studies (De Graaf, 1986, Crook, 1997, De Graaf et al., 2000, Sullivan, 2001). The results suggest that encouraging young people and parents to read for pleasure and to engage in positive reading activities are beneficial for the young person's GCSE attainment. Reading for pleasure may have strong effects in GCSE attainment due to the direct application to scholastic work, such as increasing vocabulary and improving general literacy. The significance of parents' reading behaviours over and above the effects of the young person's own reading behaviours might suggest that positive reading behaviours within the household have wider educational benefits.

There are broader policy implications for these findings. The schools' inspectorate, Ofsted, introduced the concept of cultural capital into its inspection framework for the first time from September 2019. Ofsted's definition of cultural capital is rather vague, and it is not currently apparent how schools will be assessed for providing their pupils with cultural capital. The analyses in Chapter 5 provide new empirical evidence to support the older distinction that positive parent and child reading behaviours are

more convincingly associated with higher GCSE attainment compared with visiting theatres, museums, galleries, or historical places. In the absence of clear guidance around operationalising and measuring cultural capital, as illustrated in this thesis, policymakers, schools, and teachers may benefit from avoiding the confusing language of cultural capital and using the more forthright, and ultimately more implementable, terminology of reading.

The results in Chapter 5 suggested that reading behaviours are important, but that other indicators of cultural capital are not. The results did not, however, provide any clear or convincing evidence that cultural capital mediates the socio-economic background effect in GCSE attainment. Most candidate cultural capital measures demonstrated significant associations with socio-economic background. The young people and their parents from more advantaged socio-economic positions engaged in cultural activities more, and more frequently, than their less advantaged peers. The substantive conclusions about the role of parental socio-economic background in GCSE attainment remained the same despite the inclusion of cultural capital measures. There is therefore a largely unexplained socio-economic effect which persists over and above the effects of cultural capital. These results are more consistent with cultural mobility theory, whereby cultural capital has effects additional to social origin, rather than explaining the effect (DiMaggio, 1982, De Graaf et al., 2000).

Sullivan (2002) argued that cultural capital is a vague concept which is often ill-defined, and that there are sizeable operationalisation challenges to measuring cultural capital. An original and unique aspect of this thesis has been the attention to sensitivity analyses of alternative cultural capital measures within the same study. Sensitivity analyses were undertaken to directly address the operationalisation challenge of cultural capital in empirical work. The overall trends with regards to the effects of cultural capital on GCSE attainment, and the potential mediatory effect on parental socio-economic background, remained consistent regardless of the measure, or combination of measures, of cultural capital used. This could not have been known *a priori*. Due to the operationalisation challenges, particularly comparability across studies, it would be good practice for future research to

undertake similar sensitivity analyses of all potential candidate measures available within a dataset.

2.4 The Role of Educational Aspirations

'Raising aspirations' has been a key political concern in recent decades. However, the analyses in Chapter 6 do not support the theoretical concept of a 'poverty of aspirations'. Educational aspirations are high for young people from all socio-economic backgrounds and tend to be relatively static over the period of the secondary school years. This adds to a recent body of literature which has found that young people's aspirations are generally high (St Clair and Benjamin, 2011, Baker et al., 2014, Treanor, 2017). Where statistical differences in aspirations existed between young people from more and less advantaged backgrounds, the differences were small in effect size. Furthermore, most of the social class effect in predicting university aspirations is mediated by attainment at age 11. This would support the idea that a young person's attainment influences their more realistic ambitions for continuing in education. These analyses provide further support to the findings in Chapter 4, that socio-economic differences in prior attainment are consequential in explaining later GCSE attainment.

There was evidence of a strong effect of university aspirations on GCSE attainment. Positive educational attitudes on behalf of the parents and children were similarly significantly associated with higher GCSE attainment. The effects of parental education level and parental social class persisted over and above the effects of positive educational aspirations and attitudes. The results of the analyses demonstrated that social stratification by parental socio-economic background is clearly observed by the age of 11. This has consequences for a young person's educational aspirations over the course of their secondary schooling. The focus of potential interventions which aim to close the attainment gap should be targeted earlier in the educational life course, which could have beneficial consequences for young people's attainment and aspiration formation over the course of their secondary schooling.

A useful methodological contribution of this work was the use of panel data models to examine the socio-economic determinants of aspirations. Gorard et al. (2012) strongly cautioned against interpreting a causal link between aspirations and attainment. The analyses in Chapter 6 do not make firm claims about causality. Nonetheless, the temporal ordering of the data has been leveraged to better understand the interdependencies of aspirations and attainment. The random effects logistic regression panel models of university aspirations can therefore be understood as the effect of attainment at the end of primary school on aspirations, with attainment occurring before aspiration formation. The cross-sectional linear regression models of GCSE attainment can be understood as the effects of having aspirations to go to university during secondary school, given attainment at the end of primary school.

2.5 Substantive Reflections

The focus of this thesis has been the relationship between parental socio-economic background and children's school GCSE attainment. The particular focus in Part 2 was on potential sociological explanations for inequalities in GCSE attainment. Studies examining social stratification and education within the discipline of sociology have tended to focus on the role of social, cultural or economic factors. The work in this thesis has directly contributed to this tradition.

In other academic disciplines, alternative explanations have been explored. For example, examining the role of genetics across a range of outcomes is routine in the fields of developmental psychology, psychometrics and behavioural genetics (Feinstein, 2003). Freese et al. (2003) highlighted the potential relevance of biology in social inquiry, and the need for greater awareness of the relationship between sociological and biological explanations. Harris and Schorpp (2018) provided an overview of the use of biomarkers and biological mechanisms for sociologists working in social stratification and health.

Analysing biomarkers and epigenetic data in the Avon Longitudinal Study of Parents and Children, Jerrim et al. (2015) investigated the role of three candidate genes and gene-environment interactions in reading scores and found weak associations

between the candidate genes and reading scores. The evidence provided limited support for the candidate genes explaining the socio-economic gaps in reading scores. Recent research examining the role of genetics in GCSE outcomes using twin studies have found strong associations between genetic factors and educational attainment (for example, Shakeshaft et al., 2013, Krapohl et al., 2014, Rimfeld et al., 2015,). Shakeshaft et al. (2013) analysed GCSE results using a national twin study and found that genetic factors explained more of the variance in mean GCSE scores than both family and school environment factors. Krapohl et al. (2014) noted that differences in children's educational outcomes are highly heritable from the early school years. The authors found that intelligence and genetically influenced traits such as personality and psychopathology contributed to heritability when analysing GCSE attainment. Rimfeld et al. (2015) examined the effects of genes in GCSE outcomes and found evidence of high heritability across academic subjects, even after controlling for intelligence.

An advancement in large-scale social science surveys is the collection of biomarkers and epigenetic information. Recently, the UKHLS began to collect a range of biomarker and epigenetic information for a sub-sample of adults, to enable both social scientific and biological analyses (see Benzeval et al., 2016). There was no biological data collected in the BHPS or for young people in the UKHLS households. Therefore, it has not been possible to examine the role of genetics in the analyses in this thesis. The advancement in collecting genetic information could, however, provide potentially fruitful future research opportunities.

3 Methodological Reflections

3.1 Large-Scale Social Science Surveys

The UK has an impressive set of birth cohort studies which were commissioned at 12-yearly intervals, with the National Survey of Health and Development cohort in 1946 (Wadsworth et al., 2006), the National Child Development Study in 1958 (Power and Elliott, 2006), and the British Birth Cohort Study in 1970 (Elliott and Shepherd, 2006). The birth cohort studies can be suitably utilised to study a broad range of

research outcomes, including educational outcomes. A birth cohort study was not commissioned for young people growing up in the 1990s or 2000s. Participants in the later Millennium Cohort Study (MCS) were born between 2000 and 2002, meaning that the latest British birth cohort members are currently aged 17-19 (Connelly and Platt, 2014). There is therefore a lack of suitable birth cohort data available to study young people's educational outcomes between the 1990s and early 2010s.

Household panel surveys are universal social science resources which can be utilised for a number of different research purposes and outcomes. The timing of the British Household Panel Survey and UK Household Longitudinal Study partially fill the data gap between the portfolio of birth cohort studies. The BHPS and UKHLS provide adequate analysis of educational outcomes for young people growing up in the 1990s and 2000s, and into the early 2010s. The analyses in this thesis have contributed to academic and policy knowledge about socio-economic background effects in GCSE attainment for these cohorts of young people growing up in contemporary England.

The UKHLS has provided a great opportunity to study educational attainment using valid and reliable education data from administrative sources. The youth questionnaires of the UKHLS have enabled the use of a rich set of measures of cultural capital and educational aspirations. There have, however, been a number of limitations. The National Pupil Database has only been linked to parents consenting in Wave 1 of the UKHLS. This has restricted the number of cases linked to administrative data. A further restriction of using these data is the rotating questionnaire design of the UKHLS, meaning that potential measures of cultural capital or parent educational aspirations are not available in all waves. Some analyses in Part 2 have had considerably reduced sample sizes as a consequence of these two issues.

The use of administrative NPD data is subject to undertaking SURE (Secure Users of Research data Environments) researcher training and working in the Secure Lab environment. This introduces practical research challenges. There is generally a delay between applying for data access, undertaking and passing the required training, and gaining access to the Secure Lab server. The data cannot be

downloaded to researchers' desktops, and strict procedures around safe uses of data must be followed. All data management, data analyses, and the writing up of results in this thesis had to take place within the Secure Lab environment. This was a fixed location using an institutional desktop with remote access to the Secure server. All of the final outputs were subject to statistical disclosure control from a dedicated team at the UK Data Service. Connecting to the Secure Lab server is more restrictive and less convenient than using most large-scale social science surveys which are readily downloadable under a standard End User license. Access is, however, less restrictive and more convenient than that required by many administrative datasets, which must be analysed in Safe Havens.

Looking to future research opportunities, the MCS dataset with linked GCSE examinations was released in July 2019, which will provide the research community with new opportunities to analyse the educational outcomes of young people growing up in the 2010s. There is a potential future 'data crisis'. A new birth cohort study was announced in 2011 with research council funding. The new study, 'Life Study', was due to recruit over 80,000 babies born between 2014 and 2018.[20] Funding was removed in late 2015 and the study subsequently closed due to recruitment challenges. Household panel surveys may take on even greater data significance for cohorts of young people growing up in the 2020s and beyond.

3.2 Changes to Curriculum and Assessment

The education system in England has changed since the young people in the preceding analyses sat their GCSE examinations. The structure of GCSE assessment in England was updated in 2015 for implementation from 2017.[21] First, a numeric rather than alphabetic grading system for GCSEs was phased in, starting with English and Maths. The new grading system ranges from 9 (highest) to 1 (lowest). Second, linear GCSEs have been introduced to replace modular GCSEs, with examinations at the end of the two years of study. This replaces coursework and modular

[20] See https://www.lifestudy.ac.uk [accessed 18.08.19].
[21] See https://www.gov.uk/government/publications/get-the-facts-gcse-and-a-level-reform/get-the-facts-gcse-reform [accessed 22.05.19].

examinations throughout school Years 10 and 11. Future research into GCSE outcomes for young people in England will need to be mindful of the changes in the assessment structure of GCSEs. In particular, the changes to the curriculum, the modular assessments, and the grading system will have consequences for comparability with earlier analyses.

Accountability measures have also changed with a move from absolute to relative assessment for pupils' attainment. The key policy benchmark of attainment for young people and their schools was the attainment of 5 or more GCSEs at grades A*-C (Leckie and Goldstein, 2009). This has been a benchmark measure employed in the analyses in Part 1. From 2016, secondary school accountability measures were updated, including Progress 8 and Attainment 8 benchmark attainment measures (see Department for Education, 2019). The relative measures identify progress from Key Stage 2 to GCSE. The work throughout this thesis has demonstrated the challenge of measuring GCSE attainment. The change in the policy benchmark measure may provide opportunities and further empirical challenges for social researchers. For example, age 11 test scores are not usually available in social surveys, and are not normally known by the young person or their parents, and therefore researchers may increasingly need to use administrative data to study school educational attainment, with the associated restrictions of data access and analysis.

The analyses in Chapter 4 demonstrated the cumulative nature of educational attainment, and the Government embracement of progress measures of attainment should be welcomed. The findings in Chapter 4 further highlighted the importance of socio-economic differences in attainment at age 16 which persist over and above attainment at age 11. Politicians should therefore acknowledge the role of structural, and potentially compounding, socio-economic inequalities throughout schooling as they begin to implement the new attainment measures.

4 Final Remarks

The social stratification of education has a long research tradition in the UK. The work presented throughout this thesis illustrates that empirical investigation into social stratification and educational attainment continues to be highly relevant for social researchers and policymakers in the 21st Century.

