Open Access

*OBM Genetics*

LIDSEN Publishing Inc.

Review

# Mapping Research on Genetics and Educational Outcomes: A Bibliometric Review (1970-2025)

Alfonso Pellegrino [1], Alessandro Stasi [2, *]

1. Columbia University, New York, NY, USA; E-Mail: ap4273@columbia.edu
2. Mahidol University International College, Mahidol University, Salaya, Thailand; E-Mail: alessandro.sta@mahidol.edu

* **Correspondence:** Alessandro Stasi; E-Mail: alessandro.sta@mahidol.edu

**Academic Editor:** Andre Megarbane

*OBM Genetics*
2026, volume 10, issue 3
doi:10.21926/obm.genet.2603355

**Received:** January 16, 2026
**Accepted:** August 26, 2026
**Published:** August 27, 2026

**Abstract**

This bibliometric review analyzes 387 Scopus-indexed publications (1970-2025) on the relationship between genetics and education, retrieved with the query ("Genetics" AND "Academic Performance") OR ("Heritability" AND "Education"). Research output was minimal until the 2000s, then expanded rapidly with the rise of twin cohorts and, after 2010, large genome-wide association studies (GWAS) and polygenic prediction. Science mapping reveals three thematic clusters: classical heritability studies of cognitive and academic traits; cohort-based developmental work linking cognition and schooling; and sociogenomic research on GWAS, polygenic scores, and gene-environment interplay. Within this query-defined corpus, publication activity is concentrated in genetics, psychology, and related fields, while education-focused journals and policy-oriented terms are less prominent. Key gaps include the lack of diverse samples, limited applied intervention research, and the absence of clear ethical and governance frameworks for potential educational uses of genetic information.

**Keywords**

Genetics; academic achievement; polygenic scores; sociogenomics; educational equity; bibliometric analysis

## 1. Introduction

The past two decades have witnessed substantial methodological advances in research on the relationship between genetics and education, driven by developments in behavioral genetics and, more recently, sociogenomics. A large body of twin and family studies has long indicated that individual differences in educational outcomes-such as academic achievement and years of schooling-are moderately to highly heritable, with genetic factors accounting for a substantial proportion of variance alongside environmental influences [1, 2]. Landmark genome-wide association studies (GWAS) involving samples of hundreds of thousands and then millions of individuals have mapped the highly polygenic architecture of educational attainment, culminating in polygenic scores that explain over 10 percent of variance in years of education completed [3, 4]. These developments compel a reassessment of how genetics is conceptualized in educational research and, crucially, what it implies for notions of merit, fairness, and the design of schooling.

The genomic turn has moved the field beyond simple estimates of heritability toward detailed analyses of how genetic differences are embedded in social contexts. Recent work has documented that genetic influences on education operate through multiple pathways, including cognitive abilities, personality traits, motivation, and health [5], and that their expression may vary across family and institutional contexts [6]. The concept of "genetic nurture" has shown that parents' genotypes, including alleles not transmitted to the child, shape children's educational outcomes by affecting the environments that parents create [7].

These scientific advances have far-reaching normative and policy implications that are increasingly recognized but remain deeply contested. On the one hand, scholars such as Harden [8] argue that acknowledging genetic influences on educational outcomes exposes the role of moral luck in academic success and thus strengthens, rather than weakens, the case for educational justice. On the other hand, many educators and ethicists warn that introducing genetic information into educational discourse risks reifying inequalities, legitimizing tracking or selection based on presumed biological potential, and reinforcing historical patterns of exclusion if genomic tools are deployed without careful governance [9, 10]. Concerns about genetic determinism, privacy, discrimination, and the Eurocentric nature of current GWAS-whose predictive power is markedly weaker in non-European ancestry groups-have led to calls for strong safeguards and a cautious stance toward any practical use of polygenic scores in schools [4, 5].

Despite the rapidly growing empirical literature, the field remains fragmented along disciplinary lines. Much of the research linking genetics and education has been conducted by behavioral geneticists, psychologists, and genomic consortia, publishing in journals oriented toward genetics, psychiatry, and epidemiology [2, 3, 11]. While narrative reviews and theoretical contributions have begun to articulate potential implications for personalized learning, meritocracy, and equity (e.g., [5, 8, 9]), no bibliometric synthesis has yet charted the intellectual structure, temporal dynamics, and collaboration patterns of this rapidly expanding domain.

This paper provides a bibliometric review of the scholarly literature at the intersection of genetics and education, focusing on publications retrieved with the query "Genetics" AND "Academic Performance" OR "Heritability" AND "Education" in Scopus through December 2025. By combining performance analysis with science mapping, we systematically characterize the evolution of

research output over time, the most influential authors, journals, and institutions, and the main thematic clusters that structure the field [12, 13]. Particular attention is devoted to tracing how normative and policy-relevant discussions on meritocracy, equity, personalized education, and governance of genetic data are embedded within or adjacent to core empirical work.

This design maps a deliberately focused corpus and does not measure the awareness, beliefs, or motivations of scholars in mainstream education, sociology, or economics. It therefore cannot establish that these disciplines have "ignored" genetics. Answering that question directly would require a separate, journal-focused content analysis examining how genetic evidence is discussed, contested, or omitted in those fields.

The review addresses three questions. First, how have publication output and disciplinary patterns in genetics and education evolved over time, and which authors, journals, and collaboration networks are most prominent? Second, which conceptual and methodological frameworks structure the empirical literature, and how are they thematically organized? Third, how prominently do normative, ethical, and policy-oriented perspectives appear in the corpus, and what gaps and future research priorities can be identified?

## 2. Literature Review

Research on the relationship between genetics and education has evolved remarkably over the past century, moving from rudimentary twin comparisons to sophisticated genomic analyses involving millions of individuals. Early behavioral genetic studies in the mid-20th century laid the empirical foundation by demonstrating that academic performance and cognitive abilities show substantial familial resemblance. Classic twin research, building on methodologies pioneered in the 1920s and refined over decades, consistently showed that monozygotic twins resembled each other more in intelligence and scholastic achievement than dizygotic twins, suggesting the presence of heritable influences on educational outcomes [5]. Subsequent twin research produced substantial but outcome-specific heritability estimates. De Zeeuw et al. [14] reported estimates ranging from 44% to 73% across primary-school achievement domains, including 66% for overall educational achievement, whereas Silventoinen et al. [2] estimated the heritability of educational attainment at 43% across 28 twin cohorts. Krapohl et al. [1] found that intelligence accounted for more of the heritability of GCSE achievement than any other single measured domain; the other measured domains collectively accounted for approximately as much heritability as intelligence.

Education-focused critiques have argued that achievement research, including PISA-based work, gives insufficient attention to cognitive ability and behavioral-genetic evidence [15, 16]. In an analysis of NLSY data, Marks and O'Connell [17] reported that adjustment for maternal cognitive ability reduced the associations between composite socioeconomic status and children's cognitive and achievement outcomes. These contributions illustrate an active disagreement over how educational inequalities should be interpreted; they do not, by themselves, establish why particular disciplines have engaged unevenly with genetics.

However, the interpretation of these early findings was deeply contested. Debates intensified following controversial claims such as those presented by Jensen [18], who argued that genetic differences accounted for much of the variance in IQ and educational outcomes and that environmental interventions had limited effect. Subsequent scholarship emphasized that heritability is a population-specific statistic and does not, by itself, establish immutability or set a

numerical ceiling on an intervention's effect. Goldberger's [19] eyeglasses analogy illustrates this logical distinction, but it does not show that education has an intervention comparable in simplicity or effect to corrective lenses.

The genomics era, beginning in the 2000s, transformed the field by enabling the direct measurement of DNA variants associated with educational outcomes. Early candidate gene studies were largely unsuccessful due to insufficient sample sizes and an underestimation of the highly polygenic architecture of complex traits. A turning point came with the first genome-wide association study (GWAS) of educational attainment conducted by the Social Science Genetic Association Consortium. The 2013 GWAS involving approximately 126,000 individuals identified only three significant loci [20], illustrating that educational attainment is influenced by thousands of tiny genetic effects. A subsequent study evaluating nearly 300,000 participants increased the number of associated loci to 74 [11]. This trend culminated in a landmark study by Lee et al. [3], which analyzed over 1.1 million individuals and identified more than 1,200 genetic variants associated with years of schooling, enabling polygenic scores explaining roughly 11-13% of variance in educational attainment. Although modest, this predictive power approximated that of major environmental factors such as family socioeconomic status, and confirmed that education-related traits are profoundly polygenic.

Even greater predictive accuracy was achieved with a 2022 GWAS of nearly 3 million individuals [4], which identified 3,952 associated genetic variants and improved polygenic score performance to approximately 12-16% explained variance. Importantly, within-family analyses showed that although genetic prediction remains significant among siblings raised in the same household, predictive accuracy is somewhat reduced, demonstrating that part of the population-level association arises through genetic nurture-the influence of parents' genes on the environments they provide [7]. Kong et al. [7] found that a polygenic score based on parents' non-transmitted alleles predicted offspring educational attainment, with an effect estimated at approximately 30% of the transmitted score's effect. This finding indicates an indirect effect of parental genotype operating through parents or other relatives; the study did not identify all of the specific environmental mediators.

