A Degree of Attraction:
Access to University and Partner Choice

Tomeu López-Nieto Veitch*[1] and Nicolau Martin-Bassols[2]

[1]Department of Statistical Sciences, University of Bologna
[2]Center for Economic and Social Research, University of Southern California

November 13, 2026

Abstract

Expanding access to higher education may reshape assortative mating by changing who individuals meet and partner with. However, establishing whether education changes who matches with whom is difficult, because observed spouse characteristics may themselves change within a given match. To isolate changes in partner composition, we use polygenic indices (PGIs) of the eventual spouse for education and cognition. These measures predict economically consequential characteristics but are fixed at conception and therefore cannot themselves respond to the focal individual's education. We exploit the post-war expansion of the British university system, linking the UK Biobank to a geocoded history of university provision, to instrument university degree attainment with the reduction in distance to the nearest university between birth and adolescence. Among individuals whose decision to obtain a degree responded to improved geographic access, we find that completing university leads to matches with spouses whose Educational Attainment, Cognitive Performance, and Highest Mathematics PGIs are 0.67, 1.02, and 0.52 standard deviations higher, respectively. It also substantially increases the probability of having a university-educated spouse and raises the spouse's occupation-based wages. By contrast, we find little evidence of systematic changes in within-couple similarity. Our results provide causal evidence that obtaining a university degree changes whom individuals match with, as revealed by the genetic characteristics of their spouses.

JEL Classification: I23, I26, J12, J24, D13.
Keywords: Assortative mating, higher education, university expansion, UK Biobank, genetic similarity, polygenic scores.

*Corresponding author: tomeu.lopeznieto@unibo.it. Department of Statistical Sciences, University of Bologna. All errors are our own. This research was supported by the European Social Science Genetics Network (ESSGN) under the European Union's Horizon 2020 research and innovation programme's Marie Skłodowska-Curie grant agreement (ESSGN 101073237). The views expressed in this publication are those of the authors and do not necessarily reflect those of the European Union, MSCA Horizon Europe, or ESSGN. Neither the European Union, nor the granting authority or ESSGN can be held responsible for them.

1 Introduction

Individuals tend to form partnerships with others who resemble them in characteristics such as education, earnings, and other dimensions of socioeconomic advantage. This phenomenon, known as assortative mating, is also documented across a much broader range of behavioural, cognitive, physical, and health-related traits (Horwitz et al., 2023). In economics, particular attention has focused on its role in shaping inequality. By concentrating similar levels of socioeconomic advantage within households, assortative mating can increase inequality across families (Greenwood et al., 2014; Eika, Mogstad, and Zafar, 2019).[1] These differences can then persist across generations, as households transmit economic and social advantages and disadvantages to their children (Fernández, Guner, and Knowles, 2005; Gayle, Golan, and Soytas, 2022; Collado, Ortuño-Ortín, and Stuhler, 2023). Assortative mating can therefore shape inequality both within and across generations, making it important to understand whether policies that change individual opportunities also reshape who matches with whom.

In this paper, we ask whether expanding access to university changes whom individuals subsequently partner with. However, establishing whether policy actually reshapes assortative mating is not straightforward. As formalised by Chiappori et al. (2025), changes in the distributions of characteristics such as education and income complicate comparisons of assortative mating. For instance, changes in spouses' education or earnings can make existing couples appear more or less alike, even when who partners with whom remains unchanged.[2] To distinguish changes in partner composition from changes within a match, we use the eventual spouse's polygenic indices (PGIs) for Educational Attainment, Cognitive Performance, and Highest Mathematics. These indices aggregate the predictive information contained in many genetic variants for a particular trait (or phenotype) and capture related but distinct dimensions of education and cognition, predicting broader socioeconomic outcomes beyond the traits they target.[3] Because PGIs are fixed at conception, an effect of obtaining a university degree on the spouse's PGI must reflect a change in whom the individual partners with rather than a change in characteristics generated after treatment. We combine these outcomes with quasi-experimental variation from the post-war expansion of the British university system. Linking the UK Biobank to a geocoded history of university provision, we use the reduction in distance to the nearest university between birth and age 15 as an instrument for obtaining a university degree.

We find that obtaining a university degree substantially changes the characteristics of the eventual spouse. Our estimates indicate that completing university leads individuals to match with spouses whose Educational Attainment, Cognitive Performance, and Highest Mathematics PGIs are 0.67, 1.02, and 0.52 standard deviations higher, respectively. Since these indices cannot themselves respond to the focal individual's education, the estimates provide direct evidence of a change in partner composition along education- and cognition-related dimensions. The resulting matches also differ sharply in realised socioeconomic characteristics: obtaining a university degree increases the probability that the spouse holds a university degree by 86 percentage points, raises the spouse's school-leaving age by 4.43 years, and

[1] Recent work by Fagereng, Guiso, and Pistaferri (2026) documents assortative mating on both premarital wealth and returns to wealth, highlighting how partner selection shapes the distribution of resources at marriage and their subsequent accumulation.

[2] Chiappori et al. (2026) illustrate this concern for income, showing that analyses using married couples' current incomes can yield misleading conclusions about sorting compared with analyses using premarital incomes.

[3] Education-related PGIs predict educational attainment and broader socioeconomic outcomes (Okbay et al., 2022; Papageorge and Thom, 2020; Barth, Papageorge, and Thom, 2020; Belsky et al., 2016), while cognitive-performance indices also predict educational and related socioeconomic outcomes (Lee et al., 2018; Becker et al., 2021). These applications form part of a broader literature in "genoeconomics" using genetic data to study socioeconomic outcomes, selection, heterogeneity, and gene–environment interplay (Benjamin et al., 2012, 2026; Biroli et al., 2026). We describe the PGIs in more detail in Section 4.3.

increases the spouse's occupation-based hourly wage by approximately 42 percent. By contrast, estimates of within-couple similarity based on education, wages, PGIs, and genome-wide kinship are generally imprecise, so we find little evidence that these substantial changes in partner composition translate into systematic changes in assortative resemblance.

These effects are large, and their local nature is important for interpretation. Our estimates apply to individuals whose decision to obtain a university degree responds to expansion-driven improvements in geographic access rather than to the average university graduate. We find that university completion responds most strongly to improved access among individuals born in higher-SES districts and those in the middle tertile of the EA PGI. Complementary analyses using alternative distance cutoffs suggest that those responding to improved access had relatively favourable observable prospects for university completion but grew up with less university provision nearby. Although these profiles are suggestive, they point to geographic barriers facing individuals otherwise well placed to pursue higher education. For these individuals, attending university may substantially alter the environment in which partnerships are formed by changing peers and social networks, geographic mobility, and the set of potential partners encountered. In a marriage market already strongly sorted by education, this provides a plausible interpretation of the large local effects on partner composition, although we do not separately identify this mechanism.

As a complementary benchmark, we exploit the 1972 Raising of the School Leaving Age (ROSLA), which increased compulsory schooling from age 15 to 16 and shifted education at a substantially lower margin than the university expansion. Geruso and Royer (2018) show that its marriage-market effects were concentrated around relatively low qualifications. We instead examine whether ROSLA affects spouses' university attainment, school-leaving age, wages, or our principal genetic outcomes. We find no effects comparable to those associated with university completion. Because the two reforms affect different educational margins and different complier populations, we do not interpret this contrast as a formal comparison of treatment effects. Rather, it suggests that the marriage-market consequences of additional education depend on the margin at which schooling is shifted.

Our first contribution is to the literature on education and marriage markets. A long-standing sociological literature emphasises that organised settings, and educational institutions in particular, shape whom people meet and therefore whom they are able to match with (Kalmijn and Flap, 2001; Blossfeld and Timm, 2003). Economic models likewise recognise that schooling can generate marriage-market as well as labour-market returns.[4] Causal evidence shows that education changes the educational characteristics of eventual spouses (Lefgren and McIntyre, 2006), that compulsory schooling can shift matching at lower qualification margins (Geruso and Royer, 2018), and that access to particular fields, institutions, and elite programmes changes subsequent partner formation (Artmann et al., 2021; Kirkebøen et al., 2025; Barrios-Fernández, Neilson, and Zimmerman, 2026). Moreover, Abramitzky, Pérez, and Price (2026) show that higher-education institutions can themselves operate as marriage markets, with greater exposure to opposite-sex classmates and longer overlap on campus increasing the probability of marrying a fellow student.

Hence, expanding access to higher education may also reshape patterns of partnership formation.[5]

[4] In Becker's assignment framework, individuals sort on characteristics that affect the surplus generated by a match (Becker, 1973, 1974). Chiappori, Iyigun, and Weiss (2009) show explicitly that schooling can carry a marriage-market return, while Chiappori, Costa Dias, and Meghir (2018) formalise how education choices may reflect both labour- and marriage-market opportunities. Complementary evidence shows that educational qualifications and pre-existing ability can jointly shape equilibrium matching (Anderberg and Zhu, 2010).

[5] A broader literature studies the consequences of improved university access through expansions in university sites and programmes. Examples include the expansion of university sites and programmes in Italy (Bratti, Cottini, and Ghinetti, 2022),

Recent studies document such responses. Suhonen and Karhunen (2019) find that improved university access in Finland increased the likelihood that women formed families with university-educated men, Nybom et al. (2023) show that Swedish higher-education expansions increased sorting between partners by skill, and Fang and Luo (2026) find that China's expansion increased educational assortative matching. We extend this evidence to the education- and cognition-related genetic characteristics of eventual spouses. Because these characteristics are fixed before treatment, they also allow us to distinguish changes in partner composition from changes in outcomes realised within an existing match.

By examining the genetic characteristics of eventual spouses, we also contribute to the literature on genetic assortative mating and partner selection. Existing genetic evidence shows that spouses are more genetically similar than randomly paired individuals and that assortative mating is reflected in correlations in trait-associated genetic variation between partners (Domingue et al., 2014; Robinson et al., 2017; Yengo et al., 2018). More recently, Hofmeister et al. (2026) show that genetic assortment on education has strengthened across more recent generations, demonstrating that patterns of genetic sorting between partners can change over time. Closest to our causal question, Abdellaoui et al. (2023) show that earlier-born siblings tend to match with spouses with higher Educational Attainment PGIs and provide evidence that this relationship operates through socioeconomic status, particularly education and income. Barban et al. (2021) also combine genetic data with spouse outcomes, using the EA PGI as an instrument for own schooling to estimate its effect on spouse education. We instead place the spouse's PGIs on the outcome side of the analysis. To the best of our knowledge, we provide the first causal evidence that an education policy changes whom individuals partner with along genetic dimensions related to education and cognition.

Taken together, our findings connect these strands of the literature. Work on education and marriage markets, including studies of higher-education expansions, shows that schooling and university access can shape matching opportunities and partner outcomes; the genetic-assortment literature documents sorting on genetic characteristics. We bring these insights together by showing that, for individuals who complete university as a result of improved access, the degree changes the genetic and socioeconomic characteristics of the eventual spouse. Because the genetic outcomes are fixed before treatment, they reveal re-sorting in who becomes the spouse rather than changes generated within an otherwise unchanged match. More broadly, the results identify household formation as an additional margin through which expansions in higher education can shape the distribution and intergenerational transmission of socioeconomic advantage.

The remainder of the paper is organised as follows. Section 2 formalises how genetic spouse outcomes identify re-sorting in partner composition and distinguishes this from changes in within-couple similarity. Section 3 describes the post-war expansion of British higher education and the Raising of the School Leaving Age. Section 4 presents the UK Biobank data, couple construction, genetic measures, and university-access data. Section 5 sets out the empirical strategy. Section 6 reports the results, and Section 7 concludes.

the geographic expansion of higher education in Finland (Toivanen and Väänänen, 2016), the West German higher-education expansion (Westphal, Kamhöfer, and Schmitz, 2022), higher-education openings and local skill supply in Norway (Carneiro, Liu, and Salvanes, 2023), and changes in university accessibility in Finland (Dang, Haapanen, and Suhonen, 2026).

2 Education, Partner Composition, and Genetic Outcomes

We now formalise how genetic outcomes allow us to distinguish changes in *who becomes the spouse* from changes in outcomes realised by a given partner.

Let $D_i \in \{0, 1\}$ denote individual i's university attainment, and let $P_i(d)$ denote the partner with whom i would form a couple under treatment state d. Partner identity is therefore itself potentially treatment-dependent. For a spouse characteristic T, let $T_{ij}(d)$ denote the value of that characteristic for potential partner j if individual i were in treatment state d and the two formed a couple. For instance, if T denotes earnings, $T_{ij}(d)$ allows the earnings of the same potential partner j to differ depending on whether i attended university, even when i and j remain matched. The causal effect of individual i's university attainment on the characteristic of their eventual spouse is therefore

$$\tau_i^T = T_{iP_i(1)}(1) - T_{iP_i(0)}(0) \tag{1}$$

where τ_i^T can reflect either a change in who individual i partners with or a change in the value of T realised by a given potential partner, or a combination of both. We can therefore decompose τ_i^T as

$$\tau_i^T = \underbrace{T_{iP_i(1)}(1) - T_{iP_i(0)}(1)}_{\text{partner re-sorting}} + \underbrace{T_{iP_i(0)}(1) - T_{iP_i(0)}(0)}_{\text{partner-outcome response}}. \tag{2}$$

For outcomes such as education or earnings, the two components in Equation (2) cannot generally be separated. University may change who becomes the spouse, but it may also affect the education or earnings ultimately realised by a given partner. An effect on spouse education or earnings therefore shows that university changes the socioeconomic characteristics of the resulting match, but does not by itself establish that a different individual becomes the spouse.[6]

Genetic characteristics provide a useful special case.[7] In the empirical analysis, we observe them through polygenic indices (PGIs), which summarise genetic variation associated with particular traits.[8] Let G_j denote the PGI of potential partner j. Then, because genotype is fixed at conception, we have that

$$G_{ij}(d) = G_j \qquad \text{for every } i, j, d. \tag{3}$$

Therefore, the partner-outcome component of Equation (2) disappears, leaving

$$\tau_i^G = G_{P_i(1)} - G_{P_i(0)}. \tag{4}$$

[6] Equation (2) uses $d = 1$ as the reference state for the re-sorting component. An equivalent decomposition can be written using $d = 0$, so the allocation of the total effect between the two components is not unique for a generic realised outcome. This ambiguity disappears for genetic characteristics.

[7] The same logic applies to other pre-match characteristics that cannot be affected by the focal individual's university attendance.

[8] Section 4.3 describes the construction and interpretation of the PGIs used in the analysis.

Equation (4) is the key object underlying our empirical analysis. University cannot change the genetic characteristic of an otherwise unchanged potential spouse. A causal effect on the spouse's PGI must therefore arise because treatment changes which individual becomes the spouse, and hence reveals re-sorting along that genetically indexed dimension of partner composition.

This interpretation does not require PGIs to measure causal genetic effects, nor does it imply that individuals observe genetic information or choose partners on the basis of genotype. PGIs are useful here because they predict economically consequential characteristics but, being fixed at conception, cannot themselves respond to the focal individual's education. Our Educational Attainment, Cognitive Performance, and Highest Mathematics PGIs therefore allow us to ask whether university changes whom individuals match with along dimensions associated with education and cognition. Spouse education and wages provide complementary evidence on the socioeconomic characteristics of the matches that result, although effects on these outcomes need not arise solely through partner re-sorting.

However, a change in who becomes the spouse need not make the resulting couple more or less similar. University could lead an individual to match with a different type of partner while leaving conventional measures of within-couple resemblance largely unchanged. We therefore treat within-couple similarity as a separate, secondary outcome, asking whether the re-sorting identified through partner characteristics is accompanied by a systematic change in resemblance between spouses.

In what follows, we first use spouse PGIs to study whether university changes who becomes the spouse, and spouse education and wages to characterise the resulting matches. We then examine whether these changes are accompanied by differences in within-couple similarity. The next two sections describe the institutional setting and data before we turn to the empirical strategy.

3 Institutional Background

Two major education-policy changes affected the birth cohorts studied in this paper. The first, and central to our research design, was the post-war expansion of higher education. This expansion increased the number and geographic dispersion of universities, generating variation across birthplaces and birth cohorts in access to university during childhood. The second was the 1972 Raising of the School Leaving Age (ROSLA), which increased compulsory schooling from age 15 to age 16 for pupils born on or after 1 September 1957.

3.1 The Post-War Expansion of British Higher Education

In the years following the Second World War, British higher education remained small, selective, and geographically concentrated (Anderson, 2006; Carpentier, 2006). In 1962, only around 4 percent of each generation entered university, with access varying substantially by socioeconomic background, sex, and region. For many prospective students, university attendance required moving away from home and bearing accommodation and travel costs. In 1961–62, only around 20 percent of full-time university students lived at home, while most relied on private lodgings or university accommodation (Robbins, 1963, pp. 194–195). In this initially sparse system, a nearby institution could reduce the cost of attendance and widen the range of feasible university options.

University provision expanded progressively from the 1950s and accelerated during the 1960s. Al-

though closely associated with the 1963 Robbins Report, which called for higher education to be available to all suitably qualified applicants, expansion was already under way before its publication (Robbins, 1963; Willetts, 2014). Growth came both from new universities and from existing higher-education institutions acquiring university status, notably the Colleges of Advanced Technology (CATs). Together with growth at established universities, these developments substantially increased the scale of higher education. University enrolment rose from around 150,000 in the mid-1960s to nearly 300,000 by the early 1980s (Pratt, 1997), while annual first degrees awarded more than doubled between 1960 and 1970 (Bolton, 2012; Calhoun, 2014).

The geographic expansion was uneven. The first wave of new universities favoured cities without existing university provision, while subsequent plans placed greater emphasis on larger population centres (Robbins, 1963, pp. 162–163). The availability of suitable sites and existing educational institutions also shaped where the sector grew. As a result, university opportunities expanded at different times and to different degrees across the country.

3.2 The Raising of the School Leaving Age (ROSLA)

Alongside the expansion of higher education, Britain also increased compulsory schooling at the lower end of the education distribution. The minimum school-leaving age had been 15 since 1947, but the 1972 Raising of the School Leaving Age (ROSLA) reform increased it to 16 from the beginning of the 1972–73 school year (United Kingdom Parliament, 1944, 1972). Pupils born before 1 September 1957 could therefore leave school at 15, whereas those born on or after that date were required to remain for another year. Because the school year ran from September to August and grade repetition or acceleration was uncommon, birth date closely determined when pupils could legally leave school (Clark and Royer, 2013; Sharp, Hutchison, and Keys, 2002). Before the reform, approximately 35 percent of pupils left school at the minimum legal age, and ROSLA substantially reduced this share (Geruso and Royer, 2018; Banks, Carvalho, and Perez-Arce, 2019).

The reform primarily affected pupils who would otherwise have left school at 15, particularly those from less advantaged backgrounds. Pupils already intending to remain in education beyond that age were largely unaffected. ROSLA therefore operated at a markedly different educational margin from the university expansion. Consistent with this distinction, evidence indicates that the reform increased schooling without affecting university completion (Barcellos, Carvalho, and Turley, 2018; Geruso and Royer, 2018). This makes ROSLA a useful benchmark for examining whether changes in partner composition also arise when education increases at the compulsory-schooling margin.

4 Data and Measures

4.1 Data Sources

4.1.1 UK Biobank

Our main data source is the UK Biobank (UKB), a large-scale biomedical database and research resource containing genetic, health, and socioeconomic information for roughly half a million participants aged

40–69 at recruitment between 2006 and 2010 (Sudlow et al., 2015). The UKB is particularly well suited to our setting for two reasons. First, it is one of the largest publicly available resources that contain genome-wide genetic data linked to rich information on education, labour-market outcomes, and place of birth. Second, its participants belong to cohorts born during the period of post-war higher-education expansion, allowing us to link individuals' early-life exposure to local university openings to their later-life partnership outcomes.

Our outcomes of interest fall into three groups. First, we measure the educational attainment of the spouse. UKB participants were asked whether they held a college or university degree and reported the age at which they completed full-time education. We therefore use two spouse-level education outcomes: an indicator for whether the spouse holds a university degree and the spouse's school-leaving age (SLA), as recorded in the UKB.[9] The university indicator captures attainment at the higher-education margin, while school-leaving age provides a continuous measure of educational attainment. Second, we use genetic outcomes to capture dimensions of partner composition and within-couple similarity constructed from characteristics that are fixed at conception. Our partner-composition outcomes are the spouse's Educational Attainment PGI and the Cognitive Performance and Highest Mathematics PGIs. We additionally examine within-couple similarity in these indices and spouse-pair genetic kinship. We describe the construction and interpretation of these measures in Section 4.3. Third, we examine the spouse's wages. The UKB does not report wages directly, so we construct an occupation-based wage proxy. Participants report their current occupation, or their most recent occupation if retired, and these occupations are coded into more than 400 Standard Occupational Classification (SOC) categories. Following Kweon et al. (2020), we impute log hourly wages from each participant's current or most recent occupation (4-digit SOC), using their model that combines occupation-specific mean and median hourly wages from the Annual Survey of Hours and Earnings (ASHE) with sex, age and year. Because the measure is assigned at the occupation level, it should be interpreted as a proxy for individual wages rather than as directly observed wages.