This thesis has taken a systematic approach to investigating socio-economic inequalities in school attainment and exploring potential explanations for the persisting effect using contemporary, nationally representative data. Governmental changes to the education system, types of schools, curriculum, assessment style and grades, and numerous policy interventions have not managed to weaken the persisting association between parental socio-economic background and children's school attainment. This thesis has made an original contribution to knowledge by developing a contemporary sociological understanding of the role of parental socio-economic background in GCSE attainment for young people into the 21st Century. The central findings of this thesis provide considerable evidence that socio-economic inequalities in school-level attainment persisted throughout the 1990s and 2000s, and continue to persist into the 2010s. The most convincing explanation for socio-economic inequalities in GCSE attainment is the role of attainment at age 11. Prior attainment is similarly stratified by socio-economic background and therefore the evidence suggests that not only is stratification occurring much earlier in a young person's education, but it is also having a deterministic effect on later educational outcomes. This thesis further provides new empirical evidence to suggest that conventional sociological and policy explanations do not reasonably explain the observed socio-economic differences in GCSE attainment.

Appendix 1: Additional Tables and Figures

Table A1. 1: Logistic regression models of different threshold measures of GCSE attainment

	1+ A*-C			5+ A*-Cs			9+ A*-Cs		
Logistic Regression Models	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-0.73	(0.38)		-0.90	(0.24)	***	-0.93	(0.21)	***
School-level education	-0.88	(0.39)	*	-1.19	(0.26)	***	-1.28	(0.23)	***
Below school-level education	-1.41	(0.44)	**	-1.87	(0.32)	***	-2.11	(0.35)	***
Parental NS-SEC									
1.1 Large employers/higher man.	-0.63	(0.65)		-0.51	(0.36)		-0.61	(0.32)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower managerial/professional	-0.79	(0.47)		-0.57	(0.30)		-1.08	(0.21)	***
3 Intermediate	-0.66	(0.53)		-0.63	(0.34)		-0.89	(0.27)	**
4 Small employers/own account	-1.42	(0.48)	**	-0.97	(0.33)	**	-1.37	(0.31)	***
5 Lower supervisory/technical	-1.23	(0.53)	*	-1.06	(0.33)	**	-1.44	(0.34)	***
6 Semi-routine	-1.61	(0.52)	**	-1.30	(0.35)	***	-1.35	(0.32)	***
7 Routine	-1.36	(0.53)	*	-1.01	(0.35)	**	-1.59	(0.35)	***
Housing tenure									
Owned or privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-1.23	(0.22)	***	-1.13	(0.23)	***	-0.90	(0.28)	**
Gender									
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	0.68	(0.16)	***	0.58	(0.12)	***	0.73	(0.12)	***
Constant	2.64	(0.61)	***	1.45	(0.44)	**	-0.39	(0.42)	
Observations	1624			1624			1624		
McFadden's Adjusted Pseudo R^2	0.109			0.115			0.139		
Cox-Snell Pseudo R^2	0.123			0.173			0.196		
Nagelkerke Pseudo R^2	0.214			0.236			0.270		
BIC (d.f.)	1407.392 (31)			2065.612 (31)			1982.690 (31)		
BIC (deviance)	-10598.267			-9940.048			-10022.970		
BIC (likelihood ratio chi square)	8.447			-87.450			-133.248		

*Source: BHPS synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Models also include ethnicity and wave year. * p<.05, ** p<.01, *** p<.001.*

Table A1. 2: Multinomial logistic regression model of GCSE attainment brackets

	0 A*-C v 9+ A*-C			1-4 v 9+ A*-Cs			5-8 v 9+ A*-Cs			9+ A*-Cs (Ref.)	
Multinomial Regression	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**
Parental Education											
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)
Further education	1.16	(0.40)	**	1.19	(0.30)	***	0.75	(0.24)	**	Ref.	(.)
School-level education	1.56	(0.40)	***	1.64	(0.34)	***	1.02	(0.25)	***	Ref.	(.)
Below school-level education	2.73	(0.51)	***	2.55	(0.44)	***	1.53	(0.41)	***	Ref.	(.)
Parental NS-SEC											
1.1 Large employers/higher man.	0.93	(0.69)		0.64	(0.43)		0.53	(0.37)		Ref.	(.)
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)
2 Lower managerial and professional	1.36	(0.48)	**	0.92	(0.34)	**	1.13	(0.26)	***	Ref.	(.)
3 Intermediate occupations	1.11	(0.55)	*	0.91	(0.39)	*	0.85	(0.33)	*	Ref.	(.)
4 Small employers and own account	2.13	(0.51)	***	1.23	(0.44)	**	1.25	(0.36)	***	Ref.	(.)
5 Lower supervisory and technical	2.03	(0.58)	***	1.44	(0.43)	**	1.26	(0.40)	**	Ref.	(.)
6 Semi-routine occupations	2.29	(0.55)	***	1.38	(0.45)	**	0.95	(0.38)	*	Ref.	(.)
7 Routine occupations	2.27	(0.59)	***	1.45	(0.47)	**	1.49	(0.38)	***	Ref.	(.)
Housing Tenure											
Owned or privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)
Social housing	1.72	(0.35)	***	0.94	(0.31)	**	0.26	(0.32)		Ref.	(.)
Gender											
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)
Female	-1.15	(0.19)	***	-0.73		***	-0.58	(0.15)	***	Ref.	(.)
Constant	-1.42	(0.73)		-1.04	(0.55)		-0.23	(0.45)		Ref.	(.)
Observations	1624										
McFadden's Adjusted Pseudo R^2	0.077										
Cox-Snell Pseudo R^2	0.276										
Nagelkerke Pseudo R^2	0.296										
BIC (d.f.)	4525.524 (93)										

Source: BHPS synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Models also include ethnicity and wave year.
** p<.05, ** p<.01, *** p<.001.*

Table A1. 3: Stereotype logistic regression model of GCSE attainment brackets

Stereotype Logistic Regression	**Coef.**	**S.E.**	
Parental Education			
Higher education	Ref.	(.)	
Further education	-1.46	(0.33)	***
School-level education	-1.93	(0.34)	***
Below school-level education	-3.13	(0.43)	***
Parental NS-SEC			
1.1 Large employers and higher managerial	-0.93	(0.50)	
1.2 Higher professional occupations	Ref.	(.)	
2 Lower managerial and professional occupations	-1.47	(0.34)	***
3 Intermediate occupations	-1.28	(0.41)	**
4 Small employers and own account workers	-2.10	(0.46)	***
5 Lower supervisory and technical occupations	-2.13	(0.46)	***
6 Semi-routine occupations	-2.36	(0.48)	***
7 Routine occupations	-2.25	(0.49)	***
Housing Tenure			
Owned or privately rented	Ref.	(.)	
Social housing	-1.91	(0.40)	***
Gender			
Male	Ref.	(.)	
Female	1.13	(0.17)	***
Φ1	1.00	(.)	
Φ2	0.70	(0.05)	***
Φ3	0.48	(0.05)	***
Θ1	-2.02	(0.58)	***
Θ2	-0.98	(0.40)	*
Θ3	-0.46	(0.27)	
Observations	1624		
McFadden's Adjusted Pseudo R^2	0.090		
Cox-Snell Pseudo R^2	0.249		
Nagelkerke Pseudo R^2	0.267		
BIC (d.f.)	4156.037 (35)		
BIC (based on deviance)	-7849.623		
BIC (based on likelihood ratio chi square)	-242.449		

Source: BHPS synthetic cohorts of English Year 12 pupils.
Adjusted for complex survey design. Models also include ethnicity and wave year.
** p<.05, ** p<.01, *** p<.001.*

Table A1. 4: Comparison of count regression models of number of GCSEs attainment at grades A*-C

	Poisson			Negative binomial			Zero-inflated Poisson			Zero-inflated negative binomial		
Count regression: Number of A*-C	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education												
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-0.18	(0.04)	***	-0.19	(0.04)	***	-0.14	(0.03)	***	-0.14	(0.03)	***
School-level education	-0.28	(0.05)	***	-0.29	(0.06)	***	-0.22	(0.04)	***	-0.23	(0.04)	***
Below school-level education	-0.60	(0.09)	***	-0.62	(0.09)	***	-0.41	(0.07)	***	-0.41	(0.07)	***
Parental NS-SEC												
1.1 Large employers and higher managerial	-0.10	(0.06)		-0.10	(0.06)		-0.07	(0.05)		-0.07	(0.05)	
1.2 Higher professional occupations	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower managerial and professional	-0.14	(0.04)	***	-0.14	(0.04)	**	-0.11	(0.03)	**	-0.11	(0.04)	**
3 Intermediate occupations	-0.13	(0.06)	*	-0.13	(0.06)		-0.12	(0.05)	*	-0.12	(0.05)	*
4 Small employers and own account workers	-0.28	(0.07)	***	-0.29	(0.07)	***	-0.17	(0.05)	**	-0.17	(0.06)	**
5 Lower supervisory and technical	-0.30	(0.07)	***	-0.31	(0.07)	***	-0.22	(0.06)	***	-0.23	(0.06)	***
6 Semi-routine occupations	-0.36	(0.08)	***	-0.38	(0.08)	***	-0.21	(0.06)	***	-0.21	(0.06)	***
7 Routine occupations	-0.31	(0.08)	***	-0.31	(0.09)	***	-0.21	(0.07)	**	-0.21	(0.07)	**
Housing Tenure												
Owned or privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-0.46	(0.10)	***	-0.47	(0.10)	***	-0.18	(0.06)	**	-0.18	(0.06)	**
Gender												
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	0.19	(0.03)	***	0.21	(0.03)	***	0.12	(0.02)	***	0.12	(0.02)	***

Table A1.4 continued overleaf.

Table A1.4 cont.	Poisson		Negative binomial		Zero-inflated Poisson			Zero-inflated negative binomial		
Logistic regression: Zero A*-Cs	**Coef.**	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E**		**Coef.**	**S.E.**	
Parental Education										
Higher education					Ref.	(.)		Ref.	(.)	
Further education					0.73	(0.38)		0.73	(0.39)	
School-level education					0.87	(0.40)	*	0.87	(0.41)	*
Below school-level					1.39	(0.44)	**	1.38	(0.45)	**
Parental NS-SEC										
1.1 Large employers and higher managerial					0.63	(0.66)		0.64	(0.67)	
1.2 Higher professional occupations					Ref.	(.)		Ref.	(.)	
2 Lower managerial and professional					0.79	(0.48)		0.80	(0.49)	
3 Intermediate occupations					0.65	(0.53)		0.65	(0.55)	
4 Small employers and own account workers					1.42	(0.49)	**	1.43	(0.50)	**
5 Lower supervisory and technical					1.21	(0.53)	*	1.22	(0.55)	*
6 Semi-routine occupations					1.61	(0.52)	**	1.62	(0.53)	**
7 Routine occupations					1.36	(0.53)	*	1.36	(0.54)	*
Housing Tenure										
Owned or privately rented					Ref.	(.)		Ref.	(.)	
Social housing					1.23	(0.22)	***	1.23	(0.23)	***
Gender										
Male					Ref.	(.)		Ref.	(.)	
Female					-0.67	(0.16)	***	-0.68	(0.17)	***
Alpha			3.45	(0.02)				0.05	(0.01)	
Observations	1624		1624		1624			1624		
McFadden's Adjusted Pseudo R2	0.093		0.029		0.049			0.036		
Cox-Snell Pseudo R2	0.473		0.183		0.289			0.237		
Nagelkerke Pseudo R2	0.474		0.183		0.290			0.238		
BIC (d.f.)	9741.483 (31)		9009.561 (32)		8680.351 (62)			8637.312 (63)		

Source: BHPS synthetic cohorts of English Year 12 pupils. Adjusted for complex survey design. Models also include ethnicity and wave year.
** p<.05, ** p<.01, *** p<.001.*

Table A1. 5: Total, direct, and indirect effects of socio-economic background on GCSE attainment, accounting for English score at Key Stage 2, standardised[22]

	Direct Effects			Indirect Effects			Total Effects		
English Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-0.08	(0.03)	**	-0.06	(0.02)	**	-0.14	(0.04)	***
School-level education	-0.21	(0.04)	***	-0.10	(0.03)	***	-0.31	(0.04)	***
Below school-level education	-0.15	(0.03)	***	-0.14	(0.03)	***	-0.29	(0.04)	***
Parental NS-SEC									
1.1 Large employer/higher man.	-0.04	(0.03)		-0.01	(0.02)		-0.05	(0.04)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower man./professional	-0.05	(0.04)		-0.04	(0.03)		-0.09	(0.05)	
3 Intermediate	-0.04	(0.03)		-0.04	(0.03)		-0.08	(0.04)	*
4 Small employer/own account	-0.02	(0.04)		-0.07	(0.03)	**	-0.09	(0.04)	
5 Lower supervisory/technical	-0.09	(0.03)	**	-0.05	(0.02)	*	-0.14	(0.04)	***
6 Semi-routine	-0.10	(0.04)	*	-0.09	(0.03)	**	-0.19	(0.05)	***
7 Routine	-0.05	(0.03)		-0.09	(0.03)	**	-0.14	(0.04)	***
8 Not in employment	-0.12	(0.04)	**	-0.12	(0.03)	***	-0.23	(0.05)	***

Table A1. 6: Total, direct, and indirect effects of socio-economic background on GCSE attainment, accounting for Maths score at Key Stage 2, standardised

	Direct Effects			Indirect Effects			Total Effects		
Maths Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-0.07	(0.03)	*	-0.08	(0.02)	***	-0.14	(0.04)	***
School-level education	-0.20	(0.04)	***	-0.11	(0.02)	***	-0.31	(0.04)	***
Below school-level education	-0.16	(0.03)	***	-0.13	(0.03)	***	-0.29	(0.04)	***
Parental NS-SEC									
1.1 Large employer/higher man.	-0.05	(0.03)		0.01	(0.02)		-0.05	(0.04)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower man./professional	-0.07	(0.04)		-0.02	(0.03)		-0.09	(0.05)	
3 Intermediate	-0.06	(0.03)		-0.02	(0.02)		-0.08	(0.04)	*
4 Small employer/own account	-0.04	(0.04)		-0.05	(0.02)	*	-0.09	(0.05)	
5 Lower supervisory/technical	-0.10	(0.03)	**	-0.05	(0.02)	*	-0.14	(0.04)	***
6 Semi-routine	-0.14	(0.04)	**	-0.06	(0.03)	*	-0.19	(0.05)	***
7 Routine	-0.06	(0.03)	*	-0.08	(0.02)	**	-0.14	(0.04)	***
8 Not in employment	-0.16	(0.04)	***	-0.08	(0.03)	**	-0.24	(0.05)	***

[22] The effects might not add exactly due to rounding to 2 decimal places.