As genomic research expanded, scholars examined whether genetic variance differs across socioeconomic contexts. Turkheimer et al. [21] reported SES moderation of genetic variance in full-scale and performance IQ, but not verbal IQ, in one US sample. A later meta-analysis found a moderately sized interaction across US studies but zero or reversed effects in Western Europe and Australia [22], while a large Florida study found no such moderation for test scores [23]. The evidence therefore indicates heterogeneity rather than a general pattern. Among participants aged 30 or older, Silventoinen et al. [2] reported lower heritability of educational attainment in cohorts born in 1950-1989 than in cohorts born in 1900-1949. However, the study did not establish that expanded educational opportunity caused this difference and found no systematic regional pattern. More generally, greater equality of opportunity need not reduce relative heritability and may increase genotype-phenotype correlations by reducing environmental variation [24]. More recently, studies of national policy differences, such as variation in school tracking systems, showed that delayed tracking-keeping students in comprehensive systems longer-tends to reduce the influence of genetic differences on ultimate educational attainment, whereas early tracking amplifies them [25]. These findings suggest that institutional context may matter, but the direction and magnitude of moderation cannot be assumed to be uniform.

In parallel with empirical advances, normative debates intensified. The genomic prediction of education stimulated philosophical and ethical questions about merit, fairness, and the meaning of educational opportunity. Harden [8] argued that genetic differences represent a form of moral luck: individuals vary in genetically influenced propensities that affect their academic success independently of personal effort. On this account, genetic evidence complicates the view that educational outcomes reflect effort alone. This is a normative argument for attending to unchosen disadvantage, not a proposal to award extra marks, alter admissions decisions, or allocate educational support according to an individual polygenic score. Given the limitations of current individual-level prediction, educational support should remain based on observed educational needs rather than genomic ranking [4, 8]. At the same time, critics warn of potential misuse. Concerns include the risk of deterministic labeling, stigmatization, privacy violations, and the uneven predictive accuracy of polygenic scores across ancestral populations, which may deepen inequalities if genetic tools are introduced prematurely [4]. These concerns have led research bodies and ethicists to caution against applying genetic predictors for high-stakes educational decisions, noting that current predictive power, although meaningful for population-level research, is too imprecise for individual-level placement or tracking [10].

Despite these cautions, genetic research has begun informing discussions in educational theory, particularly regarding personalized or “precision education”. Asbury and Plomin [9] argued that acknowledging genetic diversity in learning propensities may eventually help tailor educational interventions, much as precision medicine aims to match treatments to biological profiles. For instance, polygenic indicators of reading difficulty might motivate early, targeted literacy support. Yet, even proponents emphasize that genetic information should complement-not replace-attention to psychological, social, and pedagogical factors. Although educational interventions may target traits associated with educational outcomes, the cited evidence does not establish the magnitude or durability of achievable change [5].

Taken together, twin and family studies indicate substantial genetic contributions to variation in educational achievement and attainment, although estimates differ by outcome, population, and period [2, 14]. GWAS have identified many loci statistically associated with educational attainment, but their individual effects are very small, and the biological and social mechanisms connecting these associations to student performance remain incompletely understood [3, 4]. Research also documents indirect parental genetic effects, while evidence that socioeconomic context moderates genetic influence varies across populations and outcomes [7, 22, 23]. These findings support studying genetic and environmental influences jointly and evaluating interventions on their demonstrated effects; they do not show that environmental changes can produce equivalent outcomes regardless of genetic predisposition.

## 3. Research Methodology

### *3.1 Bibliometric Approaches*

To address the objectives of this review on the interface between genetics and education, we employ a combination of performance analysis and science mapping, two complementary bibliometric strategies widely used to evaluate scientific fields [12, 13]. Performance analysis was used to describe annual publication output, document types, and the most productive and cited authors, publication sources, and countries. The reported indicators comprise publication counts,

total citations, a simple citations-per-year measure for authors, and source-level h-index and quartile values where available.

Science mapping was used to examine two relational structures reported in this review: author co-citation and keyword co-occurrence. Author co-citation identifies authors who are cited together and should not be interpreted as a map of co-authorship or collaboration. Keyword co-occurrence identifies terms that appear together in the bibliographic metadata and was used to summarize the principal thematic concentrations in the corpus.

VOSviewer was used to generate two-dimensional network layouts and identify clusters using association-strength normalization [26, 27]. The archived author co-citation map records a threshold value of 20, but the original threshold definition was not preserved.

### *3.2 Data Collection and Screening*

The bibliographic data were sourced from Scopus, a multidisciplinary abstracting and indexing database that is widely used in bibliometric studies and offers broad coverage of peer-reviewed journals and conference proceedings across psychology, genetics, epidemiology and the social sciences [28]. No a priori lower bound was imposed on the publication year in order to capture the earliest twin and family studies on educational outcomes; all records indexed up to 2 December 2025 were considered. Because the focus of this review is the intersection between genetic or genomic factors and educational or academic performance, we implemented a Boolean query in the Scopus title, abstract, and keyword fields that combined two conceptual conjunctions with a logical OR. Specifically, the final search string was:

TITLE-ABS-KEY ((“genetics” AND “academic performance”) OR (“heritability” AND “education”))

The search was deliberately designed to retrieve a focused set of publications in which genetics or heritability was explicitly connected with academic performance or education. We recognize that this strategy may omit relevant publications using only terms such as “educational attainment”, “academic achievement”, “school achievement”, “learning”, “GWAS”, “polygenic score”, “genetic nurture”, “gene-environment interaction”, or “sociogenomics”. The original query was retained because broadening it after screening would define a substantially different corpus and require all bibliometric analyses to be repeated. The results should therefore be interpreted as a map of a query-defined core literature rather than as a comprehensive census of all genetics-education research. The search was conducted on 2 December 2025 and returned 421 records.

All available bibliographic metadata were exported from Scopus in CSV format, including authors, affiliations, titles, abstracts, author keywords, index keywords, source titles, cited references, document types, languages, and citation counts. Screening followed the PRISMA 2020 framework [29]. The search returned 421 records. During the preliminary relevance assessment, 12 records focused solely on clinical or medical traits, or on general genetics without an analytical connection to educational outcomes, and were removed, leaving 409 records for title-and-abstract screening. A further 22 records were excluded because they were not in English or did not meet the eligibility criteria for document type. The remaining 387 reports were retrieved and assessed, and no additional reports were excluded at the full-text stage. The final corpus therefore comprised 339 research articles, 43 reviews, and 5 conference papers, as detailed in Figure 1.

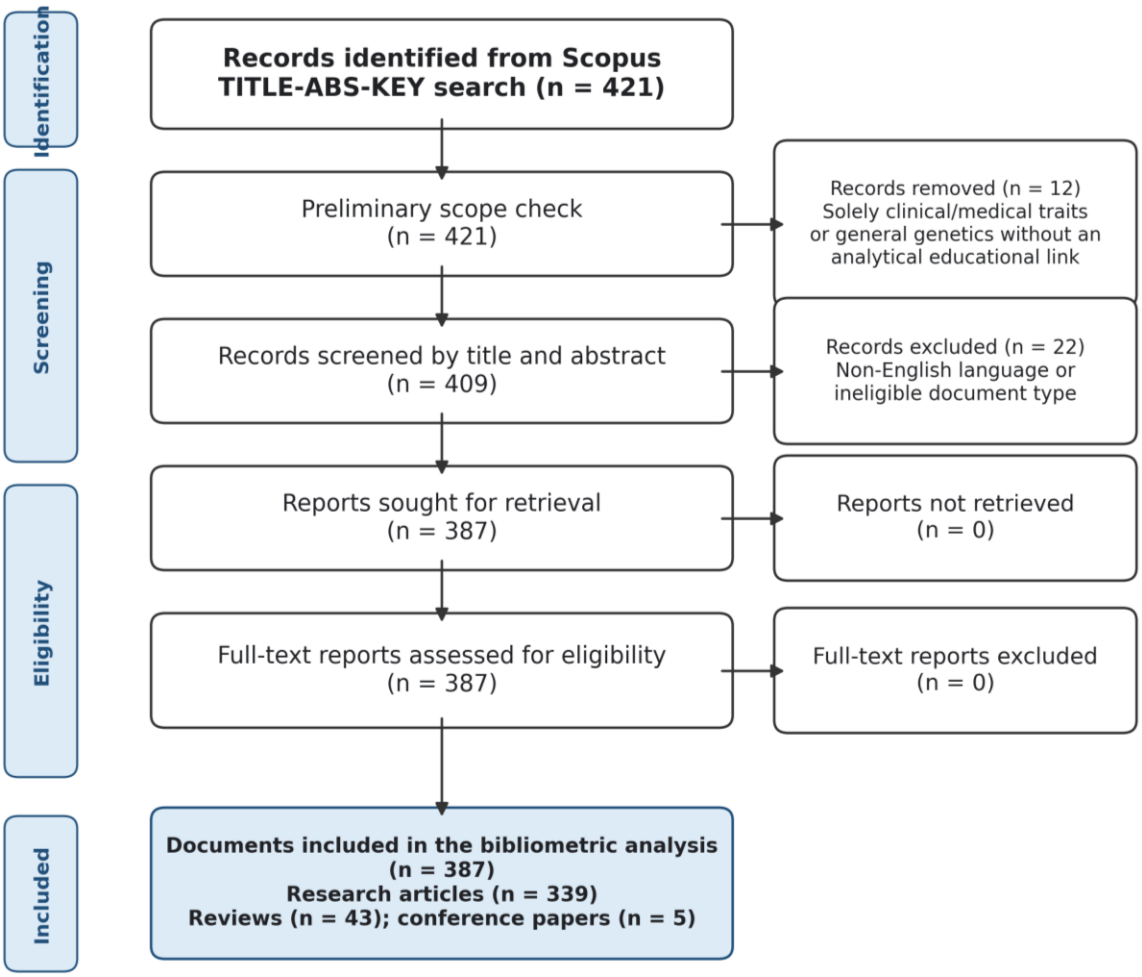


**Figure 1** PRISMA 2020 flow diagram for identification and screening of the bibliometric corpus. The Scopus search yielded 421 records; 12 were removed during preliminary relevance assessment, and 22 were excluded because of language or document-type criteria, leaving 387 documents for analysis. Source: Authors' compilation from the Scopus search and screening records.

