Research Matters / 34
A Cambridge University Press & Assessment publication

ISSN: 1755-6031

Journal homepage: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/

What's in a name? Are surnames derived from trades and occupations associated with lower GCSE scores?

Joanna Williamson and Tom Bramley (Research Division)

To cite this article: Williamson, J., & Bramley, T. (2022). What's in a name? Are surnames derived from trades and occupations associated with lower GCSE scores? *Research Matters: A Cambridge University Press & Assessment publication, 34,* 76–97

To link this article: https://www.cambridgeassessment.org.uk/Images/research-matters-34-whats-in-a-name.pdf

Abstract:
In England, there are persistent associations between measures of socio-economic advantage and educational outcomes. Research on the history of names, meanwhile, confirms that surnames in England – like many other countries – were highly socially stratified in their origins. These facts prompted us to wonder whether educational outcomes in England might show variation by surname origin, and specifically, whether surnames with an occupational origin might be associated with slightly lower average GCSE scores than surnames of other origins. Even though surnames do not measure an individual's socio-economic position, our hypothesis was that in aggregate, the educational outcomes of a group defined in this way might still reflect past social history.

In line with the research hypothesis, the results showed that the mean GCSE scores of candidates with occupational surnames were slightly lower than the mean GCSE scores of candidates with other surnames. The difference in attainment was a similar size to the difference expected between candidates half a year apart in age, and much smaller than the "gap" between male and female candidates. The explanation for the identified effect was beyond the scope of the current research, but surname effect mechanisms proposed in the literature include the psychological (e.g., implicit egotism), sociological and socio-genetic.

What's in a name? Are surnames derived from trades and occupations associated with lower GCSE scores?

Joanna Williamson and Tom Bramley (Research Division)

Introduction

For many readers, the image conjured up by a character called Mr or Miss Darcy will be different from the image associated with a Mr or Miss Tinker. In Thomas Hardy's social-realist novels, similarly, plots can plausibly turn on the distinction implied by d'Urberville as opposed to Durbeyfield. These expectations and associations reflect socio-cultural knowledge of surname origins, even if we understand it only implicitly. In England, Darcy and Turberville are examples of a "distinctive class of surnames" belonging to Norman, Breton and Flemish estate-owners who arrived in Britain with the Norman Conquest (Clark & Cummins, 2014, p. 525). The Darcy and Turberville families, along with others such as the Montgomery and Mandeville families, were major landowners in the Domesday Book. The hypothetical Tinker family, meanwhile, have a surname derived from the occupation of mending pots and pans.

Educational outcomes in England today vary according to socio-economic advantage: there are persistent associations between measures of educational attainment and socio-economic indicators such as parental education level and entitlement to free school meals (Sutherland et al., 2015). Research literature on the history of names, meanwhile, confirms that surnames in England – as well as many other countries – were highly socially stratified in their origins (Hanks & Parkin, 2016). The above facts prompted us to wonder whether the educational achievements of school students in England (as captured, for instance, in GCSE results) might show variation according to surname or family name origin.

Economic historians have used surnames to link cross-sectional data on socio-economic status (e.g., enrolment lists for elite universities from different centuries, as used by Clark & Cummins, 2014), and thereby measure long-term social mobility. What was not clear, was whether the origin of a surname itself (i.e., without linking it to a separate index of socio-economic status) would carry information to the extent that it would be reflected in today's GCSE results. To the best of our knowledge, this had not previously been researched. Our hypothesis was that

surnames with an occupational origin would be associated with slightly lower average GCSE scores than surnames of other origins. While we know that a surname is not an empirical measure of an individual's socio-economic position, our hypothesis was that in aggregate, the educational outcomes of a group defined in this way might still reflect past social history.

Surname origins

Across the world, there exist three main systems of naming. In a binomial system, individuals are known by a given name (or sometimes several) together with a family name or surname that is inherited and subsequently passed on between generations. In patronymic naming systems, given names are instead accompanied by a name that describes parentage, for example Jakobsson or Jakobsdóttir for the son or daughter of someone named Jakob. In the Arabic naming system, meanwhile, an individual's name consists of up to five elements: besides a given name (*ism*), other possible elements include names with nickname, patronymic and locative meanings[1].

In most European countries, a binomial naming system was established between the twelfth and fourteenth centuries, as non-hereditary names began to be fixed within families and handed down from generation to generation. In England specifically, it was rare to see individuals recorded with more than one name prior to the Domesday Book of 1086. As in other European countries, non-hereditary "by-names" appeared before surnames, which added descriptive details to distinguish an individual from others with the same given name. These descriptive names fell into four main categories: reference to a trade or occupation, reference to a person's geographical location or origin, description of a relationship to another person (e.g., patronymic names), or reference to some physical characteristic or behaviour (Hanks & Parkin, 2016, p. 3). These four categories provide the most common classification for surname origin in England today (Table 1).

1 For further details of naming systems and their history, see Hanks & Parkin (2016). On the Arabic naming system in particular, Hanks and Parkin note that individuals moving from Arabic-speaking societies to countries using a binomial naming system have adopted different elements of the possible five as their surname.

Table 1: Typology of surnames in Britain and Ireland.

Surname origin	Description	Sub-types	Examples	Less obvious examples
Relationship	Reference to a relationship to another person	Patronymics	Peterson, Michaels	Brightman
Locative	Reference to a person's geographical location or origin	Topographical, toponymic	Churchill, Oppenheimer	Dubois
Occupational	Reference to a trade or occupation	Other	Archer, Fowler	Palfreyman
		Status	Laird	Pasha, Villain
Nickname	Reference to a physical characteristic or behaviour	_	King	Mordaunt

The "less obvious examples"[2] in Table 1 include surnames whose origins require a little more explanation than others. The name "Brightman" indicates a relationship to someone whose given name involved the Old English stem word "beorht" (meaning "bright"), while "Dubois" is a locative name translating literally to "of the woods", and "Palfreyman" refers to the occupation of maintaining saddle horses. The surname "Pasha" derives from the (high) rank of this name within the Ottoman empire, and, at the opposite extreme, "Villain" derives from the Anglo-Norman word *villein* (meaning *serf*). The name "Mordaunt", meanwhile, originated as a Norman nickname for someone with a sharp tongue.

Although "status" surnames are usually considered a subset of occupational surnames, they can reflect status in different ways. A surname that appears to describe a high-status role is more likely to indicate a servant to the high-status individual, or a nickname based on personal qualities. For instance, "Baron" is listed in the *Oxford Dictionary of Family Names in Britain and Ireland* as both a nickname and status name, while "Knight" is listed as a status name, occupational name, nickname and relationship name (Hanks et al., 2016). Similar explanations apply in other languages – in German, the surnames Kaiser (emperor) and König (King) originated as nicknames and indications of subordinate roles (Silberzahn & Uhlmann, 2013). As a broader point, status-origin names illustrate why research on the origin and history of names is important, even when the semantic meaning of the name seems unambiguous.