Table A1. 7: Rotated and unrotated factor loadings of cultural capital measures

	Variable	Factor 1 loading		Factor 2 loading		Factor 3 loading		Uniqueness
		Un.	R.	Un.	R.	Un.	R.	
Parent	Theatre	0.71	**0.74**	-0.45	-0.11	-0.08	0.26	0.29
	Museum	0.80	**0.75**	-0.34	0.05	-0.11	0.22	0.23
	Historic place	0.73	**0.63**	-0.25	0.10	-0.08	0.21	0.40
	Opera	0.55	**1.00**	-0.48	-0.18	-0.51	-0.23	0.20
	Classical music	0.72	**0.74**	-0.23	0.12	-0.24	0.03	0.36
	Reads	0.62	0.08	-0.36	-0.13	0.65	**0.97**	0.06
	Library	0.50	0.17	0.05	0.27	0.11	0.24	0.74
	Discuss books	0.62	-0.06	-0.22	0.01	0.71	**0.99**	0.06
Child	Theatre	0.62	**0.44**	0.15	**0.45**	-0.24	-0.11	0.54
	Museum	0.70	0.20	0.52	**0.83**	-0.16	-0.11	0.23
	Historic place	0.75	0.33	0.36	**0.70**	-0.15	-0.05	0.30
	Reads	0.45	-0.10	0.55	**0.74**	0.03	-0.00	0.49
	Library	0.42	-0.21	0.73	**0.90**	0.00	-0.09	0.30
	Discuss books	0.46	-0.03	0.57	**0.74**	0.25	0.23	0.40

Table A1. 8: Linear regression model building process of GCSE point score including cultural capital factor scores

	Model 1			Model 2			Model 3			Model 4		
Linear Regression: GCSE Point Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level												
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-7.86	(2.94)	**	-6.86	(2.94)	*	-6.40	(2.83)	*	-6.54	(2.82)	*
School-level education	-12.88	(2.42)	***	-10.89	(2.45)	***	-10.53	(2.42)	***	-10.29	(2.36)	***
Below school-level education	-16.21	(3.51)	***	-13.59	(3.71)	***	-12.62	(3.64)	**	-11.49	(3.70)	**
Parental NS-SEC												
1.1 Large employers/higher managerial	-9.01	(4.70)		-8.89	(4.50)	*	-8.16	(4.61)		-8.50	(4.70)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower management and professional	-9.55	(3.24)	**	-9.16	(3.23)	**	-9.05	(3.19)	**	-9.28	(3.20)	**
3 Intermediate	-9.51	(3.56)	**	-8.15	(3.75)	*	-8.09	(3.81)	*	-8.36	(3.87)	*
4 Small employers and own account	-10.20	(4.12)	*	-9.02	(4.22)	*	-8.74	(4.18)	*	-8.93	(4.21)	*
5 Lower supervisory and technical	-14.01	(4.32)	**	-12.26	(4.43)	**	-11.29	(4.26)	**	-10.93	(4.25)	*
6 Semi-routine	-12.94	(4.53)	**	-11.67	(4.61)	*	-12.15	(4.45)	**	-12.31	(4.47)	**
7 Routine	-14.74	(4.40)	**	-12.49	(4.68)	**	-13.54	(4.53)	**	-13.28	(4.56)	**
8 Not in employment	-16.87	(4.08)	***	-14.90	(4.23)	***	-14.91	(4.05)	***	-14.67	(4.13)	***
Housing Tenure												
Owned/privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-11.70	(2.54)	***	-10.68	(2.43)	***	-10.47	(2.40)	***	-9.81	(2.32)	***
Gender												
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	7.05	(1.82)	***	6.74	(1.82)	***	5.36	(1.86)	**	5.05	(1.84)	**
Parent Highbrow Factor Score				3.09	(1.14)	**	1.51	(1.21)		1.10	(1.19)	
Child Factor Score							4.06	(1.04)	***	3.64	(1.09)	**
Parent Reading Factor Score										2.57	(1.15)	*
Constant [output omitted]												
Observations	736			736			736			736		
Adjusted R^2	0.267			0.282			0.311			0.316		
BIC (d.f.)	6550.585 (22)			6541.742 (23)			6516.249 (24)			6516.644 (25)		

Source: Wave 2 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design.
*Model also includes ethnicity, age, academic year. * p<.05, ** p<.01, *** p<.001.*

Table A1. 9: Linear regression model building process of GCSE point score including cultural capital summed scales

	Model 1			Model 2			Model 3			Model 4			Model 5		
Linear Regression: GCSE Point Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level															
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-7.86	(2.94)	**	-7.56	(2.92)	*	-7.64	(2.84)	**	-7.05	(2.89)	*	-6.60	(2.86)	*
School-level education	-12.88	(2.42)	***	-11.72	(2.41)	***	-11.50	(2.39)	***	-10.79	(2.50)	***	-10.79	(2.49)	***
Below school-level education	-16.21	(3.51)	***	-14.23	(3.73)	***	-13.07	(3.71)	***	-10.88	(3.77)	**	-11.23	(3.81)	**
Parental NS-SEC															
1.1 Large employers/higher managerial	-9.01	(4.70)		-9.77	(4.61)	*	-8.81	(4.63)		-8.11	(4.56)		-8.06	(4.55)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower management and professional	-9.55	(3.24)	**	-9.60	(3.17)	**	-8.99	(3.16)	**	-9.03	(3.17)	**	-9.21	(3.14)	**
3 Intermediate	-9.51	(3.56)	**	-8.81	(3.66)	*	-8.25	(3.72)	*	-8.73	(3.77)	*	-8.90	(3.76)	*
4 Small employers and own account	-10.20	(4.12)	*	-9.73	(4.10)	*	-8.89	(4.11)	*	-9.29	(4.27)	*	-9.44	(4.30)	*
5 Lower supervisory and technical	-14.01	(4.32)	**	-12.89	(4.34)	**	-11.26	(4.38)	*	-10.21	(4.40)	*	-10.69	(4.31)	*
6 Semi-routine	-12.94	(4.53)	**	-11.92	(4.57)	*	-11.61	(4.50)	*	-11.82	(4.54)	*	-12.00	(4.45)	**
7 Routine	-14.74	(4.40)	**	-12.60	(4.68)	**	-12.41	(4.61)	**	-13.70	(4.69)	**	-14.30	(4.57)	**
8 Not in employment	-16.87	(4.08)	***	-15.23	(4.09)	***	-14.56	(3.98)	***	-14.74	(4.03)	***	-14.83	(3.98)	***
Housing Tenure															
Owned/privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-11.70	(2.54)	***	-10.73	(2.47)	***	-10.35	(2.42)	***	-9.71	(2.33)	***	-9.88	(2.32)	***
Gender															
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	7.05	(1.82)	***	6.78	(1.81)	***	6.27	(1.87)	**	5.50	(1.87)	**	4.77	(1.87)	*

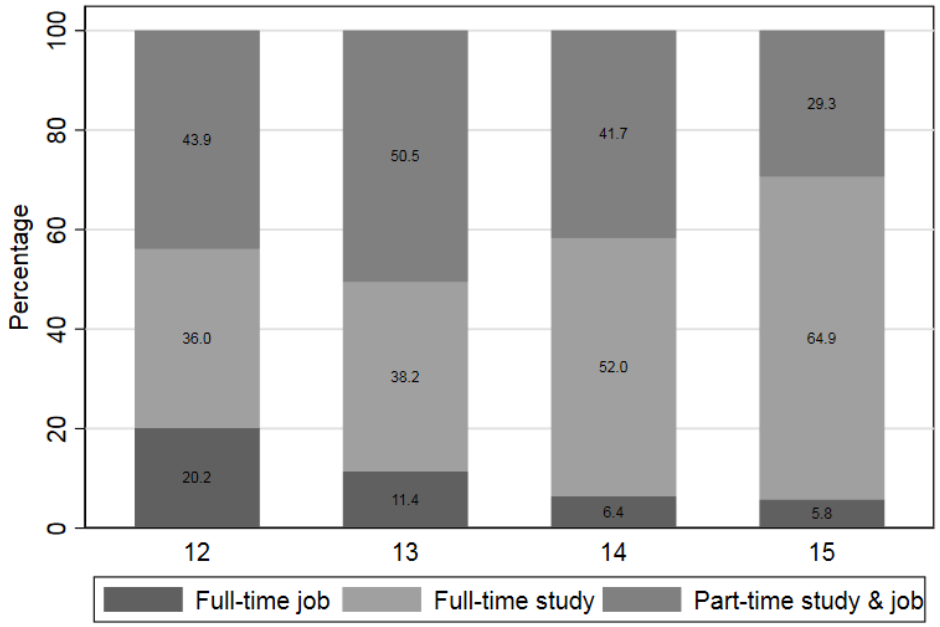
Table A1.9 continued overleaf.

Table A1.9 continued

	Model 1		Model 2			Model 3			Model 4			Model 5		
	Coef.	**S.E.**	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parent Highbrow Scale														
0			Ref.	(.)		Ref.	(.)		Ref.	(.)		Ref.	(.)	
1			2.28	(3.23)		1.76	(3.24)		0.23	(3.10)		0.10	(3.16)	
2			7.35	(2.90)	*	5.59	(2.94)		3.06	(2.81)		2.89	(2.82)	
3			6.83	(2.96)	*	4.21	(3.12)		0.42	(3.18)		0.68	(3.19)	
Child Highbrow Scale														
0						Ref.	(.)		Ref.	(.)		Ref.	(.)	
1						2.46	(2.47)		2.79	(2.52)		1.99	(2.47)	
2						5.84	(2.74)	*	5.87	(2.67)	*	4.48	(2.67)	
3						7.12	(2.71)	**	6.93	(2.68)	*	4.21	(2.83)	
Parent Reading Scale														
0									Ref.	(.)		Ref.	(.)	
1									7.23	(3.50)	*	6.87	(3.56)	
2									11.19	(3.48)	**	10.10	(3.62)	**
3									13.20	(5.37)	*	12.07	(5.43)	*
Child Reading Scale														
0												Ref.	(.)	
1												3.00	(2.23)	
2												6.02	(2.68)	*
3												6.65	(3.12)	*
Constant [output omitted]														
Observations	736		736			736			736			736		
Adjusted R^2	0.267		0.278			0.293			0.304			0.312		
BIC (d.f.)	6550.585 (22)		6556.199 (25)			6557.570 (28)			6562.945 (31)			6570.960 (34)		
BIC (based on deviance)	1692.080		1697.694			1699.065			1704.439			1712.455		
BIC (likelihood ratio chi square)	-111.719		-106.104			-104.734			-99.359			-91.343		

Source: Wave 2 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design.
*Model also includes ethnicity, age, and academic year. * p<.05, ** p<.01, *** p<.001.*

Figure A1. 1: Post-16 destinations by age of respondent, pooled data



Source: UKHLS Youth Questionnaire with linked NPD. Pooled cross-sectional data, n=2100.
Note: Ages 11/12 and 15/16 combined due to statistical disclosure control.