Studies were included when they explicitly engaged genetic or genomic factors, such as heritability estimates, twin or family designs, genome-wide association studies, polygenic scores, candidate genes or related constructs, and when they featured a substantive educational or academic component. Educational components were defined as outcomes such as test scores, grades or standardized measures of academic performance; indicators of educational attainment such as years of schooling or highest qualification; learning or cognitive performance specifically framed in an educational or schooling context; or explicit discussion of implications for education, schooling or educational policy grounded in empirical results. Only peer-reviewed journal articles, reviews and conference papers written in English were retained for analysis. Records were excluded when they were editorials, letters, notes, short non-peer-reviewed commentaries or errata; when they focused exclusively on clinical or psychiatric genetics, ophthalmologic traits or other biomedical outcomes with education used solely as a covariate or background descriptor; or when education appeared only as a generic proxy for socioeconomic position without any analytic focus on schooling or academic performance. Non-English items were also removed at this stage. The application of

these criteria resulted in a core corpus of 387 documents, consisting of 339 research articles, 43 reviews and 5 conference papers, published between 1970 and 2025. Seminal pre-1970 works and non-indexed sources were consulted, where necessary, as backward citations to contextualize intellectual roots, but did not enter the quantitative analyses.

### *3.3 Data Preprocessing and Analysis*

Data preprocessing was performed prior to performance analysis and science mapping to harmonize metadata and reduce obvious duplication. Author names, institutional affiliations, and journal titles were standardized by resolving spelling variants, initials, and evident name changes. Keyword labels were converted to lowercase and normalized for punctuation, spacing, singular and plural forms, and acronym or expanded-form variants. For example, “GWAS” was grouped with “genome-wide association study”, while “PGS”, “polygenic score”, and “polygenic index” were treated as one keyword family. Conceptually distinct terms were not assumed to be exact synonyms. Slash-separated labels indicate related keyword families created for summary presentation rather than claims that the terms are interchangeable.

Following preprocessing, performance analysis was conducted using full counting. For the country analysis, a country was credited once per publication when at least one author reported an affiliation in that country; consequently, multinational publications contribute to more than one country total. Citation indicators were calculated at the document, author, and source levels. The reported science-mapping outputs comprise an author co-citation network and a keyword co-occurrence analysis.

VOSviewer generated the network layout using association-strength normalization. Version 1.6.20 of VOSviewer was used with full counting and association-strength normalization. The author co-citation threshold of at least 20 citations per cited author was set; an author-keyword co-occurrence threshold of at least five occurrences; layout attraction = 2 and repulsion = 1; clustering resolution = 1.00; minimum cluster size = 5; and merging of small clusters were the settings used. The co-citation analysis used cited authors, while the keyword analysis used normalized author keywords only; Scopus index keywords were not combined with author keywords, and a manually checked thesaurus file was used to harmonize variants.

## 4. Results

### *4.1 Volume, Growth Trajectory, and Geographic Dispersion*

The final corpus for this review consists of 387 Scopus-indexed documents in English-339 research articles, 43 reviews, and 5 conference papers-that explicitly link genetics or heritability to educational or academic outcomes, as identified by the query (“Genetics” AND “Academic Performance”) OR (“Heritability” AND “Education”) [28]. Figure 2 plots the annual publication counts reported below. The earliest article in the dataset dates to 1970, but output remained extremely sparse for several decades: between 1970 and 1999 only 24 eligible documents were published, typically one or two per year, with many years showing no publications at all. Scholarly production began to increase in the early 2000s, with 65 documents appearing between 2000 and 2009. Even in this period, annual output rarely exceeded ten papers, reflecting a still-nascent research area dominated by twin and family designs in behavioral genetics [1, 2].

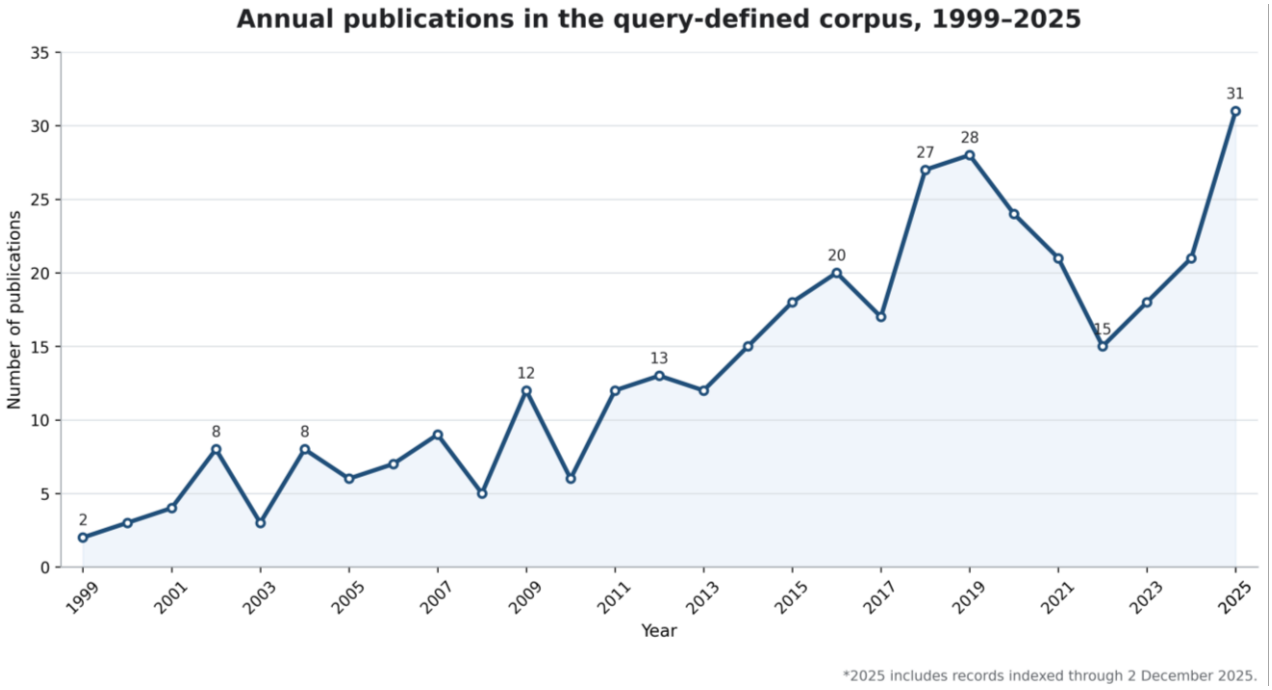


**Figure 2** Annual number of Scopus-indexed publications in the query-defined corpus, 1999-2025 (n = 365 displayed). Twenty-two additional documents were published before 1999. The 2025 count includes records retrieved through 2 December 2025. Source: Scopus data retrieved on 2 December 2025.

A marked acceleration occurs from 2010 onwards, coinciding with the rise of large-scale consortia and genome-wide association studies of educational attainment [3, 4, 11]. Between 2010 and 2019, the field produced 168 documents-over 43 per cent of the entire corpus-with annual counts rising from 6 papers in 2010 to peaks of 27 and 28 in 2018 and 2019, respectively. This period encompasses the first GWAS identifying hundreds of loci for education and the initial wave of polygenic prediction studies [3, 5]. The pattern continues into the 2020s: from 2020 to 2025 the dataset contains 130 documents, with yearly output fluctuating between 15 and 31 publications. The partial count for 2025 already stands at 31 papers, the highest single-year figure in the series, suggesting that activity in genetics-education research is still expanding rather than plateauing. If one considers the interval 2001-2024, the annual number of publications grows from 4 to 21, corresponding to an approximate compound annual growth rate of 7-8 per cent, consistent with the broader expansion of sociogenomic research over this period [5, 12]. Overall, roughly three-quarters of all documents in the corpus have appeared since 2010, underscoring how strongly the field is shaped by recent genomic developments.

Geographically, the literature is concentrated in a relatively small set of high-income, research-intensive countries (Figure 3), a pattern typical of bibliometric profiles in the social and life sciences [12, 28].