Participation in the UKB was voluntary and concentrated over a short recruitment window, resulting in a sample that is not nationally representative. In particular, respondents are disproportionately drawn from higher-education, higher-socioeconomic-status, and healthier populations (Swanson, 2012; Fry et al., 2017). To mitigate selection bias arising from voluntary participation, we use inverse probability weights based on UK Census microdata. These weights are designed to make the UKB sample more representative of the underlying population and have been shown to reduce selection-related bias by almost 90% (van Alten et al., 2024).

Identifying Couples. The UKB was not designed around household structure, so couples must be inferred from the available covariates. We follow Abdellaoui et al. (2023) and classify two UKB participants as a couple when they satisfy the following joint conditions: (i) they report the same home postcode on at least one assessment occasion[10]; (ii) they report the same homeownership or renting status, (iii) the same length of time at the address, and (iv) the same number of children; (v) they attended the same UKB assessment centre on the same day; (vi) both report living with their spouse; and (vii) the pair consists of one male and one female. We also eliminate all pairs where either spouse appeared more than once in the data. These restrictions minimise the risk of spuriously matching co-resident non-partners, such as flatmates or relatives, and yield a sample of couples for whom we can measure partner outcomes

[9] We follow Davies et al. (2018) and code university graduates as having left full-time education at 21.

[10] A UK postcode typically covers approximately 15 addresses (Office for National Statistics, 2023).

and within-couple similarity both phenotypically and genetically. In Appendix A, we assess the validity of this matching procedure using genetically identified children and examine the direction of bias from any remaining false positives. The results strongly support the accuracy of the procedure and suggest that any residual false couples would, if anything, attenuate our estimates.

4.1.2 Historical District-Level Data

We complement the individual-level UKB data with district-level measures from the *Labour Gazette*, the 1951 Census, and the Great Britain Historical Database (Southall and Dorling, 2021). These measures capture the socioeconomic context of the district of birth and serve two purposes. First, the UKB contains limited information on individuals' childhood family background and early-life socioeconomic environment. District-level measures therefore provide proxies for the conditions under which individuals grew up, including local human capital, labour-market conditions, population density, and socioeconomic composition. Second, university openings were not randomly assigned across space. New institutions were often located in areas selected for regional need, under-provision, local demand, or broader development considerations. These same factors may be correlated with educational trajectories, marriage markets, and later-life outcomes. Predetermined district-level controls therefore help account for both early-life background differences across individuals and the non-random geography of university expansion.

To construct these controls, we combine several sources of district-level historical data.[11] We digitise unemployment rates from the *Labour Gazette*, which provide a measure of local labour-market conditions in the district of birth before or around the period of university expansion. We use the 1951 Census to measure the district's baseline educational, socioeconomic, and urban structure. Educational composition is measured by the share of adult males who remained in full-time education until at least age 15, the statutory school-leaving age at the time.[12] Urban structure is captured by the district's urban/rural classification. Socioeconomic composition is measured using the Registrar General's occupational social-class scheme reported in the 1951 Census. We define high-SES composition as the share of individuals in Social Classes I and II, corresponding to professional, managerial, and related occupations, and low-SES composition as the share in Social Classes IV and V, corresponding to partly skilled and unskilled occupations. Social Class III, which contains the bulk of skilled manual and non-manual workers, is left as the omitted middle category. Finally, we use population measures from the Great Britain Historical Database (GBHD) to capture differences in population size, population density, and demographic growth across birth districts. Together, these variables allow us to condition on pre-expansion local characteristics that may have shaped both individuals' educational and partnership outcomes and the placement of new universities.

[11] To avoid potential bad-control concerns, we measure district-level controls before the relevant expansion wherever possible. Following a similar logic to Abramitzky, Boustan, and Giuntella (2025), baseline district characteristics are fixed to 1951, and demographic growth is computed using population changes up to 1951, so that post-expansion local changes do not enter the conditioning set.

[12] The 1951 Census reports comparable education tabulations for adult males across districts, while corresponding female education tabulations are only available for towns above 50,000 inhabitants. We therefore use male educational composition as a consistent district-level proxy for the share of adults remaining in full-time education until at least age 15, incorporating female measures where comparable data are available.

4.1.3 Higher-Education Institutions Data

Alongside the individual-level UKB records, we assemble an original dataset covering higher-education institutions in England and Wales. The backbone of this dataset is the official UK government registry of bodies entitled to award degrees.[13] We manually digitise the registry and complement it with supplementary historical sources. For each institution, we record its institutional category, the year in which it acquired university status, the year of closure where applicable, and the geographic coordinates of its principal campus.

The empirical analysis focuses on institutions with university status, since these correspond most closely to the tertiary-education category recorded in the UKB. We therefore date an institution's entry into the university-access measure by the year in which it acquired university status, rather than by its original founding year. An institution is defined as active in year t if it had acquired university status by year t and had not closed by that year. This coding allows us to reconstruct the stock of operating universities in each calendar year.

Figure 1 summarises the spatial and temporal distribution of university openings in England and Wales during the period under study. Prior to 1945, universities were concentrated in historic university towns and a limited number of large industrial cities, leaving substantial parts of the country without nearby provision. The post-war expansion partially redressed this imbalance. New openings between 1945 and 1980 reached previously underserved areas, including East Anglia, the South Coast, the South West, parts of the North East, and the outer ring around London. The number of active universities increased from roughly 95 at the end of the 1940s to just over 150 by 1980. This staggered and geographically diffuse expansion provides the institutional variation used below to measure changes in access to university during childhood.

4.2 Measuring Changes in University Access

We construct an individual-level measure of the expansion-induced change in geographic access to university during childhood.[14] To construct the instrument, we combine our newly assembled dataset of higher-education institutions with geocoded birthplace coordinates from the UK Biobank. For each individual, we compute the distance from the birthplace to the nearest university operating in the year of birth and the distance from the same birthplace to the nearest university operating by the year in which the individual turned 15.[15] Formally, we define the reduction in distance to the nearest university as

$$\Delta Dist_i = Dist_i^{\text{birth}} - Dist_i^{\text{age 15}} \tag{5}$$

where $Dist_i^{\text{birth}}$ is the Euclidean distance from individual i's birthplace to the nearest university operating in the year of birth, and $Dist_i^{\text{age 15}}$ is the distance from the same birthplace to the nearest

[13] *Source:* GOV.UK guidance on recognised degree-awarding bodies. Further details on the construction of the dataset, institutional classifications, geocoding procedure, and treatment of closures and mergers are reported in Appendix F.

[14] This approach builds on studies exploiting higher-education expansions as a source of variation in university access (e.g. Currie and Moretti, 2003; Toivanen and Väänänen, 2016; Suhonen and Karhunen, 2019; Bratti, Cottini, and Ghinetti, 2022; Braccioli et al., 2024; Ichino, Rustichini, and Zanella, 2026), extending earlier work using pre-existing proximity to universities (e.g. Card, 1995; Kling, 2001; Carneiro, Heckman, and Vytlacil, 2011).

[15] We use age 15 because it is close to the timing of educational decisions relevant for higher-education trajectories while avoiding direct overlap with the Raising of the School Leaving Age (ROSLA). Table A.8 shows that the results are robust to alternative exposure windows.

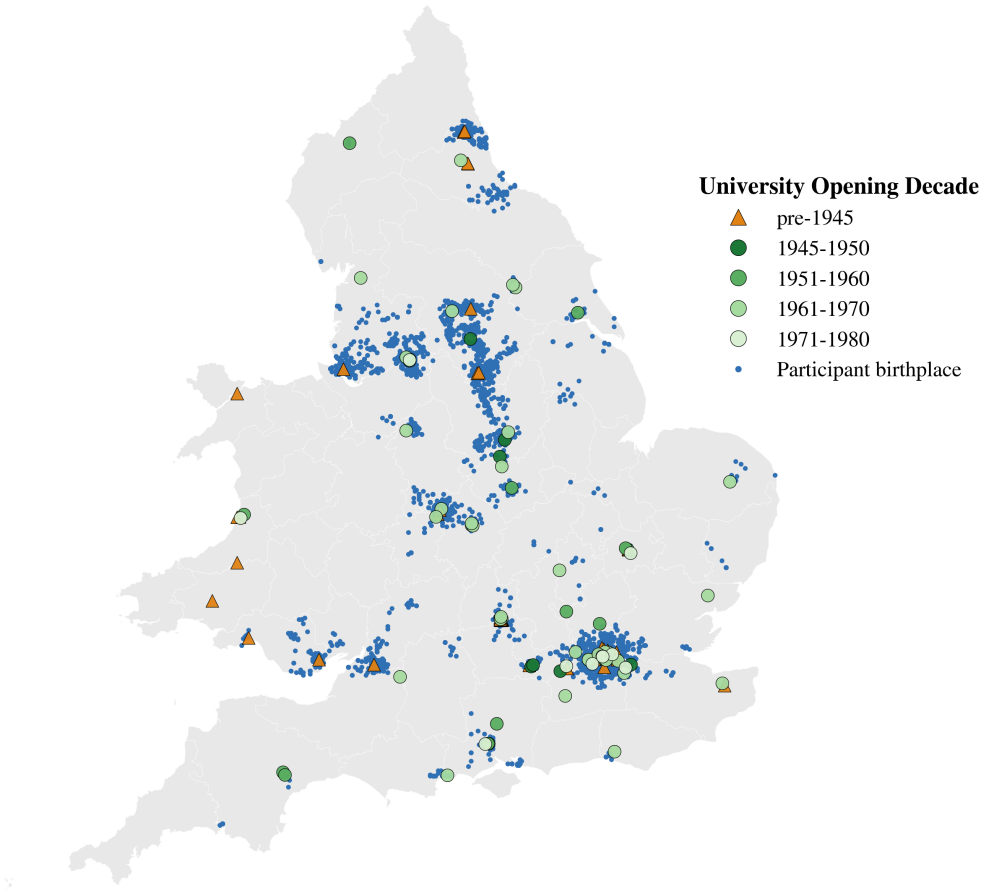


Figure 1: Spatial and Temporal Distribution of University Openings in England and Wales. The map displays the location of each university campus, with marker style and colour indicating the decade in which the institution acquired university status. Orange triangles correspond to universities established before 1945, while circles in progressively lighter shades of green denote post-war openings in 1945–1950, 1951–1960, 1961–1970, and 1971–1980, respectively. Small blue points represent the birth locations of UKB participants included in the analysis sample, illustrating the spatial overlap between the population of interest and the geography of higher education provision. Light grey boundaries trace the 1951 county geography (*source*: author's calculations from HEI data).

university operating by the year in which individual i turned 15. A value of zero indicates that no closer university became available during childhood. Positive values indicate that the expansion reduced the distance to the nearest university, with larger values corresponding to larger improvements in local access to higher education. Negative values indicate that the distance to the nearest operating university increased during childhood, for example because the nearest university ceased operating.

This measure has two advantages over a conventional proximity instrument. First, it exploits changes in access generated by the expansion rather than permanent cross-sectional differences in proximity to universities. Second, relative to discrete opening or availability indicators, it captures the intensity of the access shock: an opening that reduces the distance to the nearest university by 5 kilometres represents a smaller improvement in access than one that reduces it by 50 kilometres. Hence, the

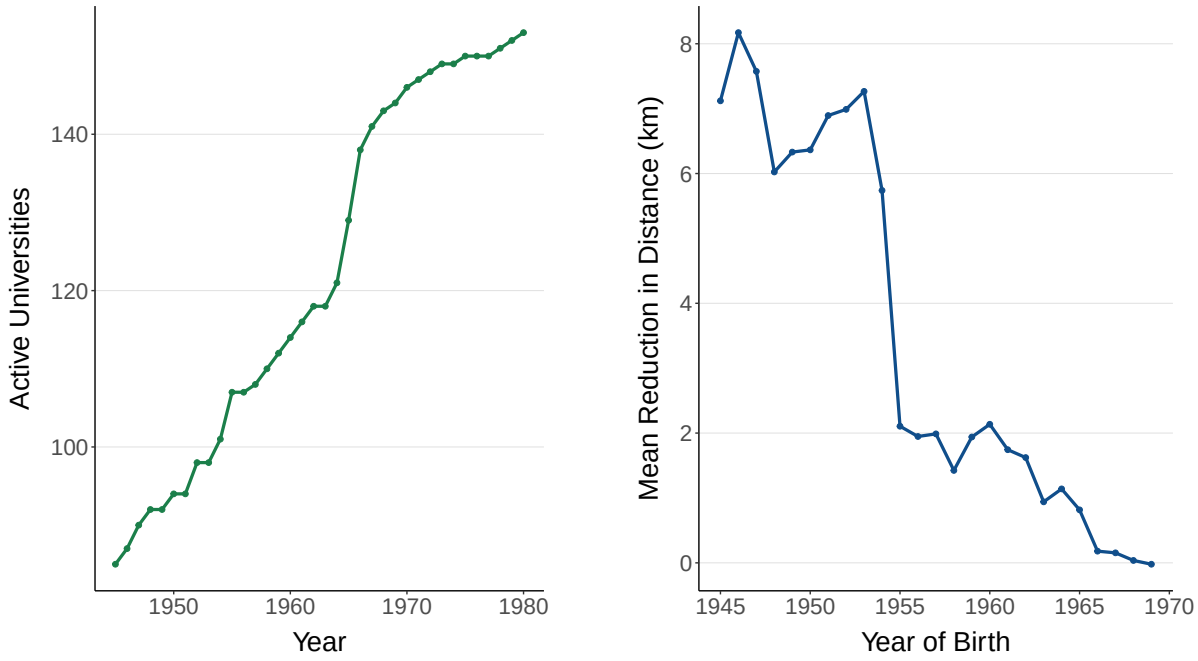


Figure 2: University Expansion and Reduction in Distance to University. The left panel shows the number of active universities in England and Wales by year (source: HEI). The right panel shows the average reduction in distance to the nearest university by birth year in the UKB analysis sample.

economic interpretation of the instrument is that a larger reduction in distance lowers the cost of attending university.

The UKB records individuals' place of birth but not their place of residence during adolescence. We therefore assign university access using birthplace coordinates as a proxy for childhood location. This measure may misclassify exposure for individuals who moved before reaching university-entry age, since the change in access experienced at their actual residence may have differed from that assigned at their birthplace. However, historical mobility patterns suggest that this source of measurement error is limited for the cohorts in our sample.[16] Moreover, if childhood mobility is not systematically related to the expansion-induced change in university access, conditional on the controls in our specification, birthplace-based assignment should primarily make the instrument a noisier measure of the access individuals actually experienced. The resulting measurement error would tend to weaken the first stage rather than generate a systematic bias in a particular direction.

[16] Our calculations from the 1961 Census suggest that 4.7% of children aged 0–14 in England and Wales moved across districts in the preceding year, a proportion similar to that among adults. Using UK Biobank siblings, von Hinke and Vitt (2024) estimate annual childhood moving rates of 1.2% for distances of at least one kilometre, 0.9% across districts, and 0.6% beyond 50 kilometres. These sibling-based estimates capture moves between births and may therefore understate overall childhood mobility.

4.3 Genetic Measures

4.3.1 Polygenic Indices (PGIs)

To measure predetermined education- and cognition-related characteristics of spouses, we use polygenic indices (PGIs). A PGI summarises predictive information contained in a large number of genetic variants into a single measure for a particular trait or outcome (phenotype). Complex behavioural and socioeconomic traits are highly polygenic: rather than being associated with a small number of variants, they are predicted by the combined information contained in many variants whose individual associations are typically small. Genome-wide association studies (GWAS) estimate these associations in large discovery samples, and PGIs combine them using the resulting association estimates as weights (Becker et al., 2021). We standardise each index to have mean zero and unit variance in the analysis sample.[17]

PGIs are predictive measures rather than estimates of the causal effect of an individual's genotype on an outcome. The associations used to construct them may reflect direct genetic effects, but also indirect genetic effects operating through relatives and the environments they provide, assortative mating, and residual population structure.[18] The indices should therefore be interpreted as predictors associated with their respective outcomes rather than as measures of innate ability or direct genetic causation.

These interpretive limitations do not affect the property of PGIs that is central to our design. An individual's genotype, and therefore their PGI, is fixed at conception. Obtaining a university degree cannot change the PGI of a given potential spouse. Conditional on the validity of the empirical design, an effect on the spouse's PGI must therefore operate through a change in which individual becomes the spouse rather than through a change in the genetic characteristics of an otherwise unchanged partner. PGIs consequently allow us to observe re-sorting along predetermined dimensions of partner composition.

Our main analysis uses three PGIs capturing related but distinct dimensions of education and cognition: Educational Attainment, Cognitive Performance, and Highest Mathematics. The Educational Attainment PGI predicts years of completed education. The Cognitive Performance PGI is based on fluid intelligence, measured in the UK Biobank by the number of correct answers to 13 verbal and numerical reasoning questions in a two-minute test, while the Highest Mathematics PGI is anchored to the highest mathematics class taken and therefore captures progression through mathematics coursework rather than performance on a particular test (Lee et al., 2018).[19] Together, the three indices allow us to assess whether changes in partner composition extend beyond a single education-related genetic measure.

To reduce confounding from population structure, we restrict the analysis to participants of European ancestry and condition on genetic principal components in all specifications involving genetic outcomes (Price et al., 2006). Appendix B.1 describes the construction of each index and the underlying discovery samples.

[17] Formally, a conventional PGI can be written as $PGI_i = \sum_{j=1}^{J} \hat{\beta}_j G_{ij}$, where G_{ij} denotes individual i's allele count at variant j and $\hat{\beta}_j$ is the corresponding association estimated in the discovery GWAS. Appendix B.1 provides further details on GWAS and score construction.

[18] See Kong et al. (2018) on indirect genetic effects, Robinson et al. (2017) and Yengo et al. (2018) on assortative mating, and Price et al. (2006) on population stratification.

[19] None of our couples or their relatives contributes to the discovery samples of these indices. The Educational Attainment PGI is based on the EA3 GWAS (Lee et al., 2018) re-estimated without UK Biobank participants; the Cognitive Performance PGI on our own GWAS of fluid intelligence within the UK Biobank, which excludes our couples and their relatives; and the Highest Mathematics PGI on the single-trait index of the PGI Repository (Becker et al., 2021), whose discovery sample contains no UK Biobank participants. Appendix B.1 describes their construction and validation.

4.3.2 Genome-Wide Genetic Similarity: Kinship

To capture a broader form of spousal genetic similarity, we use the *kinship* coefficient. While PGIs measure genetically indexed association with specific traits, the kinship coefficient measures overall genome-wide relatedness between two individuals. It is a standard measure of genetic relatedness in genome-wide data and has been used in prior studies to quantify genetic similarity between spouses.[20]

For a pair of individuals, the kinship coefficient is the probability that two alleles, one drawn at random from each individual at the same locus, are identical by descent (Lynch and Walsh, 1998; Manichaikul et al., 2010). Two alleles are said to be identical by descent (IBD) when they are copies of the same ancestral allele inherited through a common ancestor. Put differently, the kinship coefficient is a scalar measure of genome-wide genetic relatedness, with larger values indicating greater expected sharing of alleles inherited from common ancestors. As a result, unrelated individuals have expected kinship close to zero, whereas first-degree relatives, such as parent–child pairs or full siblings, have expected kinship of 0.25. Second-degree relatives have expected kinship of 0.125. Values above zero therefore indicate that two individuals are more genetically related than would be expected for two unrelated members of the same reference population.[21]

In our setting, kinship complements the similarity measures constructed from the PGIs. The PGI measures ask whether obtaining a university degree in response to improved access changes within-couple resemblance along specific genetically indexed dimensions associated with education and cognition. Instead, kinship captures a broader margin of genetic resemblance that is not tied to any particular phenotype. Therefore, it allows us to ask whether a university degree changes the overall genetic relatedness of the resulting match, beyond similarity in the specific traits captured by the PGIs. Since education cannot change either partner's genotype, any effect on spouse-pair kinship must operate through a change in which individuals are paired together.

We estimate spouse-pair kinship from quality-controlled autosomal SNPs using the KING-robust estimator (Manichaikul et al., 2010; Chang et al., 2015). Before estimating relatedness, we restrict the genotype data to high-quality autosomal variants and prune the SNP set for linkage disequilibrium, so that the estimate is not driven by clusters of highly correlated markers. The KING-robust estimator is designed to estimate recent genetic relatedness while being relatively insensitive to population structure. Nevertheless, all specifications involving genetic outcomes control for the first 20 genetic principal components. The resulting spouse-pair kinship estimate is our measure of genome-wide genetic similarity between partners. For the empirical analysis, we standardise this measure to mean zero and unit variance so that effects are expressed in standard-deviation units. Details on the genotype filters, pruning procedure, implementation, and sample construction are reported in Appendix B.2.

[20] For instance, Domingue et al. (2014) compare kinship among observed spouses with kinship among randomly paired individuals and find greater genome-wide genetic similarity among spouses. Sebro et al. (2017) similarly use spouse-pair kinship to study structured mating and emphasise the role of ancestry-related population structure in generating observed genetic resemblance.

[21] Estimated kinship coefficients can take slightly negative values because of sampling variation. These should be interpreted as indicating no detectable excess relatedness, rather than negative biological relatedness.