Table A1. 10: Linear regression model of total GCSE point score with post-16 study aspirations

Linear Regression: GCSE Point Score	**Coef.**	**S.E.**	
Parental Education Level			
Higher education	Ref.	(.)	
Further education	-5.96	(2.03)	**
School-level education	-11.22	(1.93)	***
Below school-level education	-18.96	(2.92)	***
Parental NS-SEC			
1.1 Large employers and higher managerial	-1.67	(3.21)	
1.2 Higher professional occupations	Ref.	(.)	
2 Lower management and professional occupations	-4.16	(2.37)	
3 Intermediate occupations	-4.92	(2.69)	
4 Small employers and own account workers	-4.92	(3.10)	
5 Lower supervisory and technical occupations	-11.52	(3.23)	***
6 Semi-routine occupations	-10.28	(3.24)	**
7 Routine occupations	-10.78	(3.34)	**
8 Not in employment	-11.88	(3.19)	***
Housing Tenure			
Owned/privately rented	Ref.	(.)	
Social housing	-8.76	(2.23)	***
Gender			
Male	Ref.	(.)	
Female	3.73	(1.34)	**
Ethnicity			
White	Ref.	(.)	
Mixed	-0.98	(3.42)	
Asian/Asian British	3.91	(2.75)	
Black/Black British	-0.53	(2.92)	
Other	-5.33	(7.11)	
Post-16 Study (e.g. A' Levels)			
Yes	8.38	(1.41)	***
No	Ref.	(.)	
Importance of A' Levels (Parent)			
Very important	Ref.	(.)	
Important	-4.54	(1.80)	*
Not very/at all important	-7.30	(3.28)	*
Importance of GCSEs (Child)			
Very important	Ref.	(.)	
Important	-1.50	(1.69)	
Not very/at all important	-13.41	(5.72)	*
Constant [output omitted]			
Observations	1062		
Adjusted R^2	0.358		
BIC (d.f.)	9195.612 (28)		
BIC (based on deviance)	1795.692		
BIC (based on likelihood ratio chi square)	-309.697		

Source: Wave 1 UKHLS youth questionnaires with linked NPD records.
Adjusted for complex survey design. Model also includes age and academic year.
** p<.05, ** p<.01, *** p<.001.*

Appendix 2: Missing Data in Part 2

The following tables present models using multiple imputation by chained equations for the analyses in Part 2. Due to statistical disclosure control, it is not possible to present the missing data patterns for the UKHLS-NPD data. In the following tables in Appendix 2, Model 1 represents the complete records analysis, Model 2 represents the model after 5 imputations, and Model 3 represents the model after 10 imputations. The statistical and substantive conclusions remain the same for the complete records analyses presented throughout Part 2 and the multiple imputation models presented below. This could not have been known *a priori*.

The complete records analyses in Chapter 4 had a total sample size of 1343. The total available sample for Wave 1 participants with linked Key Stage 2 and Key Stage 4 data is 1469. Table A2.1 presents the multiple imputation by chained equation models for Key Stage 2 attainment, and Table A2.2 presents the results for GCSE attainment.

The complete records analyses in Chapter 5 were conducted on a sample of 736 young people and their parents. The total eligible sample of young people in Wave 2 of the UKHLS with linked NPD records is 863. Table A2.3 presents the multiple imputation models for GCSE attainment including cultural capital summed scales.

The complete case records in Chapter 6 had a total sample size of 1062. The total available sample for Wave 1 participants with linked Key Stage 2 and Key Stage 4 data is 1469. Table A2.4 presents the multiple imputation models of GCSE attainment including educational aspiration measures.

Table A2. 1: Linear regression models of Key Stage 2 attainment using multiple imputation by chained equations (Chapter 4)

	Complete Records			5 Imputations			10 Imputations		
Linear Regression: KS2 Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-0.37	(0.10)	***	-0.36	(0.11)	**	-0.36	(0.11)	**
School-level education	-0.41	(0.09)	***	-0.39	(0.09)	***	-0.39	(0.09)	***
Below school-level education	-0.81	(0.16)	***	-0.76	(0.14)	***	-0.76	(0.14)	***
Parental NS-SEC									
1.1 Large employers/higher man.	-0.00	(0.14)		-0.00	(0.14)		-0.00	(0.14)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower management/professional	-0.14	(0.11)		-0.13	(0.11)		-0.13	(0.11)	
3 Intermediate	-0.18	(0.13)		-0.19	(0.13)		-0.19	(0.13)	
4 Small employers/own account	-0.38	(0.14)	**	-0.39	(0.14)	**	-0.39	(0.14)	**
5 Lower supervisory/technical	-0.46	(0.17)	**	-0.45	(0.17)	**	-0.45	(0.16)	**
6 Semi-routine	-0.43	(0.14)	**	-0.44	(0.14)	**	-0.44	(0.14)	**
7 Routine	-0.69	(0.19)	***	-0.70	(0.18)	***	-0.70	(0.18)	***
8 Not in employment	-0.51	(0.14)	***	-0.54	(0.14)	***	-0.54	(0.14)	***
Housing Tenure									
Owned/privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-0.24	(0.10)	*	-0.29	(0.10)	**	-0.29	(0.10)	**
Gender									
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	0.07	(0.06)		0.06	(0.06)		0.06	(0.06)	
Ethnicity									
White	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Mixed	-0.22	(0.15)		-0.21	(0.16)		-0.21	(0.16)	
Asian/Asian British	-0.16	(0.15)		-0.21	(0.15)		-0.21	(0.14)	
Black/Black British	-0.14	(0.17)		-0.06	(0.16)		-0.06	(0.15)	
Other ethnic group	-0.21	(0.40)		-0.25	(0.33)		-0.25	(0.33)	
Constant [output omitted]									
Observations	1343			1469			1469		
Adjusted R^2	0.180								
BIC (d.f.)	3683.380 (22)								

*Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Model also includes academic year. * p<.05, ** p<.01, *** p<.001.*

Table A2. 2: Linear regression models of GCSE attainment using multiple imputation by chained equations (Chapter 4)

	Complete Records			5 Imputations			10 Imputations		
Linear Regression: GCSE Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	- 8.18	(2.11)	***	- 8.02	(2.20)	***	- 8.02	(2.19)	***
School-level education	-13.13	(1.91)	***	-12.73	(1.89)	***	-12.73	(1.89)	***
Below school-level education	-18.86	(2.74)	***	-17.90	(2.57)	***	-17.90	(2.57)	***
Parental NS-SEC									
1.1 Large employers/higher man.	- 3.52	(2.92)		- 3.29	(3.20)		- 3.29	(3.20)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower management/professional	- 4.63	(2.36)	(*)	- 4.40	(2.50)		- 4.40	(2.50)	
3 Intermediate	- 5.63	(2.55)	*	- 5.39	(2.66)	*	- 5.39	(2.65)	*
4 Small employers/own account	- 6.11	(3.03)	*	- 6.62	(3.12)	*	- 6.62	(3.12)	*
5 Lower supervisory/technical	-14.43	(3.55)	***	-14.03	(3.44)	***	-14.03	(3.43)	***
6 Semi-routine	-11.99	(2.92)	***	-12.13	(2.93)	***	-12.13	(2.93)	***
7 Routine	-12.50	(3.32)	***	-13.18	(3.39)	***	-13.18	(3.38)	***
8 Not in employment	-13.33	(2.93)	***	-14.14	(3.00)	***	-14.14	(2.99)	***
Housing Tenure									
Owned/privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	- 9.28	(1.92)	***	-10.47	(1.84)	***	-10.47	(1.84)	***
Gender									
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	4.87	(1.30)	***	4.55	(1.30)	**	4.55	(1.30)	**
Ethnicity									
White	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Mixed	- 2.33	(3.78)		- 3.22	(4.05)		- 3.22	(4.04)	
Asian/Asian British	5.33	(2.91)		3.61	(2.97)		3.61	(2.97)	
Black/Black British	4.13	(2.85)		5.81	(2.83)	*	5.81	(2.82)	*
Other ethnic group	1.73	(6.93)		3.91	(6.12)		3.91	(6.12)	
Constant [output omitted]									
Observations	1343			1469			1469		
Adjusted R^2	0.279								
BIC (d.f.)	11683.557 (22)								

*Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Model also includes academic year. * p<.05, ** p<.01, *** p<.001.*

Table A2. 3: Linear regression models of GCSE attainment using multiple imputation by chained equations (Chapter 5)

	Complete Records			5 Imputations			10 Imputations		
Linear Regression: GCSE Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	- 6.23	(2.89)	*	- 6.24	(2.64)	*	- 6.24	(2.64)	*
School-level education	-10.91	(2.52)	***	-11.38	(2.42)	***	-11.38	(2.42)	***
Below school-level education	-12.20	(3.71)	**	-14.06	(3.52)	***	-14.06	(3.52)	***
Parental NS-SEC									
1.1 Large employers/higher man.	- 8.13	(4.64)		- 8.05	(3.76)	*	- 8.05	(3.76)	*
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower management/professional	- 9.73	(3.18)	**	- 8.70	(3.24)	**	- 8.70	(3.24)	**
3 Intermediate	- 9.53	(3.64)	**	- 8.77	(3.82)	*	- 8.77	(3.82)	*
4 Small employers/own account	-10.11	(4.26)	*	- 8.35	(4.09)	*	- 8.35	(4.09)	*
5 Lower supervisory/technical	-12.13	(4.10)	**	-13.20	(3.86)	**	-13.20	(3.85)	**
6 Semi-routine	-12.93	(4.45)	**	-11.69	(4.01)	**	-11.69	(4.01)	**
7 Routine	-15.43	(4.36)	***	-14.25	(4.14)	**	-14.25	(4.14)	**
8 Not in employment	-15.97	(3.99)	***	-16.39	(3.82)	***	-16.39	(3.82)	***
Housing Tenure									
Owned/privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-10.30	(2.31)	***	-10.24	(2.11)	***	-10.24	(2.11)	***
Gender									
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	4.67	(1.84)	*	4.16	(1.64)	*	4.16	(1.64)	*
Ethnicity									
White	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Non-white	0.87	(2.71)		0.51	(2.54)		0.51	(2.54)	
Parent Reading Scale									
0	Ref.	(.)		Ref.	(.)		Ref.	(.)	
1	6.98	(3.52)	*	5.11	(2.68)		5.11	(2.68)	
2	10.43	(3.46)	**	8.40	(2.83)	**	8.40	(2.83)	**
3	12.41	(5.15)	*	12.79	(4.21)	**	12.79	(4.21)	**
Child Reading Scale									
0	Ref.	(.)		Ref.	(.)		Ref.	(.)	
1	3.96	(2.22)		4.13	(2.07)	*	4.13	(2.07)	*
2	7.78	(2.56)	**	7.43	(2.32)	**	7.43	(2.31)	**
3	9.06	(2.92)	**	10.09	(2.52)	***	10.09	(2.52)	***
Constant [output omitted]									
Observations	736			863			863		
Adjusted R^2	0.306								
BIC (d.f.)	6527.447 (25)								

*Source: Wave 2 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Model also includes age and academic year. * p<.05, ** p<.01, *** p<.001.*

Table A2. 4: Linear regression models of GCSE attainment using multiple imputation by chained equations (Chapter 6)

	Complete records			5 imputations			10 imputations		
Linear Regression: GCSE Score	**Coef.**	**S.E.**		**Coef.**	**S.E.**		**Coef.**	**S.E.**	
Parental Education Level									
Higher education	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Further education	-6.23	(2.02)	**	-5.87	(2.08)	**	-5.87	(2.08)	**
School-level education	-11.83	(1.99)	***	-10.49	(1.80)	***	-10.49	(1.80)	***
Below school-level education	-19.72	(2.94)	***	-16.76	(2.54)	***	-16.76	(2.53)	***
Parental NS-SEC									
1.1 Large employers/higher man.	-1.82	(3.22)		-2.86	(3.13)		-2.86	(3.13)	
1.2 Higher professional	Ref.	(.)		Ref.	(.)		Ref.	(.)	
2 Lower management/professional	-3.85	(2.36)		-4.17	(2.39)		-4.17	(2.38)	
3 Intermediate	-4.37	(2.67)		-5.31	(2.56)	*	-5.31	(2.56)	*
4 Small employers/own account	-3.88	(3.24)		-5.57	(2.91)		-5.57	(2.91)	
5 Lower supervisory/technical	-10.60	(3.29)	**	-12.86	(3.22)	***	-12.86	(3.21)	***
6 Semi-routine	-9.40	(3.20)	**	-10.37	(2.83)	***	-10.37	(2.83)	***
7 Routine	-8.32	(3.33)	*	-10.21	(3.15)	**	-10.21	(3.14)	**
8 Not in employment	-9.44	(3.31)	**	-12.09	(2.80)	***	-12.09	(2.80)	***
Housing Tenure									
Owned/privately rented	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Social housing	-9.36	(2.20)	***	-9.28	(1.74)	***	-9.28	(1.74)	***
Gender									
Male	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Female	2.70	(1.38)		2.81	(1.27)	*	2.81	(1.27)	*
Ethnicity									
White	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Mixed	-2.14	(3.20)		-4.54	(3.58)		-4.54	(3.58)	
Asian/Asian British	4.14	(2.79)		1.62	(2.75)		1.62	(2.73)	
Black/Black British	0.24	(3.12)		2.92	(2.49)		2.92	(2.49)	
Other ethnic group	-6.30	(5.83)		1.06	(5.73)		1.06	(5.70)	
University Aspirations									
Yes	10.41	(1.95)	***	10.83	(1.71)	***	10.83	(1.71)	***
No	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Importance of A' Levels (Parent)									
Very important	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Important	-3.81	(1.79)	*	-3.99	(1.60)	*	-3.99	(1.60)	*
Not very/at all important	-6.70	(3.24)	*	-8.40	(2.84)	**	-8.40	(2.83)	**
Importance of GCSEs (Child)									
Very important	Ref.	(.)		Ref.	(.)		Ref.	(.)	
Important	-1.94	(1.64)		-1.76	(1.40)		-1.76	(1.40)	
Not very/at all important	-12.60	(6.00)	*	-12.26	(4.67)	**	-12.26	(4.66)	**
Constant [output omitted]									
Observations	1062			1469			1469		
Adjusted R^2	0.355								
BIC (d.f.)	9201.178 (28)								

*Source: Wave 1 UKHLS youth questionnaires with linked NPD records. Adjusted for complex survey design. Model also includes age and academic year. * p<.05, ** p<.01, *** p<.001.*

Data Citation

University of Essex, Institute for Social and Economic Research. (2010) *British Household Panel Survey: Waves 1-18, 1991-2009*. [data collection]. *7th Edition*. UK Data Service. SN: 5151, http://doi.org/10.5255/UKDA-SN-5151-2.