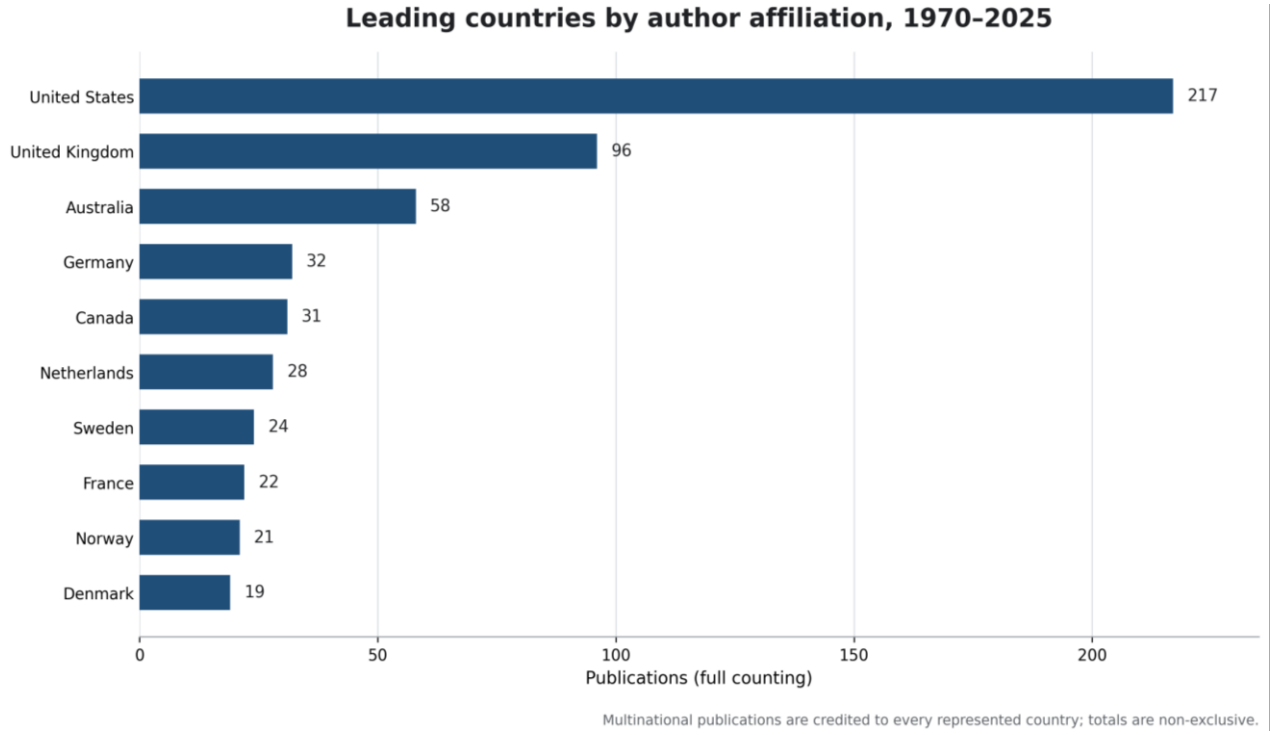


**Figure 3** Ten leading countries by author-affiliation count in the query-defined corpus, 1970-2025. Non-exclusive full counting was used: a multinational publication was credited to every country represented in its author affiliations, so totals exceed the 387-document corpus. These counts describe author affiliations, not study populations or participant ancestry. Source: Scopus data retrieved on 2 December 2025.

Using full counting at the country level (each country credited once per paper if at least one author lists an affiliation there), the United States is by far the most prolific contributor, appearing on 217 documents. The United Kingdom follows with 96 papers, then Australia with 58, Germany with 32, Canada with 31, and the Netherlands with 28. The next tier consists of Sweden (24), France (22), Norway (21), and Denmark (19). Together, these ten countries account for the majority of publications in the corpus, reflecting the strong presence of large twin registries, population cohorts, and genomic consortia in North America, Western Europe, and Australia [2, 7]. Contributions from other regions-including East Asia, Latin America, Africa, and South Asia-are comparatively sparse, despite the fact that educational inequality and debates about meritocracy are highly salient in those contexts. As in other areas of sociogenomics, the geography of research output therefore mirrors structural disparities in research infrastructure and access to large-scale genomic resources rather than the distribution of substantive educational need [4, 8].

### *4.2 Influential Authors*

Influence in the genetics-education literature, measured through a combination of publication output and normalised citation impact, is highly concentrated in a relatively small set of behaviour-geneticists and sociogenomic researchers. Table 1 reports the ten most prolific and cited authors in the corpus, based on full counting of author names across the 387 Scopus-indexed documents and aggregation of their total citations. Nicholas G. Martin leads the field with 21 documents and 2,412 citations, reflecting his long-standing role in establishing large twin registries and multivariate

genetic models that underpin much of the contemporary evidence on the heritability of educational outcomes. His work at QIMR Berghofer in Australia has provided crucial methodological and cohort infrastructure for later studies on cognition, achievement and schooling trajectories (e.g., [2]).

**Table 1** Leading authors in the genetics-education corpus (n = 387).

| Rank | Author | Country (primary affiliation) | Documents | Total citations | Citations $yr^{-1*}$ | Core contribution(s) |
|---|---|---|---|---|---|---|
| 1 | Martin, N.G. | Australia (QIMR Berghofer) | 21 | 2,412 | 54.8 | Long-running twin cohorts; multivariate models of cognitive and educational outcomes |
| 2 | Plomin, R. | United Kingdom (King's College London) | 20 | 2,367 | 87.7 | Behavioural genetics of achievement; twin and adoption studies; DNA-based prediction (GCTA, polygenic scores) |
| 3 | Dale, P.S. | United States (University of New Mexico) | 14 | 1,172 | 78.1 | Genetic and environmental influences on language development and early academic skills |
| 4 | Kaprio, J. | Finland (University of Helsinki) | 11 | 1,630 | 70.9 | Nordic twin registries; historical change and gene-environment interplay in educational attainment |
| 5 | Rimfeld, K. | United Kingdom (King's College London) | 10 | 849 | 65.3 | Heritability of school achievement; role of non-cognitive traits using large twin cohorts |
| 6 | Heath, A.C. | United States (Washington University in St. Louis) | 9 | 825 | 20.1 | Twin-family designs; multivariate models linking education, substance use and psychiatric traits |
| 7 | Boomsma, D. | Netherlands (Vrije Universiteit Amsterdam) | 8 | 742 | 30.9 | Dutch twin-family studies; structural models of cognitive and educational outcomes |
| 8 | Medland, S.E. | Australia (QIMR Berghofer) | 8 | 543 | 49.4 | Large-scale GWAS and meta-analyses; methodological contributions to polygenic prediction of education |
| 9 | Kremen, W.S. | United States (UC San Diego) | 8 | 259 | 16.2 | Vietnam Era Twin Study of Aging; links between education, cognitive aging and brain structure |
| 10 | Franz, C.E. | United States (UC San Diego) | 8 | 259 | 16.2 | Life-course analyses of education and cognition in twin cohorts |

*Citations $yr^{-1}$ = total citations ÷ (2025 − first genetics-education publication year for that author + 1). Note. Document and citation counts are from Scopus and were retrieved on 2 December 2025. Citations per year are a simple time-adjusted descriptive rate and are not field-normalized.

Robert Plomin follows closely with 20 documents and 2,367 citations, but he has the highest normalised citation rate in the top group, averaging 87.7 citations per year since his first genetics-education paper in this corpus. Plomin's contributions span from classical twin and adoption studies to the early application of DNA-based methods, including genome-wide polygenic scores for educational achievement and the articulation of the "new genetics" of intelligence and learning [1, 5]. Philip S. Dale, with 14 documents and 1,172 citations, anchors the literature on genetic and environmental influences on language development and early academic skills, providing some of the most detailed longitudinal evidence linking early heritable differences to later educational performance.

Jaakko Kaprio, based at the University of Helsinki, appears in 11 documents with 1,630 citations and a citations-per-year rate exceeding 70, highlighting the importance of Nordic population registers and twin cohorts in mapping genetic and environmental variance in educational attainment across historical periods [2]. Kaili Rimfeld, with 10 documents and 849 citations, represents the newer generation of UK-based researchers using large twin samples such as TEDS to examine the heritability of school achievement and the role of non-cognitive traits in educational success [1].

Andrew C. Heath (9 documents, 825 citations) and Dorret Boomsma (8 documents, 742 citations) are central figures in the development of twin-family designs and multivariate models that have been widely applied to schooling, cognitive performance and related behavioural traits, with Heath's work bridging US and Australian cohorts and Boomsma's group at Vrije Universiteit Amsterdam providing key European infrastructure. Sarah E. Medland (8 documents, 543 citations) is a leading figure in large-scale GWAS and meta-analytic work, including early efforts to aggregate educational attainment data across biobanks and consortia, which paved the way for the later 1- to 3-million person studies [3, 4, 11]. Finally, William S. Kremen and Carol E. Franz, both at the University of California San Diego, appear with 8 documents and 259 citations each; their Vietnam Era Twin Study of Aging has been instrumental in linking educational histories to mid- and late-life cognitive outcomes and neurocognitive trajectories, thereby extending the genetics-education discourse across the life course.

One community, centred on Martin and Medland in Australia and collaborators in the United States and United Kingdom, is dominated by large twin cohorts and early sociogenomic applications. A second network, anchored by Plomin, Rimfeld and colleagues in London and the Netherlands, focuses on school achievement, non-cognitive traits and the integration of twin designs with polygenic indices [1, 5]. A third, more dispersed cluster cancers on Kaprio, Boomsma and Nordic-Dutch collaborators who leverage register-based twin studies to examine historical change and gene-environment interplay in educational attainment [2]. Betweenness-centrality scores highlight figures such as Kaprio and Medland as structural brokers connecting GWAS consortia to traditional behaviour-genetic cohorts, thereby facilitating the translation of genomic discovery into education-relevant analyses [3, 4].

### *4.3 Intellectual Structure Revealed by Co-Citation*

The author co-citation analysis reveals a tripartite intellectual structure in the genetics-education literature (Figure 4). The network is organised around three densely connected clusters that

correspond to successive methodological and conceptual waves in the field, from classical twin studies through cognitive-educational psychology to contemporary sociogenomics.

