Uncovering important patterns of subject level school qualifications

A latent variable analysis of multiple cohorts of school data for England and Wales

Routledge
Taylor & Francis Group

Educational Research and Evaluation

An International Journal on Theory and Practice

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/nere20

Uncovering important patterns of subject level school qualifications: a latent variable analysis of multiple cohorts of school data for England and Wales

Christopher James Playford, Vernon Gayle & Roxanne Connelly

To cite this article: Christopher James Playford, Vernon Gayle & Roxanne Connelly (26 Nov 2024): Uncovering important patterns of subject level school qualifications: a latent variable analysis of multiple cohorts of school data for England and Wales, Educational Research and Evaluation, DOI: 10.1080/13803611.2024.2433230

To link to this article: https://doi.org/10.1080/13803611.2024.2433230

EDUCATIONAL RESEARCH AND EVALUATION
https://doi.org/10.1080/13803611.2024.2433230

Routledge
Taylor & Francis Group

Uncovering important patterns of subject level school qualifications: a latent variable analysis of multiple cohorts of school data for England and Wales

Christopher James Playford [a], Vernon Gayle [b] and Roxanne Connelly [b]

[a]Department of Social and Political Sciences, Philosophy, and Anthropology, University of Exeter, Exeter, UK; [b]School of Social and Political Science, University of Edinburgh, Edinburgh, UK

ABSTRACT

General Certificates of Secondary Education (GCSEs) are the main school qualifications in England and Wales. Analysing GCSE outcomes is challenging because pupils study many subjects, each subject is awarded an individual grade, and there is no common or compulsory set of subjects. In this study we investigate inequalities in pupils' highly individualised school GCSE profiles using latent variable models, with data from the Youth Cohort Study of England and Wales (YCS). Four latent educational groups were identified. The identification of two distinctive groups with moderate levels of GCSE outcomes but different GCSE profiles, especially in science subjects, is an important new finding. The latent variable approach remains suitable for the numeric GCSE grading scheme (i.e., grades 1–9) first implemented in 2017, because the numerical grades are also discrete ordered categories. The approach is more broadly applicable to other educational systems where there is no compulsory set of subjects.

ARTICLE HISTORY
Received 18 March 2023
Accepted 18 November 2024

KEYWORDS
Educational attainment; GCSE; sociology of youth; youth cohort study of England and Wales; latent class models

Introduction

In England and Wales, despite social and educational change, the educational qualifications that are gained at school continue to play an important role in young people's lives. Young people's outcomes in school qualifications continue to be stratified by gender and family socioeconomic circumstances (Bramley et al., 2015; Burgess et al., 2004; Connolly, 2006; Gayle et al., 2003; Gayle et al., 2016; Sammons et al., 2014; Strand, 2014; Sullivan et al., 2011; Warrington & Younger, 2000; Younger & Warrington, 2005). Quantifying school qualifications is a complex endeavour because there is no single

CONTACT Christopher James Playford c.j.playford@exeter.ac.uk; @playford_chris
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13803611.2024.2433230.

agreed upon way to measure school qualifications (Connelly et al., 2016; Prandy et al., 2004). Quantifying school qualifications in England and Wales is particularly complicated. This is because the General Certificate of Secondary Education (GCSE) is awarded for multiple subjects, a wide range of subjects are available, and historically pupils received an alphabetical grade for each GCSE subject studied. This results in pupils having highly individualised GCSE outcome profiles.

In an earlier study of a single school cohort, Playford and Gayle (2016) reported a pioneering alternative methodological approach using latent variable models as a parsimonious method of addressing the complexity of school GCSE outcomes. This approach identified two latent groups of young people with "middle" levels of overall attainment differentiated by their outcomes in GCSE science subjects. Membership of these groups was heavily stratified by sex and ethnicity. These findings were masked by more conventional methods of measuring school GCSE outcomes. Playford et al. (2016) replicated this approach using data on Scottish pupils.

There is increasing concern across a wide range of academic disciplines that empirical results cannot reliably be reproduced (Baker, 2016). This study is a genuine attempt to test the robustness and generalisability of the findings reported in Playford and Gayle (2016). In this paper we assay the benefits of the latent variable modelling approach. We extend the single cohort analyses in Playford and Gayle (2016) through an examination of pooled school-year cohorts in the decade following the introduction of GCSEs.

Context

GCSEs are the standard qualifications that are undertaken by pupils in Year 11 (age 15-16) in England and Wales since the late 1980s[1] (Department for Education, 1985; Mobley et al., 1986; North, 1987). GCSEs have historically been a mixture of assessed coursework and examinations (see Ashford et al., 1993). Generally, each subject is assessed separately and a subject specific GCSE is awarded. It is usual for pupils to study about nine subjects (Jin et al., 2011). GCSEs are graded into discrete ordered categories, historically the highest being A, and the lowest G. From 1994 a higher grade of A* was introduced[2] (Yang & Woodhouse, 2001).

The GCSE subjects offered by individual schools vary greatly. Whilst the National Curriculum stipulates that schools must make courses available which enable pupils to study in a range of areas, there is no requirement that pupils must undertake specific GCSE qualifications.[3] English, Mathematics and Science subjects are a "core" component of the National Curriculum and are offered by all schools (Jin et al., 2011). The provision of GCSE Science subjects differs between schools. Some schools offer Science as a double award GCSE (worth 2 GCSEs), others offer 3 separate GCSEs in Biology, Chemistry and Physics.

Historically, there has been substantial variation in the patterns of optional or "non-core" subjects studied (Benn & Chitty, 1997; Salisbury et al., 1999). Geography and History are provided by nearly all schools, whereas subjects such as Sociology are far less common (see Claessen, 2005). Pupils were given a large degree of choice over which subjects are studied within the constraints of the subjects offered by their school. Consequently, there are inevitably a large variety of possible patterns of GCSE outcomes. Carroll and Gill (2017) identified from the National Pupil Database that there were a total of 49 different GCSE subjects studied by at least 1% of pupils. Therefore, attempts to comprehensively compare pupils' highly individualised GCSE profiles require a statistically informed method of simplification.

In addition to variation in the subjects studied by pupils, both the overall number of grades A*-C attained, and gaining an A*-C grade in particular GCSE subjects, have been consequential for progression in the education system (Abrahams, 2018). Admission to General Certificate of Education A' level courses, the usual qualifications required to enter university (UCAS, 2016), has historically been conditional on gaining 5 or more GCSEs at grades A*-C, often including English and Mathematics (Banerjee, 2017). For many subjects (particularly sciences and languages) students are often also required to have an A*-C grade in a relevant subject at GCSE. Gaining an A*-C grade in GCSE English and Mathematics is also important to employers as evidence of literacy and numeracy (Wolf, 2011). A wide range of vocational courses also require at least a C grade in GCSE English and Maths (UCAS, 2014). Therefore, overall attainment in GCSE qualifications, subject choice, and attainment in particular GCSE subjects all have marked implications for a young person's future educational and employment opportunities (Jin et al., 2011).