The social stratification of surnames in England included both which surnames were carried by which individuals, but also whether an individual had a surname at all (Hanks & Parkin, 2016, p. 4). While some wealthy landowners arrived in England with hereditary surnames, or adopted the use of hereditary surnames soon after the Norman conquest, others in society were still known by bynames many centuries later. Elsewhere in Britain and Ireland, research has identified varying patterns of surname establishment and development. In Ireland, surnames were established early and there has been "considerable exchange of surnames

2 All examples in Table 1 (and their classifications and explanations) come from the *Oxford Dictionary of Family Names in Britain and Ireland* (Hanks et al., 2016).

between Britain and Ireland for almost a millennium" (p. 6). In Scots-speaking regions, meanwhile, surnames began to be used at a similar date as in England, but their development towards widespread use was slower (p. 7). In Wales, the development of hereditary surnames occurred much later than in England, Ireland and Scotland – hereditary surnames were rare in Wales as late as 1500.

Surnames in research

Administrative records provide highly extensive (sometimes population-wide) data sets linking surnames to marriages and births, residency, education and ownership. In combination with their heritability, this makes surnames a valuable information source, and surnames have consequently been used by researchers in a wide range of disciplines, including genetic and demographic studies (e.g., Relethford, 1992), geography (e.g., Longley et al., 2007), sociology (e.g., Jackson, 2009), and economics and development studies (e.g., Dasar, 2019).

Economic historians have demonstrated that surnames can offer an innovative source of data for researching social mobility, not only in England but countries as diverse as Spain, Sweden, Chile, China and Korea (Clark & Cummins, 2014; Clark et al., 2015; Guell et al., 2014)[3]. In this research, surnames have been used to trace families over multiple generations, but surnames themselves have not been treated as informative. Guell et al. (2014, p. 694) emphasise this in quite strong terms, explaining that surnames are "intrinsically irrelevant" except for the fact that "they get passed from one generation to the next, alongside other characteristics that *do* matter". The argument for this is that most surnames tend towards more even distribution across social strata over time. Among long-established surnames in England, those that were commonly found in the population by 1800 were by that point associated with average levels of social status (Clark & Cummins, 2014, p. 525). Guell et al. state firmly that "We cannot learn anything from the name Smith" (p. 695). In our research, we wondered whether we could in fact learn something from the name Smith – or at least, detect differences in the GCSE outcomes of the group of students that includes Smiths, Tinkers, Bakers and Butchers, relative to students with other surnames.

Research that is concerned with the information captured by surnames themselves is rare. An example is the research by Voracek et al. (2015) into the relative physical strengths of men with the surnames Tailor and Smith. This research aimed to replicate earlier work by Bäumler (1980), who put forward a "genetic-social hypothesis" for the association of different body types with certain occupational surnames. The logic of this hypothesis was that (1) both surnames and important physical characteristics are inherited, (2) many trades (including blacksmiths) were historically organised around guilds and showed high levels of within-group marriage and apprenticeships, and (3) physical characteristics that are useful or prerequisite for a given trade will be selected for in those joining the trade. In

3 By making use of surnames that are relatively rare within a society, researchers have been able to track the social status of families across far longer time periods than in conventional studies on social mobility, which have typically been restricted to studying adjacent generations (parent–child relationships, or at most grandparent–parent–child relationships).

combination, these factors point to "above-chance preservation of any heritable traits within lineages", and might have resulted in "discernible physical differences" between men in contemporary society with the surnames Tailor and Smith (Voracek et al., 2015, p. 2).

Earlier studies found that Smiths were indeed heavier than Tailors, considered themselves more suited for activities requiring strength, and were overrepresented in sports requiring strength (Bäumler, 1980; Stemmler & Bäumler, 2003). The two new studies reported by Voracek et al. (2015) gave a mixed picture: Smiths did rate themselves more highly in strength-related activities, and Tailors rated themselves more highly in dexterity-related activities, but the effects were small and only the "Smith" effect significant. The findings also showed an increasing prevalence of Smiths from "light-stature to medium-stature to heavy-stature sports" (p. 8), but no pattern in the prevalence of Tailors in different sports, and no differences between Tailors and Smiths in their basic physical characteristics.

These findings could be explained by adaptations to Bäumler's hypothesis: for example, a stronger occupational selection effect for blacksmiths than for tailors, or fewer opportunities for assortative mating for tailors. Voracek et al. (2015) point out that the findings could also be explained by an entirely different explanatory mechanism, namely the psychological effect of implicit egotism. Implicit egotism is the idea that people have an (unconscious) preference for people, places and things that they associate with themselves, and in this case would mean that "Smiths, merely because of their surname, would feel more inclined to weight-train" and take part in strength-related sports (p. 9). Previous studies have shown that people are more likely to undertake careers and move to locations that resemble their surname. For example, Canadians with surnames beginning Tor-, Cal- and Win- are disproportionately highly represented in Toronto, Calgary and Winnipeg respectively (Pelham et al., 2002; Pelham et al., 2003).

While implicit egotism is about the automatic associations that individuals make in relation to themselves, surnames may also have impact through the associations made by others. A study by Silberzahn and Uhlmann (2013), for example, showed that Germans with noble-sounding names such as König (King) are over-represented in management positions, compared to Germans with surnames that refer to ordinary occupations such as Becker/Bäcker (baker) or no social role. Silberzahn and Uhlmann (2013) hypothesise that this is due to associative cognition on the part of others: the high status associated with the noble-sounding name "may implicitly spill over to its bearer, influencing the status accorded to that person and consequential life outcomes" (p. 2441). For instance, the status may lead to more positive interpretations of traits and performance in the workplace. The influence of a noble-sounding name on an individual's self-perception may also be a factor – perhaps by encouraging them to pursue high-status roles.

A final area of surname research to consider is on "alphabetism", that is, the impact of having a surname that starts with a letter ranked later in the alphabet. Cauley and Zax (2018), for example, showed that surname initials ranked further from the beginning of the alphabet were associated with lower educational attainment, as well as other poor life outcomes, for men (women were excluded from the study).

Names and social biases

Name effects have been more extensively researched in relation to negative social discrimination based on race, class and culture. English-language studies in this area have focused more frequently on given names than surnames, but it is worth outlining some notable findings.

Experimental studies have used manipulation of names on student work or application letters to test the effect of (inferred) race, class or culture on the judgements made about an individual by others – particularly teachers and employers. As summarised by Fryer and Levitt (2004, p. 771), these studies have repeatedly shown that "resumes with traditional names are substantially more likely to lead to job interviews than are identical resumes with distinctively minority-sounding names". An important contribution by Fryer and Levitt was to show that the association between a distinctively minority-sounding name and life outcomes is no longer found once socio-economic circumstances at birth are controlled for.