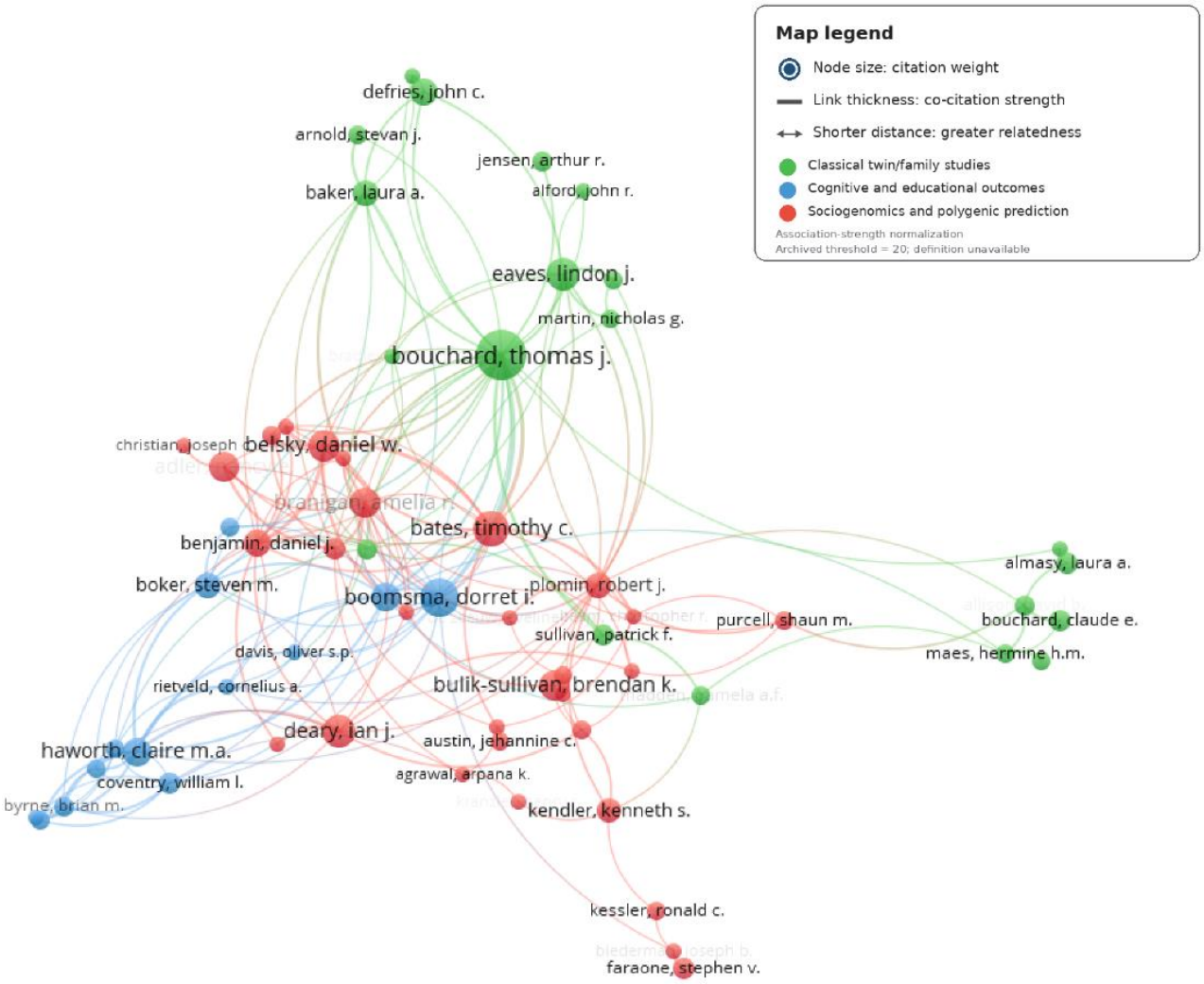


**Figure 4** Author co-citation network of the query-defined genetics-education corpus. Node size represents citation weight; links join authors cited together, with thicker links indicating stronger co-citation; shorter distances indicate greater relatedness. Colours denote the three interpreted clusters: green, classical behaviour genetics; blue, cognitive and educational outcomes in twin and cohort studies; and red, sociogenomics and polygenic prediction. Association-strength normalization was used. Source: Authors' analysis of Scopus data retrieved on 2 December 2025 using VOSviewer.

The first cluster (green), dominated by names such as Thomas J. Bouchard Jr., Lindon J. Eaves, John C. DeFries, Hermine H. Maes, Nicholas G. Martin and Claude Bouchard, represents the classical behaviour-genetic tradition. Seminal contributions here include the Minnesota Study of Twins Reared Apart, which reported that around 70% of the variance in IQ among separated monozygotic twins could be attributed to genetic variation [30] and the methodological corpus on structural models for twin and family data developed by Eaves and colleagues. This cluster also encompasses early quantitative-genetic analyses of reading, maths and school performance in twin samples led by DeFries and Martin, which provided some of the first robust estimates of the heritability of

specific academic skills. Collectively, these works established the canonical toolkit of twin, adoption and extended-kinship designs and articulated the core empirical claim that individual differences in cognitive ability and educational outcomes are substantially heritable [30, 31]. The dense co-citation ties within this cluster reflect the continued reliance of later studies on these foundational designs and estimates, which motivate new research questions or benchmark genomic findings.

The second cluster (blue) is organised around Dorret I. Boomsma, Ian J. Deary, Cornelius A. Rietveld, Claire M.A. Haworth, Brian M. Byrne, Steven M. Boker and collaborators, and can be characterised as cognitive and educational outcomes in twin and cohort studies. Boomsma's influential reviews on twin resources and multivariate genetic modelling highlight the creation of large twin registries and their application to behavioural and educational traits [31]. Deary's longitudinal work linking psychometric intelligence at age 11 to performance in national examinations at age 16 in more than 70,000 Scottish children [32] is a central node, repeatedly co-cited when authors discuss the tight coupling between intelligence and educational achievement. Within the same cluster, Haworth, Byrne and colleagues contribute genetically informed studies of literacy, numeracy and cognitive development across childhood and adolescence, often using cross-national twin datasets. Rietveld and co-authors mark the bridge from behaviour genetics to genomics: their 2013 GWAS of 126,559 individuals was the first to identify specific genetic variants associated with educational attainment [20], and is heavily co-cited with both twin-based heritability papers and later sociogenomic work. Conceptually, this cluster consolidates the view that intelligence and education are tightly correlated but not identical, and that the genetic architecture of educational achievement likely overlaps with, yet extends beyond, that of general cognitive ability [1, 33]. The co-citation pattern suggests that this group of authors provides an integrative bridge between classical twin methods and the first generation of DNA-based studies of schooling.

The third cluster (red) brings together Robert Plomin, Brendan K. Bulik-Sullivan, Patrick F. Sullivan, Daniel J. Benjamin, Daniel W. Belsky, Kenneth S. Kendler, Stephen V. Faraone and collaborators, and can be labelled sociogenomics and polygenic prediction. Plomin's programmatic papers on the "new genetics of intelligence" and the use of DNA in education [5] are co-cited with large consortial GWAS and methodological work on linkage disequilibrium score regression and polygenic scores led by Bulik-Sullivan and colleagues [34]. Benjamin and co-authors link Rietveld's initial education GWAS to broader questions in social-science genetics, including the design of large-scale consortia and the interpretation of small-effect loci in complex social traits [20, 35]. Belsky's work on polygenic scores and social mobility, including analyses of how education-linked scores predict intergenerational class movement [36], introduces explicitly normative and policy-oriented themes into the cluster and is frequently co-cited with Harden's [8] The Genetic Lottery in more recent publications. Kendler, Faraone and Sullivan contribute the psychiatric-genetics backbone-multi-trait GWAS, cross-disorder correlations and methods for dissecting shared genetic architectures-that sociogenomic studies draw on when relating education to mental health, personality and other outcomes. Overall, the co-citation ties within this red cluster reflect a shift from estimating aggregate heritability to using genome-wide data to model fine-grained genetic overlap across traits and to construct polygenic indices with predictive and interpretive ambitions [3, 4, 11].

Taken together, the three clusters trace a coherent intellectual trajectory. The green cluster codifies the classical twin-study paradigm and its claim that educational outcomes are substantially heritable. The blue cluster refines this insight by embedding intelligence and achievement in large

longitudinal cohorts, integrating psychometrics with genetically sensitive designs and introducing the first genome-wide analyses of schooling. The red cluster extends the enterprise into full sociogenomics, developing statistical tools and consortia to perform very large GWAS, derive polygenic scores and connect genetic findings to questions of social mobility, inequality and policy. The sparse but visible bridges between clusters-particularly authors like Boomsma, Deary, Rietveld and Plomin, who co-locate in more than one community-suggest that the field has evolved cumulatively rather than through sharp breaks, with each generation of methods re-using, re-interpreting and sometimes contesting the assumptions of its predecessors.

### *4.4 Journals and Venues Shaping the Discourse*

Publication outlets in the genetics-education corpus display a markedly skewed distribution, with a small number of specialised journals and high-impact general science venues concentrating a large share of both output and citations, a pattern consistent with other fields in the social and life sciences [12, 28]. As shown in Table 2, the top ten sources collectively publish just over half of all documents in the dataset. The three leaders-Behavior Genetics, Twin Research and Human Genetics, and Nature Genetics-alone account for approximately 28 per cent of the total output and attract a substantial portion of the field's citations. Behavior Genetics alone accounts for forty-eight papers, reflecting its role as the flagship outlet for classical twin and family studies on the heritability of cognitive abilities, academic achievement and educational attainment (e.g., [1, 31]). Twin Research and Human Genetics similarly serves as the primary venue for methodological work and cohort reports from large twin registries that underpin many education-relevant analyses in Nordic, Dutch and Australian samples [2].