4.4 Sample Construction and Descriptive Statistics

We restrict the analysis to UKB participants born between 1945 and 1969 in England and Wales.[22] Additionally, we restrict the sample to individuals of European ancestry to reduce concerns about population stratification. Participants must have genome-wide genetic data, a geocoded place of birth, and the information required to construct exposure to university openings during childhood. We also require the predetermined individual- and district-level covariates included in the empirical specifications. These covariates account for baseline differences across birth locations that may be associated with both university placement and subsequent educational and partnership outcomes.

Furthermore, we restrict the sample to participants for whom we can identify a spouse within the UKB using the matching procedure described in Section 4.1. The resulting sample consists of complete male–female spouse pairs for whom both partners have the genetic, socioeconomic, geographic, and outcome information required for the analysis. We use this common complete-couple sample throughout, so that estimates for spouse-level outcomes and within-couple similarity outcomes are directly comparable and are not driven by changes in sample composition across outcomes. The final sample contains 12,494 individuals, corresponding to 6,247 couples.

We report the descriptive statistics for the main variables in our analysis in Table 1. In our sample, 32.9 percent of individuals are university graduates, the average school-leaving age is 18.41 years, and the average occupation-based hourly wage is £14.90. The three PGIs are standardised measures and are thus centred around zero, with standard deviations close to one. Because each member of a complete couple appears once as the focal individual and once as the spouse, the pooled spousal characteristics in column (1) also describe the corresponding marginal distributions of the analysis sample. Appendix Table A.4 reports a broader set of descriptive statistics for the analysis sample and the covariates used in the empirical specifications.

Panel B describes our instrument. Distance to the nearest university fell by 3.6 km on average between birth and age 15, with a standard deviation of 11.09 km, and 27.6 percent of individuals experienced a strictly positive reduction. Among those who experienced a reduction in distance, the average reduction was 13.1 km, with values ranging from just above 0 to 92 km (see Table A.5). Moreover, consistent with the timing of openings, earlier cohorts experienced larger reductions in distance on average, as shown in Figure 2. The university expansion therefore generated substantial variation in access in our sample.

Columns (2) to (4) reveal substantial differences in partner characteristics by university-completion status. Among non-graduates, 20.4 percent of spouses complete university, compared with 58.9 percent among university graduates, a difference of 38.5 percentage points. Spouses of university graduates also leave full-time education 1.81 years later and earn £2.02 more per hour than spouses of non-graduates. The same pattern also extends to genetic characteristics. Spouses of university graduates score 0.294 standard deviations higher on the Educational Attainment PGI, 0.218 standard deviations higher on the Cognitive Performance PGI, and 0.286 standard deviations higher on the Mathematics PGI. University graduates also experienced larger improvements in university access, with reductions in distance 2.27 km greater on average and a 5.0-percentage-point higher probability of experiencing any reduction. However, these descriptive differences cannot distinguish the causal effect of obtaining a university degree on partner outcomes from selection into university, motivating the instrumental-variables analysis that follows.

[22] Scotland is excluded because comparable pre-expansion district-level covariates are not consistently available for Scottish birth locations. Its distinct secondary-education, university-entry, and degree structure during the period studied also limits comparability with England and Wales.

Table 1: Descriptive Statistics by University-Completion Status

	(1) Mean (SD)	(2) Non-Graduates	(3) University Graduates	(4) Difference (3) - (2)
Panel A: Spousal Characteristics				
Spouse University Degree	0.329 (0.470)	0.204 (0.403)	0.589 (0.492)	0.385 (0.016)
Spouse School-Leaving Age (SLA)	18.41 (2.49)	17.83 (2.40)	19.63 (2.22)	1.81 (0.07)
Spouse Hourly Wage (£)	14.90 (5.09)	14.25 (4.87)	16.27 (5.26)	2.02 (0.14)
Spouse EA PGI	−0.012 (0.988)	−0.107 (0.968)	0.187 (1.000)	0.294 (0.025)
Spouse Cog PGI	−0.008 (1.004)	−0.079 (1.001)	0.139 (0.995)	0.218 (0.025)
Spouse Math PGI	−0.018 (0.997)	−0.110 (0.974)	0.176 (1.017)	0.286 (0.026)
Panel B: Exposure to University Openings				
$\Delta Dist$	3.60 (11.09)	2.86 (9.50)	5.13 (13.70)	2.27 (0.26)
$\mathbb{I}\{\Delta Dist > 0\}$	0.276 (0.447)	0.259 (0.438)	0.309 (0.462)	0.050 (0.012)
N	12,494	8,263	4,231	12,494

Notes: The table reports descriptive statistics for the pooled complete-couple analysis sample. Panel A reports characteristics of the focal individual's spouse, and Panel B reports the focal individual's exposure. Columns (1)–(3) report weighted means, with standard deviations in parentheses below. Column (1) pools the sample, while columns (2) and (3) divide focal individuals according to whether they hold a university degree. Column (4) reports the mean among focal university graduates minus the mean among focal non-graduates, with spouse-pair-clustered standard errors in parentheses. Means, standard deviations, and differences are weighted using the UKB inverse-probability weights employed in the main analysis. Hourly wage is the occupation-specific median hourly wage assigned from the Annual Survey of Hours and Earnings and is expressed in 2015 pounds. Reduction in distance, $\Delta Dist$, is measured in kilometres; $\mathbb{I}\{\Delta Dist > 0\}$ equals one for a strictly positive reduction. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. The unit of observation is the focal individual, and each of the 6,247 couples contributes two observations.

5 Empirical Strategy

We examine whether obtaining a university degree changes whom individuals partner with and how closely the resulting couples resemble one another. Our primary focus is on the effect of university completion on partner composition, that is, whether obtaining a university degree leads individuals to partner with spouses who differ in characteristics such as educational attainment, wages, and education- and cognition-related polygenic indices (PGIs). We then examine whether these changes in partner composition translate into greater or lesser within-couple similarity. Finally, we use the 1972 Raising of the School Leaving Age (ROSLA) reform as a complementary analysis.

5.1 Estimating the Effect of University on Partner Outcomes and Couple Similarity

We start by estimating the causal effects of university attainment on two sets of outcomes: partner composition and couple similarity. To address the endogeneity of selection into university, we instrument university attainment with the reduction in distance to the nearest university, as described in Section 4.2. Throughout, we let C_i denote an indicator for whether individual i holds a university degree and let $\Delta Dist_i$ denote the reduction in distance, expressed in tens of kilometres.

Control Variables Set. To address predetermined differences across individuals and the non-random placement of universities, we condition on a vector X_i containing individual characteristics and characteristics of the local environment at birth. The individual-level controls include sex; the number of siblings and the individual's EA PGI, which capture complementary dimensions of predetermined early-life advantage;[23] and the first twenty genetic principal components, which account for ancestry-related population structure (Price et al., 2006). Local controls include baseline university provision, measured as the number of operating universities within 10 kilometres of the birthplace at birth, and the historical district characteristics described in Section 4.1.2. These are the urban status, educational composition, log population size and density, pre-expansion population growth, historical unemployment, and socioeconomic composition of the birth district. We also include fixed effects for region of birth interacted with academic birth cohort, λ_{rc}, where r indexes 11 broad regions and c indexes one-year birth cohorts running from September to August. These fixed effects absorb permanent differences across regions, common changes across cohorts, and any cohort patterns that differ across regions.

Because the instrument is defined by geographic distance, we combine broad geographic fixed effects with the predetermined local characteristics described above rather than using more disaggregated fixed effects, which would absorb an increasing share of the geographic variation in expansion-induced access. Becker, Tommasi, and Zhang (2025) show that this can substantially reduce identifying variation and power in distance-based IV designs, and argue that directly controlling for economically relevant confounders can be more efficient when these are observed. In Section 6.4, we examine the sensitivity of the estimates to replacing the baseline fixed effects with substantially finer county-by-cohort fixed effects to validate our choice.

Effect of university on partner composition. Our primary outcomes are characteristics of the eventual spouse: whether they hold a university degree, their school-leaving age, their Educational Attainment PGI, and their wages. These outcomes allow us to examine whether obtaining a university degree changes whom individuals ultimately partner with and the socioeconomic characteristics of the resulting match.

We let i denote the focal individual and $j(i)$ their eventual spouse. We instrument individual i's university attainment with their exposure to university expansion and examine its effect on spouse characteristics, $Y_{j(i)}$. Conditioning on the focal individual's predetermined controls, X_i, and birth-region-

[23] Following Barcellos, Carvalho, and Turley (2021), we use the number of siblings as a proxy for family circumstances, motivated by the literature on the quantity–quality trade-off and the relationship between sibship size and children's educational outcomes (Steelman et al., 2002; Chan, Henderson, and Stuchbury, 2019). The EA PGI provides an individual-level genetic correlate of educational attainment, which is also associated with family background.

by-cohort fixed effects, we estimate the following two-stage least squares specification:

$$C_i = \pi_0 + \pi_1 \Delta Dist_i + X_i'\delta + \lambda_{rc} + \nu_i \tag{6}$$

$$Y_{j(i)} = \beta_0 + \beta_1 \widehat{C}_i + X_i'\theta + \lambda_{rc} + \eta_i \tag{7}$$

Thus, the coefficient of interest is β_1, which captures the effect of obtaining a university degree on the characteristics of the eventual spouse among individuals whose university completion responds to improved geographic access. For example, when the outcome is the spouse's university completion, the estimate asks whether obtaining a degree changes the probability of partnering with someone who also holds a university degree. When the outcome is the spouse's EA PGI, it asks whether obtaining a degree leads individuals to partner with spouses who have different EA PGIs.

Effect of university on couple similarity. We also examine whether obtaining a university degree changes similarity between partners. These outcomes are defined at the couple level and ask whether university attainment changes the resemblance between spouses, rather than only the characteristics of the spouse.

Let ℓ index couples, and let $m(\ell)$ and $f(\ell)$ denote the male and female partner in couple ℓ. For continuous traits T, such as education, wages, and the PGIs, we define similarity as the negative squared within-couple difference:

$$S_\ell^T = -\left(T_{m(\ell)} - T_{f(\ell)}\right)^2 \tag{8}$$

Hence, larger values of S_ℓ^T correspond to greater similarity between partners. For genome-wide genetic relatedness, we use spouse-pair kinship directly, denoted by S_ℓ^K, since kinship is already increasing in genetic similarity between spouses. For comparability across outcomes, we standardise each similarity measure to have mean zero and unit standard deviation in the analysis sample.

Given that similarity is defined at the couple level, university attainment may be endogenous for both members of the couple. For $s \in \{m, f\}$, let $C_{s(\ell)}$ denote spouse s's university attainment, $\Delta Dist_{s(\ell)}$ their reduction in distance to the nearest university, and $X_{s(\ell)}$ the corresponding vector of predetermined individual- and birthplace-level controls defined above. The term $\lambda_{r(s)c(s)}$ denotes the region-of-birth-by-cohort fixed effect for spouse s. For compactness, let W_ℓ collect the predetermined controls of both spouses, $X_{m(\ell)}$ and $X_{f(\ell)}$, together with their respective region-of-birth-by-cohort fixed effects, $\lambda_{r(m)c(m)}$ and $\lambda_{r(f)c(f)}$. We therefore instrument the university attainment of each spouse with both spouses' exposure to university expansion.

Thus, the two first stages are:

$$C_{m(\ell)} = \pi_{m0} + \pi_{mm}\Delta Dist_{m(\ell)} + \pi_{mf}\Delta Dist_{f(\ell)} + W_\ell'\delta_m + \nu_{m(\ell)}, \tag{9}$$

$$C_{f(\ell)} = \pi_{f0} + \pi_{fm}\Delta Dist_{m(\ell)} + \pi_{ff}\Delta Dist_{f(\ell)} + W_\ell'\delta_f + \nu_{f(\ell)}. \tag{10}$$

The second stage is given by:

$$S_\ell = \alpha + \beta_m \widehat{C}_{m(\ell)} + \beta_f \widehat{C}_{f(\ell)} + W_\ell'\theta + u_\ell. \tag{11}$$

The coefficients β_m and β_f capture whether obtaining a university degree in response to improved geographic access changes within-couple similarity through the male and female side of the couple, respectively.

5.2 Threats to Identification

Our design exploits differences in how post-war university expansion changed access across birthplaces and cohorts. Conceptually, we compare individuals born in the same region and cohort who experienced different reductions in the distance to the nearest university during childhood. Thus, to interpret the resulting differences in partner outcomes as causal effects of university attainment, we require that, conditional on the baseline controls and fixed effects, expansion-induced improvements in access were unrelated to other determinants of partner outcomes and affected those outcomes only through the focal individual's university attainment. We also assume monotonicity, that is, that a larger reduction in distance to the nearest university would not lead anyone to forgo a university degree they would have obtained with a smaller reduction.

These identifying requirements give rise to three principal concerns. First, given that, as described in Section 3.1, openings were not randomly located, they may have been related to unobserved characteristics or trajectories that also influenced subsequent partner formation. Although the baseline specification controls for observed individual characteristics, proxies for early-life advantages such as family background, and predetermined differences across birth districts, it cannot fully account for unobserved spatial differences in parental resources or educational aspirations, or for local changes in educational demand, economic opportunities, migration, and social composition. If such factors were correlated with where and when universities opened, the instrument could compare individuals and places who already followed different educational and marriage-market trajectories. Second, university openings may themselves have changed local economic and social conditions, thereby violating the exclusion restriction. By creating employment, attracting new residents, or altering the local pool of potential partners, an opening could affect subsequent partner outcomes even when the focal individual's own education remained unchanged (Bharati, Chang, and Li, 2023), although Rogne, Fauske, and Hart (2025) find little evidence of such effects for Norwegian college openings. In that case, the instrument would capture both the effect of improved access on university attainment and the broader local consequences of university expansion. Third, partners originating from the same area may have experienced similar changes in university access. The focal individual's reduction in distance could therefore be correlated with the eventual spouse's own access shock. In that case, the instrument might predict the spouse's education or earnings through the spouse's own exposure to university expansion, rather than through the effect of the focal individual's university attainment on whom they match with. Thus, this concern is particularly relevant when couples formed within the same local marriage market or had already met before university entry.

To address these concerns, we present several complementary exercises in Section 6.4. We investigate selection into university placement using predetermined district characteristics and individual-level balance tests, and allow cohort shocks to vary across counties rather than broad regions. We use post-entry (placebo) openings to assess whether the first stage is specific to access changes occurring during the relevant educational period and whether these placebo openings nevertheless predict partner outcomes through broader local channels. Finally, we examine the geographic origins of the resulting couples and re-estimate the effects among spouses born in different districts or sufficiently far apart. These latter exercises provide suggestive evidence on whether the estimates are driven by partners sharing the same local access shock, although such restrictions may introduce selection if university access itself affects how far apart eventual spouses were born.

5.3 Complementary Design: ROSLA as a Compulsory-Schooling Benchmark

The university-expansion design identifies effects at the university-completion margin. As a complementary benchmark, we use the 1972 Raising of the School Leaving Age (ROSLA) reform to examine whether similar changes in partner composition arise at the compulsory-schooling margin. The reform raised the minimum school-leaving age from 15 to 16 for pupils born on or after 1 September 1957. We exploit this birth-date cutoff in a fuzzy regression-discontinuity design, using reform eligibility as an instrument for remaining in full-time education until at least age 16.

Let R_i denote the focal individual's date of birth relative to 1 September 1957, normalised so that $R_i = 0$ at the cutoff, and define $ROSLA_i = \mathbb{1}\{R_i \geq 0\}$. Because reform eligibility shifted, but did not perfectly determine, school-leaving behaviour, we estimate the following two-stage local-linear model:

$$Edu16_i = \alpha_0 + \alpha_1 ROSLA_i + \alpha_2 R_i + \alpha_3 R_i \times ROSLA_i + \varepsilon_i, \tag{12a}$$

$$Y_{j(i)} = \beta_0 + \beta_1 \widehat{Edu16}_i + \beta_2 R_i + \beta_3 R_i \times ROSLA_i + u_i. \tag{12b}$$

where $Edu16_i$ indicates that individual i remained in full-time education until at least age 16, and $Y_{j(i)}$ denotes the outcome of their eventual spouse. The coefficient α_1 captures the discontinuity in remaining in education at the ROSLA cutoff, while β_1 captures the corresponding effect on the spouse outcome for individuals whose school-leaving behaviour was changed by the reform.

We estimate local linear regressions within a bandwidth of three years on either side of the cutoff, allowing the birth-date slope to differ before and after the reform[24]. Observations are weighted using a triangular kernel together with the UKB inverse-probability weights employed throughout the analysis, and standard errors are clustered by spouse pair. The identifying assumption is that, in the absence of ROSLA, potential schooling choices and potential spouse outcomes would have evolved continuously through the September 1957 cutoff. Together with the exclusion and monotonicity conditions, this allows β_1 to be interpreted as the local average treatment effect of remaining in education until at least age 16 for individuals whose school-leaving behaviour was changed by the reform.

The ROSLA and university-expansion estimates apply to different complier populations and different educational margins. We therefore use the ROSLA results as a complementary benchmark rather than as an alternative estimate of the effect identified by the university-expansion design.

6 Results

6.1 From University Access to Partner Composition

We begin by examining whether improved university access increased university completion and whether these access shocks are reflected in the characteristics of eventual spouses. Table 2 reports the first-stage estimates of Equation (6) in Panel A and the corresponding reduced-form effects on spousal outcomes in Panel B.

[24] Appendix Figure A.3 examines the sensitivity of the estimates to alternative bandwidth choices.

Panel A shows that reductions in distance to the nearest university increased university completion. In the pooled sample, a 10-km reduction in distance raises the probability of completing university by 3.1 percentage points, an increase of approximately 9.4 percent relative to the sample mean of 32.9 percent. The corresponding cluster-robust Wald F-statistic is 42.10. The first-stage coefficients are 3.5 percentage points for men and 2.6 percentage points for women, with F-statistics of 41.14 and 13.73, respectively. University expansion therefore generated substantial variation in university completion in both sex-specific samples, although the weaker first stage among women implies less precise female-specific 2SLS estimates. The positive average response is also reflected in the relationship across exposure levels. Appendix Figure A.1 and Table A.6 show that university completion generally increases with larger reductions in distance.

Panel B reports the reduced-form effects on spousal outcomes. In the pooled sample, a 10-km reduction in distance increases the spouse's probability of completing university by 2.7 percentage points, school-leaving age by 0.138 years, and log hourly wages by 0.011 log points, or approximately 1.1 percent. All three estimates are positive and statistically significant. Thus, the same access shocks that increased the focal individual's university completion also changed the realised education and earnings of their eventual spouse.

The polygenic-index estimates provide complementary evidence of a change in partner composition. A 10-km reduction in distance increases the spouse's Educational Attainment, Cognitive Performance, and Mathematics PGIs by 0.021, 0.032, and 0.016 standard deviations, respectively. The similarity of the estimates across the three indices suggests that the shift is not confined to a single measure of educational or cognitive propensity. The sex-specific reduced forms are positive for both men and women, but the point estimates for all three spousal PGIs are larger among female focal individuals. The effects on all three PGIs are imprecisely estimated among men, but larger and statistically significant among women. Although the point estimates are larger among women, formal tests provide no clear evidence that the effects differ by sex for any of the three PGIs.

Taken together, the first-stage and reduced-form estimates establish a link from local university expansion to both university completion and subsequent partner composition. The next subsection scales the reduced-form changes by the first-stage response to estimate the effect of completing university on the characteristics of eventual spouses.

6.2 Effects on Partner Composition

We now estimate how completing university affects the educational, labour-market, and genetically indexed characteristics of eventual spouses. Table 3 reports the 2SLS estimates for the pooled sample in Panel A and separately by the sex of the focal individual in Panels B and C.