Also in the US context, Figlio (2005) has argued that teacher expectations are sensitive to the perceived socio-economic status as well as racial status of school students' given names, with effects on school achievement. Students whose given names are associated with low socio-economic status achieve lower test results than their siblings and fellow students whose names are less class-identifiable (pp. 21–22), and this striking finding is evident even between twins. Consistent with Fryer and Levitt (2004), Figlio (2005) found that the perceived class status of names had a larger effect than the perceived racial status of a name. The explanatory mechanism suggested was that teachers "may use a child's name as a signal of unobserved parental contributions to that child's education, and expect less from children with names that 'sound' like they were given by uneducated parents" (p. 1). Data on teachers' referrals to "gifted" programmes, as well as promotions of students to the next grade, were consistent with this proposed explanation.

In a UK experiment, meanwhile, Jackson (2009) found that (fictitious) job applicants whose name, school type and interests were all associated with high social status were more likely to receive a reply from employers than candidates whose equivalent characteristics were associated with low social status. Jackson notes that first names in England can offer "extremely strong signals of class origin" (examples used in the study were "Camilla Bevans-Brown" and "Donna Taylor"). No difference was found in the rate of positive replies, however, and, most interestingly, Jackson concluded that employers were responding to the signal implied by sets of characteristics, and that no individual characteristic conferred a benefit.

Educational outcomes in England

The important context underlying this research is that educational outcomes in England (as well as in many other countries) are known to vary by socio-economic status (SES) (e.g., DfE, 2019, pp. 8–9). In an investigation of different proxies for socio-economic status, Sutherland et al. (2015) showed that after controlling for

other characteristics, free school meal (FSM) eligibility in the preceding five years "equates to the difference between a pupil gaining one grade better across seven GCSEs (e.g., moving from a C to a B) and two grades better on an eighth GCSE" (p. 8), and that parental occupation, parental education, and other household measures were even better predictors of a pupil's educational outcomes.

Educational outcomes also vary by ethnic group (e.g., DfE, 2019, p. 12), by gender (Bramley et al., 2015) and by age relative to others in the same cohort (Benton, 2014). Further, the gaps in attainment between low-SES (e.g., FSM) pupils and other pupils also vary by subgroup (DfE, 2022). For example, the attainment gap by FSM status is much larger among white pupils than among those in other ethnic groups. Low socio-economic groups include disproportionate numbers of minority ethnic students, but minority ethnic students make greater progression during secondary education than white British students, after accounting for prior attainment (Leckie & Goldstein, 2019; Wilson et al., 2009). While differences can be explained to some extent by language, cultural attitudes towards education and qualifications are also hypothesised to play a role (Hoffmann, 2018; Wilson et al., 2009).

Data and method

The research was designed to test the simple hypothesis that average GCSE results would be lower among candidates whose surname originated as an occupation than among other candidates. To give context, we were also interested in how the observed difference (if any) compared to the difference in GCSE scores by gender, and the difference in GCSE scores by birth month. Finally, we decided to look at whether the GCSE scores showed evidence of alphabetism.

Data

We obtained all results in GCSE Mathematics (A*–G) and GCSE English (A*–G) from the awarding body OCR, for the years 2012–17. We retained only those candidates for whom we had at least one GCSE Mathematics grade and at least one GCSE English grade, taken at the usual age[4] of 15–17. In this article "mean GCSE score" refers to the average of these two GCSE grades (after converting the A*–G letter grades to numbers 8–1, and taking the best result if candidates had more than one grade in either subject).

The purpose of using GCSE English and Mathematics was to obtain a large data set on educational attainment that was as free as possible from subject selection effects (since GCSE Mathematics and English are taken by almost all 16-year-olds in England). The final data set contained a mean GCSE score for just under 21 000 unique candidates. The data set also included candidate surname, gender and date of birth.

4 We excluded very early entry candidates and results from learners aged 18 or over to avoid age effects as far as possible. Restricting to candidates aged 16 exactly, however, would have reduced the available data too far (by around 25 per cent), since sitting GCSE English and/or Maths one year early, or re-sitting one of these in Year 12, was fairly common during 2012–17.

For surname classification, we used the database underpinning the *Oxford Dictionary of Family Names in Britain and Ireland* (Hanks et al., 2016), from here on abbreviated to FaNBI. This database is an up-to-date and research-based authority on UK surnames[5], listing all surnames held by 100 or more individuals in 1997. In addition, FaNBI lists surnames that have appeared in other British surname dictionaries, and "established names" (those found in both the 1881 census and 1997 electoral rolls) (Hanks & Parkin, 2016, p. 12).

As shown by the examples in Table 1, established UK surnames derive from a range of languages and cultures, reflecting many centuries of immigration history. The FaNBI database lists information on the language and/or cultural origin of listed surnames, where known. The FaNBI database also includes classification of each surname's origins according to the typology shown in Table 1. For the purposes of our analysis, we decided to record status-origin surnames as a category in their own right (rather than a subset of occupational surnames). This was for two reasons: firstly, the explanation of status-origin names (in relation to the semantic meaning of the name) is quite different from the explanation of other occupational names, and secondly, the literature suggests that status-origin names could be affected by the associative cognition effects described by Silberzahn and Uhlmann (2013). As for the language and culture field in FaNBI, the surname typology field for each surname could be blank, list one surname type, or list several types.

Data preparation and classification

We created an indicator variable for "Occupational" surname origin. Candidate surnames listed in FaNBI were flagged as "Occupational" if "Occupational" appeared in the typology list for the name, or if the surname was included in FaNBI as a variant of a listed surname which itself had "Occupational" in the typology list. Indicators for "Status", "Relationship", "Locative" and "Nickname" surname origins were created in the same way. It is important to emphasise that these surname origin indicators were a set of five independent binary indicator variables, rather than a classification variable, and that the same surname could be flagged by multiple indicators. The reason for this was that, as noted above, surnames in the FaNBI database could have zero, one, or several different origins listed in the surname typology field.

We also created an indicator variable to record whether a name derived from a British or Irish language, narrowly defined: surnames were flagged if the language/culture field in FaNBI included "English", "Welsh", "Scottish", "Irish", "Manx", "Cornish", or "Norman". Because socio-economic status, educational progression and educational outcomes can vary across ethnic and cultural groups, we were concerned about the possibility of conflating surname effects with ethnic and

5 Rather than, for instance, a re-publication, new edition or amalgamation of previous surname dictionaries. For a useful overview of the scholarship on surnames in Britain and Ireland, see Hanks and Parkin (2016). The project to create the FaNBI database and dictionary was initiated by Oxford University Press and the Arts and Humanities Research Council of Britain. The research was led by academics at the University of the West of England, in association with database experts at Brno University in the Czech Republic.

cultural group effects. Without data on candidate ethnicity or culture, we could not directly control for this, so the purpose of the British/Irish language indicator was to offer an imperfect proxy. The indicator enabled us to compare results of analyses for all FaNBI-listed surnames with results for linguistically British/Irish surnames only.