**Table 2** Leading publication venues in the genetics-education corpus (1970-2025, n = 387).

| Rank | Source title | Docs in corpus | Total citations | Journal h-index | Best SJR quartile (2024) | Publisher | Predominant contribution |
|---|---|---|---|---|---|---|---|
| 1 | Behavior Genetics | 48 | 2,450 | 106 | Q1 | Springer Nature | Core outlet for twin and family studies on cognitive and educational outcomes |
| 2 | Twin Research and Human Genetics | 36 | 1,820 | 95 | Q3 | Cambridge Univ. Press | Reports from major twin registries; methodological work on heritability of education |
| 3 | Nature Genetics | 24 | 5,300 | 660 | Q1 | Springer Nature | Landmark GWAS and polygenic-score studies of educational attainment |
| 4 | Psychological Science | 18 | 1,450 | 331 | Q1 | SAGE | Integrative studies linking cognitive psychology, genetics and educational performance |
| 5 | Proceedings of the National Academy of Sciences (PNAS) | 16 | 3,000 | 896 | Q1 | National Academy of Sciences | High-impact reports on intelligence, schooling and life-course outcomes |
| 6 | Molecular Psychiatry | 14 | 1,900 | 261 | Q1 | Springer Nature | Genetic overlaps between education, mental health and neurocognitive traits |
| 7 | Intelligence | 12 | 620 | 113 | Q1 | Elsevier | Psychometric and genetically informed studies of IQ and school achievement |
| 8 | Developmental Psychology | 11 | 710 | 256 | Q1 | APA | Longitudinal work on cognitive and academic development in genetically sensitive designs |
| 9 | Scientific Reports | 10 | 540 | 347 | Q1 | Springer Nature | Open-access sociogenomic and gene-environment interaction studies involving education |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 10 | Journal of Child Psychology and Psychiatry | 9 | 780 | 252 | Q1 | Wiley | Studies linking educational outcomes with psychopathology and neurodevelopment |

Note. Corpus document and citation counts were obtained from Scopus on 2 December 2025. Journal h-index and best SJR quartile values were checked against the 2024 SCImago Journal & Country Rank release on 31 July 2026. The best quartile is the highest quartile assigned across the journal's SCImago subject categories. These indicators are provided for contextual comparison only.

Nature Genetics occupies a distinctive position: although it hosts fewer articles by count, those articles-most notably the large genome-wide association studies of educational attainment by Rietveld et al. [20], Okbay et al. [11] and Lee et al. [3], and the subsequent three-million-person meta-analysis by Okbay et al. [4]-are extraordinarily highly cited, accounting for a substantial portion of the citation base of the entire field. This concentration underscores how a handful of landmark sociogenomic papers have redefined the empirical agenda and now serve as obligatory points of reference across clusters [3, 4, 11].

Beyond these core outlets, the discourse is shaped by a set of generalist but high-impact journals in psychology and multidisciplinary science. Psychological Science, Proceedings of the National Academy of Sciences of the USA (PNAS), and Molecular Psychiatry together host many of the articles that integrate genomic results with broader questions about cognitive development, mental health and social outcomes, bringing genetics-education findings to audiences beyond specialist behaviour-genetics circles [5, 21, 32]. More focused psychology journals such as Intelligence and Developmental Psychology publish work on the joint development of IQ and school achievement, as well as longitudinal studies of reading and numeracy that incorporate genetically informative designs [32, 37]. Open-access megajournals like Scientific Reports provide a venue for newer polygenic and gene-environment interaction studies, reflecting the broader trend toward open science and large consortial projects in sociogenomics [2]. Finally, Journal of Child Psychology and Psychiatry appears among the leading outlets for articles that situate educational outcomes within the broader context of child psychopathology and neurodevelopmental disorders, highlighting how learning difficulties and psychiatric traits share genetic architectures [38].

Conference proceedings play a comparatively minor role in this domain; unlike in some areas of computer science, the citation premium clearly lies with peer-reviewed journals, which reinforces the maturity and cumulative nature of the genetics-education discourse [12]. The steep Lorenz curve of sources suggests that any comprehensive understanding of the field's evolution must pay particular attention to the editorial priorities and readerships of a small set of journals that effectively gatekeep the flow of high-impact findings.

### *4.5 Thematic Concentration and Evolution*

Keyword co-occurrence mapping reinforces the diachronic patterns described above and clarifies how the genetics-education discourse has shifted from classical heritability studies toward sociogenomics and questions of equity. In the earliest period (1970-1999), the lexicon is dominated by a small set of behaviour-genetic terms-" twins", "heritability", "intelligence", "IQ"-that anchor work on the genetic architecture of cognitive ability and its association with school performance [30, 31]. Education appears largely as context: outcomes are typically described as "school achievement" or "academic performance", and the focus lies on estimating variance components rather than on the institutional features of schooling itself [21].

During the 2000s, as large longitudinal cohorts and national examinations became available, the vocabulary expanded. Terms such as "academic achievement", "reading", "mathematics", "literacy" and "numeracy" gain prominence, reflecting detailed twin and family studies of specific school subjects [32, 37]. "Educational attainment" is increasingly treated as a higher-level outcome, especially in work examining years of schooling or the highest qualification attained across adulthood [2]. Across this phase, "heritability" and "twins" remain central keywords, but they are

now tightly coupled with pedagogical constructs, indicating an emerging integration of behaviour genetics and educational research [1].

The lexicon changes most dramatically from 2010 onwards, coinciding with the genomic turn. By the late 2010s, “genome-wide association study”, “GWAS”, “polygenic score”, and “polygenic risk score” become core author keywords, particularly in papers reporting increasingly large meta-analyses of educational attainment and related traits [3, 4, 11]. The high total-link strengths of “educational attainment”, “GWAS” and “polygenic score” in the network indicate dense cross-referencing among consortial papers and secondary analyses that reuse the same discovery results. In this period, education itself-rather than intelligence-serves more often as the primary phenotype, with “educational attainment” emerging as the single most frequent keyword in the corpus.

In the most recent slice (2020-2025), the vocabulary broadens again, shifting toward gene-environment interplay and normative concerns. Terms such as “gene-environment interaction”, “socioeconomic status”, “genetic nurture”, “social mobility”, “inequality”, “equity” and “education policy” appear with increasing frequency, signalling a move from purely descriptive genetic mapping toward questions of how genetic differences are expressed in particular institutional settings and what this implies for fairness and opportunity [2, 7, 8]. The co-occurrence of “polygenic score” with “socioeconomic status” and “education policy” in the network reflects growing interest in how genomic predictors interact with structural inequalities and how they might (or might not) inform policy design. Across all time segments, “heritability”, “twins” and “intelligence” persist as connective terms, underscoring that classical behaviour-genetic concepts continue to scaffold even the most recent sociogenomic and governance-oriented work [5].

Table 3 summarises the ten most frequent normalised author keywords, their total link strengths and the thematic clusters to which they predominantly belong. Cluster A groups classical behaviour-genetic tags around intelligence, heritability and twin designs; Cluster B captures sociogenomic work focusing on educational attainment, GWAS and polygenic scores; Cluster C aggregates terms related to gene-environment interplay, socioeconomic gradients and policy or equity debates. For quantitative visualisation, the occurrence values can be directly graphed in Excel, while total-link strengths approximate each keyword’s centrality within the co-occurrence network.

**Table 3** Leading author keywords in the genetics-education corpus (1970-2025, n = 387).

| Rank | Author keyword (normalised) | Occurrence | Total link strength | Thematic cluster‡ | Indicative research focus |
|---|---|---|---|---|---|
| 1 | educational attainment | 132 | 487 | Sociogenomics (B) | Years of schooling; GWAS of education and life-course outcomes |
| 2 | intelligence/IQ | 118 | 462 | Behaviour genetics (A) | Cognitive ability as predictor and mediator of achievement |
| 3 | academic achievement/performance | 104 | 438 | Behaviour genetics (A) | School grades, test scores, subject-specific skills |
| 4 | heritability | 97 | 401 | Behaviour genetics (A) | Twin/family variance partitioning for cognitive and educational traits |
| 5 | twins/twin study | 89 | 372 | Behaviour genetics (A) | Classical twin designs; cohort and registry papers |
| 6 | genome-wide association study/GWAS | 76 | 345 | Sociogenomics (B) | Discovery of loci for education and related traits |
| 7 | Polygenic score/polygenic index | 71 | 329 | Sociogenomics (B) | DNA-based prediction of attainment and achievement |
| 8 | Gene-environment interaction | 52 | 268 | G × E & equity (C) | SES moderation; policy context; historical change |
| 9 | socioeconomic status | 48 | 247 | G × E & equity (C) | Social stratification; environmental moderation of genetic effects |
| 10 | education policy/equity | 29 | 183 | G × E & equity (C) | Normative and governance implications of genetic findings |

‡Cluster A = classical behaviour genetics of cognitive and educational outcomes; Cluster B = sociogenomics of attainment and polygenic prediction; Cluster C = gene-environment interplay, inequality and policy/justice debates. Note. Keyword occurrence and total-link-strength values are derived from Scopus metadata retrieved on 2 December 2025. Slash-separated labels group related terms and should not be interpreted as claims that the terms are exact synonyms.