In the pooled sample, obtaining a university degree increases the spouse's Educational Attainment PGI by 0.672 standard deviations, the Cognitive Performance PGI by 1.021 standard deviations, and the Mathematics PGI by 0.518 standard deviations. Because a spouse's genotype is fixed at conception, these effects cannot arise from university completion changing the characteristics of an otherwise unchanged partner. Instead, they indicate that individuals completing university in response to the expansion formed partnerships with people who differed systematically in predetermined genetic summaries associated with education, cognition, and mathematical performance. Moreover, the positive and statistically significant estimates across all three indices suggest that the change in partner composition is not specific

Table 2: Reduction in Distance, University Attainment, and Spousal Outcomes

Panel A: First-Stage Estimates

Dependent variable: Own university degree

	(1) *Pooled Sample*	(2) *Males*	(3) *Females*
$\Delta Dist$	0.031***	0.035***	0.026***
	(0.005)	(0.005)	(0.007)
Observations	12,494	6,247	6,247
First-stage Wald F-statistic	42.10	41.14	13.73

Panel B: Reduced-Form Effects on Spousal Outcomes

	Polygenic Indices			*Education and Wages*		
	(1) EA PGI	(2) Cog PGI	(3) Math PGI	(4) University	(5) SLA	(6) Log Wage
Panel B.1: Pooled Sample						
$\Delta Dist$	0.021**	0.032***	0.016*	0.027***	0.138***	0.011***
	(0.009)	(0.009)	(0.009)	(0.005)	(0.023)	(0.003)
Observations	12,494	12,494	12,494	12,494	12,494	12,494
Panel B.2: Males						
$\Delta Dist$	0.019*	0.022**	0.011	0.023***	0.139***	0.007*
	(0.012)	(0.011)	(0.011)	(0.006)	(0.024)	(0.004)
Observations	6,247	6,247	6,247	6,247	6,247	6,247
Panel B.3: Females						
$\Delta Dist$	0.031**	0.039***	0.025*	0.029***	0.122***	0.016***
	(0.014)	(0.014)	(0.015)	(0.009)	(0.039)	(0.004)
Observations	6,247	6,247	6,247	6,247	6,247	6,247
Test of equality across sexes (p-value)	0.514	0.293	0.438	0.536	0.673	0.113

Notes: Panel A reports the first-stage effect of a reduction in distance to the nearest university on the focal individual's university attainment. Panel B reports the corresponding reduced-form effects on spousal outcomes. Reduction in distance is measured per 10 km. Males and females refer to the sex of the focal individual. All specifications include the baseline controls X, birth-region-by-cohort fixed effects, the first 20 own and spouse genetic principal components, and UK Biobank inverse-probability weights. In the whole sample, standard errors are two-way clustered by spouse pair and birth district; in the sex-specific samples, they are clustered by birth district. Inference uses birth-district cluster degrees of freedom. The reported first-stage statistic is the one-restriction cluster-robust Wald F-statistic for the excluded instrument. SLA denotes school-leaving age. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. Equality tests use pooled regressions with sex-specific instrument coefficients, control coefficients, and fixed effects, with two-way clustering by spouse pair and birth district and birth-district degrees of freedom. Cluster-robust standard errors are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

to a single genetic measure. The socioeconomic estimates reinforce this interpretation. In the pooled sample, completing university increases the probability that the spouse also completed university by 86.1 percentage points and raises the spouse's school-leaving age by 4.428 years. It also increases the spouse's log hourly wage by 0.351, corresponding to an approximate wage increase of 42.1 percent. Evaluated at the pooled-sample mean hourly wage of £14.90, this corresponds to approximately £6.27 per hour.

These magnitudes are large, even against the substantial educational sorting visible in the raw data: among university graduates, 58.9 percent have a university-educated spouse, compared with 20.4 percent among non-graduates. The estimated shift towards university-educated partners is considerably larger for individuals whose own university completion responds to improvements in access. To put this large local effect in context, we examine who responds to the expansion. Appendix Table A.7 shows that reductions in distance had their largest effects on university completion among individuals born in higher-SES districts and those in the middle tertile of the EA PGI. Complementary profiles based on alternative distance cutoffs suggest that individuals responding to improved access grew up with substantially less university provision nearby (Appendix Table A.15).[25] Additionally, large improvements in access account for a disproportionate share of the identifying variation. Indeed, those whose distance to the nearest university fell by at least 30 kilometres make up just 5.0 percent of the weighted sample, yet account for 64.9 percent of the 2SLS denominator (Appendix Figure A.2).

Taken together, these patterns are consistent with the expansion affecting individuals whose university decisions were particularly sensitive to the availability of university education in their local area. For such individuals, attending university may have represented a discrete change in the environment in which partnerships were formed. In these cohorts, university attendance remained relatively uncommon, and entering higher education could expose individuals to university-based and subsequent graduate networks containing a much greater concentration of highly educated potential partners. Attending university could also change geographic mobility, occupations, and later social networks. Given the strong educational sorting already visible in the data, this change in the relevant matching environment provides one reason why the effect for individuals whose attendance was shifted by university access may be substantially larger than the average difference between graduates and non-graduates.

Heterogeneity. Panels B and C show that the estimates are positive through both the male and female university margins. For male focal individuals, university completion increases the probability of having a university-educated spouse by 64.8 percentage points, the spouse's school-leaving age by 3.938 years, and their log hourly wage by 0.199, or approximately 22 percent. The estimated effects on the spouse's Educational Attainment, Cognitive Performance, and Mathematics PGIs are 0.547, 0.614, and 0.307 standard deviations, respectively. The first two genetic estimates provide suggestive evidence of positive effects, while the Mathematics estimate is less precise. For female focal individuals, the corresponding point estimates are uniformly larger. Completing university raises the spouse's school-leaving age by 4.725 years and hourly wages by approximately 86 percent, while the spouse-university coefficient is 1.140. The effects on the spouse's Educational Attainment, Cognitive Performance, and Mathematics PGIs are 1.180, 1.508, and 0.966 standard deviations, respectively. However, the first stage is weaker among women, and the resulting 2SLS estimates are correspondingly less precise. Although the female estimates exceed the male estimates across all six outcomes, formal tests provide little evidence of

[25] These exercises characterise the populations associated with discrete improvements in university access at each distance threshold. The resulting profiles help interpret who responds to the expansion, although they need not coincide with the complier population underlying our preferred continuous-instrument specification.

systematic differences between them, with only the difference in log wages being marginally significant. We additionally test for heterogeneity by birth-district SES, urban status, birth cohort, and the focal individual's EA PGI (Appendix Figure A.4), but find no evidence of systematic differences across these groups, although several subgroup estimates are imprecise because the first stage is weak within them.

Geographic partner composition. We also examine how improvements in university access relate to the geographic origins of eventual spouses. Appendix Table A.12 shows that a 10-km reduction in distance to the nearest university is associated with a 9.8-km increase in the distance between partners' birthplaces and a higher probability of partnering with someone born at least 30 km away. The probability that both partners were born in the same district or region also falls by 2.7 and 3.6 percentage points, respectively. These patterns suggest that improvements in university access were associated with more geographically dispersed partnerships, although the reduced-form estimates do not isolate how much of this relationship operates through university completion itself.

Overall, our estimates indicate that completing university changes whom access-sensitive individuals partner with, with substantial differences in both the genetic characteristics and socioeconomic outcomes of their eventual spouses.

6.3 Effects on Couple Similarity

We next examine whether university completion changes resemblance between partners, rather than only the characteristics of the eventual spouse. As defined in Equation (8), similarity in school-leaving age, log wages, and the Educational Attainment, Cognitive Performance, and Mathematics PGIs is measured using the negative squared difference between partners, while spouse-pair kinship provides a complementary measure of genome-wide genetic similarity. We standardise each outcome and orient it so that higher values indicate greater within-couple similarity. Table 4 reports estimates from the joint couple-level specification in Equation (11), which instruments male and female university completion simultaneously, together with corresponding specifications that instrument university completion separately through the male and female side of the couple.

Columns (1)–(4) of Table 4 consider similarity in genetic characteristics. In the sex-specific specifications, the estimated effects on similarity in the EA, Cognitive Performance, and Mathematics PGIs are -0.30, -0.05, and -1.01 standard deviations for men and -0.31, -0.20, and 0.01 standard deviations for women, respectively. Although the Mathematics estimate for men suggests some reduction in similarity, most estimates are imprecise and provide little evidence that the sizeable effects of university completion on the genetic characteristics of eventual spouses documented above translate into systematic changes in similarity across these PGIs. The joint specification is considerably less precise and does not reveal a common pattern across the three measures. Kinship is the main exception. In the sex-specific specifications, the effect through male university completion is close to zero, whereas university completion among women increases spouse-pair kinship by 1.24 standard deviations and is statistically significant at the 5 percent level. The corresponding joint estimates are substantially less precise. Given the weaker first stage for female university completion, particularly in the joint specification, and the absence of a comparable pattern across the PGI measures, we regard the kinship result as suggestive rather than evidence of a general increase in genetic resemblance.

Columns (5) and (6) turn to socioeconomic similarity. In the sex-specific specifications, university

Table 3: Effects of University Completion on Spousal Outcomes

	Polygenic Indices			*Education and Wages*		
	(1)	(2)	(3)	(4)	(5)	(6)
	EA PGI	Cog PGI	Math PGI	University	SLA	Log Wage
Panel A: Pooled Sample						
Own University Degree	0.672**	1.021***	0.518**	0.861***	4.428***	0.351***
	(0.260)	(0.297)	(0.255)	(0.107)	(0.616)	(0.087)
Observations	12,494	12,494	12,494	12,494	12,494	12,494
First-stage F-statistic			42.10			
Panel B: Males						
Own University Degree	0.547*	0.614*	0.307	0.648***	3.938***	0.199*
	(0.305)	(0.314)	(0.310)	(0.161)	(0.710)	(0.102)
Observations	6,247	6,247	6,247	6,247	6,247	6,247
First-stage F-statistic			41.14			
Panel C: Females						
Own University Degree	1.180**	1.508**	0.966*	1.140***	4.725***	0.620***
	(0.564)	(0.650)	(0.571)	(0.367)	(1.584)	(0.200)
Observations	6,247	6,247	6,247	6,247	6,247	6,247
First-stage F-statistic			13.73			
Test of equality across sexes (p-value)	0.331	0.205	0.327	0.297	0.673	0.075

Notes: Each coefficient is the 2SLS estimate of the effect of the focal individual's university attainment on the listed spousal outcome. Reduction in distance to the nearest university, measured per 10 km, is used as the excluded instrument. Panel A reports estimates for the pooled sample; Panels B and C estimate the model separately for male and female focal individuals. All specifications include the baseline controls, birth-region-by-cohort fixed effects, the first 20 own and spouse genetic principal components, and UK Biobank inverse-probability weights. In Panel A, standard errors are two-way clustered by spouse pair and numerical birth-district ID. In Panels B and C, standard errors are clustered by numerical birth-district ID. Inference uses birth-district cluster degrees of freedom. SLA denotes school-leaving age. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. The equality test uses a pooled 2SLS regression with sex-specific university coefficients, instruments, control coefficients, and fixed effects, with two-way clustering by spouse pair and birth district and birth-district degrees of freedom. Cluster-robust standard errors are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

completion modestly increases similarity in school-leaving age for both men and women, but neither estimate is precisely determined. Log-wage similarity instead declines for both sexes, with estimates of -0.63 standard deviations for men and -0.99 standard deviations for women, again with substantial uncertainty. The joint estimates are less stable, particularly for school-leaving-age similarity, where the male and female coefficients move in opposite directions. Unlike the genetic outcomes, these socioeconomic similarity measures can change both because university completion alters partner composition and because it affects the focal individual's own education and earnings.

Taken together, the similarity estimates are substantially less conclusive than the partner-composition

results. The sex-specific estimates provide no indication of a broad increase in resemblance following university completion: similarity in the Cognitive Performance and Mathematics PGIs tends to decline, wage similarity also falls, while school-leaving-age similarity tends to increase and female university completion is associated with greater kinship. At the same time, these effects are generally estimated with considerable uncertainty, and identification becomes appreciably weaker when both partners' university completion is treated as endogenous simultaneously. The evidence therefore establishes much more clearly that university completion changes the characteristics of eventual partners than whether it systematically changes the degree of resemblance between them.

6.4 Validity and Robustness

We next assess the validity of our identifying assumptions and the robustness of the estimates. Specifically, we examine threats arising from endogenous university placement, unobserved local trajectories and direct local effects of university openings that could violate the exclusion restriction, and shared exposure to university expansion within couples.

We begin by asking whether university openings were systematically related to pre-existing local conditions. Appendix Table A.13 relates openings to district characteristics measured before the relevant expansion. Openings were more common in more populous districts and in districts with existing university provision nearby, with some evidence of relationships with historical unemployment and educational composition. Urban status, population density, and occupational composition, by contrast, do not systematically predict openings. These results confirm that university placement was not random, but suggest that observable targeting was concentrated along a limited set of dimensions rather than extending across a broad range of pre-expansion socioeconomic characteristics.

However, our placement regressions cannot rule out selection on unobserved local conditions or cohort-specific changes within broad regions. As discussed in Section 5, the baseline specification combines the targeted district controls with birth-region-by-cohort fixed effects. In Appendix Table A.16, we examine whether this leaves important local confounding by replacing the latter with birth-county-by-cohort fixed effects.[26] Therefore, this more demanding specification absorbs all between-county variation in expansion-induced access within each cohort. As expected for a distance-based instrument, the resulting reduction in identifying variation lowers precision. Nevertheless, all estimates remain broadly similar to the baseline, except for the Mathematics PGI, whose estimate is less stable and considerably less precise. This provides some reassurance that residual local geographic confounding is unlikely to explain the main findings.

A complementary concern is that, conditional on these geographic and local controls, changes in university access may remain correlated with predetermined individual characteristics. Therefore, in Appendix Table A.11 we test whether the reduction in distance predicts predetermined characteristics that should not themselves be affected by the instrument. Conditional on the full baseline specification, we find no relationship between the instrument and sex, birth month, birth weight, maternal smoking, or additional PGIs capturing anthropometric, cognitive, and academic characteristics. The broad balance across these predetermined characteristics supports the identifying assumption that, conditional on the baseline controls, changes in university access are unrelated to other determinants of educational attainment and partner outcomes.

[26] The estimation sample spans 42 historical counties and 11 broad regions in England and Wales.

Table 4: Effects of University Completion on Within-Couple Similarity

	Genetic Similarity				*Education and Wage Similarity*	
	(1)	(2)	(3)	(4)	(5)	(6)
	EA PGI	Cog PGI	Math PGI	Kinship	SLA	Log Wage
Panel A: Joint Model						
Male university	0.107	-0.080	-2.691	-0.937	-1.299	-0.767
	(0.841)	(0.863)	(1.757)	(1.075)	(1.682)	(1.040)
Female university	-0.164	-0.386	2.691	2.068	1.776	-0.170
	(1.051)	(1.253)	(2.362)	(1.497)	(1.911)	(1.569)
First-stage F, male university			7.14			
First-stage F, female university			3.94			
Panel B: Sex-Specific Models						
Male focal individuals	-0.298	-0.053	-1.011*	0.124	0.290	-0.627*
	(0.329)	(0.316)	(0.546)	(0.311)	(0.336)	(0.323)
First-stage F-statistic			41.14			
Female focal individuals	-0.306	-0.197	0.012	1.238**	0.654	-0.993
	(0.578)	(0.649)	(0.418)	(0.539)	(0.476)	(0.636)
First-stage F-statistic			13.73			
Couples	6,247	6,247	6,247	6,247	6,247	6,247

Notes: Each coefficient is a weighted 2SLS estimate of the effect of university attainment on the listed within-couple similarity outcome. Outcomes are standardised and oriented so that higher values indicate greater similarity: for the PGIs, school-leaving age, and log wages, similarity is the negative squared difference between partners, as defined in equation (8); kinship is spouse-pair kinship. Panel A reports the joint couple-level estimates $\widehat{\beta}_m$ and $\widehat{\beta}_f$ from equation (11), in which male and female university attainment are instrumented simultaneously with each partner's $\Delta Dist$, as in equations (9) and (10); standard errors are two-way clustered by the male and female birth districts, and the first-stage F-statistics are the excluded-instrument Wald statistics for each endogenous variable. Panel B reports sex-specific models in which only the focal partner's university attainment is instrumented, with their own $\Delta Dist$; standard errors are clustered by birth district. $\Delta Dist$ is measured per 10 km. Panel A includes both partners' baseline controls; Panel B includes the focal individual's baseline controls, as in the main specification, so its first stages coincide with those in Table 3. All specifications include birth-region-by-cohort fixed effects, the first 20 own and 20 spouse genetic principal components, and UK Biobank inverse-probability weights. Inference uses birth-district cluster degrees of freedom. SLA denotes school-leaving age. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. Cluster-robust standard errors are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

As discussed in Section 5.2, the first-stage estimates could reflect persistent differences across places and cohorts rather than changes in access during childhood, while university openings could themselves have altered local economic and social conditions in ways that affected partner outcomes. To assess these possibilities, Appendix Table A.10 reconstructs the distance measure using openings that occurred only after the usual university-entry age. Because these openings postdated the focal individual's university decision, any relationship with university attainment would suggest that the baseline first stage captures persistent place–cohort differences rather than the effect of improved access. At the same time, post-

entry openings could still have created employment, attracted new residents, or altered the local pool of potential partners. A relationship with partner outcomes despite the absence of a first stage would therefore be consistent with broader local effects of university expansion. Reassuringly, the placebo measure predicts neither university completion nor the spousal outcomes. Openings occurring outside the relevant educational window therefore do not reproduce either component of the main empirical relationship.

A narrower concern is that both spouses may have been exposed to the same local expansion. This possibility is particularly relevant for partners born near one another, who are both more likely to have experienced the same access shock and more likely to have met before university. In that case, the focal individual's instrument could predict the spouse's outcomes through the spouse's own exposure to expansion even if the focal individual's university attendance did not change whom they matched with. If this channel were driving the results, the estimates should weaken once couples originating from the same local area are excluded. We also examine whether the findings are specific to Greater London's unusually dense university sector and distinctive marriage market. Appendix Table A.18 shows that the estimates remain similar among couples born in different districts, when we require partners' birthplaces to be separated by at least five or ten kilometres, and when couples with either spouse born in Greater London are excluded. However, as documented in Section 6.2, university access also predicts how far apart spouses were born. Restricting the sample based on this geographic characteristic may therefore introduce selection, so these estimates cannot by themselves rule out shared exposure.

Finally, we examine the sensitivity of the estimates to how improvements in university access are measured. Appendix Table A.14 distinguishes genuinely new university openings from other sources of expansion, including university-status changes and campus openings at existing institutions. Restricting the instrument to new openings yields a larger first-stage response and positive partner-outcome estimates that are broadly consistent with the baseline. The findings are therefore robust to measuring improvements in access using new university openings alone. We additionally vary the childhood exposure window. Appendix Table A.8 reconstructs the instrument using exposure windows ending at each age from 11 to 15. The estimates remain similar across these alternative cutoffs, indicating that they are not sensitive to the precise endpoint of the childhood exposure window. We also test whether the results are robust to alternative functional forms. We find no evidence of a meaningful quadratic relationship in the first stage, and the second-stage estimates remain similar when we include a quadratic term or instead measure the change in distance in logs.

6.5 Comparison with the Compulsory-Schooling Margin

We finally use ROSLA to ask whether the changes in partner composition documented at the university margin also arise when education is increased at the compulsory-schooling margin. The first stage is clear: ROSLA increased the probability of remaining in education until at least age 16 by 7.4 percentage points in the pooled sample, 8.6 percentage points for men, and 6.1 percentage points for women.

Table 5 reports the corresponding fuzzy RDD estimates for partner characteristics from estimating Equation (12). In the pooled sample, we find no precisely estimated effect on spouse university completion, school-leaving age, log wages, or any of the three spousal PGIs. The same is true when we estimate the effects separately for men and women. Although some of the point estimates for the genetic outcomes are sizeable (e.g., the Cognitive Performance PGI), they are estimated with substantial uncertainty and do not reveal a consistent pattern across outcomes or sexes. This conclusion is consistent across alternative

Table 5: The Impact of ROSLA on Spousal Outcomes

	Polygenic Indices			*Education and Wages*		
	(1)	(2)	(3)	(4)	(5)	(6)
	EA PGI	Cog PGI	Math PGI	University	SLA	Log Wage
Panel A: Pooled Sample						
Edu16	0.386	0.903	0.010	-0.058	-0.219	0.317
	(0.865)	(0.924)	(0.907)	(0.451)	(2.273)	(0.316)
Mean dependent variable	-0.043	-0.030	-0.039	0.367	18.587	2.672
Observations	22,248	22,248	22,248	22,248	22,248	22,248
First-stage coefficient			0.074***			
			(0.016)			
Panel B: Males						
Edu16	0.005	1.082	0.346	-0.360	0.278	0.197
	(1.091)	(1.155)	(1.176)	(0.559)	(2.711)	(0.349)
Mean dependent variable	-0.036	-0.053	-0.045	0.342	18.482	2.510
Observations	11,124	11,124	11,124	11,124	11,124	11,124
First-stage coefficient			0.086***			
			(0.020)			
Panel C: Females						
Edu16	0.870	0.603	-0.619	0.339	-1.054	0.136
	(1.470)	(1.542)	(1.517)	(0.728)	(3.872)	(0.469)
Mean dependent variable	-0.049	-0.009	-0.034	0.389	18.681	2.816
Observations	11,124	11,124	11,124	11,124	11,124	11,124
First-stage coefficient			0.061***			
			(0.023)			

Notes: This table reports 2SLS estimates from Equation (12). Each coefficient is the fuzzy RDD estimate of the effect of the focal individual's completing education at age 16 or later on the listed spousal outcome. ROSLA raised the minimum school-leaving age from 15 to 16 for cohorts born from September 1957 onward; eligibility is defined by the focal individual's birth date only. The sample consists of kinship-consistent complete male–female pairs born within three years of the September 1957 ROSLA cutoff. The running variable is birth timing, measured in year-months and centred at the cutoff. All specifications use a triangular kernel, local linear trends, UK Biobank inverse-probability weights, and a bandwidth of three years. The reported first-stage coefficient is the discontinuity at the cutoff in the probability that the focal individual completed education at age 16 or later. Standard errors are `rdrobust` cluster-robust standard errors, clustered by spouse pair, and are reported in parentheses. SLA denotes school-leaving age. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

bandwidths around the September 1957 cutoff (see Appendix Figure A.3).

The absence of precisely estimated effects should not be interpreted as evidence that compulsory schooling had no consequences for partner formation. The ROSLA estimates apply to individuals who

stayed in school longer because of the reform, whereas the university-expansion estimates apply to those whose university completion responded to improved geographic access. The estimates are therefore informative about different educational margins and different complier populations. Nevertheless, the contrast is striking: whereas university completion produces broad changes in the educational, labour-market, and genetically indexed characteristics of eventual spouses, the ROSLA estimates provide no comparably systematic evidence of changes in partner composition.