Multi-part names (e.g., double-barrelled) were split, and we obtained classification details for each part. The exception to this was multi-part names including common prefix words such as "De la...", "Al..." or "Von...". In some of these cases, the entire multi-part name was listed in FaNBI as one name, allowing easy classification. If the entire name was not listed in FaNBI, we attempted a re-classification after discarding the prefix words.

The surname origin indicators were then applied to candidates' surnames: candidates were flagged if their entire surname or part (if a multi-part name) met the conditions for inclusion. So, for instance, a candidate with the surname Carter-Khan would have been flagged as having an occupational-origin name.

To make sure that no individuals would be identified, we removed given names and school names from the GCSE results data set. In the results, we do not mention or list any surnames belonging to fewer than 10 candidates in the results data set.

Analyses

We used descriptive statistics to explore the proportion of candidates in the results data set whose names we could categorise using the FaNBI database, and the surname types of those whose names were listed. We also described how average mean GCSE scores varied by surname origin, birth month and gender. We then used multilevel linear models to investigate the relationship between mean GCSE score and surname origin.

To investigate the "First letter" alphabetism hypothesis, we calculated the correlation between mean GCSE scores and surname initial letter (as a numerical rank).

Results

The results data set contained 8681 different candidate surnames[6], of which just over 80 per cent were listed in FaNBI. The candidate surnames not listed in FaNBI tended to account for smaller numbers of candidates (i.e., they were the less common surnames), so overall, over 19 000 candidates (more than 91 per cent of those in the results data set) had a surname that was listed in FaNBI.

Table 2 shows the numbers of candidates flagged by each of the surname indicators. The surname origin indicators are not mutually exclusive categories, and some candidates had names flagged by multiple indicators. Of those students whose surname appeared in FaNBI, 92 per cent had a surname of British/Irish language origin (henceforth, abbreviated to BIL). A high proportion had a surname that was a relationship name (39 per cent) or locative name (40 per

6 This total was obtained from treating each unique candidate surname at face value (i.e., Clark, Clarke, Smith-Clark and Smith would be counted as four names, even though FaNBI identifies Clarke as a variant spelling of Clark).

cent), and just under 16 per cent had a surname derived from an occupation. Examples of surnames for each surname indicator can be found in the Appendix.

Table 2: Candidates identified by each indicator, among those with surnames listed in FaNBI.

Surname indicators	Percentage of candidates (out of 19 023)
British/Irish language name	92.1
Occupation name	15.9
Status name	1.8
Relationship name	38.8
Locative name	40.2
Nickname name	20.8

GCSE results by surname

We first looked at the mean GCSE scores of common occupational surnames. Figure 1 shows the average mean GCSE scores for all those occupational surnames that appeared at least 30 times in the results data set. For reference, it also shows the three surnames with the highest[7] average mean GCSE scores (Chan, Alexander and Jennings), the three surnames with the lowest average mean GCSE scores (Wheeler, Weaver and Mellor), and the average mean GCSE score for all candidates in the result data set. Unsurprisingly, the average mean GCSE scores for the commonly found occupational surnames are clustered around the average.

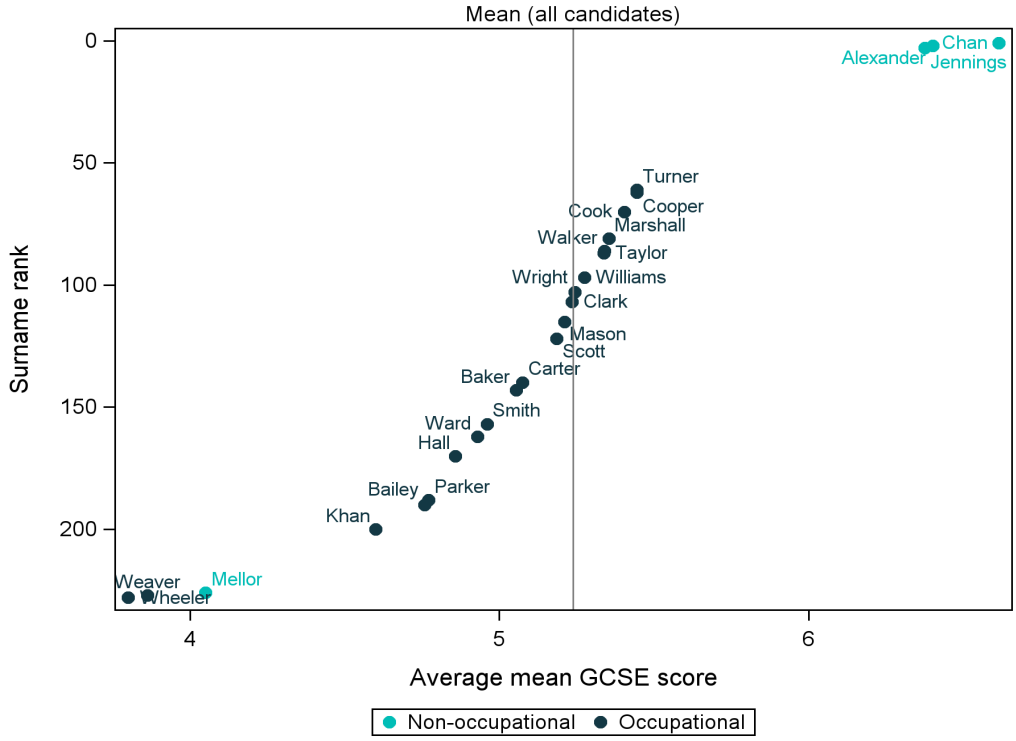

Figure 1: Average mean GCSE scores for common occupational surnames in the results data set.

7 Considering only surnames with at least 10 occurrences in the results data set. Note that this reduced the number of different surnames from over 7000 to 289.

Analysis of candidates' mean GCSE scores by surname confirmed that after accounting for known sources of variation (gender, age and school/centre), there was significant variation between different surnames[8].

We next looked at whether there were differences in the mean GCSE scores according to each surname origin indicator (separately), as shown in Figure 2. The average mean GCSE score of candidates with an occupational-origin surname was slightly lower than for those without an occupational-origin surname, in line with the research hypothesis. There was a much larger difference, however, according to whether candidates had a status-origin surname.