## 5. Discussion

The bibliometric findings describe a field with strong growth in publication output but a marked concentration in behavioral genetics, psychology, and neuroscience. Research linking genetics and education has expanded in volume, but much of this work appears in psychology, behavioral genetics, and neuroscience venues rather than in mainstream education journals. Behavioral-genetic studies report substantial genetic influences on educational outcomes, although estimates differ between academic achievement and educational attainment and across study designs. The "precision education" paradigm, analogous to precision medicine, has nevertheless scarcely taken root. Instead, the topic remains siloed: the intellectual core is dominated by a few influential research groups (e.g., behavioral geneticists in the UK and the large international GWAS consortia) whose work on twin studies, genome-wide associations, and gene-environment interplay forms the backbone of the literature. For instance, key contributors like Robert Plomin and the SSGAC have elucidated how DNA differences shape achievement: twin studies show ~50-60% heritability of test scores [1, 39] and polygenic scores now predict ~10-15% of variance in educational attainment. These findings underscore that "genes matter" for educational performance. These estimates are not small and challenge explanations of educational inequality that treat differences in achievement or attainment as wholly environmental [15, 16]. The concentration of publication activity outside education-focused venues suggests disciplinary separation within the retrieved corpus. Still, it does not establish that education scholars have ignored genetics or explain why education journals are less prominent.

Geographical representation is another important limitation of the literature captured in this review. The country analysis shows a marked concentration in a small group of high-income research systems. The United States appears in 217 publications, followed by the United Kingdom with 96 and Australia with 58; Germany, Canada, the Netherlands, Sweden, France, Norway, and Denmark complete the ten leading countries. Because full counting was used, multinational publications were credited to every country represented in the authors' affiliations, and these totals therefore overlap. Nevertheless, the results show that the visible publication base is concentrated in North America, Western Europe, and Australia. At the same time, countries in Africa, Latin America, South and Southeast Asia, and the Middle East do not appear among the ten leading contributors. This imbalance may shape the research questions examined and limit the extent to which findings can be generalized across educational systems and socioeconomic contexts. Importantly, these figures describe author affiliations, not the locations or ancestry of study participants. The present bibliometric data therefore cannot establish how geographically or ancestrally diverse the underlying study samples were. Future research should report these dimensions separately and expand data collection and research partnerships in currently underrepresented regions.

The present results should therefore not be interpreted as a direct test of whether mainstream educational research has ignored genetics. Such a test would require a separate sample of education, sociology, and economics journals, followed by a content analysis of how scholars in those fields engage with behavioral-genetic evidence.

Several cultural explanations have been proposed, although the present bibliometric analysis cannot determine which, if any, explains the observed publication pattern. First is the shadow of eugenics and determinism. Both researchers and practitioners voice concern that acknowledging

genetic influences could lead to labeling students by "predetermined" ability, reinforcing inequalities or even reviving harmful genetic determinism. The idea of testing children's DNA and grouping them by aptitude evokes understandable fears of a dystopian track where those deemed low-genetic-potential receive reduced support. This concern may have contributed to limited engagement with genetics in some education circles, although its prevalence and influence were not measured in this review. A second, related interpretation is the dominance of environmental accounts, including what Pinker [40] calls the "blank slate" view, together with the "growth mindset" ethos [41]. For decades, policy and practice have emphasized that with effort and the right environment, all students can achieve, whereas attributing outcomes to DNA is often seen as fatalistic or at odds with the ethos of equity. Some educators worry that introducing genetics undermines the message that abilities are malleable. Critics of predominantly environmental accounts argue that they may understate the contributions of cognitive ability and genetic differences to educational achievement [15, 16]. These perspectives may help explain the disciplinary separation observed in the corpus, but they remain interpretive hypotheses rather than conclusions established by the bibliometric evidence. Accordingly, observations concerning publication venues, authors, citations, countries, and keywords are treated as bibliometric findings, whereas explanations involving disciplinary culture, ethics, or policy are presented as interpretive commentary.

Despite these headwinds, the field shows tentative signs of maturation toward a more nuanced view. A small but growing body of scholarship is attempting to bridge the gap by reframing genetic knowledge as part of a holistic, equity-driven toolkit rather than a tool of exclusion. Schools, however, already possess strong predictors, particularly prior achievement. The practical question is therefore whether a polygenic score provides reliable, actionable information beyond existing academic records and assessments [15]. The concept of "precision education" captures this reframing: just as precision medicine seeks to use all relevant individual data (including genomics) to optimize care, precision education would use insights about a learner-potentially including genetic propensities-to tailor instruction and interventions. Crucially, proponents stress that incorporating genetics should never mean denying opportunities to a child with a disadvantageous genetic profile; rather, it should signal where more help is required. Our analysis found that recent articles explicitly emphasize safeguards against misinterpretation: for example, Lovett et al. [42] note that knowing a child has a higher genetic risk for a reading disorder must not lead to lowered expectations, but instead prompt early interventions. Similarly, ethicists and legal scholars argue that if we acknowledge DNA as one driver of educational outcomes, then fairness demands we channel resources to counteract the luck of the "genetic lottery" [8]. Indeed, some have provocatively asked whether the state has an obligation to "compensate" for bad genetic luck-for instance, by providing extra support to students with below-average polygenic scores, much as health policy tries to support those with elevated genetic health risks. While consensus on such policies is far off, the conversation marks an evolution from the earlier silence. It moves the discourse beyond bibliometric trends and citation clusters into the realm of normative implications: what should we do with this knowledge? In short, the field stands at a crossroads. Evidence that genetic differences contribute to variation in educational outcomes is substantial, but its interpretation and educational relevance remain contested. Further research should first establish whether genetic information adds educational value beyond existing measures and, if so, under what safeguards.

### *5.1 Future Research Avenues*

A direct assessment of disciplinary engagement should sample leading education, sociology, and economics journals and examine whether twin, extended-twin, pedigree, GWAS, and polygenic-score evidence is cited, interpreted, challenged, or omitted. Such a study could test whether engagement differs by method, but authors' motivations should not be inferred without direct evidence. The evidence from this review points to several urgent directions for future research and policy, if the promise of genetic insights is to be reconciled with the values of education. First, we need large-scale, longitudinal studies and open benchmarks to truly test "precision education" interventions. To date, most empirical findings on genetics and achievement come from observational twin registries or genome-wide association studies, rather than from applied trials in classrooms. As a result, we know a great deal about what genes predict in general (e.g., years of schooling, test scores), but very little about how to use that information to improve individual student outcomes. Before any classroom application is considered, research should test whether polygenic information adds meaningful predictive or intervention value beyond prior achievement and other routinely available measures. Given current limitations in individual prediction and cross-ancestry portability, classroom disclosure of students' polygenic scores or their use for placement would be premature. Crucially, these studies must report outcomes in terms that matter to educators and families (skill gains, reduced dropout rates, etc.), not just statistical effect sizes. As with medical trials, transparency and reproducibility will be key: shared datasets (with privacy protections) and cross-site collaborations can ensure that results are not simply one-off successes on an "ideal" sample. In short, a future research priority is to move from laboratory correlations to real-world experiments, establishing whether incorporating genetic information meaningfully enhances educational practice-and under what conditions it does so.

Research involving identifiable genetic data in education should be governed by safeguards proportionate to the data's sensitivity, participants' ages, and the intended use. These safeguards, including consent or assent, privacy protection, data security, and attention to ancestry-related validity, should enable responsible research rather than operate as a blanket prohibition [43]. The problem of supporting low-achieving students while maintaining opportunity predates the use of polygenic indices. Genetic information does not resolve that dilemma, although it introduces distinct concerns concerning familial data, privacy, cross-ancestry validity, and deterministic labeling. For example, studies should explore how to present genetic risk information to teachers and parents in a way that avoids stigma or self-fulfilling prophecies, drawing on communication research in personalized medicine. Legal research, too, has a role: scholars could draft model policies or even legislation that prohibit misuse of genetic data (such as tracking students into rigid ability groups) while permitting its use to flag needs for additional support. Importantly, explainability and transparency must be central: if an algorithm or assessment incorporates genetic factors, educators should be able to understand and justify any resulting decisions. This echoes calls in other domains (e.g., criminal justice risk assessments) that "black-box" tools are unacceptable when individual rights are at stake. In education, this translates to ensuring that if a student is identified as at risk by a genetic-informed tool, there is a clear, evidence-based explanation and a human oversight mechanism. Research can help by developing user-friendly tools (for instance, interfaces that show how a student's genetic and environmental profiles together inform a recommendation) and by testing these tools in mock scenarios with educators to refine their

usability and ethical alignment. In sum, future work must create the socio-technical infrastructure-from best-practice guidelines to validated decision aids-that would allow any introduction of genetics into education to be done safely, fairly, and with public trust.

Third, the field should explicitly tackle the global and equity dimensions of genomics in education. Thus far, the genetic studies of education outcomes have overwhelmingly been based on European-ancestry populations and have been led by a handful of high-income countries. This raises two concerns. One is scientific generalizability: polygenic scores constructed in European cohorts lose considerable accuracy in other ancestral groups, potentially exacerbating achievement gaps if applied universally. The other is a participation gap: developing nations, and even disadvantaged communities within wealthy nations, have not been equally involved in-or benefited from-research at this nexus. Future research agendas should prioritize diversity. This means expanding behavioral genetic cohorts to include underrepresented populations, studying gene-environment interplay in varied socioeconomic contexts, and examining how different education systems (e.g., highly tracked vs. comprehensive schools) modulate genetic influences. Such comparative work can inform whether certain policies mitigate or magnify inherent differences-for example, recent evidence suggests that high-quality schools can compensate for genetic disadvantages in reading ability. We need more data on whether interventions like extra tutoring or individualized learning plans have different effects on students with different genetic profiles, especially in resource-limited settings. Additionally, international collaboration can help ensure that ethical standards are globally informed. Just as organizations like the Nuffield Council on Bioethics in the UK have begun examining the use of polygenic scores in education policy, UNESCO or other international bodies could convene experts to develop principles applicable across diverse cultures and education systems (for instance, guarding against genetic discrimination in schooling is likely to be a universal concern). Ultimately, a future in which genomic insights are used for the benefit of learners everywhere-not just in wealthy, high-tech school districts-will require capacity building: training educational researchers in genomics, providing open-access tools and data, and fostering cross-cultural dialogues about what personalization should (and should not) entail in different societies.