This pattern is also consistent with earlier evidence on ROSLA and marital sorting. Geruso and Royer (2018) find that the reform shifted spouse qualifications at the lower end of the education distribution, primarily away from spouses with no qualifications and towards those with intermediate credentials, but find little evidence of increased matching with spouses holding two or more A-levels or higher qualifications. The evidence therefore suggests that the partner-composition effects in our main analysis are more closely associated with the university margin than with an additional year of compulsory schooling. One possible explanation is that university attendance changes dimensions of the educational and social environment—including peer groups, geographic mobility, and the set of potential partners encountered—that need not change to the same extent when schooling is extended from age 15 to 16.

7 Discussion

This paper asks whether expanding access to higher education changes the partnerships that individuals subsequently form. We find that obtaining a university degree substantially changes the characteristics of the eventual spouse among individuals whose university completion responds to improved geographic access. In particular, these individuals partner with spouses who have higher Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices (PGIs), and who are more likely to hold a university degree, remain in education longer, and have higher occupation-based wages. While previous work has shown that higher education can affect partner characteristics, we provide novel evidence that higher-education policy can shift partner composition even along genetically indexed dimensions associated with socioeconomic advantage. Because these genetic characteristics are fixed at conception, the effects establish that obtaining a university degree changes who becomes the spouse rather than merely outcomes realised within an otherwise unchanged match.

Our results therefore complement a growing literature showing that university attendance, institutions, fields of study, and peer environments can affect partner outcomes (Artmann et al., 2021; Kirkebøen et al., 2025; Barrios-Fernández, Neilson, and Zimmerman, 2026; Abramitzky, Pérez, and Price, 2026). We show that the university expansion brought into university individuals who would otherwise not have obtained a degree, and those individuals subsequently matched with a substantially different set of partners. Moreover, the geography of the resulting matches is consistent with this interpretation. Larger expansion-induced improvements in access predict spouses being born farther apart and lower probabilities that partners originate from the same geographical area. Although we do not interpret these geographic estimates as causal effects of university attainment, their common direction is consistent with the expansion exposing affected individuals to a geographically broader marriage market. By contrast, our estimates of within-couple similarity are considerably less precise. We therefore cannot establish whether university systematically increases or decreases resemblance between spouses in education, wages, PGIs, or genome-wide kinship. Furthermore, we find no comparable effects on our principal spouse outcomes at the lower schooling margin shifted by the 1972 Raising of the School Leaving Age. This complements Geruso and Royer (2018), who show that ROSLA affected matching around the lower

qualification margin directly targeted by the reform, and suggests that the marriage-market consequences of education depend on the particular educational margin shifted by policy.

Our evidence speaks most directly to the access-sensitive individuals identified by the university expansion, rather than to the average effect of university attainment in the population. The linked UK Biobank couples are also not representative of the British population as a whole, and spouse relationships are inferred rather than directly reported, although our genetic validation provides strong evidence that the identified pairs are genuine couples. Finally, our data do not reveal whether the observed re-sorting occurs through peers encountered at university, geographic mobility, later occupational and social networks, or changes in marriage-market position. These features limit how broadly the estimates can be generalised and prevent us from distinguishing the mechanisms through which crossing the university margin changes partner formation.

Individual genetic characteristics are fixed, but their distribution across the families that form is not. More broadly, previous work shows that the relationship between socioeconomic circumstances and genetic characteristics can itself be shaped by social institutions (Abdellaoui et al., 2023). Our results provide a policy counterpart to this idea. By bringing into university individuals who would otherwise not have obtained a degree, the expansion changed not only the education and earnings of their eventual spouses, but also whom they partnered with along genetically indexed dimensions. Educational institutions can therefore shape how socioeconomic resources and inherited predictors are combined within the families that form. Because partners pool resources, jointly shape children's environments, and transmit genetic characteristics across generations, such re-sorting provides a potentially persistent channel through which educational opportunity can shape the distribution and intergenerational transmission of advantage.

References

Abdellaoui, Abdel, Oana Borcan, Pierre-André Chiappori, David Hugh-Jones, Fartein Ask Torvik, and Eivind Ystrøm. 2023. "Trading Social Status for Genetics in Marriage Markets: Evidence from Great Britain and Norway." Working paper, Masaryk University Economic Seminars.

Abramitzky, Ran, Leah Platt Boustan, and Osea Giuntella. 2025. "Enclaves and Assimilation in the Age of Mass Migration: Evidence from Ethnic Catholic Churches." NBER Working Paper 33362, National Bureau of Economic Research.

Abramitzky, Ran, Santiago Pérez, and Joseph Price. 2026. "Colleges, Assortative Mating, and Intergenerational Mobility in Twentieth-Century America." NBER Working Paper 35602, National Bureau of Economic Research.

Anderberg, Dan and Yu Zhu. 2010. "The Effect of Education on Marital Status and Partner Characteristics: Evidence from the UK." CESifo Working Paper 3104, CESifo. Available at SSRN: https://ssrn.com/abstract=1633360.

Anderson, Robert. 2006. *British Universities Past and Present*. London: Hambledon Continuum.

Artmann, Elisabeth, Nadine Ketel, Hessel Oosterbeek, and Bas van der Klaauw. 2021. "Field of Study and Partner Choice." *Economics of Education Review* 84:102149.

Banks, J., L. S. Carvalho, and F. Perez-Arce. 2019. "Education, decision making, and economic rationality." *Review of Economics and Statistics* 101 (3):428–441.

Barban, Nicola, Elisabetta De Cao, Sonia Oreffice, and Climent Quintana-Domeque. 2021. "The effect of education on spousal education: A genetic approach." *Labour Economics* 71:102023.

Barcellos, Silvia H., Leandro S. Carvalho, and Patrick Turley. 2018. "Education can reduce health differences related to genetic risk of obesity." *Proceedings of the National Academy of Sciences* 115 (42):E9765–E9772.

———. 2021. "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages, and Later-Life SES." NBER Working Paper 28750, National Bureau of Economic Research.

Barrios-Fernández, Andrés, Christopher Neilson, and Seth Zimmerman. 2026. "Elite Universities and the Intergenerational Transmission of Human and Social Capital." *American Economic Review* 116 (6):2120–2165.

Barth, Daniel, Nicholas W. Papageorge, and Kevin Thom. 2020. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy* 128 (4):1474–1522.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4):813–846.

———. 1974. "A Theory of Marriage: Part II." *Journal of Political Economy* 82 (2, Part 2):S11–S26.

Becker, Joel, Casper A. P. Burik, Grant Goldman, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Daniel W. Belsky et al. 2021. "Resource profile and user guide of the Polygenic Index Repository." *Nature Human Behaviour* 5 (12):1744–1758.

Becker, Sascha O., Denni Tommasi, and Lina Zhang. 2025. "Killing Distance: How Fixed Effects Drain Power in Spatial Designs." Working paper. Available at https://sites.google.com/site/dennitommasi/.

Belsky, D. W., T. E. Moffitt, D. L. Corcoran, B. Domingue, H. Harrington, S. Hogan, R. Houts et al. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development." *Psychological Science* 27 (7):957–972.

Benjamin, Daniel J., David Cesarini, Christopher F. Chabris, Edward L. Glaeser, David I. Laibson, Vilmundur Guðnason, Tamara B. Harris et al. 2012. "The promises and pitfalls of genoeconomics." *Annual Review of Economics* 4:627–662.

Benjamin, Daniel J., David Cesarini, Patrick Turley, and Alexander Strudwick Young. 2026. "Social-Science Genomics: Progress, Challenges, and Future Directions." *Journal of Economic Literature* Forthcoming.

Bharati, Tushar, Simon Chang, and Qing Li. 2023. "Does Tertiary Education Expansion Affect the Fertility of Women Past the College-Entry Age?" *Journal of Economic Behavior & Organization* 212:1029–1055.

Biroli, Pietro, Titus Galama, Stephanie von Hinke, Hans van Kippersluis, Cornelius A. Rietveld, and Kevin Thom. 2026. "The Economics and Econometrics of Gene–Environment Interplay." *Review of Economic Studies* 93 (1):144–180.

Blossfeld, Hans-Peter and Andreas Timm, editors. 2003. *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*. Dordrecht: Kluwer Academic Publishers.

Bolton, Paul. 2012. "Education: Historical Statistics." Standard Note SN/SG/4252, House of Commons Library.

Braccioli, Federica, Paolo Ghinetti, Simone Moriconi, Costanza Naguib, and Michele Pellizzari. 2024. "Education Expansion, College Choice and Labour Market Success." BSE Working Paper 1419, Barcelona School of Economics.

Bratti, Massimiliano, Elena Cottini, and Paolo Ghinetti. 2022. "Education, Health and Health-Related Behaviors: Evidence from Higher Education Expansion." IZA Discussion Paper 15035, IZA Institute of Labor Economics.

Calhoun, Craig. 2014. "The Robbins Report and British Higher Education Past and Future." In *Shaping Higher Education: 50 Years After Robbins*, edited by Nicholas Barr. London: London School of Economics and Political Science, 65–86.

Card, David. 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, edited by L. N. Christofides, E. K. Grant, and R. Swidinsky. University of Toronto Press, 201–222.

Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6):2754–2781.

Carneiro, Pedro, Kai Liu, and Kjell G. Salvanes. 2023. "The Supply of Skill and Endogenous Technical Change: Evidence from a College Expansion Reform." *Journal of the European Economic Association* 21 (1):48–92.

Carpentier, Vincent. 2006. "Funding in higher education and economic growth in France and the United Kingdom, 1921-2003." *Higher Education Management and Policy* 18 (3):1–26.

Chan, Tak Wing, Morag Henderson, and Rachel Stuchbury. 2019. "Family Size and Educational Attainment in England and Wales." *Population Studies* 73 (2):165–178.

Chang, Christopher C., Carson C. Chow, Laurent C. A. M. Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. 2015. "Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets." *GigaScience* 4:7.

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir. 2018. "The Marriage Market, Labor Supply, and Education Choice." *Journal of Political Economy* 126 (S1):S26–S72.

Chiappori, Pierre-André, Monica Costa Dias, Costas Meghir, and Hanzhe Zhang. 2025. "Changes in Marital Sorting: Theory and Evidence from the United States." *Journal of Political Economy* 133 (10):3045–3077.

Chiappori, Pierre-André, Carlo Fiorio, Alfred Galichon, and Stefano Verzillo. 2026. "Assortative Matching on Income." *Econometrica* 94 (3):957–989.

Chiappori, Pierre-André, Murat Iyigun, and Yoram Weiss. 2009. "Investment in Schooling and the Marriage Market." *American Economic Review* 99 (5):1689–1713.

Clark, Damon and Heather Royer. 2013. "The Effect of Education on Adult Mortality and Health: Evidence from Britain." *American Economic Review* 103 (6):2087–2120.

Collado, M. Dolores, Ignacio Ortuño-Ortín, and Jan Stuhler. 2023. "Estimating intergenerational and assortative processes in extended family data." *Review of Economic Studies* 90 (3):1195–1227.

Currie, Janet and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Quarterly Journal of Economics* 118 (4):1495–1532.

Dang, Thang, Mika Haapanen, and Tuomo Suhonen. 2026. "The Long-Term Health Consequences of Expanding Access to Higher Education." *Economics of Education Review* 112:102794.

Davies, Neil M., Matt Dickson, George Davey Smith, Gerard J. van den Berg, and Frank Windmeijer. 2018. "The Causal Effects of Education on Health Outcomes in the UK Biobank." *Nature Human Behaviour* 2 (2):117–125.

Domingue, Benjamin W., Jason Fletcher, Dalton Conley, and Jason D. Boardman. 2014. "Genetic and Educational Assortative Mating among US Adults." *Proceedings of the National Academy of Sciences* 111 (22):7996–8000.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2019. "Educational Assortative Mating and Household Income Inequality." *Journal of Political Economy* 127 (6):2795–2835.

Fagereng, Andreas, Luigi Guiso, and Luigi Pistaferri. 2026. "Marriage, Assortative Mating and Wealth Inequality." *Review of Economic Studies* Advance Access.

Fang, Ming and Sai Luo. 2026. "More than a Degree: China's College Expansion and the Marriage Market." *Journal of Comparative Economics* 54 (3):709–731.

Fernández, Raquel, Nezih Guner, and John Knowles. 2005. "Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality." *Quarterly Journal of Economics* 120 (1):273–344.

Fry, Anna, Thomas J. Littlejohns, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, and Naomi E. Allen. 2017. "Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population." *American Journal of Epidemiology* 186 (9):1026–1034.

Gayle, George-Levi, Limor Golan, and Mehmet Soytas. 2022. "What is the Source of the Intergenerational Correlation in Earnings?" *Journal of Monetary Economics* 129:24–45.

Geruso, Michael and Heather Royer. 2018. "The Impact of Education on Family Formation: Quasi-Experimental Evidence from the UK." NBER Working Paper 24332, National Bureau of Economic Research.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review: Papers and Proceedings* 104 (5):348–353.

Hofmeister, Robin J., Davide Marnetto, Théo Cavinato, Leona Knüsel, Giorgia Modenini, Erwan Pennarun, Andres Metspalu et al. 2026. "Parental haplotype reconstruction in up to 440,209 UK and Estonian individuals reveals recent assortative mating dynamics." *Nature Human Behaviour* .

Horwitz, Tanya B., Jared V. Balbona, Katie N. Paulich, and Matthew C. Keller. 2023. "Evidence of Correlations between Human Partners Based on Systematic Reviews and Meta-Analyses of 22 Traits and UK Biobank Analysis of 133 Traits." *Nature Human Behaviour* 7 (9):1568–1583.

Ichino, Andrea, Aldo Rustichini, and Giulio Zanella. 2026. "College, Cognitive Ability, and Socioeconomic Disadvantage: Policy Lessons from the U.K. in 1960–2004." *Review of Economic Studies* Advance online publication.

Kalmijn, Matthijs and Henk Flap. 2001. "Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices." *Social Forces* 79 (4):1289–1312.

Kirkebøen, Lars J., Edwin Leuven, Magne Mogstad, and Jack Mountjoy. 2025. "College as a Marriage Market." NBER Working Paper 28688, National Bureau of Economic Research. Revised May 2025.

Kling, Jeffrey R. 2001. "Interpreting Instrumental Variables Estimates of the Returns to Schooling." *Journal of Business & Economic Statistics* 19 (3):358–364.

Kong, Augustine, Gudmar Thorleifsson, Michael L. Frigge, Bjarni J. Vilhjalmsson, Alexander I. Young, Thorgeir E. Thorgeirsson, Stefania Benonisdottir et al. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359 (6374):424–428.

Kweon, Hyeokmoon, Casper Burik, Richard Karlsson Linnér, Ronald de Vlaming, Aysu Okbay, Daphne Martschenko, Kathryn Harden, Thomas A. DiPrete, and Philipp Koellinger. 2020. "Genetic Fortune: Winning or Losing Education, Income, and Health." Tinbergen Institute Discussion Paper 2020-053/V, Tinbergen Institute.

Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet et al. 2018. "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals." *Nature Genetics* 50 (8):1112–1121.

Lefgren, Lars and Frank McIntyre. 2006. "The Relationship between Women's Education and Marriage Outcomes." *Journal of Labor Economics* 24 (4):787–830.

Lynch, Michael and Bruce Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.

Manichaikul, Ani, Josyf C. Mychaleckyj, Stephen S. Rich, Kathy Daly, Michele Sale, and Wei-Min Chen. 2010. "Robust Relationship Inference in Genome-Wide Association Studies." *Bioinformatics* 26 (22):2867–2873.

Mbatchou, Joelle, Leland Barnard, Joshua Backman, Anthony Marcketta, Jack A. Kosmicki, Andrey Ziyatdinov, Christian Benner et al. 2021. "Computationally Efficient Whole-Genome Regression for Quantitative and Binary Traits." *Nature Genetics* 53 (7):1097–1103.

Nybom, Martin, Erik Plug, Bas van der Klaauw, and Lennart Ziegler. 2023. "Skills, Parental Sorting, and Child Inequality." Working Paper 2023:12, Institute for Evaluation of Labour Market and Education Policy (IFAU).

Office for National Statistics. 2023. "Postal Geographies." Updated 3 January 2023.

Okbay, Aysu, Yeda Wu, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Seyed Moeen Nehzati, Julia Sidorenko et al. 2022. "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals." *Nature Genetics* 54 (4):437–449.

Papageorge, Nicholas W. and Kevin Thom. 2020. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." *Journal of the European Economic Association* 18 (3):1351–1399.

Pratt, John. 1997. *The Polytechnic Experiment: 1965–1992*. Buckingham: Society for Research into Higher Education and Open University Press.

Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. "Principal components analysis corrects for stratification in genome-wide association studies." *Nature Genetics* 38 (8):904–909.

Robbins, Lionel. 1963. *Higher Education: Report of the Committee Appointed by the Prime Minister Under the Chairmanship of Lord Robbins, 1961-63*. London: Her Majesty's Stationery Office (HMSO). Cmnd. 2154.

Robinson, Matthew R., Aaron Kleinman, Mariaelisa Graff, Anna A. E. Vinkhuyzen, David Couper, Michael B. Miller, Wouter J. Peyrot et al. 2017. "Genetic Evidence of Assortative Mating in Humans." *Nature Human Behaviour* 1 (1):0016.

Rogne, Adrian Farner, Agnes Fauske, and Rannveig Kaldager Hart. 2025. "Educational Expansions and Fertility: Evidence from Norwegian College Reforms." *European Journal of Population* 41 (1):14.

Sebro, Ronnie, Gina M. Peloso, Josée Dupuis, and Neil J. Risch. 2017. "Structured Mating: Patterns and Implications." *PLOS Genetics* 13 (4):e1006655.

Sharp, C., D. Hutchison, and W. Keys. 2002. "Comparing school systems to explain enduring birth date effects: a response to McDonald (2001)." *Compare: A Journal of Comparative and International Education* 32 (2):239–241.

Southall, H. R. and D. Dorling. 2021. "Great Britain Historical Database: Census Data: Education Statistics, 1951–1961."

Steelman, Lala Carr, Brian Powell, Regina Werum, and Scott Carter. 2002. "Reconsidering the Effects of Sibling Configuration: Recent Advances and Challenges." *Annual Review of Sociology* 28 (1):243–269.

Sudlow, Cathie, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey et al. 2015. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLoS Medicine* 12 (3):e1001779.

Suhonen, Tuomo and Hannu Karhunen. 2019. "The intergenerational effects of parental higher education: Evidence from changes in university accessibility." *Journal of Public Economics* 176:195–217.

Swanson, John M. 2012. "The UK Biobank and selection bias." *Lancet* 380 (9837):110.

Toivanen, Otto and Lotta Väänänen. 2016. "Education and Invention." *Review of Economics and Statistics* 98 (2):382–396.

Trampush, Joey W., M. L. Z. Yang, J. Yu, E. Knowles, G. Davies, D. C. Liewald, J. M. Starr et al. 2017. "GWAS Meta-Analysis Reveals Novel Loci and Genetic Correlates for General Cognitive Function: A Report from the COGENT Consortium." *Molecular Psychiatry* 22 (3):336–345.

United Kingdom Parliament. 1944. "Education Act 1944." Public General Acts. 7 & 8 Geo. 6 c. 31. Set the minimum school-leaving age at 15 (implemented 1947).

———. 1972. "The Raising of the School Leaving Age Order 1972." Statutory Instruments. Statutory Instrument 1972 No. 444. Raised the minimum school-leaving age to 16 from 1 September 1972.

van Alten, S., B. W. Domingue, J. Faul, T. Galama, and A. T. Marees. 2024. "Reweighting UK Biobank corrects for pervasive selection bias due to volunteering." *International Journal of Epidemiology* 53 (3):dyae054.

von Hinke, Stephanie and Nicolai Vitt. 2024. "An analysis of the accuracy of retrospective birth location recall using sibling data." *Nature Communications* 15 (1):2665.

Westphal, Matthias, Daniel A. Kamhöfer, and Hendrik Schmitz. 2022. "Marginal College Wage Premiums Under Selection Into Employment." *Economic Journal* 132 (646):2231–2272.

Willetts, David. 2014. "Introduction." In *Shaping Higher Education: 50 Years After Robbins*, edited by Nicholas Barr. London: London School of Economics and Political Science, 1–12.

Yengo, Loic, Matthew R. Robinson, Matthew C. Keller, Kathryn E. Kemper, Yuanhao Yang, Maciej Trzaskowski, Jacob Gratten et al. 2018. "Imprint of Assortative Mating on the Human Genome." *Nature Human Behaviour* 2 (12):948–954.

A Assessing False Positives in Spouse Matching

Because spouses are not directly identified in the UK Biobank, the matching procedure may produce false positives, that is, pairs that satisfy all seven matching criteria described in Section 4.1 but are not spouses. Following Abdellaoui et al. (2023), we begin by asking whether the members of an imputed spouse pair share a genetically identified child, inferred using the kinship coefficient described in Section 4.3.2. Among the 540 pairs for which the UK Biobank contains a sampled child of at least one partner, 476 share the same genetically identified child. Thus, 88.2% of the genetically informative pairs are confirmed as the joint biological parents of a sampled participant.