University of Essex. Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2018). *Understanding Society: Waves 1-7, 2009-2016 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. *10th Edition.* UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-11.

Department for Education, University of Essex. Institute for Social and Economic Research and NatCen Social Research. (2015). *Understanding Society: Wave 1, 2009-2011: Linked National Pupil Database: Secure Access* [computer file]. *2nd Edition*. Colchester, Essex: UK Data Archive [distributor]. SN: 7642, http://dx.doi.org/10.5255/UKDA-SN-7642-2.

Bibliography

ACKER, J. 1973. Women and social stratification: A case of intellectual sexism. *American Journal of Sociology,* 78**,** 936-945.

ACOCK, A. C. 2013. *Discovering structural equation modeling using Stata,* College Station, Texas, Stata Press.

ALLEN, M. 2017. Path Analysis. *The SAGE encyclopedia of communication research methods.* Thousand Oaks, California: SAGE Publications.

ALLISON, P. D. 1999. Comparing logit and probit coefficients across groups. *Sociological Methods & Research,* 28**,** 186-208.

ALLISON, P. D. 2009. *Fixed effects regression models,* Thousand Oaks, California, SAGE Publications.

ANDERS, J. 2017. The influence of socioeconomic status on changes in young people's expectations of applying to university. *Oxford Review of Education,* 43**,** 381-401.

ANGRIST, J. D. & PISCHKE, J. R.-S. 2009. *Mostly harmless econometrics: An empiricist's companion,* Princeton, Princeton University Press.

ARCHER, L., DEWITT, J. & WONG, B. 2014. Spheres of influence: What shapes young people's aspirations at age 12/13 and what are the implications for education policy? *Journal of Education Policy,* 29**,** 58-85.

ASCHAFFENBURG, K. & MAAS, I. 1997. Cultural and educational careers: The dynamics of social reproduction. *American Sociological Review,* 62**,** 573-587.

BABB, P. 2005. A summary of focus on social inequalities. London: Office for National Statistics.

BAKER, W., SAMMONS, P., SIRAJ-BLATCHFORD, I., SYLVA, K., MELHUISH, E. C. & TAGGART, B. 2014. Aspirations, education and inequality in England: Insights from the Effective Provision of Pre-School, Primary and Secondary Education project. *Oxford Review of Education,* 40**,** 525-542.

BALL, S. J. 2003. *Class strategies and the education market: The middle classes and social advantage,* London, Routledge.

BARON, R. & KENNY, D. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology,* 51**,** 1173-1182.

BATHMAKER, A.-M., INGRAM, N. & WALLER, R. 2013. Higher education, social class and the mobilisation of capitals: Recognising and playing the game. *British Journal of Sociology of Education,* 34**,** 723-743.

BELL, A., FAIRBROTHER, M. & JONES, K. 2018. Fixed and random effects models: Making an informed choice. *Quality and Quantity***,** 1-24.

BELLER, E. 2009. Bringing intergenerational social mobility research into the twenty-first century: Why mothers matter. *American Sociological Review,* 74**,** 507-528.

BENNETT, T. 2009. *Culture, class, distinction,* Abingdon, England, Routledge.

BENZEVAL, M., KUMARI, M. & JONES, A. M. 2016. How do biomarkers and genetics contribute to Understanding Society? *Health Economics*, 25, 1219-1222.

BERGMAN, M. M. & JOYE, D. 2005. Comparing social stratification schemas: CAMSIS, CSP-CH, Goldthorpe, ISCO-88, Treiman, and Wright. University of Cambridge: Cambridge Studies in Social Research, No. 10, available at http://www.sociology.cam.ac.uk/research/srg/cs10.

BERNSTEIN, B. 1964. Elaborated and restricted codes: Their social origins and some consequences. *American Anthropologist,* 66**,** 55-69.

BERRIDGE, D. M. 1992. Fitting the continuation ratio model using GLIM4. *In:* FAHRMEIR, L., FRANCIS, B., GILCHRIST, R. & TUTZ, G. (eds.) *Advances in GLIM and statistical modelling. Lecture notes in statistics, vol 78.* New York: Springer.

BERRINGTON, A., ROBERTS, S. & TAMMES, P. 2016. Educational aspirations among UK young teenagers: Exploring the role of gender, class and ethnicity. *British Educational Research Journal,* 42**,** 729-755.

BLANDEN, J., GREGG, P. & MACHIN, S. 2003. Changes in educational inequality. *IDEAS Working Paper Series from RePEc.* St. Louis: Federal Reserve Bank of St. Louis.

BLANDEN, J. & MACHIN, S. 2004. Educational inequality and the expansion of UK higher education. *Scottish Journal of Political Economy,* 51**,** 230-249.

BLAU, P. M. & DUNCAN, O. D. 1967. *The American occupational structure,* New York, Wiley.

BOLIVER, V. 2015. Are there distinctive clusters of higher and lower status universities in the UK? *Oxford Review of Education,* 41**,** 1-20.

BOTTERO, W. 2005. *Stratification social division and inequality,* Abingdon, Oxon, Routledge.

BOTTERO, W. & PRANDY, K. 2003. Social interaction distance and stratification. *British Journal of Sociology,* 54**,** 177-197.

BOUDON, R. 1974. *Education, opportunity and social inequality: Changing prospects in Western society,* New York, Wiley.

BOURDIEU, P. 1973. Cultural reproduction and social reproduction. *In:* BROWN, R., K. (ed.) *Knowledge, education, and cultural change: Papers in the sociology of education.* London: Tavistock.

BOURDIEU, P. 1984. *Distinction: A social critique of the judgement of taste,* London, Routledge & Kegan Paul.

BOURDIEU, P. 1986. The forms of capital. *In:* RICHARDSON, J. (ed.) *Handbook of theory and research for the sociology of education.* Westport, CT.: Greenwood.

BOURDIEU, P. & PASSERON, J.-C. 1979. *The inheritors: French students and their relation to culture,* Chicago, University of Chicago Press.

BOURDIEU, P. & PASSERON, J. C. 1990. *Reproduction in education, society and culture,* London, SAGE Publications.

BRADLEY, H. 2014. Class descriptors or class relations? Thoughts towards a critique of Savage et al. *Sociology,* 48**,** 429-436.

BRANT, R. 1990. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics,* 46**,** 1171-1178.

BRAUNS, H., SCHERER, S. & STEINMANN, S. 2003. The CASMIN educational classification in international comparative research. *In:* HOFFMEYER-ZLOTNIK, J., P., H. & WOLF, C. (eds.) *Advances in cross-national comparison: A European working book for demographic and socio-economic variables.* London: Kluwer.

BREEN, R. & GOLDTHORPE, J. H. 1997. Explaining educational differentials. *Rationality and Society,* 9**,** 275-305.

BREEN, R. & YAISH, M. 2006. Testing the Breen – Goldthorpe model of educational decision making. *In:* MORGAN, S., L., GRUSKY, D., B. & FIELDS, G., S. (eds.) *Mobility and inequality: Frontiers of research in sociology and economics.* Stanford, California: Stanford University Press.

BROECKE, S. & HAMED, J. 2008. Gender gaps in higher education participation: An analysis of the relationship between prior attainment and young participation by gender, socio-economic class and ethnicity. MPRA Paper 35595, University Library of Munich, Germany.

BROWN, P. 1990. The 'third wave': Education and the ideology of parentocracy. *British Journal of Sociology of Education,* 11**,** 65-86.

BUCK, N. & MCFALL, S. 2011. Understanding Society: Design overview. *Longitudinal and Life Course Studies,* 3**,** 5-17.

BUIS, M. L. 2007. SEQLOGIT: Stata module to fit a sequential logit model. Statistical Software Components, Boston College Department of Economics.

BUKODI, E. & GOLDTHORPE, J. H. 2013. Decomposing 'social origins': The effects of parents' class, status, and education on the educational attainment of their children. *European Sociological Review,* 29**,** 1024-1039.

BURGESS, R. G. 1986. *Key variables in social investigation,* London, Routledge & Kegan Paul.

CABINET OFFICE 2008. Aspirations and attainment amongst young people in deprived communities: Analysis and discussion paper. London: HM Government.

CABINET OFFICE 2009. Unleashing aspiration: The final report of the panel on the fair access to the professions. London: HM Government.

CABINET OFFICE 2011. Opening doors, breaking barriers: A strategy for social mobility. London: HM Government.

CAMERON, A. C. & TRIVEDI, P. K. 1998. *Regression analysis of count data,* Cambridge, Cambridge University Press.

CANNADINE, D. 1998. *Class in Britain,* London, Yale University Press.

CARPENTER, J. R. & KENWARD, M. G. 2013. *Multiple imputation and its application,* Chichester, John Wiley & Sons.

CENTRAL ADVISORY COUNCIL FOR EDUCATION 1959. Crowther report: 15 to 18. London: HM Stationery Office.

CHEN, A. 2014. QV: Stata module to compute quasi-variances. Statistical Software Components S457831: Boston College Department of Economics.

CHOWDRY, H., CRAWFORD, C., DEARDEN, L., GOODMAN, A. & VIGNOLES, A. 2013. Widening participation in higher education: Analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 176**,** 431-457.

CHOWDRY, H., CRAWFORD, C. & GOODMAN, A. 2011. The role of attitudes and behaviours in explaining socio- economic differences in attainment at age 16. *Longitudinal and Life Course Studies,* 2**,** 59-76.

CLARK, T. & LINZER, D. 2015. Should I use fixed or random effects? *Political Science Research and Methods,* 3**,** 399-408.

CLARKE, P., CRAWFORD, C., STEELE, F. & VIGNOLES, A. 2010. The choice between fixed and random effects models: Some considerations for

educational research. *IDEAS Working Paper Series from RePEc.* St. Louis: Federal Reserve Bank of St. Louis.

COLDRON, J., CRIPPS, C. & SHIPTON, L. 2010. Why are English secondary schools socially segregated? *Journal of Education Policy,* 25**,** 19-35.

CONNELLY, R. & GAYLE, V. 2016. Micro-class inequalities in early reading and mathematics skills. *Conference Paper. Social Stratification Research Seminar.* University of Cambridge, Cambridge.

CONNELLY, R. & GAYLE, V. 2019. An investigation of social class inequalities in general cognitive ability in two British birth cohorts. *The British Journal of Sociology,* 70**,** 90-108.

CONNELLY, R., GAYLE, V. & LAMBERT, P. S. 2016a. A review of educational attainment measures for social survey research. *Methodological Innovations,* 9**,** 1-11.

CONNELLY, R., GAYLE, V. & LAMBERT, P. S. 2016b. A review of occupation-based social classifications for social survey research. *Methodological Innovations,* 9**,** 1-14.

CONNELLY, R., GAYLE, V. & LAMBERT, P. S. 2016c. Modelling key variables in social science research: Introduction to the special section. *Methodological Innovations,* 9**,** 1-2.

CONNELLY, R., GAYLE, V. & LAMBERT, P. S. 2016d. Statistical modelling of key variables in social survey data analysis. *Methodological Innovations,* 9**,** 1-17.

CONNELLY, R., GAYLE, V. & PLAYFORD, C. 2019. An alternative investigation of social class inequalities in school attainment: Occupations, capitals and resources. *Conference Paper. Social Stratification Research Seminar.* Vrije Universiteit Amsterdam, Amsterdam.

CONNELLY, R., MURRAY, S. J. & GAYLE, V. 2013. Young people and school GCSE attainment: Exploring the 'middle'. *Sociological Research Online,* 18**,** 1-12.

CONNELLY, R. & PLATT, L. 2014. Cohort profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology,* 43**,** 1719-1725.

CONNOLLY, P. 2006. The effects of social class and ethnicity on gender differences in GCSE attainment: A secondary analysis of the Youth Cohort Study of England and Wales 1997–2001. *British Educational Research Journal,* 32**,** 3-21.

CONNOR, M. J. 2003. Pupil stress and Standard Assessment Tasks (SATs): An update. *Emotional and Behavioural Difficulties,* 8**,** 101-07.

COXON, A. P. M. & JONES, C. L. 1979. *Class and hierarchy: The social meaning of occupations,* London, Macmillan.

CRAWFORD, C., MACMILLAN, L. & VIGNOLES, A. 2017. When and why do initially high-achieving poor children fall behind? *Oxford Review of Education,* 43**,** 88-108.

CROLL, P. 2009. Educational participation post-16: A longitudinal analysis of intentions and outcomes. *British Journal of Educational Studies,* 57**,** 400-416.

CROLL, P. & ATTWOOD, G. 2013. Participation in higher education: Aspirations, attainment and social background. *British Journal of Educational Studies,* 61**,** 187-202.