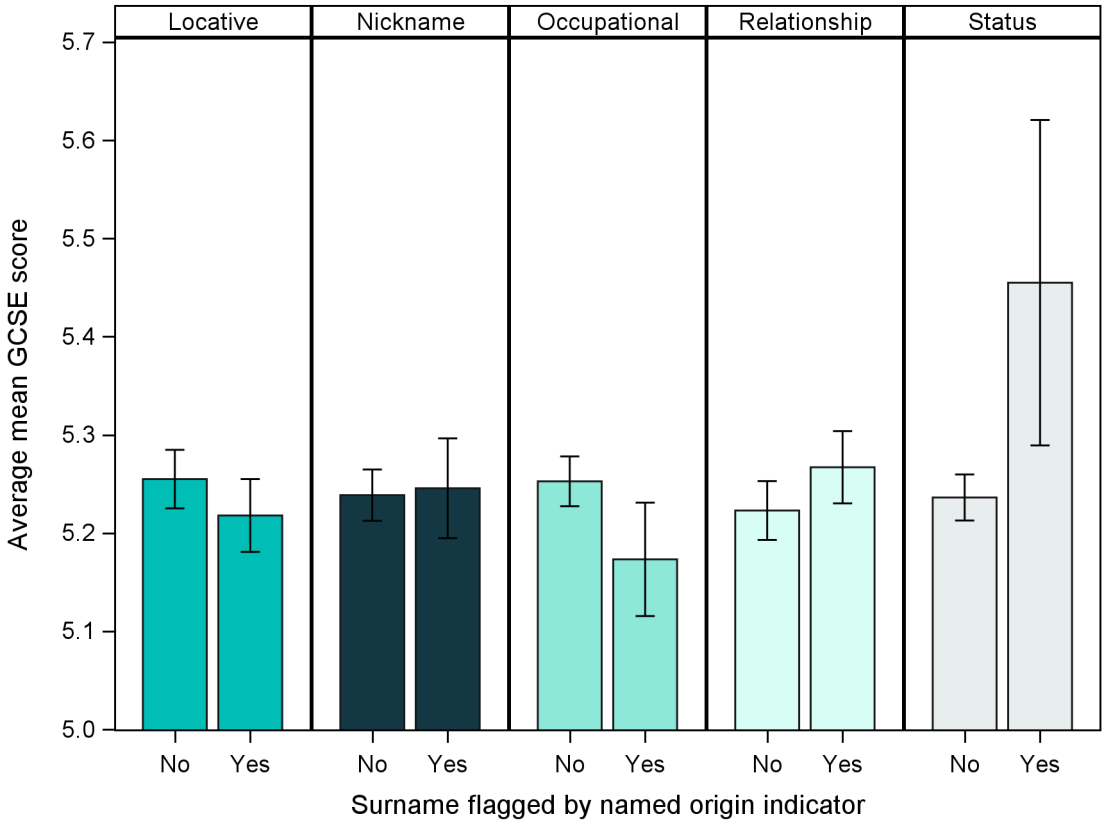

Figure 2: Differences in mean GCSE score by surname origin.

Because of the large difference in mean GCSE score according to status origin, and the fact that a surname in FaNBI could be recorded as having both an occupational and status origin, we decided to use the four-way surname classification shown in Table 3 for the remainder of the analysis. Table 3 shows that candidates with an occupational (non-status) surname had the lowest mean GCSE scores, with a mean of 5.16. At the other extreme, candidates with a status-

8 We fitted a multilevel model with gender, age and centre as fixed effects, and surname as a random effect. The variance of the random surname intercepts was small but statistically significant (0.012, p=0.038). The data set included all candidates with a surname listed in FaNBI (N=19 023), and names were analysed "as seen" (i.e., surnames listed as variants or possible variants of one another were not re-coded as the same name).

origin (non-occupational) surname had the highest mean GCSE scores, with a mean of 5.48. When the data set was restricted to consider only BIL surnames, the mean GCSE score averages for the two status-origin surname groups reduced, resulting in a smaller difference between the status-origin surname groups and the occupational (non-status) surname group.

Table 3: GCSE results by surname group.

	All candidates with surname in FaNBI			Candidates with BIL surnames		
Surname group	**N**	**N unique names**	**Mean GCSE (Std Error)**	**N**	**N unique names**	**Mean GCSE (Std Error)**
Occupational (non-status)	2879	657	5.16 (0.03)	2808	633	5.17 (0.03)
Non-status, non-occupational	15 811	6256	5.25 (0.01)	14 436	5610	5.25 (0.01)
Occupational and status	144	56	5.42 (0.13)	142	54	5.40 (0.13)
Status (non-occupational)	189	50	5.48 (0.11)	141	39	5.33 (0.13)
Total	*19 023*	*7019*		*17 527*	*6336*	

Context

To contextualise the differences in mean GCSE score by surname group, we plotted the distributions of mean GCSE scores according to birth month and gender, two other variables for which GCSE result effects are observed (Figure 3).

The average mean GCSE score for candidates with an occupational (non-status) surname was 5.16, while for candidates whose surnames had neither occupational nor status origins the average mean GCSE score was 5.25. The distributions in Figure 3 show that this difference of 0.09 GCSE points was comparable in size to the difference associated with several months' difference in birth month, and about a quarter as large as the difference in mean GCSE scores by gender.

There was a larger difference in mean GCSE scores when candidates with occupational (non-status) surnames were compared to candidates with status-origin surnames. The most extreme was the difference of 0.32 GCSE points between candidates with an occupational (non-status) surname (mean 5.16) and candidates with a status-origin (non-occupational) surname (mean 5.48). This difference was greater than the difference between September-born and August-born students, and was approximately three-quarters the size of the observed difference by gender.

Among just the BIL surnames, the largest difference by surname group was between candidates with an occupational (non-status) surname (mean GCSE 5.17) and candidates with a status-origin and occupational surname (mean GCSE 5.40), as shown in Table 3. This difference of 0.23 GCSE points was the same size as the difference between September-born and August-born students, and roughly half the size of the observed difference by gender.