To gauge what this validation rate means in practice, we compare it with the corresponding rate in a sample of "fake pairs." These are opposite-sex participants who share at least one recorded home postcode but satisfy exactly five of the seven criteria used to identify spouses. Among the 2,049 one-to-one fake pairs for which the UK Biobank contains a sampled genetic child of at least one member, only 16, or 0.8%, share the same genetically identified child. The sharp contrast between the two samples indicates that requiring all seven criteria successfully distinguishes joint-parent pairs from otherwise similar five-of-seven matches.

The 88.2% validation rate is, if anything, a conservative measure of the accuracy of the matching procedure. Both rates are conditional on the UK Biobank containing a sampled child genetically related to at least one member of the pair, and failure to observe a shared genetically identified child does not imply that a pair is incorrectly matched. Genuine spouses may have no joint biological children, may form a stepfamily, or may have joint children who are not included in the UK Biobank. Therefore, these figures cannot be used to recover the overall false-positive rate.

We also use the fake pairs to assess the direction of bias from any false positives remaining in the imputed-spouse sample. Specifically, we re-estimate the reduced-form relationship between the reduction in distance to the nearest university and partner outcomes using the five-of-seven fake partners in place of the imputed spouses. These estimates provide an empirical benchmark for the associations that plausible false-positive matches would contribute to the main results.

Table A.1 reports the results. For university attainment and school-leaving age, the fake-pair coefficients have the same sign as the imputed-spouse estimates but are substantially smaller: 0.007 rather than 0.027 for university, and 0.023 rather than 0.139 for school-leaving age. The fake-pair estimates are therefore 25% and 17% of the corresponding imputed-spouse coefficients. For the EA PGI, the contrast is stronger: the coefficient is 0.021 among imputed spouses but close to zero among fake pairs, at −0.009. The wage coefficient is also smaller among fake pairs, although the difference is modest: 0.007 compared with 0.011 among imputed spouses. Thus, if any false positives remaining in the spouse sample resemble the five-of-seven pairs, they would attenuate rather than generate the estimated relationships, with the implied attenuation being most pronounced for education and the EA PGI and considerably smaller for wages.

Table A.1: Reduced-Form Estimates Using Imputed Spouses and Fake Pairs

	Polygenic Indices			*Education and Wages*		
	(1)	(2)	(3)	(4)	(5)	(6)
	EA PGI	Cog PGI	Math PGI	University	SLA	Log Wage
Imputed spouse pairs	0.021**	0.031***	0.016*	0.027***	0.139***	0.011***
	(0.009)	(0.009)	(0.009)	(0.005)	(0.023)	(0.003)
Fake pairs	-0.009	-0.001	0.000	0.007*	0.023	0.007***
	(0.008)	(0.007)	(0.007)	(0.004)	(0.022)	(0.002)
Fake coef. relative to imputed spouses	-43%	-3%	0%	25%	17%	63%
Observations: imputed spouse pairs	12,494					
Observations: known fake pairs	18,318					

Notes: The table compares reduced-form estimates obtained using the imputed spouse pairs in the benchmark analysis with corresponding estimates obtained using fake pairs. Fake pairs are opposite-sex participants who share at least one recorded home postcode but satisfy exactly five of the seven criteria used to identify spouses, and are therefore not classified as spouses by the matching procedure. We construct a one-to-one fake-pair sample using seeded maximum-cardinality matching after capping each participant-specific candidate degree at ten. All pairs are complete male–female pairs with kinship no greater than 0.05. Reduction in distance to the nearest university is measured per 10 km. All specifications include the corresponding baseline controls, birth-region-by-cohort fixed effects, own and partner genetic principal components, and UK Biobank inverse-probability weights. In both samples, standard errors are two-way clustered by pair and numerical birth-district ID, with inference based on birth-district cluster degrees of freedom. Cluster-robust standard errors are reported in parentheses. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices used in the main analysis (Appendix B.1); both samples are scored from discovery GWASs that exclude the imputed-spouse and fake-pair samples and their close relatives, and each index is standardised within its sample. The coefficient ratios are descriptive and equal the fake-pair coefficient divided by the corresponding imputed-spouse coefficient; a negative value indicates opposite signs. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B Constructing the Genetic Measures

B.1 Construction and Validation of the Polygenic Indices

This appendix documents the sources and construction of the Educational Attainment, Cognitive Performance and Highest Mathematics polygenic indices (PGIs) used to examine whether university education changes partner composition along related but distinct dimensions.

B.1.1 Polygenic Indices Used in the Analysis

The main assortative-mating analysis uses three PGIs: Educational Attainment, Cognitive Performance, and Highest Mathematics. The Educational Attainment PGI is based on the GWAS of Lee et al. (2018), estimated without UK Biobank participants. The Cognitive Performance PGI is constructed from our own GWAS of fluid intelligence within the UK Biobank, while the Highest Mathematics PGI comes from the single-trait index provided by the Polygenic Index Repository (Becker et al., 2021). The latter draws exclusively on 23andMe data, with a discovery sample of 430,439 individuals.

An important consideration is that the individuals whose PGIs are analysed should not contribute to the GWAS used to construct them. Otherwise, overlap between the discovery and analysis samples could inflate predictive performance and introduce dependencies between the indices and the outcomes under study. For our three main PGIs, neither the analysed couples nor their close relatives contribute to the underlying discovery samples. We also examine alternative indices for educational attainment, cognitive performance, and other cognitive traits as robustness checks.[27]

All indices are oriented so that higher values correspond to higher predicted values of the relevant phenotype and standardised to mean zero and unit variance in each analysis sample.

B.1.2 Construction of the In-House Indices

To construct the Cognitive Performance PGI and the additional in-house indices, we estimate GWASs within the UK Biobank after excluding the couples sample and their close relatives. Following standard genetic quality-control restrictions, the resulting discovery sample contains 394,270 individuals. Table A.2 describes the phenotypes considered, including educational attainment and several cognitive tests. The ROSLA and fake-pair samples used in the assortative-mating analysis are scored from a second GWAS of the same design that additionally excludes those samples and their close relatives, with a discovery sample of 363,991 individuals.

We estimate the GWASs using REGENIE (Mbatchou et al., 2021), which accounts for genetic relatedness and population structure. The specifications include sex, age, age squared, genotyping array, and 40 genetic principal components. Association estimates are obtained for approximately 1.18 million imputed HapMap3 variants.

[27] Alternative measures include an in-house Educational Attainment PGI, a Cognitive Performance PGI based on the COGENT GWAS (Trampush et al., 2017), and indices constructed from additional cognitive tests. We also consider the multi-trait Highest Mathematics PGI from the Polygenic Index Repository. Unlike the main indices, the latter involves discovery-sample overlap for some couples and is therefore used only as a robustness measure (Appendix Table A.17).

Table A.2: Phenotypes Used in the In-House GWAS

Trait	UKB field(s)	Measure
Educational attainment	6138	years of education (Lee et al., 2018); standardised
Reaction time	20023	mean time to correctly identify matches (milliseconds); higher = slower
Pairs matching	399 (second round)	error count
Fluid intelligence	20016 (online: 20191)	number of correct answers
Numeric memory	4282 (online: 20240)	maximum digits recalled
Trail making A / B	6348 / 6350 (online: 20156 / 20157)	duration; higher = worse
Symbol–digit substitution	23324	correct matches
Prospective memory	20018	binary: correct at first attempt

Notes: Field numbers refer to the UK Biobank Data Showcase. For tests repeated at later visits or administered online (online field in parentheses), we use the baseline measurement where available and otherwise the earliest later measurement. Continuous cognitive measures are rank-based inverse-normal transformed within the GWAS sample. Educational attainment is mapped to years of education following Lee et al. (2018) and standardised. Trail-making times of zero indicate an incomplete test and are set to missing. Prospective memory enters a GWAS but no PGI is constructed from it.

We then construct PGIs by combining these association estimates with individual-level genetic data. Because nearby variants can be correlated through linkage disequilibrium, we use clumping to retain approximately independent variants, applying a 250 kb window and an r^2 threshold of 0.1. The procedure uses a reference sample of 20,000 individuals drawn from the GWAS discovery sample and does not impose a significance threshold on the retained variants. The resulting indices are calculated from imputed genotype dosages using PLINK2. Unlike these in-house scores, the externally sourced indices use LDpred to account for linkage disequilibrium.

B.1.3 Validation in the Couples Sample

We assess the predictive performance of the indices in the couples sample, which is independent of the discovery samples underlying the main PGIs. Table A.3 reports the correlation between each index and its corresponding phenotype, together with the additional variation explained after controlling for sex, age, age squared, and 40 genetic principal components.

The Educational Attainment PGI explains an additional 4.0 percent of the variation in university completion and in school-leaving age in the couples sample. The Cognitive Performance PGI explains an additional 4.8 percent of the variation in fluid intelligence, and the Highest Mathematics PGI 3.0 percent of the variation in university completion, which we use as its validation outcome because the corresponding mathematics measure is unavailable in the UK Biobank. Among the cognitive indices, fluid intelligence exhibits the strongest predictive performance for its corresponding phenotype, supporting its use as the main cognitive measure in the assortative-mating analysis.

These validation results establish that the indices contain predictive information in the population studied here, although their performance need not match that reported in larger or less selected samples. The restricted variation in educational attainment within the couples sample, for example, helps explain the lower predictive performance of the educational-attainment indices relative to broader validation populations.

Table A.3: Out-of-Sample Predictive Power of the PGIs

PGI	Validation phenotype	N	r	ΔR^2
Panel A: Main PGIs				
Educational attainment (EA3, no UKB)	university degree	12,494	0.204	4.0%
	school-leaving age	12,494	0.206	4.0%
Fluid intelligence (in-house)	fluid intelligence	9,656	0.225	4.8%
Highest Mathematics (single-trait)	university degree	12,494	0.178	3.0%
Panel B: Robustness PGIs				
Educational attainment (in-house)	university degree	12,494	0.199	3.7%
Cognitive performance (COGENT)	fluid intelligence	9,656	0.110	1.1%
Numeric memory (in-house)	numeric memory	5,231	0.159	2.5%
Reaction time (in-house)	reaction time	12,473	0.141	2.2%
Symbol–digit substitution (in-house)	symbol–digit	2,242	0.118	1.2%
Trail making B (in-house)	trail making B	6,672	0.117	1.2%
Pairs matching (in-house)	pairs matching	12,494	0.077	0.6%
Trail making A (in-house)	trail making A	6,953	0.068	0.3%

Notes: r is the correlation between the PGI and the validation phenotype. ΔR^2 is the incremental R^2, the increase in R^2 when the PGI is added to a regression of the phenotype on sex, age, age squared, and 40 genetic principal components. "In-house" indices are constructed from our own GWAS within the UK Biobank, which excludes the couples sample and their close relatives (Table A.2). "EA3, no UKB" denotes the Educational Attainment PGI based on Lee et al. (2018), with UK Biobank participants excluded from the discovery sample. "Single-trait" denotes the Polygenic Index Repository index estimated from a GWAS of that trait alone (here 23andMe participants only), as opposed to the Repository's multi-trait indices, which also draw on GWASs of related traits. "COGENT" denotes the cognitive-performance GWAS of Trampush et al. (2017). The main analysis sample contains 12,494 individuals; N is smaller for phenotypes measured in a subset of participants.

B.2 Construction of Spouse-Pair Kinship

We construct spouse-pair kinship using autosomal genotype data from the UK Biobank. We retain spouse pairs for whom genotype data are available for both partners and restrict the variant set to autosomal SNPs with unambiguous A/C/G/T alleles. We then apply standard variant-level quality-control restrictions, excluding SNPs with more than 2 percent missing genotypes, minor-allele frequency below 1 percent, or evidence of deviation from Hardy–Weinberg equilibrium at the thresholds used in our genetic pipeline. These restrictions leave 532,077 variants. Because nearby SNPs can be highly correlated through linkage disequilibrium, we prune the remaining set using a 200-variant window, a 50-variant step, and an r^2 threshold of 0.1. This leaves 132,690 approximately independent variants for the relatedness analysis.

Using this pruned set, we estimate genetic relatedness for each predefined spouse pair with the KING-robust estimator (Manichaikul et al., 2010), implemented in PLINK2 (Chang et al., 2015). The estimator uses patterns of allele sharing across the genome to recover pairwise relatedness and is designed to be relatively robust to population structure, making it well suited to samples in which individuals may differ in ancestry-related genetic background. We restrict the calculation to the observed spouse pairs rather than estimating relatedness among all possible pairs in the genetic sample. For each couple, we retain the resulting kinship coefficient as our raw measure of genome-wide genetic relatedness. Kinship estimates are available for 16,664 spouse pairs.

Before constructing the analysis sample, we exclude pairs with raw kinship greater than 0.05, thereby removing couples displaying levels of genetic relatedness inconsistent with an unrelated spouse pair.

Eleven of the 16,664 scored pairs exceed this threshold. Within the otherwise eligible complete-case analysis sample, the restriction removes six couples and leaves the 6,247 couples used in the main analysis. Negative kinship estimates are retained unchanged. For the couple-similarity regressions, the raw kinship coefficient is standardised to mean zero and unit variance in the corresponding estimation sample. Specifications involving genetic outcomes additionally condition on genetic principal components for both partners.

C Descriptive Statistics

Table A.4: Summary Statistics for the Complete-Couple Analysis Sample

	(1) Mean	(2) SD	(3) Non-Graduates	(4) University Graduates	(5) Difference (4) − (3)
Outcomes (*Y*)					
Spouse EA PGI	−0.01	0.99	−0.11	0.19	0.29 (0.02)
Spouse Cog PGI	−0.01	1.00	−0.08	0.14	0.22 (0.03)
Spouse Math PGI	−0.02	1.00	−0.11	0.18	0.29 (0.03)
Spouse University Degree	0.33	0.47	0.20	0.59	0.38 (0.02)
Spouse SLA	18.41	2.49	17.83	19.63	1.81 (0.07)
Spouse Log Hourly Wage	2.64	0.34	2.60	2.74	0.14 (0.01)
Treatment and Instrument					
Own University Degree	0.32	0.47			
Δ*Dist*	3.60	11.09	2.86	5.13	2.27 (0.26)
ROSLA Eligibility	0.52	0.50	0.55	0.45	−0.10 (0.01)
Controls (***X***)					
A. Individual Characteristics					
Birth Year	1957.1	6.3	1957.5	1956.3	−1.2 (0.2)
Age at Assessment	51.37	6.31	50.99	52.17	1.18 (0.18)
Male	0.52	0.50	0.51	0.54	0.03 (0.01)
EA PGI	−0.02	0.99	−0.17	0.29	0.46 (0.02)
Number of Siblings	1.94	1.55	2.00	1.81	−0.19 (0.04)
B. Birth District Characteristics					
Baseline universities within 10 km	2.84	6.55	2.78	2.97	0.18 (0.19)
Urban Residence	0.94	0.24	0.94	0.94	0.00 (0.01)
Education Level (% > 15 years)	0.23	0.08	0.23	0.24	0.01 (0.00)
Population Growth (1946–1951)	0.03	0.02	0.02	0.03	0.00 (0.00)
Log Population (1951)	4.92	1.22	4.96	4.83	−0.13 (0.03)
Log Population Density (1951)	8.03	0.96	8.04	8.00	−0.04 (0.02)
Unemployment Rate (1951)	1.04	0.94	1.06	1.00	−0.06 (0.02)
Unemployment Rate Squared (1951)	1.97	3.49	2.02	1.84	−0.18 (0.09)
Share High SES	0.16	0.06	0.16	0.17	0.01 (0.00)
Share Low SES	0.30	0.07	0.30	0.29	−0.01 (0.00)
Couple-Level Characteristics					
Kinship ($\times 10^3$)	−0.04	5.23	−0.07	0.00	0.07 (0.13)
EA PGI Diff^2	1.78	2.59	1.75	1.82	0.07 (0.07)
SLA Diff^2	8.52	21.75	9.35	6.80	−2.56 (0.49)
Wage Diff^2	0.24	0.30	0.26	0.20	−0.05 (0.01)
Birth Distance Between Spouses (km)	55.48	92.97	38.31	91.22	52.91 (3.07)
N individual observations	12,494		8,263	4,231	
N couples			6,247		

Notes: The table reports descriptive statistics for the kinship-consistent complete-couple sample used in the 2SLS analysis (complete male–female pairs with a quality-controlled KING estimate and kinship ≤ 0.05). Columns (1) and (2) report the pooled mean and standard deviation; columns (3) and (4) report means for focal individuals without and with a university degree; column (5) reports the difference between the two, with spouse-pair-clustered standard errors in parentheses. All statistics are weighted using the UK Biobank inverse-probability weights employed in the main analysis. The unit of observation is the focal individual: each couple contributes two observations, and couple-level characteristics are assigned to both partners. Δ*Dist* is measured in kilometres. ROSLA eligibility indicates birth from September 1957 onward. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. Squared differences are own value minus spouse value, squared. Kinship (multiplied by 10^3) and the squared differences are reported in raw units (with the PGIs standardised); the regressions use standardised versions, with similarity defined as the negative squared difference.

Table A.5: Distribution of the Reduction in Distance to the Nearest University

	(1) Overall	(2) Positive Reduction	(3) Negative Reduction	(4) No Reduction
Mean reduction	3.601	13.088	−0.377	0
Standard deviation	11.093	17.956	0.476	0
Minimum	−3.858	0.001	−3.858	0
25th percentile	0	1.078	−0.407	0
Median	0	2.942	−0.260	0
75th percentile	0.755	20.487	−0.136	0
Maximum	92.308	92.308	−0.012	0
Observations	12,494	3,184	130	9,180
Share of sample	100%	27.55%	1.39%	71.06%

Notes: Summary statistics for the reduction in distance to university (measured in kilometres). The reduction is calculated as the difference between the distance to the nearest university active at birth and the distance to the nearest university active by age 15. Means, standard deviations, quartiles, and frequencies use the UKB inverse-probability weights; observation counts and extrema are unweighted.

D Identification and Robustness

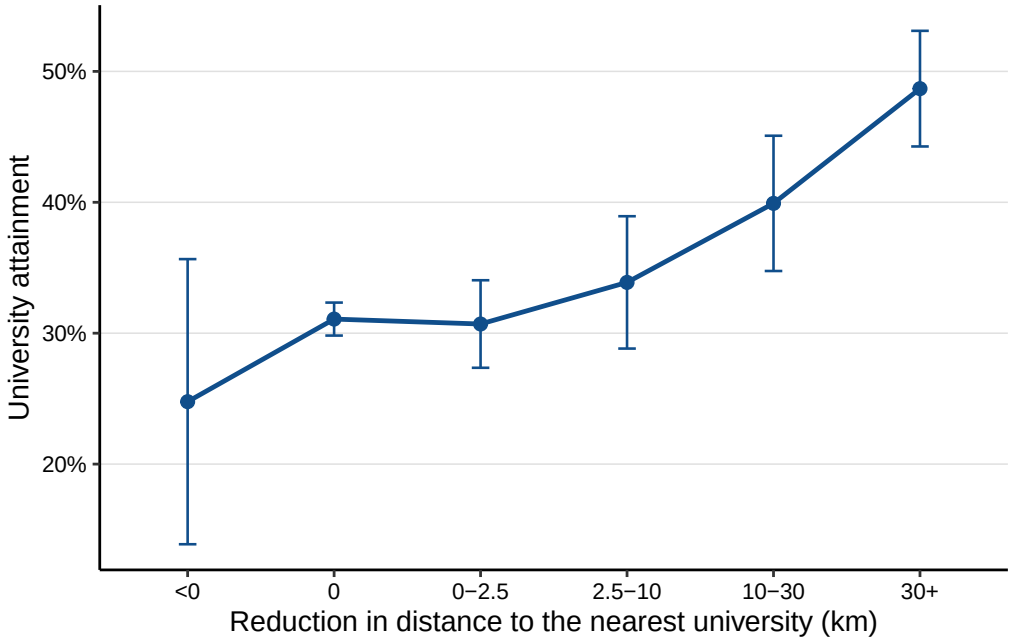


Figure A.1: University Attainment by Bins of the Reduction in Distance. The figure plots the weighted university-attainment rate across bins of the reduction in distance to the nearest university, $\Delta Dist$, in the benchmark 2SLS sample. Bins distinguish negative reductions, no change, and positive reductions of increasing size. The plotted values correspond to the "Mean University" column of Table A.6. Vertical bars report 95 percent confidence intervals based on the weighted effective sample size in each bin. Means are weighted using UK Biobank inverse-probability weights.

Table A.6: Descriptive Monotonicity Check

$\Delta Dist$ bin	(1) Mean $\Delta Dist$	(2) Mean University	(3) Residualised University	(4) Δ	(5) Δ_0	(6) Observations
< 0	-0.38	0.248	-0.045	—	-0.039	130
0	0.00	0.311	-0.006	0.039	—	9,184
$(0, 2.5)$	1.11	0.307	-0.022	-0.017	-0.017	1,159
$[2.5, 10)$	4.29	0.339	0.006	0.029	0.012	577
$[10, 30)$	17.43	0.399	0.042	0.036	0.048	615
$[30, \infty)$	46.60	0.487	0.096	0.054	0.102	829

Notes: This table provides a descriptive check of the monotonicity condition for the reduction in distance to the nearest university ($\Delta Dist$, in km). University attainment is residualised on the baseline controls, the first 20 own and 20 spouse genetic principal components, and birth-region-by-cohort fixed effects, excluding $\Delta Dist$ itself, using UK Biobank inverse-probability weights. Observations are grouped into bins that distinguish negative reductions, no change, and positive reductions of increasing size. For each bin, the table reports the weighted mean of $\Delta Dist$, the weighted university-attainment rate, the weighted mean of residualised university attainment, and the number of observations. Δ reports the change in mean residualised university attainment relative to the preceding bin, and Δ_0 the difference relative to the zero-change bin. The sample is the benchmark 2SLS sample.