Measuring school GCSE outcomes

Historically, the attainment of five or more GCSEs at grades A*–C is a standard benchmark, for example in official performance league tables (Leckie & Goldstein, 2009). This measure has been used widely in social science research (e.g., Babb, 2005; Connolly, 2006; Gayle et al., 2003; Sullivan et al., 2011; Tunstall, 2011). The main limitation of this measure is that it treats an A* in music, a B in physics and a C in sociology equally when determining whether or not a pupil has five GCSEs at grades A*–C (Gorard & Taylor, 2002). More recently, Government league tables have included a measure of the percentage of pupils gaining five or more GCSEs at grades A*–C including Maths and English (Leckie & Goldstein, 2017).

Alternative measures of pupils' GCSE attainment have been constructed to summarise aggregate or "agglomerate" performance. Some studies have simply used the number of GCSE passes at grades A*-C (see Connelly et al., 2013; Gayle et al., 2016; Schmitt & Wadsworth, 2006; Shakeshaft et al., 2013). Other studies convert

the grades attained in each GCSE into a total score. The total score for a pupil is the sum of the scores for all GCSE qualifications gained (for example, see Croxford et al., 2007; Haque & Bell, 2001; Lessof et al., 2018). The benchmark of 5+ GCSEs, the number of GCSEs, and points scores, all provide "agglomerate" or overall measures of school GCSE outcomes. However, they provide scant information on subject-specific patterns of performance.

There has been a growing interest in examining the choice of and participation in non-core GCSE subjects by the gender and social background of pupils (Bramley et al., 2015; Davies et al., 2008; McMullin & Kulic, 2016; Sullivan et al., 2010). These studies make useful contributions to understanding inequalities in school qualifications but provide fewer insights into the heterogeneous patterns of school GCSE outcomes. Further research has begun to investigate combinations or diets of GCSE subjects but have tended to focus more narrowly on the English Baccalaureate (EBacc) subjects[4] (see Anders et al., 2018; Henderson et al., 2018; Jin et al., 2011).

The central methodological aim of this paper is to go beyond analyses of either "agglomerate" GCSE outcomes or the study of groups of GCSE subjects (which are chosen a priori) and to empirically investigate the highly individualised profiles of school GCSE outcomes. The latent variable approach offers a parsimonious method which provides useful substantive insights into the underlying complexity of school GCSE outcomes.

Materials and methods

Latent variable models

Latent class models estimate the relationship between a set of observed (usually categorical) variables and a latent, or unmeasured, set of classes (Becker & Yang, 1998; Goodman, 2002; McCutcheon, 1987). Latent class models are particularly well-suited for analysing multiple categorical observed measures that are highly interrelated[5] (McCutcheon, 2002). The typical model building strategy involves the data analyst estimating a series of models with increasing numbers of latent classes and evaluating the goodness of fit of each model (see Nylund et al., 2007). A fundamental issue is choosing a model with an appropriate number of latent classes, which parsimoniously summarise the observed data (McCutcheon, 1987). In this study, the final model selected shows a number of latent classes, each of which summarises a pattern of subject-specific GCSE outcomes.

Data

The Youth Cohort Study of England and Wales (YCS) was a major longitudinal study that began in the mid-1980s. It was a large-scale nationally representative

survey funded by the government and was designed to monitor the behaviour of young people as they reached the minimum school leaving age and either remained in education or entered the labour market. The YCS has been successfully used to explore the relationship between parental and family backgrounds and filial attainment (Connolly, 2006; Demack et al., 2000; Drew, 1995; Drew et al., 1992; Gayle et al., 2003; Gayle et al., 2009; Gayle et al., 2016; Sullivan et al., 2011).

There are a number of challenges associated with analysing YCS data, most notably inadequate documentation (see Croxford, 2006b), which we have worked hard to overcome. A key attraction of the YCS is that it collected a broad range of educational information, including detailed data on GCSEs at the resolution of individual academic subjects. The YCS data on GCSEs was collected contemporaneously and is of a high-quality. A further attraction of the YCS is that it includes data on young people's personal circumstances and family background that is not available in official or administrative educational records (such as the National Pupil Database).

An important feature of the present work is that we pool multiple cohorts of the YCS spanning the 1990s, which was the first decade after GCSEs were introduced. The 1990s are a salient period for sociological inquiry because of the changes in the economic circumstances in which young people grew up (Furlong & Cartmel, 2007; Gayle et al., 2009). The British birth cohort studies have provided detailed data on young people (see Cave & von Stumm, 2021). A gap in data provision exists, and there is no birth cohort data for young people growing up in the 1990s. The Youth Cohort Study of England and Wales partially fills this gap (Gayle, 2005). The YCS provides an important data resource for education and youth research. The measures included in the survey and the observational window, render the YCS superior to all other existing UK data resources for testing the latent variable approach.[6]

In the current analysis we construct a dataset that pools YCS Cohorts 5 - 11. This provides data from Year 11 pupils in 1990, 1992, 1993, 1995, 1997, 1999 and 2001 (Courtenay, 1996a, 1996b, 2000; Fitzgerald, 2004; Johnson et al., 2006; Nice et al., 2002; Russell et al., 2004). We focus on young people who attended state funded comprehensive schools in Year 11, these are non-selective non-fee-paying schools. The YCS does not contain school-level or Local Authority-level identifiers because it is a nationally representative survey and pupils are not clustered within schools (see Croxford, 2006a). Indeed, previous studies have identified that school level variance is a minor component in understanding educational outcomes for young people, with the large majority of variation between schools explained by the composition of the school (Rasbash et al., 2010; Wilkinson et al., 2018).

A set of standard survey weights are deposited with the YCS, we used these weights to provide additional control for survey non-response (Jarvis et al., 2003). The pooled dataset provided an analytical sample of 67,937.

Results

GCSE subject outcomes

There are a large number of GCSE subjects available to pupils in England and Wales, some of which are studied by very few pupils. Within the YCS data there are approximately 80 different GCSE subjects listed. The potential number of combinations of GCSE subjects studied is also very large. Some level of subject grouping is therefore essential. To operationalise the analysis, the GCSE subjects taken by pupils has been simplified into 5 main groups based on the 17 most widely studied GCSEs (see Table 1).[7]

The rationale for this particular grouping was that English, Maths and Science were core subjects studied by nearly all pupils (Gill, 2012). Geography, History and RE are popular subject choices and were grouped into "Humanities" (for an indication of the popularity of different GCSE subjects, see Gill, 2012). All of the remaining GCSE subjects were categorised as "Other subjects" because these subjects were far less popular. An alternative subject grouping which includes the additional category of foreign languages is reported in the online supplement (see Table A2). This reduced the analytical sample and did not substantively alter the conclusions reached. A further benefit of using the 5 subject grouping is that it enabled comparison with Playford and Gayle (2016) which was an objective of this article.

The range of grades that might be awarded for each GCSE subject (A*-G) was categorised according to whether a pupil gains an A*-C (i.e., a good pass) or a D-G grade (i.e., a lower pass). This categorisation was chosen because of the importance of gaining an A*-C pass (as previously described) and because this categorisation is widely used by those in education, research, official statistics and by employers. Gaining a GCSE pass at grades C or above is a key grade threshold and is highly consequential for young people's future routes (see Jerrim, 2023).

The overall information for the latent class models is reported in Table 2. The models analyse the 5 GCSE subject grouping (described above). The GCSE outcome is whether a pupil is awarded an A*-C or a D-G grade in each

Table 1. Year 11 school GCSE subject areas, YCS cohorts 5–11.