Finally, a cross-cutting priority is privacy and data protection, which future research and development must weave into every solution. Educational genomic data is deeply personal-not only does it concern children, a vulnerable population, but it also inherently involves families (since genes are shared) and sensitive traits (learning disabilities, etc.). Any move toward precision education will face justified scrutiny over how genetic data is collected, stored, and used. Technologists and policy researchers should preempt these concerns by designing privacy-by-design approaches. For example, one avenue is to explore federated or encrypted analysis methods that allow schools to benefit from genetic insights (say, software that flags if a student might need a particular support) without ever exposing individual DNA data to misuse. Another idea is to use differential privacy techniques when reporting research findings, ensuring that no participant or school could be identified or stigmatized based on the published results. Research should also address how to obtain truly informed consent from parents (and assent from students) for any genetic testing in educational contexts-an area that intersects with psychology and sociology. By embedding privacy and equity considerations into the technical design phase, the community can avoid a scenario where a well-intentioned precision education program falters due to public backlash over perceived genetic surveillance. In other words, just as data security and ethics have become paramount in

learning analytics and student data generally, they must be front and center when the data in question is a child's genome.

### *5.2 Limitations*

This review has several limitations. It relies on a single database, Scopus, and includes only English-language publications retrieved through a deliberately focused query. Relevant research using alternative terminology, indexed in other databases, or published in other languages may therefore be absent. The results should be read as a map of a defined corpus rather than as a comprehensive account of all research connecting genetics and education. Country counts describe author affiliations rather than study populations or ancestry, and using full counting produces overlapping national totals. Citation counts and journal indicators are also time-sensitive. In addition, the archived project materials did not retain a granular screening-reason log, a complete keyword thesaurus, or all VOSviewer parameters, which limits exact replication. Finally, bibliometric patterns describe publication and citation structures; they do not assess study quality, establish causal relationships, or reveal why individual scholars or disciplines engage with genetic evidence in particular ways.

## 6. Conclusion

In conclusion, this review maps a focused Scopus corpus at the intersection of genetics and education. It documents the growth, intellectual structure, and disciplinary concentration of this literature, including the limited prominence of education-focused venues within the retrieved corpus. These patterns do not demonstrate that mainstream education, sociology, or economics scholars have deliberately ignored genetics, nor do they establish the reasons for the uneven disciplinary distribution. Future work should compare behavioral-genetic and genomic evidence across disciplines, test whether genetic information adds value beyond prior achievement and other established predictors, and examine its ethical and legal implications without presuming near-term classroom use. The longstanding challenge of supporting low-achieving students without restricting future opportunities predates polygenic indices; such indices neither create nor resolve that dilemma. A cautious research agenda should therefore acknowledge substantial genetic influence while requiring direct evidence for the effectiveness, durability, and fairness of any proposed educational application.

## Author Contributions

Alfonso Pellegrino: Conceptualization, Methodology, Data Curation, Formal Analysis, Software, Visualization, Writing – Original Draft, Validation. Alessandro Stasi: Conceptualization, Writing – Review & Editing, Supervision, Validation. Both authors: Final approval of the version to be published and accountability for all aspects of the work.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Competing Interests

The authors declare no conflicts of interest.

## Data Availability Statement

The bibliographic data analyzed in this study are sourced from Scopus and are available from the corresponding author upon reasonable request.

## AI-Assisted Technologies Statement

During the preparation of this work, the authors used Gemini in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## References

1. Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault JB, et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. Proc Natl Acad Sci USA. 2014; 111: 15273-15278.
2. Silventoinen K, Jelenkovic A, Sund R, Latvala A, Honda C, Inui F, et al. Genetic and environmental variation in educational attainment: An individual-based analysis of 28 twin cohorts. Sci Rep. 2020; 10: 12681.
3. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018; 50: 1112-1121.
4. Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. Nat Genet. 2022; 54: 437-449.
5. Plomin R, Von Stumm S. The new genetics of intelligence. Nat Rev Genet. 2018; 19: 148-159.
6. Mascheretti S, Andreola C, Scaini S, Sulpizio S. Beyond genes: A systematic review of environmental risk factors in specific reading disorder. Res Dev Disabil. 2018; 82: 147-152.
7. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: Effects of parental genotypes. Science. 2018; 359: 424-428.
8. Harden KP. The genetic lottery: Why DNA matters for social equality. Princeton, NJ: Princeton University Press; 2021.
9. Asbury K, Plomin R. G is for genes: The impact of genetics on education and achievement. Hoboken, NJ: John Wiley & Sons; 2013.
10. Meaburn E. Navigating genomics and education: Insights, opportunities and challenges. London, UK: Nuffield Council on Bioethics and Nuffield Foundation; 2025.
11. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016; 533: 539-542.
12. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. J Bus Res. 2021; 133: 285-296.

13. Zupic I, Čater T. Bibliometric methods in management and organization. Organ Res Methods. 2015; 18: 429-472.
14. De Zeeuw EL, de Geus EJ, Boomsma DI. Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. Trends Neurosci Educ. 2015; 4: 69-76.
15. Marks GN, O'Connell M. Inadequacies in the SES-achievement model: Evidence from PISA and other studies. Rev Educ. 2021; 9: e3293.
16. Marks GN, O'Connell M. Student achievement is much more about cognitive ability and genetics than SES: A response to Debouwere. Rev Educ. 2024; 12: e3483.
17. Marks GN, O'Connell M. The importance of parental ability for cognitive ability and student achievement: Implications for social stratification theory and practice. Res Soc Stratif Mobil. 2023; 83: 100762.
18. Jensen A. How much can we boost IQ and scholastic achievement. Harv Educ Rev. 1969; 39: 1-123.
19. Goldberger AS. Heritability. Economica. 1979; 46: 327-347.
20. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science. 2013; 340: 1467-1471.
21. Turkheimer E, Haley A, Waldron M, d'Onofrio B, Gottesman II. Socioeconomic status modifies heritability of IQ in young children. Psychol Sci. 2003; 14: 623-628.
22. Tucker-Drob EM, Bates TC. Large cross-national differences in gene × socioeconomic status interaction on intelligence. Psychol Sci. 2016; 27: 138-149.
23. Figlio DN, Freese J, Karbownik K, Roth J. Socioeconomic status and genetic influences on cognitive development. Proc Natl Acad Sci USA. 2017; 114: 13441-13446.
24. Scarr-Salapatek S. Race, social class, and IQ: Population differences in heritability of IQ scores were found for racial and social class groups. Science. 1971; 174: 1285-1295.
25. Lahtinen H, Martikainen P, Korhonen K, Morris T, Myrskylä M. Educational tracking and the polygenic prediction of education. Sociol Sci. 2024; 11: 186-213.
26. Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010; 84: 523-538.
27. Van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics. 2017; 111: 1053-1070.
28. Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: A comparative analysis. Scientometrics. 2016; 106: 213-228.
29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021; 372: n71.
30. Bouchard Jr TJ, Lykken DT, McGue M, Segal NL, Tellegen A. Sources of human psychological differences: The Minnesota study of twins reared apart. Science. 1990; 250: 223-228.
31. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet. 2002; 3: 872-882.
32. Deary IJ, Strand S, Smith P, Fernandes C. Intelligence and educational achievement. Intelligence. 2007; 35: 13-21.
33. Deary IJ, Johnson W. Intelligence and education: Causal perceptions drive analytic processes and therefore conclusions. Int J Epidemiol. 2010; 39: 1362-1369.

34. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015; 47: 291-295.
35. Benjamin DJ, Cesarini D, Van Der Loos MJ, Dawes CT, Koellinger PD, Magnusson PK, et al. The genetic architecture of economic and political preferences. Proc Natl Acad Sci USA. 2012; 109: 8026-8031.
36. Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A, et al. Genetic analysis of social-class mobility in five longitudinal studies. Proc Natl Acad Sci USA. 2018; 115: E7275-E7284.
37. Haworth CM, Dale P, Plomin R. A twin study into the genetic and environmental influences on academic performance in science in nine-year-old boys and girls. Int J Sci Educ. 2008; 30: 1003-1025.
38. Martin J, Taylor MJ, Lichtenstein P. Assessing the evidence for shared genetic risks across psychiatric disorders and traits. Psychol Med. 2018; 48: 1759-1774.
39. Rimfeld K, Malanchini M, Krapohl E, Hannigan LJ, Dale PS, Plomin R. The stability of educational achievement across school years is largely explained by genetic factors. NPJ Sci Learn. 2018; 3: 16.
40. Pinker S. The blank slate: The modern denial of human nature. London, UK: Penguin Books; 2003.
41. Dweck C. Mindset: The new psychology of success. New York, NY: Random House; 2006.
42. Lovett MW, Frijters JC, Wolf M, Steinbach KA, Sevcik RA, Morris RD. Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. J Educ Psychol. 2017; 109: 889-914.
43. Meyer MN, Appelbaum PS, Benjamin DJ, Callier SL, Comfort N, Conley D, et al. Wrestling with social and behavioral genomics: Risks, potential benefits, and ethical responsibility. Hastings Cent Rep. 2023; 53: S2-S49.