Table A.7: First-Stage Heterogeneity by Socioeconomic Background, Birthplace, Cohort and EA PGI

Group	$\Delta Dist$	First-stage F	Num. Obs.	Couples
High-SES birth district	0.046***	34.10	6,341	4,009
	(0.008)			
Low-SES birth district	0.016**	4.57	6,153	3,915
	(0.007)			
Urban birth district	0.033***	42.63	11,750	6,187
	(0.005)			
Rural birth district	0.042*	3.10	744	684
	(0.024)			
Born before 1960	0.028***	32.69	9,261	4,998
	(0.005)			
Born 1960 or later	0.067***	18.35	3,233	1,984
	(0.016)			
EA PGI tertile 1	0.024**	4.85	4,165	3,386
	(0.011)			
EA PGI tertile 2	0.042***	31.42	4,165	3,419
	(0.007)			
EA PGI tertile 3	0.021***	7.69	4,164	3,379
	(0.008)			

Notes: Sample: kinship-consistent complete male–female pairs. SES is defined by the sample median of district share low SES at birth; Low SES means above-median share low SES. Urban/rural is measured by the 1951 urban indicator of the birth district. EA PGI tertiles are computed within the analysis sample. Groups classify focal individuals; their spouses may fall in other groups, so couple counts need not add up across groups. The dependent variable is own university attainment. Reduction in distance is measured per 10 km. Standard errors are two-way clustered by spouse pairs and numerical birth-district ID, and inference uses birth-district cluster degrees of freedom. Cluster-robust standard errors are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. All models include controls, cohort-by-region fixed effects, the first 20 own and spouse genetic PCs, and UKB IP weights.

Table A.8: First-Stage Robustness to Alternative Childhood Windows

	(1) Birth→11	(2) Birth→12	(3) Birth→13	(4) Birth→14	(5) Birth→15
Panel A: Whole Sample					
$\Delta Dist$	0.030***	0.031***	0.031***	0.031***	0.031***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Num. Obs.	12494	12494	12494	12494	12494
Nonzero Z	2613	2819	2982	3126	3314
R2	0.114	0.115	0.115	0.115	0.115
First-stage F	38.41	38.95	39.67	39.66	42.10
Panel B: Males					
$\Delta Dist$	0.034***	0.035***	0.035***	0.035***	0.035***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Num. Obs.	6247	6247	6247	6247	6247
Nonzero Z	1310	1406	1492	1564	1667
R2	0.130	0.130	0.130	0.130	0.131
First-stage F	38.19	37.65	36.82	38.08	41.14
Panel C: Females					
$\Delta Dist$	0.024***	0.025***	0.026***	0.025***	0.026***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Num. Obs.	6247	6247	6247	6247	6247
Nonzero Z	1303	1413	1490	1562	1647
R2	0.167	0.167	0.167	0.167	0.167
First-stage F	11.90	13.57	13.91	13.23	13.73

Notes: Sample: kinship-consistent complete-couple sample used in the 2SLS estimates. The dependent variable is own university attainment. Reduction in distance is measured per 10 km for all columns. Whole-sample standard errors are two-way clustered by spouse pair and numerical birth-district ID; sex-specific standard errors are clustered by birth district. Cluster-robust standard errors are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. All specifications include controls, cohort-by-region fixed effects, the first 20 own and spouse genetic PCs, and UKB IP weights.

Table A.9: Robustness of First-Stage and Spousal-Outcome Estimates to Alternative Distance Measures

	(1) Linear	(2) Linear + Quadratic	(3) Log-Distance Change
Panel A: First Stage			
Dependent variable: Own university degree			
$\Delta Dist$	0.031***	0.045***	
	(0.005)	(0.015)	
$\Delta Dist^2$		-0.003	
		(0.002)	
$\Delta \log Dist$			0.047***
			(0.009)
p-value: Squared Term = 0		0.239	
First-stage F-statistic	42.10	23.84	26.38
Panel B: 2SLS Estimates (Spousal Outcomes)			
EA PGI	0.672**	0.703**	0.595*
	(0.260)	(0.275)	(0.303)
Cog PGI	1.021***	1.043***	0.700**
	(0.297)	(0.294)	(0.338)
Math PGI	0.518**	0.487*	0.338
	(0.255)	(0.266)	(0.353)
University	0.861***	0.883***	0.904***
	(0.107)	(0.109)	(0.133)
SLA	4.428***	4.448***	4.675***
	(0.616)	(0.641)	(0.729)
Log Wage	0.351***	0.382***	0.433***
	(0.087)	(0.095)	(0.120)
Observations	12,494	12,494	12,494

Notes: The sample is the whole-sample, kinship-consistent complete-couple sample used in the main analysis. Column (1) instruments own university attainment with the reduction in distance measured in 10 km units. Column (2) uses the reduction and its square as excluded instruments. Column (3) uses natural-log distance at birth minus natural-log distance at age 15. The reported first-stage F-statistic is the cluster-robust Wald statistic for joint significance of the excluded instruments divided by the number of excluded instruments. Column (2) reports the first-stage coefficients on both the linear and squared reduction terms. The quadratic-term test reports the first-stage p-value on the squared reduction. Cluster-robust standard errors are two-way clustered by spouse pair and birth district and are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. All models include the baseline controls, cohort-by-region fixed effects, own and spouse genetic PCs, and UKB IP weights.

Table A.10: Placebo Timing Tests Using Post-Entry Reductions in Distance

	(1)	(2)	(3)	(4)
	25 → 30	25 → 35	25 → 40	Joint p-value
Panel A: Placebo First Stage				
Dependent variable: Own university degree				
Placebo $\Delta Dist$	0.008	0.007	0.008	
	(0.011)	(0.011)	(0.012)	
Placebo F-statistic	0.53	0.36	0.42	
Panel B: Placebo Reduced Form (Spousal Outcomes)				
Spouse EA PGI	0.016	0.023	0.023	0.580
	(0.033)	(0.035)	(0.034)	
Spouse Cog PGI	0.025	0.028	0.022	0.633
	(0.026)	(0.030)	(0.027)	
Spouse Math PGI	-0.007	0.001	-0.002	0.673
	(0.023)	(0.026)	(0.024)	
Spouse University Degree	0.010	0.009	0.014	0.386
	(0.009)	(0.011)	(0.013)	
Spouse SLA	0.069	0.054	0.084	0.371
	(0.065)	(0.064)	(0.074)	
Spouse Log Wage	0.005	0.007	0.010	0.665
	(0.008)	(0.008)	(0.010)	
Observations		12,494		

Notes: Each placebo $\Delta Dist$ is the improvement in access relative to the university network at birth delivered only by universities that obtained university status strictly after age 25 and no later than the indicated age, measured at the birthplace in school years and expressed per 10 km. Openings between birth and age 25 are excluded. Panel A reports placebo regressions with the focal individual's university attainment as the outcome; Panel B reports placebo reduced-form regressions for spousal outcomes. Column (4) reports the two-way-clustered Wald-test p-value for the joint null that the coefficients on all three placebo windows equal zero, accounting for their overlap. All specifications include the baseline controls, birth-region-by-cohort fixed effects, the first 20 own and 20 spouse genetic principal components, and UK Biobank inverse-probability weights. Standard errors are two-way clustered by spouse pair and numerical birth-district ID; inference uses birth-district cluster degrees of freedom. SLA denotes school-leaving age. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. Cluster-robust standard errors are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.11: Balance of Predetermined Characteristics with Respect to University Access

	Demographic and Early-Life Characteristics			
	(1)	(2)	(3)	(4)
	Male	Birth month	Birth weight	Maternal Smoking
$\Delta Dist$	0.001	0.030	-0.014	-0.000
	(0.003)	(0.032)	(0.009)	(0.005)
Num. Obs.	12,494	12,494	7,995	10,943
	Polygenic Indices			
	(1)	(2)	(3)	(4)
	BMI PGI	Cog PGI	Math PGI	Self-Math PGI
$\Delta Dist$	-0.010	-0.000	0.004	0.010
	(0.009)	(0.011)	(0.008)	(0.009)
Num. Obs.	12,494	12,494	12,494	12,494

Notes: Each column reports a separate weighted OLS regression of the listed predetermined characteristic on the reduction in distance to the nearest university. Reduction in distance is measured per 10 km. Outcome availability is not imposed when defining the benchmark sample; each regression therefore uses all observations for which the corresponding balance outcome is available. Male and maternal smoking are indicator variables, and birth weight is measured in kilograms. All specifications condition on the baseline controls, birth-region-by-cohort fixed effects, the first 20 genetic principal components for both members of the pair, and UK Biobank inverse-probability weights. The focal participant's EA PGI and number of siblings are included in the conditioning set and are therefore not reported as balance outcomes. Because sex is itself a baseline control, it is omitted from the conditioning set in column (1), making that column a leave-one-covariate-out balance diagnostic. BMI PGI, Cog PGI, Math PGI, and Self-Math PGI denote the body mass index, Cognitive Performance, Highest Mathematics, and self-rated mathematics ability polygenic indices. Cluster-robust standard errors, two-way clustered by spouse pair and numerical birth-district ID, are reported in parentheses; inference uses birth-district cluster degrees of freedom. These regressions assess whether residual variation in the instrument is related to additional predetermined characteristics, conditional on the benchmark specification. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.12: University Access and Spouses' Birthplace and Exposure Similarity

	(1) Birth distance (km)	(2) Same birth district	(3) Same birth county	(4) Same birth region	(5) Birth distance ≥30 km	(6) Spouse Δ*Dist*	(7) Age difference (years)
Δ*Dist*	9.842***	-0.027***	-0.036***	-0.036***	0.056***	0.338***	-0.010
	(1.376)	(0.005)	(0.007)	(0.006)	(0.006)	(0.036)	(0.021)
Num. Obs.	12,494	12,494	12,494	12,494	12,494	12,494	12,494

Notes: Sample: kinship-consistent complete male–female pairs used in the 2SLS analysis. Each column reports the coefficient from a separate weighted OLS regression of the listed outcome on the focal individual's reduction in distance, measured per 10 km. Birth distance is the straight-line distance between the spouses' birthplaces in kilometres, and column (5) indicates birthplaces at least 30 km apart. Columns (2)–(4) indicate that both spouses were born in the same 1951 district, county or region. Spouse Δ*Dist* is the spouse's reduction in distance, also measured per 10 km. Age difference is the focal individual's age minus the spouse's age at assessment, in years. Cluster-robust standard errors, two-way clustered by spouse pairs and numerical birth-district ID, are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. These outcomes describe spousal sorting by birth geography and exposure, not predetermined individual balance. All specifications include controls, cohort-by-region fixed effects, the first 20 own and 20 spouse genetic PCs, and UKB IP weights; when the tested outcome is also a control, it is omitted from the control set.

Table A.13: Predetermined District Characteristics and the Geography of University Expansion

	(1)	(2)	(3)	(4)
	Any opening	Any opening	Opening count	Opening count
Baseline universities within 10 km	0.028**	0.028*	0.065**	0.079*
	(0.012)	(0.014)	(0.031)	(0.041)
Urban district	-0.007	-0.010	0.033	0.029
	(0.021)	(0.021)	(0.036)	(0.035)
Log population, 1951	0.035***	0.034***	0.068***	0.072***
	(0.009)	(0.010)	(0.020)	(0.022)
Log population density, 1951	0.004	0.006	-0.016	-0.007
	(0.009)	(0.009)	(0.023)	(0.018)
Population growth, 1946–1951	-0.006	-0.006	-0.008	-0.006
	(0.004)	(0.004)	(0.007)	(0.007)
Unemployment rate, 1951	-0.009***	-0.010	-0.011**	-0.025*
	(0.003)	(0.008)	(0.005)	(0.015)
Share educated over age 15	0.015**	0.013**	0.020*	0.019
	(0.006)	(0.007)	(0.010)	(0.012)
Low-SES share	0.008	0.007	0.009	0.010
	(0.006)	(0.006)	(0.009)	(0.009)
High-SES share	0.005	0.006	0.008	0.013
	(0.006)	(0.006)	(0.011)	(0.010)
Region fixed effects	No	Yes	No	Yes
Num. districts	1,468	1,468	1,468	1,468
R^2	0.086	0.087	0.107	0.114

Notes: The unit of observation is a 1951 England–Wales local-government district. The analysis uses the full district universe, subject only to complete data for the district characteristics included in the regressions. University events are campus-level openings or university-status acquisitions from 1945 to 1980, defined using the same university-status date field as the instrument construction and assigned to districts using campus coordinates and 1951 district boundaries. Baseline university provision is the number of universities active in 1944 within 10 km of an interior point of each district (a point guaranteed to lie within its 1951 boundary). All predictors except the urban indicator, including baseline university provision, are standardised using their means and standard deviations across the full district universe, so coefficients give the change associated with a one-standard-deviation increase. The 1951 unemployment rate uses the district rate where available and the fixed regional rate otherwise, so unemployment does not reduce the sample. District characteristics are measured at or near the beginning of the expansion period, so the table provides descriptive evidence on observable placement patterns rather than a test of instrument exogeneity. Columns (1)–(2) are linear probability models for any opening; columns (3)–(4) are OLS regressions of the number of openings. Heteroskedasticity-robust standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.14: Robustness to Alternative Definitions of University Expansion

	(1) Baseline	(2) Pure New	(3) Other Source
Panel A: 2SLS Estimates			
Spouse EA PGI	0.672**	1.419**	0.348
	(0.260)	(0.552)	(0.360)
Spouse Cog PGI	1.021***	1.148*	0.972**
	(0.297)	(0.587)	(0.398)
Spouse Math PGI	0.518**	0.785	0.422
	(0.255)	(0.577)	(0.365)
Spouse University Degree	0.861***	0.972***	0.782***
	(0.107)	(0.260)	(0.159)
Spouse SLA	4.428***	4.538***	4.364***
	(0.616)	(1.035)	(0.881)
Spouse Log Wage	0.351***	0.283*	0.346***
	(0.087)	(0.172)	(0.120)
Panel B: First Stage			
$\Delta Dist$	0.031***	0.049***	0.028***
	(0.005)	(0.012)	(0.005)
First-stage F-statistic	42.10	15.69	31.61
Observations	12,494	12,494	12,494

Notes: The sample is the kinship-consistent complete-couple sample used in the main analysis. Column (1) reports the baseline specification, in which university access is constructed using all institutions entering the university sector. Column (2) constructs the reduction in distance using only the campuses of the eight purpose-built post-war universities (Sussex, East Anglia, York, Lancaster, Essex, Kent, Warwick, and Stirling). Column (3) instead uses other sources of university expansion, including university-status changes, campus openings at established universities, and constituent-college events, but excludes the eight colleges of advanced technology (CATs) converted to universities in the 1960s (Aston, Bath, Bradford, Brunel, City, Loughborough, Salford, and Surrey); see Appendix F. In each column, the corresponding reduction in distance, measured per 10 km, is used as the excluded instrument for own university attainment. All specifications include the baseline controls, birth-region-by-cohort fixed effects, the first 20 own and spouse genetic principal components, and UK Biobank inverse-probability weights. Cluster-robust standard errors are two-way clustered by spouse pair and birth district and are reported in parentheses. The reported first-stage statistic is the cluster-robust Wald F-statistic for the excluded instrument. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.15: Access-Sensitive Profiles Across Distance-Reduction Cutoffs

	Full sample	≥ 5 km	≥ 10 km	≥ 20 km
Panel A: Instrument and outcome estimates				
Continuous-IV spouse-university estimate	0.861			
	(0.107)			
Binary-IV spouse-university estimate		0.870	0.881	0.921
		(0.143)	(0.133)	(0.128)
Binary first stage		0.093	0.111	0.126
		(0.020)	(0.020)	(0.025)
Binary first-stage F-statistic		22.40	29.66	24.46
Panel B: Predetermined characteristics				
EA PGI (SD)	−0.017	0.235	0.136	0.241
		(0.214)	(0.187)	(0.197)
Middle EA-PGI tertile	0.336	0.373	0.452	0.441
		(0.105)	(0.097)	(0.107)
Predicted university probability	0.324	0.416	0.400	0.431
		(0.034)	(0.032)	(0.032)
High-SES birth district	0.491	0.609	0.614	0.791
		(0.339)	(0.332)	(0.253)
Number of siblings	1.94	1.76	1.70	1.62
		(0.25)	(0.22)	(0.23)
Pre-birth universities within 10 km	2.842	0.754	0.624	0.860
		(0.991)	(0.894)	(0.954)
Birth year	1957.1	1954.8	1954.9	1953.9
		(1.4)	(1.4)	(1.3)
Share above cutoff		0.120	0.102	0.069
Observations	12,494	12,494	12,494	12,494

Notes: Each cutoff column uses a cumulative binary instrument equal to one when the reduction in distance to the nearest university between birth and age 15 is at least the stated cutoff and zero otherwise. The cutoff groups overlap, and each specification retains the benchmark sample. All models use the controls, cohort-by-region fixed effects, UK Biobank inverse-probability weights, and two-way clustering by spouse pair and birth district from the preferred specification. For a predetermined characteristic X, the cutoff-specific profile is obtained as the coefficient on instrumented university attainment in a 2SLS regression of $X \times$University on university attainment, using the corresponding binary access measure as the excluded instrument. These coefficients are intended as regression-adjusted, suggestive characterisations of the populations associated with each discrete access contrast rather than as unconditional complier means. In particular, they do not characterise the complier-weighted population underlying the preferred continuous instrument. Cluster-robust standard errors are reported in parentheses.

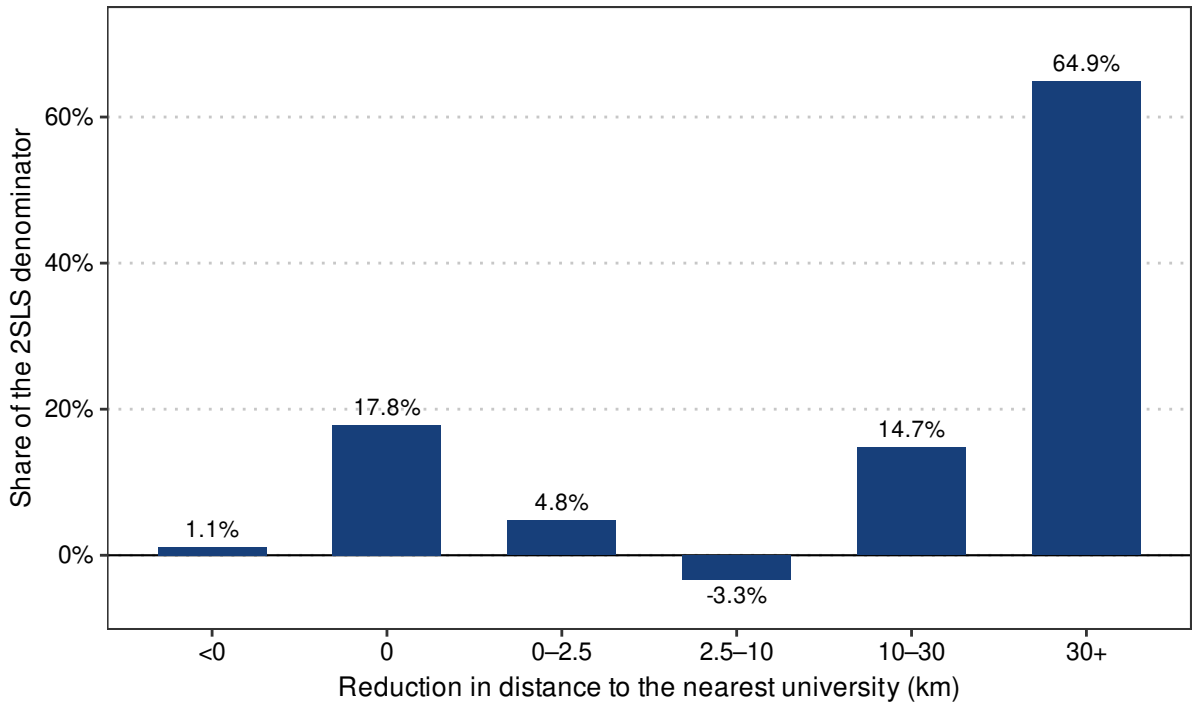


Figure A.2: Contributions to the 2SLS Denominator by Distance Reduction. The figure decomposes the residualised instrument–university-attainment cross-product in the benchmark sample of 12,494 individuals. For each distance bin b, the contribution is $\sum_{i \in b} w_i \widetilde{Z}_i \widetilde{C}_i / \sum_i w_i \widetilde{Z}_i \widetilde{C}_i$, where w_i denotes the UK Biobank inverse-probability weight and the tildes denote residuals from weighted regressions on the controls and fixed effects in the baseline specification. Contributions sum to one and may be negative; they are not shares of compliers or shares of the effect on partner outcomes. Zero denotes no change in distance. Positive bins include their lower endpoint and exclude their upper endpoint, except that the first positive bin excludes zero.