Subject Groups	GCSEs	Number of pupils gaining A*-C award	Percentage of pupils gaining A*-C award
English	English	51,668	76%
Maths	Mathematics	40,416	59%
Science	Biology; Physics; Chemistry; Double Science; Other Sciences	41,577	61%
Humanity	History; Geography; Other Humanity; Religious Education	44,277	65%
Other Subject	French; Craft Design and Technology; Other Language; Arts; Physical Education; Other GCSEs	52,112	77%
TOTAL		67,937	

Note: The 17 most frequently undertaken GCSEs; Unweighted data.

Table 2. Latent educational group models (goodness of fit statistics).

Model	χ^2	Deviance	DF	Entropy R-squared	AIC	BIC	Adj. BIC
2 Group	5784.3	333101.5	20	0.83	5806.3	5906.7	5871.8
3 Group	1080.6	328397.8	14	0.71	1114.6	1269.8	1215.7
4 Group	19.8	327337	8	0.7	65.8	275.7	202.6
5 Group	3.8	327321	2	0.69	61.8	326.5	234.3

n = 67,937; Unweighted data, YCS Cohorts 5-11.

subject. Following the convention adopted in Playford and Gayle (2016), we use the term "latent educational group" rather than "latent class" to avoid any potential confusion between latent classes, social classes, or school classes.

The goodness of fit measures reported in Table 2 indicate that the 4 group model fits the data better than the 2 or 3 group models because of the lower chi-square value and model deviance. Whilst the 5 group model is a better overall fit than the 4 group model (based on chi-square and deviance statistics) this information should be balanced against the lack of parsimony in the model (measured by the BIC and adjusted BIC statistics). The 4 group model is therefore preferred, as the model of best fit (for further discussion, see Nylund et al., 2007).

Table 3 reports the estimated probabilities of latent group membership for the 4 group model. The model reported in Table 3 was based on 67,937 observations which included pupils with complete data for all of the analytical variables. The prior probabilities of group membership relate to the probability of a randomly selected individual belonging to a particular latent group (Bartholomew et al., 2008, p. 273). The posterior probabilities are the probabilities that an

Table 3. Latent Group Model Results (Four Group Model).

Latent group	1. Poor Grades	2. Science	3. Non-Science	4. Good Grades
Prior Probabilities	*19%*	*8%*	*21%*	*52%*
Estimated Group Membership (posterior probabilities)	*20%*	*4%*	*19%*	*57%*
English A*-C	18%	49%	85%	99%
English D-G	82%	51%	15%	1%
Maths A*-C	4%	60%	26%	93%
Maths D-G	96%	40%	74%	7%
Science A*-C	5%	79%	23%	95%
Science D-G	95%	21%	77%	5%
Humanity A*-C	7%	43%	54%	95%
Humanity D-G	93%	57%	46%	5%
Other A*-C	30%	64%	78%	96%
Other D-G	70%	36%	22%	4%
n (based on posterior probabilities)	*13,595*	*2,934*	*12,602*	*38,806*

School GCSE Attainment – posterior probabilities and prior probabilities (percentages), YCS Cohorts 5-11.
All pupils gaining a GCSE pass at grades A*-G; n = 67,937.
Probabilities reported as percentages.

individual pupil with a particular pattern of GCSE subject-specific outcomes (i.e., their GCSE profile) belongs to a latent educational group, using modal assignment (see Bartholomew et al., 2008, p. 275).

For each latent educational group, the probability of pupils within each group gaining an A*-C or D-G GCSE pass in each of the subject groupings is reported below the Estimated Group Membership probabilities. Twenty percent of pupils are assigned to latent group 1 (based on the posterior probability). Pupils within this group have low probabilities of gaining A*-C grades in all 5 subject groups, particularly in Maths (4%) and Science (5%). We have ascribed this group the label "Poor Grades". In contrast, pupils in the latent group 4 have the highest probabilities of A*–C passes and have been labelled as having "Good Grades." Fifty seven percent of pupils fall into this group.

A defining feature of the other two latent groups is their performance in Maths and in Science. Pupils in latent group 2 have a 79% chance of an A*-C grade in Science, a 64% chance of passing "Other" GCSEs at grades A*-C and only a 43% chance of passing a humanities GCSE at grades A*–C. This obvious orientation towards Science GCSEs motivates us to label this latent educational group as "Science".

Pupils in latent group 3 have an 85% chance of attaining an A*–C in English, a 78% chance of attaining an A*–C in "Other Subjects" and 54% chance of attaining an A*–C in a humanities GCSE. In sharp contrast they only have a 26% chance of attaining an A*–C in Maths and a 23% chance in Science. This comparatively poor performance in Maths and Science GCSEs motivates us to label this latent educational group as "Non-Science". The benefit of the latent variable method is that reveals patterns of GCSE subject-specific outcomes that would be occluded by agglomerate measures.

We undertook a sensitivity analysis to assess the potential effects of missing data. The primary analyses were undertaken on 67,937 pupils, who had complete information on their school GCSE outcomes and the set of explanatory variables (e.g., gender, ethnicity and parental social class etc.). There were an additional 24,145 cases with incomplete information because of missing data for the explanatory variables. In order to assess the potential effect of missing data we estimated the latent variable models on the full set of 92,082 pupils (67,937 + 24,145).[8] This showed high levels of consistency with the model reported in Table 3. As a further robustness check, the latent variable models were also estimated separately for each of YCS cohorts 5–11. This sensitivity analysis confirmed that there was suitably strong consistency and verified that the results were "measurement invariant" (i.e., the latent groups were consistent across individual YCS cohorts).

Overall, these findings resonate with the results presented in Playford and Gayle (2016). The probabilities of pupils within each latent group gaining a GCSE pass in each subject grouping was remarkably consistent with Playford and Gayle (2016). There were small differences in the size of the 4 latent

Table 4. Agglomerate measures of school GCSE attainment by latent group.

Latent group	1. Poor Grades	2. Science	3. Non-Science	4. Good Grades	All
Mean Number of A*-C Passes	0.5	3.5	3.6	8.1	4.9
Mean Number of A*-F Passes	7.1	8.2	8.4	9.1	8.4
Mean GCSE Points Score (A)	24.2	35.1	36.3	51.7	40.2
Mean Number of GCSEs Studied (B)	7.9	8.1	8.4	8.7	8.4
Mean Points Score per GCSE Studied (A/B)	3.1	4.3	4.3	5.9	4.8
Grade of Mean Points Score per GCSE Studied	E	D	D	B	C

All pupils gaining a GCSE pass at grades A*-G, n = 67,937, weighted data, YCS Cohorts 5-11.

groups.[9] Playford and Gayle (2016) identified that in 1992, 23% of pupils were assigned to the "Poor Grades" group, 8% in the "Science" group, 18% in the "Arts" group, and 51% in the "Good Grades" group.

Characteristics of latent educational groups

The agglomerate measures for each latent educational group are reported in Table 4. The "poor grades" group consistently record the lowest overall GCSE attainment on all agglomerate measures. The "good grades" group consistently have the highest overall GCSE attainment on all agglomerate measures. Latent groups 2 and 3 ("Science" and "Non-Science") have moderate or middling levels of agglomerate GCSE attainment. An important empirical finding is that these two latent educational groups would not be identified through an analysis of agglomerate measures.