CROMPTON, R. 2008. *Class and stratification,* Cambridge, Polity Press.

CROOK, C. J. 1997. *Cultural practices and socioeconomic attainment: The Australian experience,* Westport, CT., Greenwood Press.

CROXFORD, L., IANNELLI, C. & SHAPIRA, M. 2007. Youth Cohort Time Series for England, Wales and Scotland, 1984-2002. *In:* SERVICE, U. D. (ed.) *National Centre for Social Research, Scottish Centre for Social Research.* Centre for Educational Sociology, University of Edinburgh.

DAVIES, S. & RIZK, J. 2017. The three generations of cultural capital research: A narrative review. *Review of Educational Research,* 88**,** 331-365.

DE GRAAF, N. D., DE GRAAF, P. M. & KRAAYKAMP, G. 2000. Parental cultural capital and educational attainment in the Netherlands: A refinement of the cultural capital perspective. *Sociology of Education,* 73**,** 92-111.

DE GRAAF, P. M. 1986. The impact of financial and cultural resources on educational attainment in the Netherlands. *Sociology of Education,* 59**,** 237-46.

DE GRAAF, P. M. 1988. Parents' financial and cultural resources, grades, and transition to secondary school in the Federal Republic of Germany. *European Sociological Review,* 4**,** 209-221.

DEARDEN, L., SIBIETA, L. & SYLVA, K. 2011. The socio-economic gradient in early child outcomes: Evidence from the Millennium Cohort Study. *Longitudinal and Life Course Studies,* 2**,** 19-40.

DEMACK, S., DREW, D. & GRIMSLEY, M. 2000. Minding the gap: Ethnic, gender and social class differences in attainment at 16, 1988-95. *Race, Ethnicity and Education,* 3**,** 117-143.

DEPARTMENT FOR EDUCATION 1985. General Certificate of Secondary Education: A general introduction. London: HM Stationery Office.

DEPARTMENT FOR EDUCATION 2010. GCSE and equivalent attainment by pupil characteristics in England 2009/10, Statistical first release 37/2010. London: Department for Education.

DEPARTMENT FOR EDUCATION 2014. School and college-level strategies to raise aspirations of high-achieving disadvantaged pupils to pursue higher education investigation. Research report. London: Department for Education.

DEPARTMENT FOR EDUCATION 2015a. Factors associated with achievement: Key stage 2. Research report. Great Britain: Department for Education.

DEPARTMENT FOR EDUCATION 2015b. Factors associated with achievement: Key stage 4. Research report. Great Britain: Department for Education.

DEPARTMENT FOR EDUCATION 2015c. Measuring disadvantaged pupils' attainment gaps over time (updated). Statistical Working Paper SFR 40/2014. London: Department for Education.

DEPARTMENT FOR EDUCATION 2019. Secondary accountability measures: Guide for maintained secondary schools, academies and free schools. London: Department for Education.

DEPARTMENT FOR EDUCATION AND SCIENCE WELSH OFFICE 1987. The National Curriculum 5-16: A consultation document. London: HM Stationery Office.

DEPARTMENT FOR EDUCATION AND SKILLS & QUALIFICATIONS AND CURRICULUM AUTHORITY 2004. The National Curriculum: Handbook for secondary teachers in England. Nottingham: Department for Education and Skills.

DEVINE, F. 1997. *Social class in America and Britain,* Edinburgh, Edinburgh University Press.

DIMAGGIO, P. 1979. On Pierre Bourdieu. *American Journal of Sociology,* 84**,** 1460-1474.

DIMAGGIO, P. 1982. Cultural capital and school success: The impact of status culture participation on the grades of U.S. high school students. *American Sociological Review,* 47**,** 189-201.

DREW, D. 1995. *"Race", education and work: The statistics of inequality,* Aldershot, Avebury.

DUMAIS, S. A. 2002. Cultural capital, gender, and school success: The role of habitus. *Sociology of Education,* 75**,** 44-68.

DUNCAN, O. D. 1966. Path analysis: Sociological examples. *American Journal of Sociology,* 72**,** 1-16.

EGERTON, M. 1997. Occupational inheritance: The role of cultural capital and gender. *Work, Employment and Society,* 11**,** 263–282.

EITLE, T. M. & EITLE, D. J. 2002. Race, cultural capital, and the educational effects of participation in sports. *Sociology of Education,* 75**,** 123-146.

ELLIOTT, J. & SHEPHERD, P. 2006. Cohort profile: 1970 British Birth Cohort (BCS70). *International Journal of Epidemiology,* 35**,** 836-843.

ERIKSON, R. 1984. Social class of men, women and families. *Sociology,* 18**,** 500-514.

ERIKSON, R. & GOLDTHORPE, J. H. 1992. *The constant flux: A study of class mobility in industrial societies,* Oxford, Clarendon Press.

ERIKSON, R., GOLDTHORPE, J. H. & HÄLLSTEN, M. 2012. No way back up from ratcheting down? A critique of the 'microclass' approach to the analysis of social mobility. *Acta Sociologica,* 55**,** 211-229.

ERIKSON, R., GOLDTHORPE, J. H., JACKSON, M., YAISH, M. & COX, D. R. 2005. On class differentials in educational attainment. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 9730-9733.

ERIKSON, R. & JONSSON, J. O. 1996. Can education be equalized? The Swedish case in comparative perspective. Oxford: Westview Press.

EVANS, G. 1992. Testing the validity of the Goldthorpe class schema. *European Sociological Review,* 8**,** 211-232.

FARKAS, G., GROBE, R. P., SHEEHAN, D. & SHUAN, Y. 1990. Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban school district. *American Sociological Review,* 55**,** 127-142.

FEINSTEIN, L. 2003. Inequality in the early cognitive development of British children in the 1970 cohort. *Economica,* 70**,** 73-97.

FEINSTEIN, L. 2004. Mobility in pupils' cognitive attainment during school life. *Oxford Review of Economic Policy,* 20**,** 213-229.

FIELD, J. 2003. *Social capital,* London, Routledge.

FIENBERG, S. & MASON, W. 1979. Identification and estimation of age-period-cohort models in the analysis of discrete archival data. *Sociological Methodology,* 10**,** 1-67.

FIRTH, D. 2003. Overcoming the reference category problem in the presentation of statistical models. *Sociological Methodology,* 33**,** 1-18.

FLOUD, J. E., ANDERSON, C. A. & HALSEY, A. H. 1961. *Education, economy, and society: A reader in the sociology of education,* New York, Free Press of Glencoe.

FREESE, J. 2007. Replication standards for quantitative social science: Why not Sociology? *Sociological Methods & Research,* 36**,** 153-172.

FREESE, J., LI, J.-C. & WADE, L. 2003. The potential relevances of biology to social inquiry. *Annual Review of Sociology*, 29, 233-256.

FREESE, J. & PETERSON, D. 2017. Replication in Social Science. *Annual Review Sociology,* 43**,** 147-165.

FRIEDMAN, S. & LAURISON, D. 2019. *The class ceiling: Why it pays to be privileged,* Bristol, Policy Press.

FURLONG, A. & BIGGART, A. 1999. Framing 'choices': A longitudinal study of occupational aspirations among 13- to 16-year-olds. *Journal of Education and Work,* 12**,** 21-35.

GALINDO-RUEDA, F., MARCENARO-GUTIERREZ, O. & VIGNOLES, A. 2004. The widening socio-economic gap in UK higher education. *National Institute Economic Review***,** 75-88.

GANZEBOOM, H. 2019. How constant is the Treiman constant? Country-specific variations in the basic similarity of occupational status hierarchies. *Conference Paper. Social Stratification Research Seminar.* Vrije Universiteit Amsterdam, Amsterdam.

GANZEBOOM, H., B. G. 1982. Explaining differential participation in high-cultural activities: A confrontation of information-processing and status seeking theories. *In:* WERNER, R. (ed.) *Theoretical models and empirical analyses.* Utrecht: E.S. Publications.

GANZEBOOM, H. B. G., DE GRAAF, P. M. & TREIMAN, D. J. 1992. A standard international socio-economic index of occupational status. *Social Science Research,* 21**,** 1-56.

GANZEBOOM, H. B. G. & TREIMAN, D. J. 1996. Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research,* 25**,** 201-239.

GAYLE, V. 1996. Modelling tabular data with an ordered outcome. *Sociological Research Online,* 1**,** 1-10.

GAYLE, V. 2005. Youth transitions. *In:* ERMISCH, J. & WRIGHT, R. E. (eds.) *Changing Scotland: Evidence for the British Household Panel Survey.* Bristol: Policy Press.

GAYLE, V., BERRIDGE, D. & DAVIES, R. B. 2003. *Econometric analysis of the demand for higher education,* Nottingham, Department for Education and Skills.

GAYLE, V. & CONNELLY, R. 2017. Analysing and comparing complex social survey data. *Workshop. Analysing and comparing large-scale and complex social survey data: Insights from the analysis of socio-economic inequalities in childhood.* Royal Statistical Society, London.

GAYLE, V., LAMBERT, P. & DAVIES, R. B. 2009. Logistic regression models in sociological research. University of Stirling: Technical Paper 2009-1 of Data Management through e-Social Science Research Node.

GAYLE, V. & LAMBERT, P. S. 2007. Using quasi-variance to communicate sociological results from statistical models. *Sociology,* 41**,** 1191-1208.

GAYLE, V. & LAMBERT, P. S. 2018. *What is quantitative longitudinal data analysis?,* London, Bloomsbury Press.

GAYLE, V., MURRAY, S. & CONNELLY, R. 2014. Young people and school General Certificate of Secondary Education attainment: Looking for the 'missing middle'. *British Journal of Sociology of Education,* 37**,** 1-21.

GLASS, D. 1954. *Social mobility in Britain,* London, Routledge & Kegan Paul.

GOLDTHORPE, J. 2003. The myth of education-based meritocracy. *New Economy,* 10**,** 234-239.

GOLDTHORPE, J. & MARSHALL, G. 1992. The promising future of class analysis: A response to recent critiques. *Sociology,* 26**,** 381.

GOLDTHORPE, J. & MCKNIGHT, A. 2004. The economic basis of social class, CASE Paper 80. London: Centre for Analysis of Social Exclusion, London School of Economics.

GOLDTHORPE, J. H. 1969. *The affluent worker in the class structure,* London, Cambridge University Press.

GOLDTHORPE, J. H. 1983. Women and class analysis: In defence of the conventional view. *Sociology,* 17**,** 465-488.

GOLDTHORPE, J. H. 1984. Women and class analysis: A reply to the replies. *Sociology,* 18**,** 491-499.

GOLDTHORPE, J. H. 2007. "Cultural capital": Some critical observations. *Sociologica,* 2**,** 1-23.

GOLDTHORPE, J. H., LLEWELLYN, C. & PAYNE, C. 1980. *Social mobility and class structure in modern Britain,* Oxford, Clarendon Press.

GOODMAN, A., GREGG, P. & WASHBROOK, E. 2011. Children's educational attainment and the aspirations, attitudes and behaviours of parents and children through childhood. *Longitudinal and Life Course Studies,* 2**,** 1-18.

GOODMAN, R. & BURTON, D. 2012. What is the nature of the achievement gap, why does it persist and are government goals sufficient to create social justice in the education system? *Education 3-13,* 40**,** 500-514.

GORARD, S. & SEE, B. H. 2009. The impact of socio-economic status on participation and attainment in Science. *Studies in Science Education,* 45**,** 93-129.

GORARD, S., SEE, B. H. & DAVIES, P. 2012. The impact of attitudes and aspirations on educational attainment and participation. York: Joseph Rowntree Foundation.

GORARD, S. & SIDDIQUI, N. 2019. How trajectories of disadvantage help explain school attainment. *SAGE Open,* 9**,** 1-14.

GORARD, S. & TAYLOR, C. 2002. Market forces and standards in education: A preliminary consideration. *British Journal of Sociology of Education,* 23**,** 5-18.

GRAETZ, B. 1988. The reproduction of privilege in Australian education. *The British Journal of Sociology,* 39**,** 358-376.

GREAT BRITAIN COMMITTEE ON HIGHER EDUCATION 1963. Higher education: Report of the Committee on Higher Education. London: HM Stationery's Office.

GREGG, P. & WASHBROOK, E. 2011. The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 11. *Longitudinal and Life Course Studies,* 2**,** 41-58.

GRUSKY, D. B. (ed.) 2014. *Social stratification: Class, race, and gender in sociological perspective,* Boulder, CO: Westview Press.

GRUSKY, D. B. & WEEDEN, K. A. 2001. Decomposition without death: A research agenda for a new class analysis. *Acta Sociologica,* 44**,** 203-218.

GUTMAN, L. & AKERMAN, R. 2008. Determinants of aspirations. London: Centre for Research on the Wider Benefits of Learning.

HALSEY, A. H., HEATH, A. F. & RIDGE, J. M. 1980. *Origins and destinations: Family, class, and education in modern Britain,* Oxford, Clarendon Press.

HAMNETT, C., RAMSDEN, M. & BUTLER, T. 2007. Social background, ethnicity, school composition and educational attainment in East London. *Urban Studies,* 44**,** 1255-1280.