Table A.16: Robustness of 2SLS Estimates to Different Fixed-Effects Specifications

	Spousal Outcomes		
	(1)	(2)	(3)
Panel A: Polygenic Indices	EA PGI	Cog PGI	Math PGI
Coefficient on own university degree			
Region × cohort FE	0.672**	1.021***	0.518**
	(0.260)	(0.297)	(0.255)
Additive FE	0.590**	0.996***	0.462*
	(0.254)	(0.297)	(0.245)
County × cohort FE	0.865*	0.883*	-0.282
	(0.524)	(0.521)	(0.610)
	(1)	(2)	(3)
Panel B: Education and Wages	University	SLA	Log Wage
Coefficient on own university degree			
Region × cohort FE	0.861***	4.428***	0.351***
	(0.107)	(0.616)	(0.087)
Additive FE	0.836***	4.194***	0.361***
	(0.124)	(0.687)	(0.092)
County × cohort FE	0.793***	4.789***	0.397**
	(0.239)	(1.374)	(0.190)
Observations		12,494	
First-stage F-statistic			
Region × cohort FE		42.10	
Additive FE		38.27	
County × cohort FE		7.66	

Notes: The sample consists of kinship-consistent complete male–female pairs used in the main 2SLS analysis. Each coefficient is the 2SLS estimate of the effect of the focal individual's university attainment on the listed spousal outcome. Reduction in distance to the nearest university, measured per 10 km, is used as the excluded instrument. All specifications include the baseline controls, the first 20 own and spouse genetic principal components, and UK Biobank inverse-probability weights. The rows distinguish the fixed-effects specification. Region × cohort FE denotes birth-region-by-cohort fixed effects; Additive FE includes additive birth-region and cohort fixed effects; and County × cohort FE denotes birth-county-by-cohort fixed effects. Cluster-robust standard errors, two-way clustered by spouse pair and numerical birth-district ID, are reported in parentheses. Inference uses birth-district cluster degrees of freedom. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. All specifications use 12,494 observations. The reported first-stage F-statistics are calculated using the corresponding fixed-effects specification. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.17: Robustness: Multi-Trait Highest Mathematics PGI

	(1) Single-Trait (Main)	(2) Multi-Trait (Overlapping)
Panel A: 2SLS Estimates		
Pooled Sample	0.518**	0.725***
	(0.255)	(0.262)
Males	0.307	0.494
	(0.310)	(0.316)
Females	0.966*	1.228**
	(0.571)	(0.564)
Panel B: Reduced-Form Estimates		
Pooled Sample	0.016*	0.023**
	(0.009)	(0.009)
Males	0.011	0.017
	(0.011)	(0.012)
Females	0.025*	0.032**
	(0.015)	(0.015)
Observations (pooled / by sex)	12,494 / 6,247	

Notes: The table compares estimates for the spouse's Highest Mathematics PGI using the single-trait index from the main analysis, whose discovery GWAS (23andMe) contains no UK Biobank data, with the PGI Repository's multi-trait index, whose construction draws on UK Biobank GWAS of related traits and therefore partly overlaps our analysis sample (see Appendix B.1). Panel A reports 2SLS estimates of the effect of own university attainment, instrumented with $\Delta Dist$, measured per 10 km. Panel B reports the corresponding reduced-form estimates. All specifications include the baseline controls, birth-region-by-cohort fixed effects, the first 20 own and 20 spouse genetic principal components, and UK Biobank inverse-probability weights. In the whole sample, standard errors are two-way clustered by spouse pair and birth district; in the sex-specific samples, they are clustered by birth district. Both PGIs are standardised in the analysis sample. Cluster-robust standard errors are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.18: Spousal-Outcome Estimates under Alternative Geographic Sample Restrictions

	Spousal Outcomes		
	(1)	(2)	(3)
Panel A: Polygenic Indices	EA PGI	Cog PGI	Math PGI
Coefficient on own university degree			
Different birth districts	0.708**	0.972***	0.529*
	(0.296)	(0.340)	(0.314)
Birth locations at least 5 km apart	0.768***	0.898***	0.436
	(0.294)	(0.299)	(0.297)
Birth locations at least 10 km apart	0.851**	1.143***	0.825**
	(0.367)	(0.419)	(0.382)
Birth locations at least 20 km apart	0.831	1.640*	1.352*
	(0.694)	(0.850)	(0.760)
Birth locations at least 30 km apart	0.950	2.051*	1.687*
	(0.841)	(1.129)	(0.993)
Both spouses born outside Greater London	0.651*	1.081***	0.244
	(0.334)	(0.412)	(0.338)

	Spousal Outcomes		
	(1)	(2)	(3)
Panel B: Education and Wages	University	SLA	Log Wage
Coefficient on own university degree			
Different birth districts	0.870***	3.913***	0.362***
	(0.129)	(0.746)	(0.116)
Birth locations at least 5 km apart	0.848***	3.981***	0.370***
	(0.118)	(0.672)	(0.105)
Birth locations at least 10 km apart	0.759***	2.876***	0.232
	(0.180)	(0.851)	(0.151)
Birth locations at least 20 km apart	0.546*	1.293	0.135
	(0.315)	(1.448)	(0.275)
Birth locations at least 30 km apart	0.527	0.330	0.031
	(0.388)	(1.672)	(0.319)
Both spouses born outside Greater London	0.824***	4.492***	0.332***
	(0.144)	(0.779)	(0.107)

Panel C: Sample and First Stage	Observations	First-stage F-statistic
Different birth districts	8,010	40.51
Birth locations at least 5 km apart	7,834	37.67
Birth locations at least 10 km apart	5,702	26.76
Birth locations at least 20 km apart	4,402	9.32
Birth locations at least 30 km apart	4,044	6.79
Both spouses born outside Greater London	9,984	24.97

Notes: The sample consists of kinship-consistent complete male–female pairs from the benchmark 2SLS sample. Each coefficient is the 2SLS estimate of the effect of the focal individual's university attainment on the listed spousal outcome. Reduction in distance to the nearest university, measured per 10 km, is used as the excluded instrument. The first row restricts the sample to couples born in different districts. The next four rows restrict the sample to couples whose birth locations are at least 5, 10, 20, and 30 km apart, respectively, based on straight-line distance between the spouses' birth coordinates. The final row excludes couples for whom either spouse was born in Greater London. Because university access predicts how far apart spouses were born (Table A.12), restrictions on birthplace separation may select on a characteristic affected by the instrument; see Section 6.4. All specifications include the baseline controls, birth-region-by-cohort fixed effects, the first 20 own and spouse genetic principal components, and UK Biobank inverse-probability weights. Cluster-robust standard errors, two-way clustered by spouse pair and numerical birth-district ID, are reported in parentheses; inference uses birth-district cluster degrees of freedom. EA PGI, Cog PGI, and Math PGI denote the Educational Attainment, Cognitive Performance, and Highest Mathematics polygenic indices. Panel C reports the estimation-sample size and first-stage F-statistic for each sample restriction. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

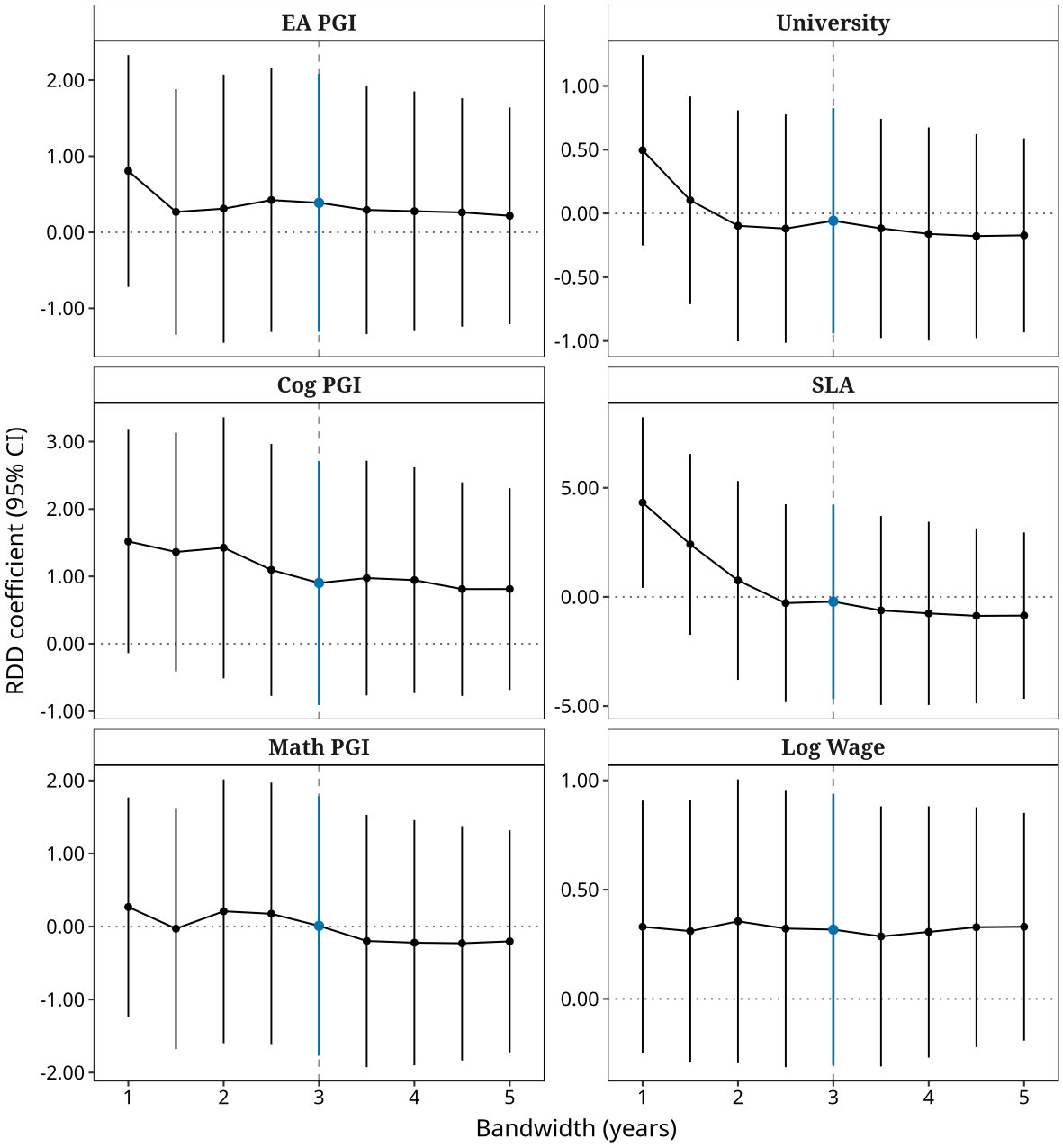


Figure A.3: ROSLA Estimates Across Alternative Bandwidths. The figure reports whole-sample fuzzy RDD estimates across alternative symmetric bandwidths. All specifications use local-linear trends, a triangular kernel, UK Biobank inverse-probability weights, and standard errors clustered by spouse pair. Vertical lines denote 95% confidence intervals. The blue estimate and dashed vertical line indicate the benchmark bandwidth of three years. The treatment is the focal individual remaining in education to age 16 or later, and the cutoff is September 1957; the benchmark estimates are in Table 5.

E Additional Results

Table A.19: Within-Couple Similarity: Reduced-Form Estimates

	Couple Similarity Outcomes					
	(1)	(2)	(3)	(4)	(5)	(6)
	EA PGI Sim. (SD)	Cog PGI Sim. (SD)	Math PGI Sim. (SD)	SLA Sim. (SD)	Wage Sim. (SD)	Kinship Sim. (SD)
Panel A: Whole Sample						
Δ*Dist* (Male)	0.001	-0.006	-0.045*	-0.017	-0.022	-0.004
	(0.017)	(0.015)	(0.025)	(0.025)	(0.017)	(0.014)
Δ*Dist* (Female)	-0.002	-0.009	0.022	0.021	-0.013	0.032**
	(0.017)	(0.019)	(0.015)	(0.019)	(0.021)	(0.016)
Num. Obs.	6,247	6,247	6,247	6,247	6,247	6,247
Panel B: Males						
Δ*Dist*	-0.011	-0.002	-0.036**	0.010	-0.022**	0.004
	(0.011)	(0.011)	(0.017)	(0.012)	(0.011)	(0.011)
Num. Obs.	6,247	6,247	6,247	6,247	6,247	6,247
Panel C: Females						
Δ*Dist*	-0.008	-0.005	0.000	0.017	-0.026*	0.032***
	(0.015)	(0.017)	(0.011)	(0.011)	(0.015)	(0.011)
Num. Obs.	6,247	6,247	6,247	6,247	6,247	6,247

Notes: Each entry reports a weighted reduced-form regression of the listed within-couple similarity outcome on Δ*Dist*, measured per 10 km. Outcomes are standardised and oriented so that higher values indicate greater similarity. Panel A includes both partners' Δ*Dist* in the same couple-level regression, the reduced form of the system in equations (9)–(11); standard errors are two-way clustered by the male and female birth districts. Panels B and C report sex-specific regressions on the focal partner's Δ*Dist*, with standard errors clustered by birth district. Panel A includes both partners' baseline controls; Panels B and C include the focal individual's baseline controls, as in the main specification. All specifications include birth-region-by-cohort fixed effects, the first 20 own and 20 spouse genetic principal components, and UK Biobank inverse-probability weights. Cluster-robust standard errors are reported in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

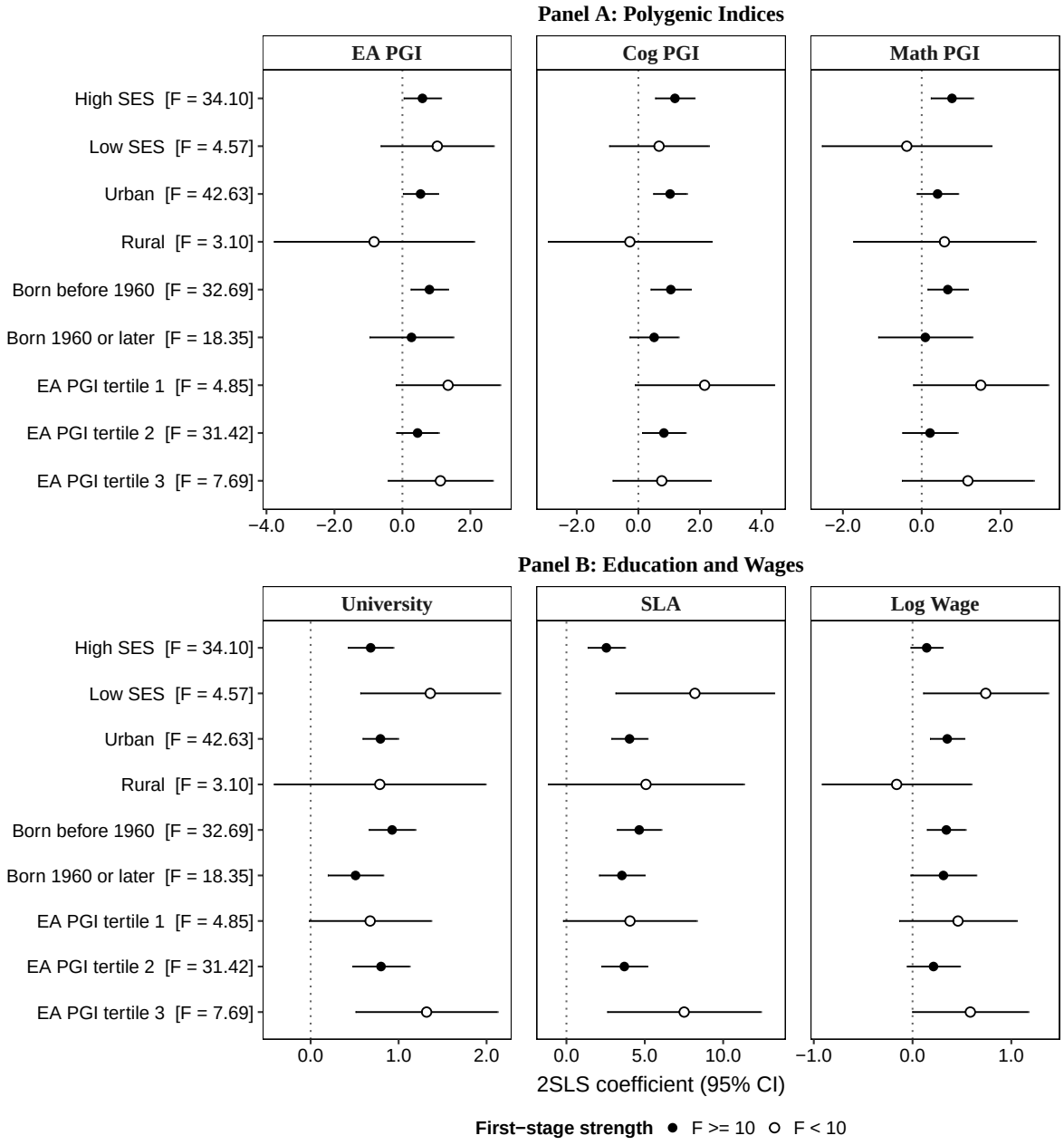


Figure A.4: Heterogeneity in the Effect of University on Spousal Outcomes. The figure reports 2SLS estimates of the effect of the focal individual's university degree on spousal outcomes, estimated separately within subgroups defined by birth-district SES (above or below the sample median of the district share of low-SES residents), the 1951 urban status of the birth district, birth cohort, and tertiles of the focal individual's EA PGI computed within the analysis sample. Each specification instruments the university degree with $\Delta Dist$, measured per 10 km, and includes the baseline controls, birth-region-by-cohort fixed effects, the first 20 own and 20 spouse genetic principal components, and UK Biobank inverse-probability weights. Horizontal lines are 95 percent confidence intervals based on cluster-robust standard errors, two-way clustered by spouse pair and birth district, with birth-district cluster degrees of freedom. Labels report the subgroup first-stage F-statistic; hollow markers indicate $F < 10$.

F Higher Education Institutions (HEI) Dataset

We construct a novel Higher Education Institution (HEI) dataset that records the timing, location, and characteristics of colleges and universities in the United Kingdom at both the institution and campus level. For each entry, we record the institution name, campus name, whether the campus is the main campus, latitude and longitude, institution type, and source information. We also collect detailed temporal information, including the institution's founding year, the year in which teaching activity began at a given campus where relevant, the year in which university status was granted, and the year of closure where applicable. In addition, the dataset contains information on student numbers in 2020, public versus private status, and curriculum. Curriculum is recorded both as free text and through a categorical classification that groups campuses into broad types such as generalist institutions, core-subject institutions, specialised institutions, theological institutions, hospitals, and campuses with no undergraduate teaching. The dataset and accompanying documentation will be made publicly available upon publication.

The dataset was assembled by combining official and institutional sources. Information on legal recognition, degree-awarding powers, and university status was checked against official UK government guidance on recognised degree-awarding institutions (GOV·UK), which directs users to the relevant official registers and lists for each UK nation.[28] Historical information on founding dates, predecessor institutions, campus openings, and institutional development was collected primarily from institutions' official history pages and related archival material hosted on university websites. Campus addresses were taken from official institution websites and, where needed, converted into geographic coordinates using mapping tools for geocoding. Since many institutions operate multiple campuses, we identify the main campus whenever possible, while retaining additional campuses when they provide relevant degree programmes. This approach ensures that our measure of proximity captures meaningful access to higher education rather than administrative definitions of institutions. These choices are important for measurement, as they affect the mapping between geographic access and actual educational opportunities. We retain the original source associated with each entry in order to ensure transparency and to facilitate future extensions and updates of the dataset.

For the empirical analyses, we use these records to reconstruct the university network available at each point in time. The relevant unit is a distinct geographic university access point, identified by institution, campus coordinates, and the year in which it enters the university network. Duplicate records representing the same access point are counted only once. A campus enters the network in its coded university-status year, which need not coincide with the institution's original founding date or the date teaching began at that location. It is considered active in a given year if it has entered the university network by that year and has not subsequently closed. The Open University is excluded. This campus-based definition underlies the calculation of distance to the nearest university and the construction of the reduction-in-distance instrument used in this paper.

We also distinguish different sources of expansion within the university network. The *Pure New* category comprises campuses belonging to eight purpose-built post-war universities: Sussex, East Anglia, York, Lancaster, Essex, Kent, Warwick, and Stirling. All remaining entries are classified as *Other Source*, including institutions that acquired university status through conversion and subsequent campus expansions. In the source-specific analyses, the eight colleges of advanced technology (CATs) converted to universities in the 1960s (Aston, Bath, Bradford, Brunel, City, Loughborough, Salford, and Surrey)

[28] For England, official information on registration, degree-awarding powers, and university title is maintained by the Office for Students (OfS) in the OfS Register.

are treated as a separate category and excluded from *Other Source*. These categories distinguish the creation of new universities from other forms of entry into the university network, which may have built on pre-existing higher-education provision. Consequently, an entry into the university network does not necessarily represent the opening of an entirely new educational institution or campus.