Table 5 reports a wider set of school GCSE measures for each latent educational group. The attainment of pupils in the "poor grades" and "good grades" groups is unsurprising. Pupils in the "Science" and "Non-Science" latent groups differ in their chances of gaining A*-C grades in GCSE English and GCSE Maths. Fifty seven percent of pupils in the "Science" latent group gain an A*-C pass in Maths but only 14% in English. In contrast, 87% of the "Non-Science" group pass English at this level but 21% gain an A*-C pass in Maths. Both latent group 2 ("Science") and latent group 3 ("Non-Science") have low probabilities of gaining an A*-C pass in both GCSE English and

Table 5. Benchmark school gcse attainment by latent group (column percentages).

Latent group	1. Poor Grades	2. Science	3. Non-Science	4. Good Grades	All
GCSE English (Grade A*-C)	12%	14%	87%	98%	68%
GCSE Maths (Grades A*-C)	2%	57%	21%	92%	52%
Both GCSE English and Maths (Grades A*-C)	0%	14%	15%	90%	47%
5+ A*-C Passes (any subject)	0%	17%	23%	98%	52%
5+ A*-C Passes (including English & Maths)	0%	4%	4%	88%	43%

All pupils gaining a GCSE pass at grades A*-G, n = 67,937, weighted data, YCS Cohorts 5-11.

Maths (14% and 15% respectively). About 4% of the "Science" group and 4% of the "Non-Science" group gain 5 or more A*-C passes including English and Maths. This suggests that outcomes in GCSE English for the "Science" group and outcomes in GCSE Maths for the "Non-Science" group are preventing pupils in these groups gaining higher levels of agglomerate GCSE attainment.

Predicting latent group membership

The next part of the analysis explores the characteristics of pupils in each latent group. We develop a two-phase mixture model, which is also known as the three-step approach (Bolck et al., 2004; Vermunt, 2010). Step one is the estimation of latent groups, step two is the assignment of pupils to latent educational groups (based on their observed pattern of GCSE outcomes), and step 3 estimates the effects of a set of (auxiliary) explanatory variables on latent group membership.[10] This approach is illustrated in Figure 1.

The results of the multinomial logistic regression model predicting latent group membership are reported in Table 6. The log odds reported are relative to being a member of the "good grades" latent group. A number of key features are observed. Pupils have lower log odds of being in any other latent group than the "good grades" group for more recent cohorts than in 1990. This reflects and controls for a pattern of rising GCSE attainment over time (Farquharson et al., 2022). The log odds of being in the "poor grades" group follows a conventional pattern of educational disadvantage. Pupils with parents in lower supervisory

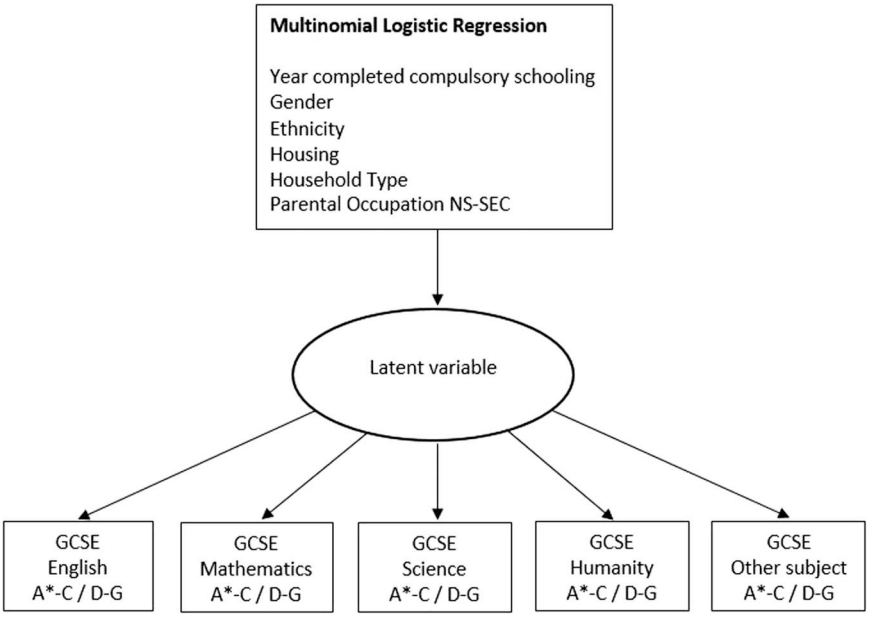

Figure 1. General schema of Latent Educational Group model and Multinomial Logistic Regression model predicting latent group membership.

Table 6. Multinomial logistic regression model results – latent educational group membership (modal assignment).

Year completed compulsory schooling	"Good Grades" Coefficient	"Poor Grades"			"Science"			"Non-Science"		
		Coefficient	SE		Coefficient	SE		Coefficient	SE	
1990	0.00	0.00	0.00		0.00	0.00		0.00	0.00	
1992	0.00	−0.07	0.04		−0.16	0.07	*	−0.03	0.04	
1993	0.00	−0.25	0.04	***	−0.33	0.08	***	−0.18	0.04	***
1995	0.00	−0.32	0.04	***	−0.43	0.08	***	−0.30	0.04	***
1997	0.00	−0.48	0.04	***	−0.56	0.08	***	−0.42	0.04	***
1999	0.00	−0.55	0.05	***	−0.55	0.09	***	−0.44	0.05	***
2001	0.00	−0.74	0.05	***	−0.69	0.08	***	−0.50	0.04	***
Gender										
Female	0.00	0.00	0.00		0.00	0.00		0.00	0.00	
Male	0.00	0.54	0.02	***	1.46	0.05	***	−0.23	0.02	***
Ethnicity										
White	0.00	0.00	0.00		0.00	0.00		0.00	0.00	
Black	0.00	0.21	0.11	*	−0.07	0.22		0.57	0.09	***
Indian	0.00	−0.35	0.08	***	−0.14	0.14		−0.16	0.07	*
Pakistani	0.00	0.25	0.10	*	−0.02	0.19		0.36	0.10	***
Bangladeshi	0.00	−0.32	0.20		−0.19	0.32		0.18	0.18	
Other Asian	0.00	−0.98	0.14	***	−0.51	0.24	*	−0.62	0.13	***
Other Ethnicity	0.00	−0.10	0.14		−0.13	0.23		0.24	0.11	*
Housing tenure										
Owned	0.00	0.00	0.00		0.00	0.00		0.00	0.00	
Rented	0.00	0.94	0.03	***	0.55	0.07	***	0.60	0.04	***
Other Housing	0.00	0.27	0.10	**	0.63	0.16	***	0.19	0.10	
Household Type										
Lives with both parents	0.00	0.00	0.00		0.00	0.00		0.00	0.00	
Only lives with Mother	0.00	0.03	0.04		0.09	0.07		0.08	0.04	*
Only lives with Father	0.00	0.48	0.07	***	0.46	0.12	***	0.39	0.07	***
Other Household	0.00	0.53	0.10	***	0.19	0.20		0.44	0.10	***
Parental NS-SEC										
1.1 Large employers and higher managerial	0.00	0.56	0.07	***	0.15	0.12		0.44	0.06	***
1.2 Higher professional	0.00	0.00	0.00		0.00	0.00		0.00	0.00	
2 Lower managerial and professional	0.00	0.78	0.06	***	0.34	0.09	***	0.52	0.05	***
3 Intermediate	0.00	1.24	0.06	***	0.70	0.09	***	0.86	0.05	***
4 Small employers and own account	0.00	1.58	0.06	***	1.06	0.09	***	1.01	0.05	***
5 Lower supervisory and technical	0.00	1.89	0.07	***	1.11	0.11	***	1.17	0.06	***
6 Semi-routine	0.00	2.10	0.06	***	1.24	0.09	***	1.34	0.05	***
7 Routine	0.00	2.28	0.07	***	1.31	0.11	***	1.37	0.06	***
Constant	0.00	−1.91	0.06	***	−3.61	0.10	***	−1.45	0.05	***
Observations	67937									
AIC	136984.4				Log Likelihood			−68414.2		
BIC	137696.2				Pseudo R Squared			0.067		