HARRIS, K. & SCHORPP, K. 2018. Integrating biomarkers in social stratification and health research. *Annual Review of Sociology*, 44, 361-386.

HARTAS, D. 2016. Young people's educational aspirations: Psychosocial factors and the home environment. *Journal of Youth Studies,* 19**,** 1145-1163.

HAWKES, D. & PLEWIS, I. 2006. Modelling non- response in the National Child Development Study. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 169**,** 479-491.

HEATH, A. & BRITTEN, N. 1984. Women's jobs do make a difference: A reply to Goldthorpe. *Sociology,* 18**,** 475-490.

HEATH, A. F. & CLIFFORD, P. 1990. Class inequalities in education in the twentieth century. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 153**,** 1-16.

HERRNSON, P. 1995. Replication, verification, secondary analysis, and data collection in Political Science. *PS, Political Science and Politics,* 28**,** 452-455.

HILL, T. D., DAVIS, A. P., ROOS, J. M. & FRENCH, M. T. 2019. Limitations of fixed-effects models for panel data. *Sociological Perspectives,* 0**,** 1-13.

HOGGART, R. 1957. *The uses of literacy: Aspects of working-class life, with special reference to publications and entertainments,* London, Chatto and Windus.

HOLM, A. & JAEGER, M. M. 2008. Does relative risk aversion explain educational inequality? A dynamic choice approach. *Research in Social Stratification and Mobility,* 26**,** 199-219.

HOOPER, D., COUGHLAN, J. & MULLEN, M. 2008. Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods,* 6**,** 53-60.

HOPE, K. & GOLDTHORPE, J. H. 1974. *The social grading of occupations: A new approach and scale,* Oxford, Clarendon Press.

HOSKINS, K. & BARKER, B. 2016. Aspirations and young people’s constructions of their futures: Investigating social mobility and social reproduction. *British Journal of Educational Studies,* 65**,** 45-67.

HOUSE OF COMMONS EDUCATION COMMITTEE 2014. Underachievement in education by white working class children. First report of 2014-15. London: HM Stationery Office.

HOUT, M. & DIPRETE, T. A. 2006. What we have learned: RC28's contributions to knowledge about social stratification. *Research in Social Stratification and Mobility,* 24**,** 1-20.

JACKSON, B. 1962. *Education and the working class: Some general themes raised by a study of 88 working-class children in a northern industrial city,* London, Routledge & Kegan Paul.

JACKSON, M. V. 2013. *Determined to succeed? Performance versus choice in educational attainment,* Stanford, California, Stanford University Press.

JAEGER, M. M. 2009. Equal access but unequal outcomes: Cultural capital and educational choice in a meritocratic society. *Social Forces,* 87**,** 1943-1971.

JAEGER, M. M. & BREEN, R. 2016. A dynamic model of cultural reproduction. *American Journal of Sociology,* 121**,** 1079-1115.

JAEGER, M. M. & HOLM, A. 2003. Which background factors matter more in intergenerational educational attainment: Social class, cultural capital or cognitive ability? A random effects approach. *IDEAS Working Paper Series from RePEc.* St. Louis: Federal Reserve Bank of St. Louis.

JAEGER, M. M. & HOLM, A. 2007. Does parents' economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime? *Social Science Research,* 36**,** 719-744.

JANN, B. 2013. COEFPLOT: Stata module to plot regression coefficients and other results. Statistical Software Components S457686, Boston College Department of Economics.

JANZ, N. 2016. Bringing the gold standard into the classroom: Replication in university teaching. *International Studies Perspectives,* 17**,** 392-407.

JENCKS, C. 1973. *Inequality: A reassessment of the effect of family and schooling in America,* London, Allen Lane.

JENKINS, A. & SABATES, R. 2007. The classification of qualifications in social surveys. London: Centre for Longitudinal Studies, Institute of Education, University of London.

JENKINS, R. 1992. *Pierre Bourdieu,* London, Routledge.

JERRIM, J., VIGNOLES, A., LINGAM, R. & FRIEND, A. 2015. The socio-economic gradient in children's reading skills and the role of genetics. *British Educational Research Journal*, 41, 6-29.

JONES, J., JOYCE, M. & THOMAS, J. 2003. Non-employment and labour availability. Bank of England Quarterly Bulletin, Autumn 2003.

JONSSON, J. O. 1987. Class origin, cultural origin, and educational attainment: The case of Sweden. *European Sociological Review,* 3**,** 229-242.

JONSSON, JAN O., GRUSKY, DAVID B., DI CARLO, M., POLLAK, R. & BRINTON, MARY C. 2009. Microclass mobility: Social reproduction in four countries. *American Journal of Sociology,* 114**,** 977-1036.

KALMIJN, M. & KRAAYKAMP, G. 1996. Race, cultural capital, and schooling: An analysis of trends in the United States. *Sociology of Education,* 69**,** 22-34.

KAPLAN, D. 2009. *Structural equation modeling foundations and extensions,* London, SAGE Publications.

KARABEL, J. & HALSEY, A. H. 1977. *Power and ideology in education,* New York, Oxford University Press.

KATSILLIS, J. & RUBINSON, R. 1990. Cultural capital, student achievement, and educational reproduction: The case of Greece. *American Sociological Review,* 55**,** 270-79.

KHATTAB, N. 2015. Students' aspirations, expectations and school achievement: What really matters? *British Educational Research Journal,* 41**,** 731-748.

KIM, J.-O. 1978. *Factor analysis statistical methods and practical issues,* London, SAGE Publications.

KING, G. 1995. Replication, replication. *PS: Political Science and Politics,* 28**,** 541-559.

KNIES, G. E. 2018. Understanding Society The UK Household Longitudinal Study Waves 1-8 User Guide. Colchester, Essex Institute for Social and Economic Research, University of Essex.

KOHLER, U. & KREUTER, F. 2012. *Data analysis using Stata,* College Station, Texas, Stata Press.

KORUPP, S., GANZEBOOM, H. & VAN DER LIPPE, T. 2002. Do mothers matter? A comparison of models of the influence of mothers' and fathers' educational and occupational status on children's educational attainment. *Quality & Quantity,* 36**,** 17-42.

KRAAYKAMP, G. & EIJCK, K. V. 2010. The intergenerational reproduction of cultural capital: A threefold perspective. *Social Forces,* 89**,** 209-231.

KRAPOHL, E., RIMFELD, K., SHAKESHAFT, N., G., TRZASKOWSKI, M., MCMILLAN, A., PINGAULT, J.-B., ASBURY, K., HARLAAR, N., KOVAS, Y., DALE, P., S. & PLOMIN, R. 2014. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences*, 111, 15273- 15278.

LACEY, C. 1970. *Hightown Grammar: The school as a social system,* Manchester, Manchester University Press.

LAMBERT, P., S. 2012. Comparative scaling of educational categories by homogamy – Analysis of UK data from the BHPS. *DAMES Node, Technical Paper 2012-1***,**

http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=37EDDC215C76893F60C3C4BC629C6892?doi=10.1.1.458.9067&rep=rep1&type=pdf.

LAMBERT, P. S. & BIHAGEN, E. 2014. Using occupation-based social classifications. *Work, Employment & Society*, 28, 481-494.

LAMBERT, P. S., CONNELLY, R., BLACKBURN, R. M. & GAYLE, V. (eds.) 2012. *Social stratification: Trends and processes,* Surrey: Ashgate Publishing Company.

LAMBERT, P., S. & GRIFFITHS, D. 2013. Seven new social classes? *Pulling Apart*, Available at: http://pullingapartproject.wordpress.com/2013/04/04/seven-new-social-classes/.

LAMONT, M. & LAREAU, A. 1988. Cultural capital: Allusions, gaps and glissandos in recent theoretical developments. *Sociological Theory,* 6**,** 153-168.

LAREAU, A. & WEININGER, E. 2003. Cultural capital in educational research: A critical assessment. *Theory and Society,* 32**,** 567-606.

LECKIE, G. & GOLDSTEIN, H. 2009. *The limitations of using school league tables to inform school choice,* Bristol, Centre for Market and Public Organisation.

LEE, V. E. & BURKAM, D. T. 2002. *Inequality at the starting gate: Social background differences in achievement as children begin school*, Washington D.C., Economic Policy Institute.

LEWIS-BECK, M. S., BRYMAN, A. & FUTING LIAO, T. 2004. Pseudo-R-Squared in The SAGE encyclopedia of social science research methods. Thousand Oaks, CA: SAGE Publications.

LIU, X. 2014. Fitting stereotype logistic regression models for ordinal response variables in educational research. 13**,** 528-545.

LOCKWOOD, D. 1958. *The blackcoated worker: A study in class consciousness,* London, Allen & Unwin.

LONG, J. S. 1997. *Regression models for categorical and limited dependent variables,* Thousand Oaks, London, SAGE Publications.

LONG, J. S. & FREESE, J. 2014. *Regression models for categorical dependent variables using Stata,* College Station, Texas, Stata Press.

LONGHI, S. & NANDI, A. 2015. *A practical guide to using panel data,* Los Angeles, SAGE Publications.

LUCAS, SAMUEL R. 2001. Effectively maintained inequality: Education transitions, track mobility, and social background effects. *American Journal of Sociology,* 106**,** 1642-1690.

LUNT, M. 2001. Stereotype ordinal regression. *Stata Technical Bulletin,* 61**,** 12-18.

MACHIN, S. & VIGNOLES, A. 2004. Educational inequality: The widening socio-economic gap. *Fiscal Studies,* 25**,** 107-128.

MACHIN, S. & VIGNOLES, A. 2006. Education policy in the UK. *IDEAS Working Paper Series from RePEc.* London: Centre for the Economics of Education, London School of Economics.

MARCENARO-GUTIERREZ, O., GALINDO-RUEDA, F. & VIGNOLES, A. 2007. Who actually goes to university? *Empirical Economics,* 32**,** 333-357.

MARKS, G. N. 2008. Accounting for the gender gaps in student performance in reading and mathematics: Evidence from 31 countries. *Oxford Review of Education,* 34**,** 89-109.

MARSHALL, G. 1988. *Social class in modern Britain,* London, Hutchinson Education.

MCCULLAGH, P. 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B (Methodological),* 42**,** 109-127.

MCCULLOCH, A. 2017. Educational aspirations trajectories in England. *British Journal of Educational Studies,* 65**,** 69-85.

MCKIBBIN, R. 1998. *Classes and cultures England, 1918-1951,* Oxford, Oxford University Press.

MEHMETOGLU, M. & JAKOBSEN, T. G. 2017. *Applied statistics using Stata: A guide for the social sciences,* London, SAGE Publications.

MENARD, S. W. 2002. *Applied logistic regression analysis,* London, SAGE Publications.

MILLS, C. 2014. The Great British class fiasco: A comment on Savage et al. *Sociology,* 48**,** 437-444.

MOOD, C. 2010. Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review,* 26**,** 67-82.

MORRIS, T., DORLING, D. & DAVEY SMITH, G. 2016. How well can we predict educational outcomes? Examining the roles of cognitive ability and social position in educational attainment. *Contemporary Social Science,* 11**,** 154-168.

MUNDLAK, Y. 1978. On the pooling of time series and cross section data. *Econometrica,* 46**,** 69.

MURRAY, S., GAYLE, V. & CONNELLY, R. 2012. Exploring educational attainment between the elite and the NEET:A contemporary analysis of British Household Panel Survey data.

NEISSER, U., BOODOO, G., BOUCHARD, T., BOYKIN, A., BRODY, N., CECI, S., HALPERN, D., LOEHLIN, J., PERLOFF, R., STERNBERG, R. & URBINA, S. 1996. Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77.

O'CONNELL, A., A. 2006. *Logistic regression models for ordinal response variables,* Thousand Oaks, California, SAGE Publications.

OFFICE FOR NATIONAL STATISTICS 2010. Standard Occupational Classification 2010, Volume 3, The National Statistics Socio-Economic Classification: (Rebased on the SOC2010) User Manual. Southampton: Palgrave Macmillan.

OFFICE FOR NATIONAL STATISTICS 2013. Ethnic group: Harmonised concepts and questions for social data sources. London: ONS.

OFFICE FOR STANDARDS IN EDUCATION CHILDREN'S SERVICES AND SKILLS 2019. School inspection update. January 2019. Special edition. Manchester: Ofsted.

PAHL, R. E. 1989. Is the emperor naked? Some questions on the adequacy of sociological theory in urban and regional research. *International Journal of Urban and Regional Research,* 13**,** 709-720.

PAKULSKI, J. 1996. *The death of class,* London, SAGE Publications.

PARKIN, F. 1971. *Class inequality and political order: Social stratification in capitalist and Communist societies*, London, MacGibbon & Kee.

PATERSON, L. & IANNELLI, C. 2007. Social class and educational attainment: A comparative study of England, Wales, and Scotland. *Sociology of Education,* 80**,** 330-358.

PAYNE, G. 2013a. Models of contemporary social class: the Great British Class Survey. *Methodological Innovations Online,* 8**,** 3-17.

PAYNE, G. 2013b. *Social divisions*, Basingstoke, Hampshire : Palgrave Macmillan.

PAYNE, J. 1995a. Qualifications between 16 and 18: A comparison of achievements on routes beyond compulsory schooling. Moorfoot: Employment Department.