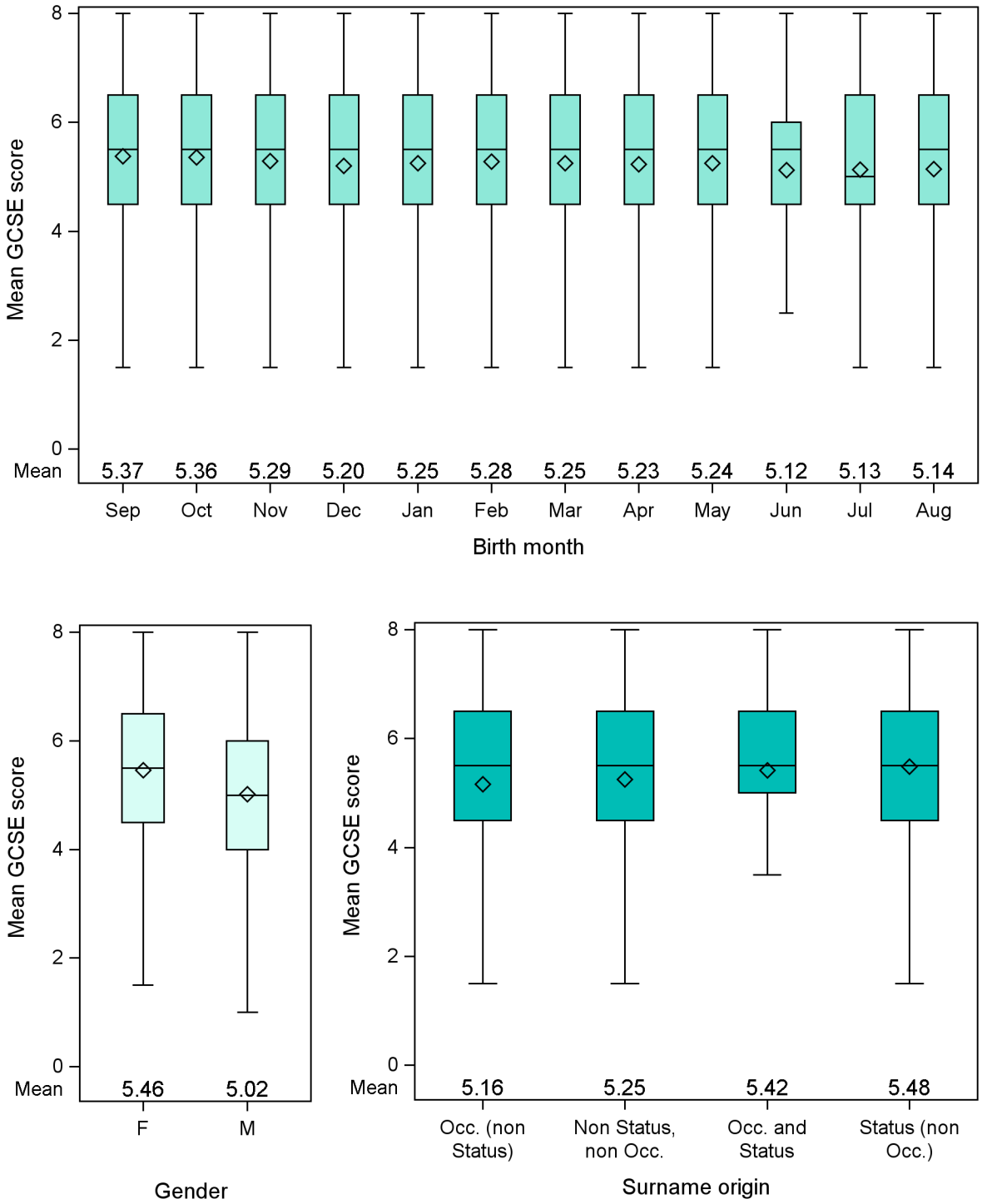

Figure 3: Distributions of mean GCSE score, by birth month, gender and surname origin.

Models

We estimated multilevel linear regression models for mean GCSE score, for candidates with a surname listed in FaNBI. A multilevel model structure was used in order to account for the clustering of students within schools.

The simplest model (Model 1) used the following structure:

> mean GCSE score = surname group + random intercept for centre.

In Model 2, we included the additional contextual variables of gender and birth month:

> mean GCSE score = surname group + gender + birth month + random intercept for centre.

In Model 1, the estimated effect of an occupational-origin (non-status) surname, in comparison with the baseline category of a non-status non-occupational surname, was -0.06 GCSE points (p=0.028). The estimated effect of having a status-origin (non-occupational) surname was 0.16 (p=0.091).

When the additional variables of gender and birth month were added (Model 2), the estimated effects associated with an occupational-origin (non-status) surname and non-status non-occupational surname changed very little from Model 1. The size of the estimated gender effect was larger than either surname effect (-0.28, p<.0001), while the estimated effect of birth month was smaller (-0.01, p<.0001).

When the models were re-estimated using only the candidates with BIL surnames (Model 3 and Model 4) the size of the effect associated with an occupational-origin (non-status) surname remained unchanged (-0.05), but there was an increase in p-value, so that the effect was no longer statistically significant at the 5 per cent level. The estimated effects for gender and birth month remained very similar to those found in Model 2.

Table 4: Estimated model parameters.

	Model 1*		Model 2*		Model 3†		Model 4†	
Effect	**Estimate (SE)**	**Pr > |t|**	**Estimate (SE)**	**Pr > |t|**	**Estimate (SE)**	**Pr > |t|**	**Estimate (SE)**	**Pr > |t|**
Intercept	4.27 (0.10)		4.52 (0.10)	<.0001	4.22 (0.10)		4.47 (0.10)	<.0001
Occupational (non-status)	-0.06 (0.03)	0.028	-0.05 (0.03)	0.031	-0.05 (0.03)	0.061	-0.05 (0.03)	0.065
Occupational and status	0.13 (0.10)	0.202	0.14 (0.10)	0.178	0.12 (0.11)	0.258	0.13 (0.10)	0.221
Status (not occupational)	0.16 (0.09)	0.091	0.17 (0.09)	0.069	0.12 (0.11)	0.261	0.13 (0.11)	0.235
[Non-status, non-occupational]	0	.	0	.	0	.	0	.
Gender M			-0.28 (0.02)	<.0001			-0.28 (0.02)	<.0001
Gender F			0	.			0	.
Birth month			-0.01 (0.00)	<.0001			-0.02 (0.00)	<.0001

*Model estimated using all candidates in results data set with a surname listed in FaNBI, and both gender and birth month data available (n=19 022)

†Model estimated using only candidates with British/Irish language surname, and both gender and birth month data available (n=17 526)

Alphabetism

Figure 4 shows the average mean GCSE score for surnames beginning with each letter, together with 95 per cent confidence intervals (results for X are not plotted, as there were fewer than 10 candidates with X-surnames). In contrast to the alphabetism hypothesis, mean GCSE scores did not decrease for surnames with initials farther from the beginning of the alphabet. Calculating correlations confirmed that there was no association between mean GCSE score and the initial letters of candidate surnames (r = 0.01, p = 0.193, N = 19 023). This remained the case when the analysis was repeated for BIL surnames only (r = 0.0003, p = 0.973, N=17 527).