For raw percentages of students assigned to latent groups, see Table A5 in the online supplement.
*$p<.05$.
** $p<.01$.
*** $p<.001$.

and technical occupations (NS-SEC 5), semi-routine occupations (NS-SEC 6) and routine occupations (NS-SEC 7) have higher log odds of being assigned to the "poor grades" latent group. Pupils who live in rented accommodation or those who live in a household which does not include both their father and mother have higher log odds of being in the "poor grades" group. Female pupils are less likely to be in the "poor grades" group. This is consistent with

the overall pattern of girls outperforming boys (Sammons et al., 2014). The effect of ethnicity is mixed with some minority ethnic groups performing better than white pupils and some others performing less well.

The effects of gender on membership of the "Science" and "Non-Science" latent groups differ. Males have higher log odds of being in the "Science" group than females, but lower log odds of being in the "Non-Science" group. Membership of the "Science" and "Non-Science" latent groups also differs by pupils' ethnic group. Black, Pakistani and pupils in the "other" ethnicity category have higher log odds than white pupils of being in the "non-Science" latent group. In contrast, the log odds of being in the "Non-Science" group are lower for Indian and other Asian pupils. These findings would be obscured by models estimating agglomerate GCSE outcomes.

There is less differentiation between the "Science" and "non-Science" groups on the basis of other variables in the model. Pupils with parents in Lower Supervisory and Technical (NS-SEC 5), Semi-Routine (NS-SEC 6) and Routine (NS-SEC 7) occupations have higher log odds of being in any of the latent groups other than the "Good Grades" group. There is no significant difference in the effect of parental social class on membership of the "Science" and "Non-Science" latent educational groups. This is clearly demonstrated in Figure 2 where quasi-variance comparison intervals are displayed (see Gayle & Lambert, 2007). Pupils with parents in Semi-Routine (NS-SEC 6) and Routine (NS-SEC 7) occupations have higher log odds of being in the "Poor Grades" group but are equally likely to be in either of the "Science" and "Non-Science" groups.

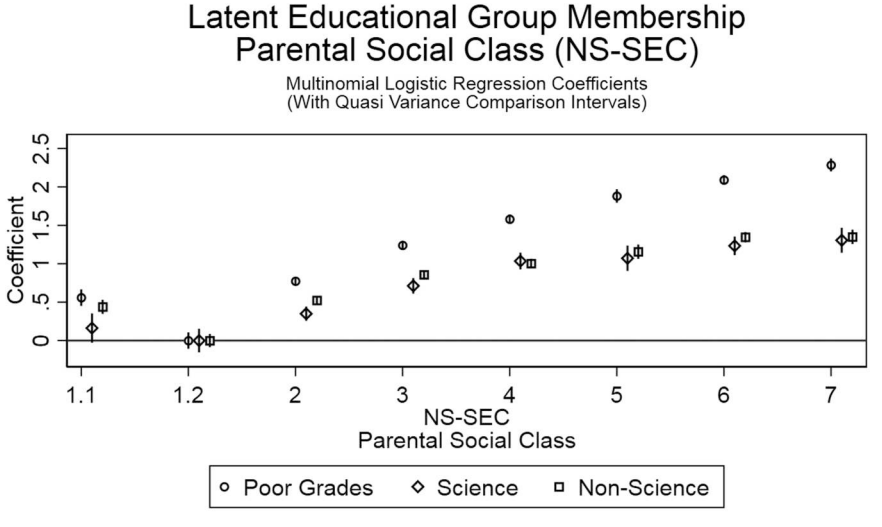

Figure 2. Latent educational group membership by parental social class (NS-SEC).

Discussion

There have been a number of salient changes in school education in England and Wales in the last half century. The qualifications that pupils gain at school continue to be important determinants of their pathways, and influence choices and participation in education, training and the labour market. Despite the long history of comprehensive (i.e., non-selective) secondary schooling, the establishment of the National Curriculum, and the implementation of GCSEs as a single tier set of qualifications, outcomes in school GCSEs are still stratified by a pupil's social background.

Measuring inequalities in school qualifications in England and Wales is especially challenging. This is because school children do not study for a single school leaving certificate or high school diploma; nor is their attainment evaluated with a single metric such as a grade point average (GPA). Using "agglomerate" measures it is straightforward to detect either "good" or "poor" overall GCSE outcomes. "Agglomerate" measures are subject agnostic and therefore they occlude important subject-specific GCSE profiles. The study of groups of GCSE subjects (which are chosen a priori) suffers the similar limitation of missing the opportunity to study inequalities in the nuanced patterns of subject-specific GCSE profiles. This is important because the range of choices available to pupils at different schools differs. The latent variable approach taken can assist in understanding educational outcomes across a range of GCSE subjects whilst not overlooking the complexity of individualised GCSE profiles.

The central aim of this paper has been to extend the methodological investigation of pupils' highly individualised GCSE profiles. The latent variable approach that we have employed has identified 4 latent educational groups which represent distinctive profiles of GCSE outcomes. This provides convincing evidence of the robustness of the earlier findings reported in Playford and Gayle (2016) and establishes the reliability of the approach using data spanning a decade. The modelling identifies two latent educational groups, one with "good" and one with "poor" GCSE outcomes. Whilst these two groups could have been detected with "agglomerate" measures, two further groups have been identified which have moderate levels of "agglomerate" GCSE outcomes, but they have distinctly different outcomes in science subjects. Through the application of a multinomial logistic regression model predicting latent class membership we identify the characteristics of pupils assigned to each of the latent groups. There were significant gender, ethnicity and social class differences in latent group membership. The identification of two distinctive groups with moderate levels of GCSE outcomes but qualitatively different GCSE profiles, especially in science subjects, is an important new finding.

The grading system for GCSEs was revised in 2016. Under the revised system grade 9 is the highest grade and grade 1 is the lowest (Ofqual, 2018). The change from an alphabetical to a numeric grading scheme may tempt data

analysts to treat individual subject grades as a "score". For analyses that require an agglomerate measure of school GCSE outcomes this will be a functional approach, in a similar manner to producing an overall GCSE points score. The issue of appropriately representing a pupil's profile of GCSE outcomes remains unchanged however. This is because pupils continue to study mixed and individualised diets of GCSE subjects.

The numeric nature of the revised grading system may beguile data analysts into assuming that an alternative latent variable modelling approach, such as factor analysis, might be preferable to a latent class model. A factor analysis approach will rest on the strong assumption that GCSE grades are based on a metric scale. In practice each GCSE subject outcome is one of nine (discrete) ordered categories that have been assigned a score. The specific form of latent variable model which is most appropriate for the numerical GCSE grading system will best be evaluated empirically.