PAYNE, J. 1995b. Routes beyond compulsory schooling. Moorfoot: Employment Department.

PAYNE, J. 2001a. Patterns of participation in full-time education after 16: An analysis of the England And Wales Youth Cohort Study. Research report No. 307. Great Britain: Department for Education and Skills.

PAYNE, J. 2001b. Work-based training for young people: Data from the England and Wales Youth Cohort Study. Research report No. 276. Great Britain: Department for Education and Skills.

PETERSON, R. A. & KERN, R. M. 1996. Changing highbrow taste: From snob to omnivore. *American Sociological Review,* 61**,** 900-907.

PETT, M. A., LACKEY, N. R. & SULLIVAN, J. J. 2003. *Making sense of factor analysis,* Thousand Oaks, California, SAGE Publications.

PLATT, L. 2010. *Understanding inequalities: Stratification and difference,* Oxford, Polity Press.

PLATT, L. & PARSON, S. 2018. Occupational aspirations of children from primary school to teenage years across ethnic groups. London: Centre for Longitudinal Studies, UCL Institute of Education and the Runnymede Trust.

PLATT, L. & PARSONS, S. 2017. Is the future female? Educational and occupational aspirations of teenage boys and girls in the UK. Working paper 2017/17. London: Centre for Longitudinal Studies, UCL Institute of Education.

PLAYFORD, C. J. & GAYLE, V. 2015. The concealed middle? An exploration of ordinary young people and school GCSE subject area attainment. *Journal of Youth Studies,* 19**,** 149-168.

POWER, C. & ELLIOTT, J. 2006. Cohort profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology,* 35**,** 34-41.

PRANDY, K. 1999. Class, stratification and inequalities in health: A comparison of the Registrar-General's social classes and the Cambridge Scale. *Sociology of Health & Illness,* 21**,** 466-484.

PRANDY, K. & LAMBERT, P. 2003. Marriage, social distance and the social space: An alternative derivation and validation of the Cambridge Scale. *Sociology,* 37**,** 397-411.

PRIEUR, A. & SAVAGE, M. 2013. Emerging forms of cultural capital. *European Societies,* 15**,** 246-267.

RAFTERY, A. 1995. Bayesian model selection in social research. *Sociological Methodology,* 25**,** 111-195.

RAFTERY, A. E. & HOUT, M. 1993. Maximally maintained inequality: Expansion, reform, and opportunity in Irish education, 1921-75. *Sociology of Education,* 66**,** 41-62.

REAY, D. 2004. Education and cultural capital: The implications of changing trends in education policies. *Cultural Trends,* 13**,** 73-86.

REAY, D. 2017. *Miseducation: Inequality, education and the working classes,* Bristol, Policy Press.

REAY, D., DAVIES, J., DAVID, M. & BALL, S. J. 2001. Choices of degree or degrees of choice? Class, race and the higher education choice process. *Sociology,* 35**,** 855-874.

REEVES, D. J., BOYLE, W. F. & CHRISTIE, T. 2001. The relationship between teacher assessments and pupil attainments in Standard Test Tasks at Key Stage 2, 1996-98. *British Educational Research Journal,* 27**,** 141-160.

RIMFELD, K., KOVAS, Y., DALE, P., S. & PLOMIN, R. 2015. Pleiotropy across academic subjects at the end of compulsory education. *Scientific Reports*, 5: 11713.

ROSCIGNO, V. J. & AINSWORTH-DARNELL, J. W. 1999. Race, cultural capital, and educational resources: Persistent inequalities and achievement returns. *Sociology of Education,* 72**,** 158-178.

ROSE, D. & PEVALIN, D. J. 2003. *A researcher's guide to the National Statistics Socio-economic Classification,* London, SAGE Publications.

ROSE, D. & REILLY, K. 1997. *Constructing classes: Towards a new social classification for the UK,* Swindon, ESRC/ONS.

ROTHON, C. 2007. Can achievement differentials be explained by social class alone? An examination of minority ethnic educational performance in England and Wales at the end of compulsory schooling. *Ethnicities,* 7**,** 306-322.

SAMMONS, P. 1995. Gender, ethnic and socio-economic differences in attainment and progress: A longitudinal analysis of student achievement over 9 years. *British Educational Research Journal,* 21**,** 465-485.

SAMMONS, P. & SMEES, R. 1998. Measuring pupil progress at Key Stage 1: Using baseline assessment to investigate value added. *School Leadership & Management,* 18**,** 389-407.

SAMMONS, P., WEST, A. & HIND, A. 1997. Accounting for variations in pupil attainment at the end of Key Stage 1. *British Educational Research Journal,* 23**,** 489-511.

SAVAGE, M. 2015. *Social class in the 21st century*, London : Pelican, an imprint of Penguin Books.

SAVAGE, M. 2016. The fall and rise of class analysis in British sociology, 1950-2016. *Tempo Social,* 28**,** 57-72.

SAVAGE, M., DEVINE, F., CUNNINGHAM, N., TAYLOR, M., LI, Y., HJELLBREKKE, J., LE ROUX, B., FRIEDMAN, S. & MILES, A. 2013. A new model of social class? Findings from the BBC's Great British Class Survey experiment. *Sociology,* 47**,** 219-250.

SAVAGE, M., WARDE, A. & DEVINE, F. 2005. Capitals, assets, and resources: Some critical issues. *British Journal of Sociology,* 56**,** 31-47.

SCHNEIDER, S. L. 2011. Measuring educational attainment. Survey question bank: Topic overview 6, https://www.ukdataservice.ac.uk/media/262853/discover_sqb_education_schneider.pdf.

SCHREIBER, J. B., NORA, A., STAGE, F. K., BARLOW, E. A. & KING, J. 2006. Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research,* 99**,** 323-338.

SCHRÖDER, H. & GANZEBOOM, H. B. G. 2014. Measuring and modelling level of education in European societies. *European Sociological Review,* 30**,** 119-136.

SCOTT, J. 2004. Family, gender, and educational attainment in Britain: A longitudinal study. *Journal of Comparative Family Studies,* 35**,** 565-589.

SEWELL, W. & SHAH, V. 1968a. Parents' education and children's educational aspirations and achievements. *American Sociological Review,* 33**,** 191-209.

SEWELL, W. & SHAH, V. 1968b. Social class, parental encouragement, and educational aspirations. *American Journal of Sociology,* 73**,** 559-572.

SEWELL, W. H., HALLER, A. O. & PORTES, A. 1969. The educational and early occupational attainment process. *American Sociological Review,* 34**,** 82-92.

SHAKESHAFT, N. G., TRZASKOWSKI, M., MCMILLAN, A., RIMFELD, K., KRAPOHL, E., HAWORTH, C. M. A., DALE, P. S. & PLOMIN, R. 2013. Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLoS ONE*, 8, e80341.

SHAVIT, Y. & BLOSSFELD, H.-P. 1993. *Persistent inequality: Changing educational attainment in thirteen countries,* Oxford, Westview Press.

SMITH, E. 2015. Can higher education compensate for society? Modelling the determinants of academic success at university. *British Journal of Sociology of Education,* 37**,** 1-23.

SMITHSON, M. 2003. *Confidence intervals,* Thousand Oaks, California, SAGE Publications.

SORENSEN, A. 1994. Women, family and class. *Annual Review of Sociology,* 20**,** 27-47.

ST CLAIR, R. & BENJAMIN, A. 2011. Performing desires: The dilemma of aspirations and educational attainment. *British Educational Research Journal,* 37**,** 501-517.

STACEY, M. 1969. *Comparability in social research,* London, Heinemann Educational Books for the British Sociological Association and the Social Science Research Council.

STANWORTH, M. 1984. Women and class analysis: A reply to John Goldthorpe. *Sociology,* 18**,** 159-170.

STATACORP. 2017a. *Stata 15 Base Reference Manual*. College Station, Texas: Stata Press.

STATACORP. 2017b. *Stata Statistical Software: Release 15*. College Station, Texas: StataCorp LLC.

STEELE, F., VIGNOLES, A. & JENKINS, A. 2007. The effect of school resources on pupil attainment: A multilevel simultaneous equation modelling approach. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 170**,** 801-824.

STEWART, A., PRANDY, K. & BLACKBURN, R. M. 1973. Measuring the class structure. *Nature,* 245**,** 415-417.

STEWART, A., PRANDY, K. & BLACKBURN, R. M. 1980. *Social stratification and occupations,* London, Macmillan.

STRAND, S. 1999. Ethnic group, sex and economic disadvantage: Associations with pupils' educational progress from baseline to the end of Key Stage 1. *British Educational Research Journal,* 25**,** 179-202.

STRAND, S. 2014a. Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'getting it' for the white working class. *Research Papers in Education,* 29**,** 131-171.

STRAND, S. 2014b. School effects and ethnic, gender and socio- economic gaps in educational achievement at age 11. *Oxford Review of Education,* 40**,** 223-245.

SULLIVAN, A. 2001. Cultural capital and educational attainment. *Sociology,* 35**,** 893-912.

SULLIVAN, A. 2002. Bourdieu and education: How useful is Bourdieu's theory for researchers? *Netherlands Journal of Social Sciences,* 38**,** 144-166.

SULLIVAN, A. 2009. Academic self-concept, gender and single-sex schooling. *British Educational Research Journal,* 35**,** 259-288.

SULLIVAN, A. & BROWN, M. 2015. Reading for pleasure and progress in vocabulary and mathematics. *British Educational Research Journal,* 41**,** 971-991.

SULLIVAN, A., HEATH, A. & ROTHON, C. 2011. Equalisation or inflation? Social class and gender differentials in England and Wales. *Oxford Review of Education,* 37**,** 215-240.

SULLIVAN, A., KETENDE, S. & JOSHI, H. 2013. Social class and inequalities in early cognitive scores. *Sociology,* 47**,** 1187-1206.

SZRETER, S. 1984. The genesis of the Registrar-General's social classification of occupations. *The British Journal of Sociology,* 35**,** 522-546.

TAYLOR, C. 2018. The reliability of free school meal eligibility as a measure of socio-economic disadvantage: Evidence from the Millennium Cohort Study in Wales. *British Journal of Educational Studies,* 66**,** 29-51.

TAYLOR, M. F. E., WITH BRICE, J., BUCK, N. & PRENTICE-LANE, E. 2010. British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices. Colchester: University of Essex.

TREANOR, M. 2017. Can we put the 'poverty of aspirations' myth to bed now? Research Briefing 91. University of Edinburgh: Centre for Research on Families and Relationships.

TREIMAN, D. J. 1977. *Occupational prestige in comparative perspective,* New York, Academic Press.

TREIMAN, D. J. 2009. *Quantitative data analysis doing social research to test ideas,* San Francisco, California, Jossey-Bass.

UEBERSAX, J. S. 2015. The tetrachoric and polychoric correlation coefficients. Statistical methods for rater agreement.

UNESCO 2012. International Standard Classification of Education: ISCED 2011. Montreal, QC, Canada: UNESCO Institute for Statistics.

UNIVERSITY OF ESSEX 2015. Linked Understanding Society - National Pupil Database Wave 1 Linkage User Manual Colchester, Essex: Institute for Social and Economic Research, University of Essex.

VAN DE WERFHORST, H. G. & ANDERSEN, R. 2005. Social background, credential inflation and educational strategies. *Acta Sociologica,* 48**,** 321-340.

VUONG, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica,* 57**,** 307-333.

WADSWORTH, M., KUH, D., RICHARDS, M. & HARDY, R. 2006. Cohort profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology,* 35**,** 49-54.

WEEDEN, K. A. & GRUSKY, D. B. 2004. Are there any big classes at all? *Research in Social Stratification and Mobility,* 22**,** 3-56.

WHETTON, C. 2009. A brief history of a testing time: National Curriculum assessment in England 1989-2008. *Educational Research,* 51**,** 137-159.

WILLIAMS, R. 2009. Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research,* 37**,** 531-559.

WILLIAMS, R. 2016. Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology,* 40**,** 7-20.

WILLIS, P. E. 1977. *Learning to labour: How working class kids get working class jobs,* Aldershot, Ashgate.

WINTERTON, M. T. & IRWIN, S. 2012. Teenage expectations of going to university: The ebb and flow of influences from 14 to 18. *Journal of Youth Studies,* 15**,** 858-874.

WITHEY, P. & TURNER, S. 2015. The analysis of SATS results as a measure of pupil progress across educational transitions. *Educational Review,* 67**,** 23-34.

WOLFE, R. 1998. OCRATIO: Stata modules to fit continuation-ratio models on ordinal response data. Statistical Software Components S334801: Boston College Department of Economics.

WRIGHT, S. 1960. Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics,* 16**,** 189-202.

WYSE, D. & TORRANCE, H. 2009. The development and consequences of National Curriculum assessment for primary education in England. *Educational Research,* 51**,** 213-228.

YANG, M. & WOODHOUSE, G. 2001. Progress from GCSE to A and AS Level: Institutional and gender differences, and trends over time. *British Educational Research Journal,* 27**,** 245-267.

YUAN, K.-H. 2005. Fit indices versus test statistics. *Multivariate Behavioral Research,* 40**,** 115-148.