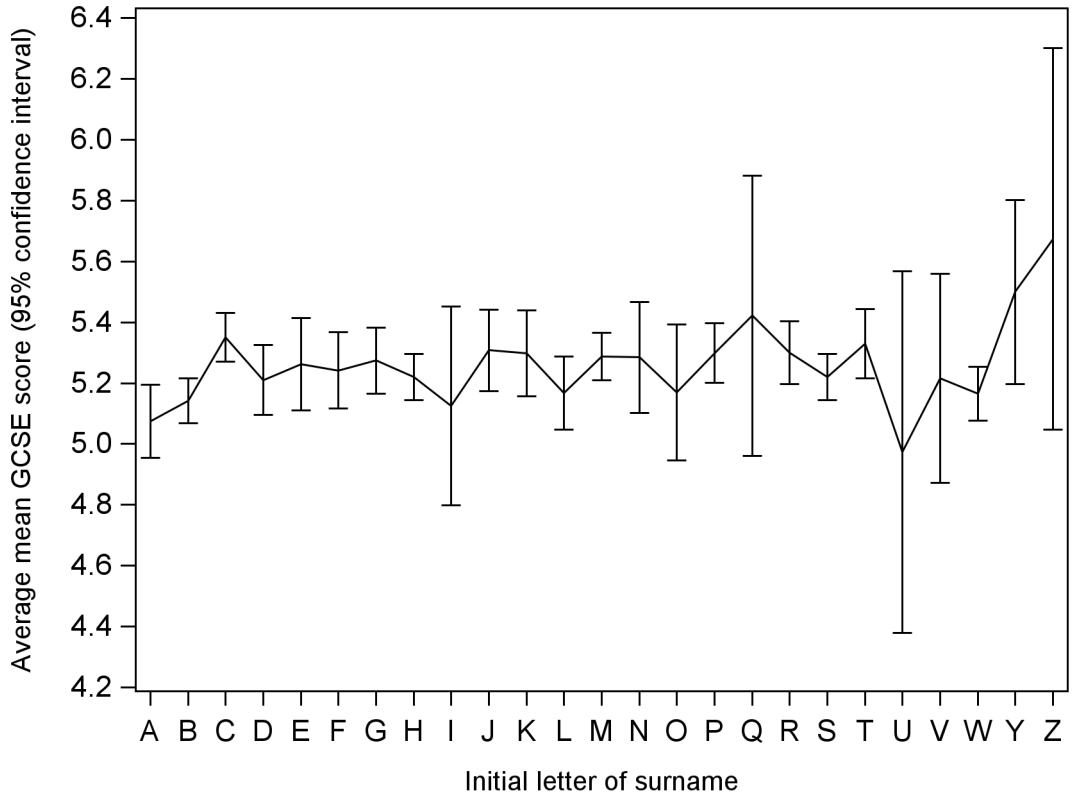

Figure 4: Average mean GCSE scores (with 95 per cent confidence intervals) by initial surname letter.

Discussion

The results of this simple study showed that the mean GCSE scores of candidates with occupational surnames were slightly lower than the mean GCSE scores of candidates with other surnames. This is in line with the research hypothesis, but the difference in GCSE attainment was not large: it was a similar size to the average difference expected between candidates half a year apart in age, and much smaller than the well-known "gap" between male and female GCSE candidates.

The size of the estimated occupational surname effect was consistent across all models estimated. The associated p-value, however, increased when the data set was restricted to candidate surnames from British or Irish languages, moving from

~0.03 to ~0.06 and consequently over the conventional threshold for statistical significance. Taken together, the regression model outcomes indicate a result that is unlikely to have arisen purely by chance. At the same time, we emphasise that the British/Irish language indicator was an imperfect proxy for investigating the impact of ethnic or cultural group differences on the surname investigation, and further research, using data on candidate ethnicity and ideally family immigration background, would be needed in order to better understand this.

This study identified a small negative effect associated with occupational non-status surnames. The explanation for this effect was beyond the scope of the current research, but surname mechanisms proposed in the literature include the psychological (e.g., implicit egotism, associative cognition), sociological (e.g., reading surnames as information signals about social class) and socio-genetic (e.g., Bäumler's (1980) "genetic-social" explanation for Tailor–Smith differences). This study hoped to offer a novel look at educational attainment and social inequalities. Ultimately, the findings are a reminder that these are highly complex matters, and that caution is needed to avoid over-interpreting small differences.

Acknowledgements

We would like to thank Professors Richard Coates, Patrick Hanks and Peter McClure and Dr Adam Rambousek for their generous assistance accessing the *Oxford Dictionary of Family Names in Britain and Ireland* (FaNBI) data base.

References

Bäumler, G. (1980). Differences in physique in men called 'Smith' or 'Tailor' considered as results of a genetic effect dating back over several centuries. *Personal and Individual Differences, 1*, 308–310.

Benton, T. (2014). *Should we age-standardise GCSEs?* Cambridge Assessment Research Report. https://www.cambridgeassessment.org.uk/Images/285155-should-we-age-standardise-gcses-.pdf

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

Cauley, A. S., & Zax, J. S. (2018). *Alphabetism: The effects of surname initial and the cost of being otherwise undistinguished*. SSRN. https://ssrn.com/abstract=3272556 or http://dx.doi.org/10.2139/ssrn.3272556

Clark, G., & Cummins, N. (2014, Dec). Surnames and social mobility in England, 1170-2012. *Human Nature, 25*(4), 517–537. https://doi.org/10.1007/s12110-014-9219-y

Clark, G., Cummins, N., Hao, Y., & Vidal, D. D. (2015). Surnames: A new source for the history of social mobility. *Explorations in Economic History, 55*, 3–24. https://doi.org/10.1016/j.eeh.2014.12.002

Dasar, B. N. S. (2019). A methodology to track economic mobility of castes in south Indian village, circa 1866–2001. *Journal of Rural Development, 38*(4), 580–6074.

DfE. (2019). *Key stage 4 performance, 2019 (revised)*. Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/863815/2019_KS4_revised_text.pdf

DfE. (2022, March 18). *GCSE results (Attainment 8)*. UK Government. https://www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training/11-to-16-years-old/gcse-results-attainment-8-for-children-aged-14-to-16-key-stage-4/latest

Figlio, D. N. (2005). *Names, Expectations and the Black-White Test Score Gap* (NBER Working Paper No. w11195). https://ssrn.com/abstract=684721

Fryer, R. G., & Levitt, S. D. (2004). The Causes and Consequences of Distinctively Black Names. *The Quarterly Journal of Economics, 119*(3), 767–805.

Guell, M., Rodriguez Mora, J. V., & Telmer, C. I. (2014). The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating. *The Review of Economic Studies, 82*(2), 693–735. https://doi.org/10.1093/restud/rdu041

Hanks, P., Coates, R., & McClure, P. (Eds.). (2016). *The Oxford Dictionary of Family Names in Britain and Ireland*. Oxford University Press.