The latent variable approach furnishes a clear set of typologies that help to summarise pupils' school GCSE outcomes. Such typologies are important because they can directly inform discussions about the effectiveness of educational policies and practices for addressing inequalities in education. Typologies of GCSE outcomes may also explain young people's immediate pathways at the end of school and provide insights into their choices and chances in occupational training or more advanced forms of education. Failure to gain a GCSE pass in English or Maths at grades A*-C is barrier to smooth progression to post-16 educational courses (Lupton et al., 2021). Individually, GCSE English, Mathematics and Science subjects are also a requirement for studying these subjects at GCE A' level and progression to prestigious universities in the UK (see Dilnot, 2016; Moulton et al., 2018). For schools, recognising that there may be distinct groups of young people, particularly with regard to gender and ethnicity, who may need additional assistance in either English or Maths and Science is very valuable, especially in understanding social inequalities in continuing educational participation and ultimately employment. Developing empirically grounded typologies of school qualifications may be instrumental in understanding the longer-term effects that school education has on occupational outcomes later in the lifecourse.

Ethics

Permission was granted by the University of Exeter College of Social Science and International Studies Research Ethics Officer.

Notes

1. Approximately 600,000 candidates were entered for GCSE qualifications in 2014/15 (Gill 2016). In comparison, the International Baccalaureate diploma was completed by 5,000 students in 2016 (UCAS 2017).

2. Following reforms from 2016, the GCSE grading system has changed in England. Grade 9 is the new highest grade and grade 1 (Ofqual 2018). Wales has retained the A*-G grading system (see https://www.qualificationswales.org/english/qualifications/gcses-and-a-levels/gcses accessed 26.10.20). In Northern Ireland, the grading system used varies by exam board (see https://www.nidirect.gov.uk/articles/gcses accessed 26.10.20).
3. The Key Stage 4 curriculum (for pupils aged 14-16) is organised into core and foundational subjects. The core subjects have usually been English, Mathematics and Science, and the foundation subjects have been computing, physical education and citizenship. In addition, schools must offer arts, design and technology, humanities, and modern foreign languages, and they must provide some form of religious education and sex education (see https://www.gov.uk/national-curriculum/key-stage-3-and-4 accessed 11.09.19).
4. Pupils are considered to have achieved the English Baccalaureate if they gain a grade C or more in: English; Maths; a Science; Geography or History; and a modern or ancient foreign language (Jin et al., 2011).
5. The correlation between GCSE subject group outcomes is reported in table A1 in the online supplement.
6. For example, the National Pupil Database only contains Key Stage 4 attainment data from 2001 onwards and does not include a parental social class measure.
7. For an example of the GCSE subject coding frame and the subject groupings used in YCS cohort 7, see pages 48–51 of https://doc.ukdataservice.ac.uk/doc/3533/mrdoc/pdf/3533userguide.pdf accessed 02.07.24.
8. The estimates from this model are reported in table A3 in the online supplement.
9. For comparison, the Latent Group Model Results (Four Group Model) from Playford and Gayle (2016) is reported in table A4 in the online supplement.
10. Additional sensitivity analyses have been conducted based on alternative estimation methods which are reported in the online supplement. These include posterior probability assignment (see table A6) and a one-step structural equation model (see tables A7 and A8). These sensitivity analyses provide addition detailed information that persuades us that the analyses that are the focus of this paper are most suitable.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the ESRC under [Grant ES/R004978/2].

Data availability statement

The Youth Cohort Study data used in this article have been cited appropriately in the "Materials and Methods" section. The supporting Stata code used in the article is also available with Jupyter notebooks for all tables and figures presented in the article and online-only supplement. This is available from: https://github.com/ChrisPlayford/ycs_subject_analysis

Notes on contributors

Dr Christopher James Playford I am a quantitative sociologist working in the fields of social stratification and the sociology of education. My work has focused on modelling the role of family background on educational attainment with a substantive interest in inequality and disadvantage. I specialise in the secondary analysis of large-scale survey and administrative data. I have methodological interests in a range of statistical techniques including generalised linear and mixed models, latent class analysis and multiple imputation of missing data. In a previous role I researched child development and emotional well-being. I have also published work on research reproducibility.

Prof Vernon Gayle My work involves the statistical analysis of large-scale and complex social science datasets. These datasets include both social surveys and administrative data resources. The analysis of longitudinal (i.e., repeated contacts) data is an area in which I specialise. The main substantive focus of my work is social stratification. I have particular interests in the sociology of youth and youth transitions, education and sport. I also have interests in demography, with a focus on migration, and to a lesser extent fertility. I have also undertaken work in the area of digital social research. My methodological research focuses on a range of challenges, which include topics such as quasi-variance estimation, missing data and multiple imputation, and the graphical representation of data. I am attempting to promote the "Public Awareness of Social Statistics".

Dr Roxanne Connelly I am a Senior Lecturer in Sociology and Quantitative Methods. My research falls in the interconnected fields of Social Stratification and the Sociology of Education. My methodological interests include techniques for the analysis of complex social survey data, longitudinal data analysis, administrative data, and open social science.

ORCID

Christopher James Playford http://orcid.org/0000-0002-6069-4898
Vernon Gayle http://orcid.org/0000-0002-1929-5983
Roxanne Connelly http://orcid.org/0000-0002-3886-1506

References

Abrahams, J. (2018). Option blocks that block options: Exploring inequalities in GCSE and A level options in England. *British Journal of Sociology of Education*, *39*(8), 1143–1159. https://doi.org/10.1080/01425692.2018.1483821

Anders, J., Henderson, M., Moulton, V., & Sullivan, A. (2018). The role of schools in explaining individuals' subject choices at age 14. *Oxford Review of Education*, *44*(1), 75–93. https://doi.org/10.1080/03054985.2018.1409973

Ashford, S., Gray, J., & Tranmer, M. (1993). *The introduction of GCSE exams and changes in post-16 participation (23)*. Employment Department Research Series YCS Reports.

Babb, P. (2005). *A summary of focus on social inequalities*. Office for National Statistics. http://www.statistics.gov.uk/articles/nojournal/FOSI_summary_article.pdf

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Banerjee, P. A. (2017). Is informal education the answer to increasing and widening participation in STEM education? *Review of Education*, *5*(2), 202–224. https://doi.org/10.1002/rev3.3093

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2008). *Analysis of multivariate social science data*. CRC Press.