Hanks, P., & Parkin, H. (2016). Family Names. In C. Hough (Ed.), *The Oxford Handbook of Names and Naming*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199656431.013.9

Hoffmann, N. I. (2018). Cognitive achievement of children of immigrants: Evidence from the Millennium Cohort Study and the 1970 British Cohort Study. *British Educational Research Journal, 44*(6), 1005–1028. https://doi.org/10.1002/berj.3476

Jackson, M. (2009, Dec). Disadvantaged through discrimination? The role of employers in social stratification. *British Journal of Sociology, 60*(4), 669–692. https://doi.org/10.1111/j.1468-4446.2009.01270.x

Leckie, G., & Goldstein, H. (2019). The importance of adjusting for pupil background in school value added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal, 45*(3), 518–537. https://doi.org/10.1002/berj.3511

Longley, P. A., Webber, R., & Lloyd, D. (2007). The Quantitative Analysis of Family Names: Historic Migration and the Present Day Neighborhood Structure of Middlesbrough, United Kingdom. *Annals of the Association of American Geographers, 97*(1), 31–48. https://doi.org/10.1111/j.1467-8306.2007.00522.x

Pelham, B. W., Carvallo, M., DeHart, T., & Jones, J. T. (2003, Nov). Assessing the validity of implicit egotism: a reply to Gallucci (2003). *Journal of Personality and Social Psychology, 85*(5), 800–807. https://doi.org/10.1037/0022-3514.85.5.800

Pelham, B. W., Mirenberg, M. C., & Jones, J. T. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology, 82*(4), 469–487. https://doi.org/10.1037/0022-3514.82.4.469

Relethford, J. (1992). Analysis of Marital Structure in Massachusetts Using Repeating Pairs of Surnames. *Human Biology, 64*(1), 25–33.

Silberzahn, R., & Uhlmann, E. L. (2013, Dec). It pays to be Herr Kaiser: Germans with noble-sounding surnames more often work as managers than as employees. *Psychological Science, 24*(12), 2437–2444. https://doi.org/10.1177/0956797613494851

Stemmler, M., & Bäumler, G. (2003). Testing the statistical association between family names and success in certain athletic disciplines in men called 'Smith' or 'Tailor'. *Psychological Test and Assessment Modeling, 45*(2), 254.

Sutherland, A., Ilie, S., Vignoles, A., RAND Europe, & the University of Cambridge. (2015). *Factors associated with achievement: key stage 4 (DFE- RR407)*. Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/473673/RR407_-_Factors_associated_with_achievement_-_key_stage_4.pdf

Voracek, M., Rieder, S., Stieger, S., & Swami, V. (2015). What's in a Surname? Physique, Aptitude, and Sports Type Comparisons between Tailors and Smiths. *PLoS One, 10*(7), e0131795. https://doi.org/10.1371/journal.pone.0131795

Wilson, D., Burgess, S., & Briggs, A. (2009). The dynamics of school attainment of England's ethnic minorities. *Journal of Population Economics, 24*(2), 681–700. https://doi.org/10.1007/s00148-009-0269-0

Appendix

The examples in Tables 5 and 6 are taken from the FaNBI database; some but not all appeared in the results data set. Asterisks indicate names **not** listed in FaNBI as deriving from a British or Irish language (i.e., the Language/Culture field contained none of the following: "English", "Welsh", "Scottish", "Irish", "Manx", "Cornish", or "Norman"). The names with asterisks were therefore excluded in the analyses restricted to British/Irish language names.

Table 5: Examples of surnames with each surname indicator (not mutually exclusive).

Locative	Relationship		Occupational	Nickname	Status
Allen	Adams	Matthews	Bailey	Bell	Butler
Bailey	Allen	Mills	Baker	Brown	Chowdhury*
Bell	Anderson	Moore	Carter	Campbell	Fry
Benton	Anning	Morgan	Chamberlain	Chamberlain	Gentleman
Botham	Bell	Morris	Chapman	Cox	Guest
Bramley	Bennett	Owen	Clarke	Fry	Knight
Burgh	Brown	Pearson	Cook	Gray	Laird
Cox	Collins	Phillips	Cooper	Green	Lehmann*
Crawley	Cox	Powell	Cox	Guest	McIntosh
Darcy	Davies	Price	Fisher	Jenkins	Patel*
Fisher	Dickens	Richard	Foster	King	Prior
Graham	Edwards	Richardson	Grainger	Knight	Senior
Gray	Ellis	Robertson	Hunt	Lloyd	Stewart
Green	Evans	Robinson	Jagger	Mitchell	Tarrier
Hall	Foster	Roger	Knight	Moore	Tennant
Hill	Gibson	Russell	Marshall	Morris	Yoke
Holmes	Green	Scott	Mason	Palmer	
Jones	Griffiths	Simpson	Miller	Price	
Kelly	Harris	Swift	Parker	Prior	
Lee	Harrison	Thomas	Phillips	Reid	
Mills	Harvey	Thompson	Potter	Russell	
Moore	Hill	Watson	Prior	Senior	
Murray	Hughes	White	Smith	Shakespeare	
Newton	Jackson	Wilkinson	Stewart	Swift	
Russell	James	Williams	Taylor	Tarrier	
Shaw	Jenkins	Williamson	Tinker	Turner	
Simpson	Johnson	Wilson	Turner	White	
Smith	King		Walker	Wood	
Turner	Knight		Ward	Young	
White	Lewis		Webb		
Wood	Martin		Williams		
			Wright		

Table 6: Examples of surnames in each surname group (mutually exclusive categories).

Occupational (non-status)	Occupational and status	Status (non-occupational)	Non-status, non-occupational
Bailey	Ackerman	Agha*	Ahmed*
Baker	Batchelor	Alderman	Begum*
Carter	Fentiman	Baron	Brown
Clarke/Clark	Hackman	Bond	Davis/Davies
Cohen	Henman	Butler	Edwards
Cook	Hodgman	Chowdhury*	Evans
Cooper	Holder	Fouracre	Green
Fisher	Knight	Franklin	Harris
Fletcher	Maidman	Freeman	Hill
Foster	Master	Fry	Jackson
Harper	Monkman	Gentleman	Johnson
Hunter	Nutman	Headman	Jones
Kantor*	Paxman	Heritage	King
Mason	Parson	Laird	Lee
Parker	Prior	Le Maistre*	Lewis
Potter	Richter*	Pasha*	Martin
Slater	Sargent	Patel*	Moore
Smith	Servant	Portman	Morris
Spencer	Squire	Rabin	Roberts
Taylor	Stewart	Schultz*	Robinson
Turner	Swain	Tennant	Thomas
Walker	Swan	Vassall	Thompson
Ward	Tillman	Vavasour	White
Williams	Tubman	Villain	Wilson
Wright	Waterman	Yeoman	Wood