Becker, M. P., & Yang, I. (1998). Latent class marginal models for cross-classifications of counts. *Sociological Methodology*, *28*(1), 293–325. https://doi.org/10.1111/0081-1750.00050

Benn, C., & Chitty, C. (1997). *Thirty years on: Is comprehensive education alive and well or struggling to survive?* Penguin Books.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*(1), 3–27. https://doi.org/10.1093/pan/mph001

Bramley, T., Vidal Rodeiro, C. L., & Vitello, S. (2015). *Gender differences in GCSE*. Cambridge Assessment Research Report, Issue. Cambridge Assessment. https://www.cambridgeassessment.org.uk/Images/gender-differences-in-gcse.pdf

Burgess, S., McConnell, B., Propper, C., & Wilson, D. (2004). Girls rock, boys roll: An analysis of the age 14–16 gender gap in English schools. *Scottish Journal of Political Economy*, *51*(2), 209–229. https://doi.org/10.1111/j.0036-9292.2004.00303.x

Carroll, M., & Gill, T. (2017). *Uptake of GCSE subjects 2016*. Statistics Report Series No. 114. Cambridge Assessment. http://www.cambridgeassessment.org.uk/Images/420406-uptake-of-gcse-subjects-2016.pdf

Cave, S. N., & von Stumm, S. (2021). Secondary data analysis of British population cohort studies: A practical guide for education researchers. *British Journal of Educational Psychology*, *91*(2), e12386. https://doi.org/10.1111/bjep.12386

Claessen, M. J. A. (2005). *Provision of GCSE subjects* (2). (Statistics Report Series No. 2, Issue. Cambridge Assessment. http://www.cambridgeassessment.org.uk/images/109896-provision-of-gcse-subjects.pdf

Connelly, R., Gayle, V., & Lambert, P. S. (2016). A review of educational attainment measures for social survey research. *Methodological Innovations*, *9*. https://doi.org/10.1177/2059799116638001

Connelly, R., Murray, S., & Gayle, V. (2013). Young people and school GCSE attainment: Exploring the 'middle'. *Sociological Research Online*, *18*(1), 6. https://doi.org/10.5153/sro.2880

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*(1), 3–21. https://doi.org/10.1080/01411920500401963

Courtenay, G. (1996a). *Youth cohort study of England and Wales, 1991-1993; cohort five, sweep one to three [data collection]*. UK Data Service. Retrieved October 10, 2019. SN: 3531.

Courtenay, G. (1996b). Youth cohort study of England and Wales, 1992-1994; cohort six, sweep one to three [data collection]. *UK Data Service*. Retrieved October 10 2019. SN: 3532.

Courtenay, G. (2000). Youth cohort study of England and Wales, 1994-1996; Cohort seven, sweep one and two, [data collection]. *UK Data Service*, 2nd Edition. Retrieved October 10, 2019. SN: 3533.

Croxford, L. (2006a). *The youth Cohort survey data*. Working Paper 5, ESRC Research project on education and youth transitions in England, Wales and Scotland, 1984-2002. Centre for Educational Sociology.

Croxford, L. (2006b). *The youth Cohort surveys - How Good is the Evidence?* Special CES Briefing No. 38. Edinburgh: Centre for Educational Sociology.

Croxford, L., Iannelli, C., & Shapira, M. (2007). *Documentation of the youth cohort time-series datasets*. Centre for Educational Sociology (Ed.), UK Data Archive Study Number 5765 -

Youth Cohort Time Series for England, Wales and Scotland, 1984-2002. Edinburgh: UK Data Archive.

Davies, P., Telhaj, S., Hutton, D., Adnett, N., & Coe, R. (2008). Socioeconomic background, gender and subject choice in secondary schooling. *Educational Research*, *50*(3), 235–248. https://doi.org/10.1080/00131880802309358

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*(2), 117–143. https://doi.org/10.1080/13613320050074005

Department for Education. (1985). *General certificate of secondary education: A general introduction*. Her Majesty's Stationery Office.

Dilnot, C. (2016). How does the choice of A-level subjects vary with students' socio-economic status in English state schools? *British Educational Research Journal*, *42*(6), 1081–1106. https://doi.org/10.1002/berj.3250

Drew, D. (1995). *'Race', education and work: The statistics of inequality*. Avebury.

Drew, D., Gray, J., & Sime, N. (1992). *Against the odds: The education and labour market experiences of black young people*. Youth Cohort Series 19 Employment Department Training Research and Development Series, 68, 1-56.

Farquharson, C., McNally, S., & Tahir, I. (2022). *Education inequalities* (IFS deaton review of inequalities, Issue. https://ifs.org.uk/inequality/wp-content/uploads/2022/08/Education-inequalities.pdf

Fitzgerald, R. (2004). *Youth Cohort study of England and Wales, 2000-2002; Cohort ten, sweep one, two and three, [data collection]*. UK Data Service, 2nd Edition, Retrieved October 10 2019. SN: 4571.

Furlong, A., & Cartmel, F. (2007). *Young people and social change: New perspectives* (2nd ed.). Open University Press.

Gayle, V. (2005). Youth transitions. In J. Ermisch, & R. E. Wright (Eds.), *Changing Scotland: Evidence from the British Household Panel Survey* (pp. 33–46). Policy Press.

Gayle, V., Berridge, D., & Davies, R. B. (2003). *Econometric analysis of the demand for higher education (472)*. DfES Publications.

Gayle, V., & Lambert, P. S. (2007). Using quasi-variance to communicate sociological results from statistical models. *Sociology*, *41*(6), 1191–1208. https://doi.org/10.1177/0038038507084830

Gayle, V., Lambert, P. S., & Murray, S. (2009). School-to-work in the 1990s: Modelling transitions with large-scale datasets. In R. Brooks (Ed.), *Transitions from education to work: New perspectives from Europe and beyond* (pp. 17–41). Palgrave Macmillan.

Gayle, V., Murray, S., & Connelly, R. (2016). Young people and school general certificate of secondary education attainment: Looking for the 'missing middle'. *British Journal of Sociology of Education*, *37*(3), 350–370. https://doi.org/10.1080/01425692.2014.935292

Gill, T. (2012). *Uptake of GCSE subjects 2010* (35). (Research division assessment statistics Report Series No.35 - Revised, Issue. Cambridge Assessment. http://www.cambridgeassessment.org.uk/Images/109925-uptake-of-gcse-subjects-2010.pdf

Gill, T. (2016). *Uptake of level 2 qualifications in English schools 2015* (103). Cambridge Assessment Statistics Report Series No.103, Issue. Cambridge Assessment. https://www.cambridgeassessment.org.uk/Images/306929-uptake-of-level-2-qualifications-in-english-schools-2015.pdf

Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In J. A. Hagenaars, & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 3–55). Cambridge University Press.

Gorard, S., & Taylor, C. (2002). Market forces and standards in education: A preliminary consideration. *British Journal of Sociology of Education*, *23*(1), 5–18. https://doi.org/10.1080/01425690120102827

Haque, Z., & Bell, J. F. (2001). Evaluating the performances of minority ethnic pupils in secondary schools. *Oxford Review of Education*, *27*(3), 357–368. https://doi.org/10.1080/03054980120067401

Henderson, M., Sullivan, A., Anders, J., & Moulton, V. (2018). Social class, gender and ethnic differences in subjects taken at age 14. *The Curriculum Journal*, *29*(3), 298–318. https://doi.org/10.1080/09585176.2017.1406810

Jarvis, L., Exley, S., & Tipping, S. (2003). *Youth cohort study: Survey of 16 year olds (Cohort 11 Sweep 1) technical report.* http://doc.ukdataservice.ac.uk/doc/5452/mrdoc/pdf/5452userguide.pdf

Jerrim, J. (2023). The benefits of meeting key grade thresholds in high-stakes examinations. New evidence from England.. *British Journal of Educational Studies*, *71*(1), 5–28. https://doi.org/10.1080/00071005.2022.2033692

Jin, W., Muriel, A., & Sibieta, L. (2011). *Subject and course choices at ages 14 and 16 amongst young people in England: Insights from Behavioural Economics* (Research Report DFE-RR160, Issue. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/182677/DFE-RR160.pdf

Johnson, M., Exley, S., Robinson, C., Park, A., Jarvis, L., & Phillips, M. (2006). Youth Cohort study of England and Wales, 2002-2005; Cohort eleven, sweep one to four, [data collection]. UK Data Service. Retrieved October 10 2019. SN: 5452.

Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *172*(4), 835–851. https://doi.org/10.1111/j.1467-985X.2009.00597.x

Leckie, G., & Goldstein, H. (2017). The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8. *British Educational Research Journal*, *43*(2), 193–212. https://doi.org/10.1002/berj.3264

Lessof, C., Ross, A., Brind, R., Harding, C., Bell, E., & Kyriakopoulos, G. (2018). *Understanding KS4 attainment and progress: evidence from LSYPE2 (831)*. Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/748514/Understanding_KS4_LSYPE2_research-report.pdf

Lupton, R., Thomson, S., Velthuis, S., & Unwin, L. (2021). *Moving on from initial GCSE 'failure': Post-16 transitions for 'lower attainers' and why the English education system must do better* https://www.nuffieldfoundation.org/wp-content/uploads/2021/02/Post16-transitions-for-lower-attainers-Final-report.pdf

McCutcheon, A. L. (1987). *Latent class analysis*. Sage Publications.

McCutcheon, A. L. (2002). Basic concepts and procedures in single- and multiple-group latent class analysis. In J. A. Hagenaars, & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 56–85). Cambridge University Press.

McMullin, P., & Kulic, N. (2016). Onwards or upwards? – The role of subject choice and schools in there production of educational inequality in England. In H.-P. Blossfeld, S. Buchholz, J. Skopek, & M. Triventi (Eds.), *Models of secondary education and social inequality: An international comparison* (pp. 213–232). Edward Elgar Publishing. https://doi.org/10.4337/9781785367267.00024

Mobley, M., Emerson, C., Goddard, I., Goodwin, S., & Letch, R. (1986). *All about GCSE*. Heinemann Educational.

Moulton, V., Sullivan, A., Henderson, M., & Anders, J. (2018). Does what you study at age 14–16 matter for educational transitions post-16? *Oxford Review of Education*, *44*(1), 94–117. https://doi.org/10.1080/03054985.2018.1409975

Nice, D., McAleese, I., Russell, N., & Finch, S. (2002). *Youth Cohort study of England and Wales, 1996-2000; Cohort eight, sweep one to three, [data collection]*. UK Data Service, 3rd Edition. Retrieved October 10 2019. SN: 3805.

North, J. (1987). *The GCSE: An examination*. Claridge Press.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Ofqual. (2018). *Grading new GCSEs*. Retrieved November 27 2018 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719124/Grading_new_GCSEs25.6.2018.pdf

Playford, C. J., & Gayle, V. (2016). The concealed middle? An exploration of ordinary young people and school GCSE subject area attainment. *Journal of Youth Studies*, *19*(2), 149–168. https://doi.org/10.1080/13676261.2015.1052049

Playford, C. J., Gayle, V., Connelly, R., & Murray, S. (2016). Parental socioeconomic influences on filial educational outcomes in Scotland: Patterns of school-level educational performance using administrative data. *Contemporary Social Science*, *11*(2-3), 183–202. https://doi.org/10.1080/21582041.2016.1172728

Prandy, K., Unt, M., & Lambert, P. (2004). *Not by degrees: Education and social reproduction in twentieth-century Britain* International Sociological Association Research Committee 28 on Social Stratification and Mobility, Neuchatel, Switzerland. http://www.camsis.stir.ac.uk/papers/not_by_degrees_1.pdf

Rasbash, J., Leckie, G., Pillinger, R., & Jenkins, J. (2010). Children's educational progress: Partitioning family, school and area effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(3), 657–682. https://doi.org/10.1111/j.1467-985X.2010.00642.x

Russell, N., Finch, S., McAleese, I., Nice, D., Fitzgerald, R., & La Valle, I. (2004). Youth Cohort study of England and Wales, 1998-2000; Cohort nine, sweep one to four, [data collection]. UK Data Service, 5th Edition, Retrieved October 10 2019. SN: 4009.

Salisbury, J., Rees, G., & Gorard, S. (1999). Accounting for the differential attainment of boys and girls at school. *School Leadership & Management*, *19*(4), 403–426. https://doi.org/10.1080/13632439968943

Sammons, P., Sylva, K., Melhuish, E., Siraj, I., Taggart, B., Toth, K., & Smees, R. (2014). *Influences on students' GCSE attainment and progress at age 16* (352). Department for Education. http://193.61.4.225/web-files/our-staff/academic/edward-melhuish/documents/3.RR352_-_Influences_on_Students_GCSE_Attainment_and_Progress_at_Age_16.pdf

Schmitt, J., & Wadsworth, J. (2006). Is there an impact of household computer ownership on children's educational attainment in Britain? *Economics of Education Review*, *25*(6), 659–673. https://doi.org/10.1016/j.econedurev.2005.06.001

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*(12), e80341. https://doi.org/10.1371/journal.pone.0080341

Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'getting it' for the white working class. *Research Papers in Education*, *29*(2), 131–171. https://doi.org/10.1080/02671522.2013.767370

Sullivan, A., Heath, A., & Rothon, C. (2011). Equalisation or inflation? Social class and gender differentials in England and Wales. *Oxford Review of Education*, *37*(2), 215–240. https://doi.org/10.1080/03054985.2011.559380

Sullivan, A., Zimdars, A., & Heath, A. (2010). The social structure of the 14-16 curriculum in England. *International Studies in Sociology of Education*, *20*(1), 5–21. https://doi.org/10.1080/09620211003655572

Tunstall, R. (2011). *Social housing and social exclusion 2000-2011* (153). (Case Paper 153, Issue. CASE LSE.

UCAS. (2014). *UK qualifications: For entry to university or college in 2015*. UCAS. https://www.ucas.com/sites/default/files/2015-uk-qualifications.pdf

UCAS. (2016). *Progression pathways*. UCAS. https://www.ucas.com/sites/default/files/progression_pathways_report_final_v2_0.pdf

UCAS. (2017). *Qualification information profiles: International Baccalaureate (IB) Diploma*. Retrieved November 16, 2020 from https://qips.ucas.com/qip/international-baccalaureate-ib-diploma

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450–469. https://doi.org/10.1093/pan/mpq025

Warrington, M., & Younger, M. (2000). The other side of the gender gap. *Gender and Education*, *12*(4), 493–508. https://doi.org/10.1080/09540250020004126

Wilkinson, D., Bryson, A., & Stokes, L. (2018). *DP No. 11372: Assessing the variance in pupil attainment. How important is the school attended?* IZA - Institute of Labor Economics https://www.iza.org/publications/dp/11372/assessing-the-variance-in-pupil-attainment-how-important-is-the-school-attended

Wolf, A. (2011). *Review of vocational education: The Wolf Report*. Department for Education. http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/eOrderingDownload/Wolf-Report.pdf

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*(3), 245–267. http://www.informaworld.com/10.1080014119201200482296

Younger, M., & Warrington, M. (2005). *Raising boys' achievement (636)*. Department for Education and Skills.