

Mentored research experiences bridge the wealth gap in PhD admissions

Wenhao Sun

whsun@umich.edu

University of Michigan <https://orcid.org/0000-0002-8416-455X>

Nicholas David

University of Michigan <https://orcid.org/0000-0002-5949-4581>

Misha Teplitskiy

University of Michigan <https://orcid.org/0000-0002-2839-7228>

Sidney Xiang

University of Michigan

Daniel Romero

University of Michigan <https://orcid.org/0000-0002-8351-3463>

Dallas Card

University of Michigan <https://orcid.org/0000-0001-5573-8836>

Article

Keywords:

Posted Date: August 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10084422/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Mentored research experiences bridge the wealth gap in PhD admissions

Authors: Nicholas David¹, Misha Teplitskiy², Sidney Xiang², Daniel Romero^{2,3,4}, Dallas Card², Wenhao Sun^{1*}

Affiliations:

¹ Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI, USA

² School of Information, University of Michigan, Ann Arbor, MI, USA

³ Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI, USA

⁴ Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

*Corresponding Author: whsun@umich.edu

First Paragraph:

PhD admissions decide which aspiring researchers gain access to scientific careers. Socioeconomic background has long shaped academic opportunity, influencing entry into elite undergraduate institutions¹⁻³ and the professorate⁴, yet its role in doctoral admissions has remained difficult to measure directly. Here we analyze 52,888 PhD applications submitted between 2013-2023 to one of the largest PhD-granting institutions in the United States, combining applicant credentials and recommendation letters with faculty evaluations and socioeconomic measures—including family home equity, parental education and reported financial hardship. We uncover a distinct wealth gap in doctoral admissions: applicants from higher socioeconomic backgrounds are both more likely to apply and to be admitted. Although faculty do not explicitly evaluate wealth, socioeconomic advantage systematically propagates through key admissions criteria including grades, test scores and undergraduate institutional prestige. In contrast, we find that mentored research experiences, and their resulting recommendations, provide an influential and wealth-agnostic signal of research potential. Expanding access to scientific training therefore calls for a stronger ecosystem of mentored research experiences, broader faculty participation in mentoring, and earlier encouragement for aspiring scholars to seek out these opportunities.

Graduate students are the lifeblood of science. As the primary workhorse behind cutting-edge research, they form the next generation of knowledge creators and shape the future of scientific innovation. Universities therefore hold a great responsibility to admit PhD applicants with the strongest potential to perform scientific research. In principle, a fair and effective doctoral admissions system would select candidates based on research potential alone. However, socioeconomic background has been linked to academic access for both undergraduates and faculty. Nearly half of professors originate from the top income quintile⁵, and access to elite undergraduate institutions is strongly associated with familial wealth through non-academic admissions criteria such as legacy admissions^{6,7}, athletics^{7,8}, and extracurricular activities^{8,9}. These wealth-linked pathways have contributed to broader societal tensions over who gains access to scientific training, and how academic institutions should address inequities^{10–13}, with calls for more holistic admissions^{14–16}.

To better inform graduate admissions policy, we need a more empirical and data-driven understanding of how socioeconomic status influences the competitiveness of PhD applicants. Because doctoral admissions are typically conducted by professors without formal training, when faced with large applicant pools and limited time they may rely on simple screening heuristics like grades, standardized test scores, and institutional prestige¹⁷. These factors have been shown to imperfectly predict long-term research success¹⁸, and also reflect socioeconomic disparities^{19,20}. On the other hand, mentored research experiences could provide a wealth-agnostic signal of research potential, as recommendation letters can testify to personal characteristics such as motivation, perseverance, and independence. In the United States, the National Science Foundation’s Research Experience for Undergraduates (REU) program is one prominent example of such a mentored research experience; other examples include research assistantships, senior theses, and master’s projects.

Historical dataset of doctoral applications

Here, we perform data-driven modeling of the relationships between PhD application features, socioeconomic status and admissions. To do so, we collate eleven years (2013–2023) of digitized PhD applications submitted to a large public R1 university in the U.S., totaling 175,702 applications across 154 doctoral programs. Each application includes a research statement, personal statement, curriculum vitae, and recommendation letters (typically three), as well as basic biographical and demographic information, grades, transcripts, and (in some cases) standardized test scores. For the engineering college, we further analyzed 22,968 faculty evaluations with numerical ratings and textual comments, providing direct insight into the faculty decision-making process. Details of this database construction, including the text-mining of original PDF applications, are in **S1**.

We further augmented this dataset with external data to better understand the associations of institutional prestige, recommender prestige, and socioeconomic status with PhD admission. For all applicants, we matched undergraduate institutions to global rankings²¹, and recommenders to bibliometric prominence²², both of which are associated with academic selection^{17,23}. To evaluate the socioeconomic status of domestic (United States) applicants, we link their applicant-provided permanent addresses to residential parcel-level property records, which contains information to assess their home value and home equity (see **S2** and **S3**). We assume that most doctoral applicants have not yet accumulated significant personal wealth, and that their listed permanent address is their primary familial residence. Housing wealth is a well-established proxy for intergenerational socioeconomic

position, even more so than income^{24–26}. We supplement this housing-wealth metric with application-provided indicators of socioeconomic status including parental education attainment, self-reported financial hardship, and prior Pell Grant recipient status—core indicators of socioeconomic status in higher education^{27–29}.

This dataset was designed to support a broad range of quantitative studies of doctoral admissions, and not every application will have all the features mentioned above. Therefore, our analyses in this work are performed on subsets of the data for which all the relevant features are available. Notably, to leverage housing data, we perform our wealth analysis only on applicants who provided a permanent address in the United States. This necessarily excludes most international applicants, limiting the downstream socioeconomic analysis mostly to U.S. citizens and permanent residents. In total there are 52,888 applications with complete application features, and 12,727 applications with validated housing wealth values.

The wealth gap in doctoral admissions

First, we identify a wealth gap in doctoral admissions, where applicants from wealthier backgrounds are both more likely to apply to PhD programs and more likely to be admitted. This finding addresses an empirical gap between two well-documented phenomena: the preferential admission of wealthy students to elite undergraduate institutions^{1–3} and the disproportionate representation in the professoriate of individuals from upper-income backgrounds⁴.

Figure 1A shows that domestic college graduates from wealthy families are more likely to apply for PhDs than the typical U.S. homeowner. Across every application term between 2013–2023, the applicant pool came from families with 18.6–48.7% higher home values than the U.S. median (Fig. 1A). In 2023, for instance, while the national median home price was \$432,950, the typical applicant came from a home valued \$614,536 or 41.9% more. This gap is especially pronounced for applicants applying to STEM majors, consistent with prior work showing that students from higher-socioeconomic status are more likely to enroll in STEM related fields^{30–32}. We also find that doctoral applicants tend to have parents of higher educational attainment compared to the typical U.S. family (Fig. 1B). PhD applicants disproportionately come from families in which at least one parent holds a graduate/professional degree ($\Delta = 33\%$), and are underrepresented among those whose parents' highest education is an associate degree, high school diploma, or less ($\Delta = -5\%$, -9% , -5%). These differences indicate that the typical PhD applicant grows up in a more educated and more socioeconomically-advantaged household than the typical U.S. family—suggesting that the pathway into doctoral education may be shaped long before students consider applying.

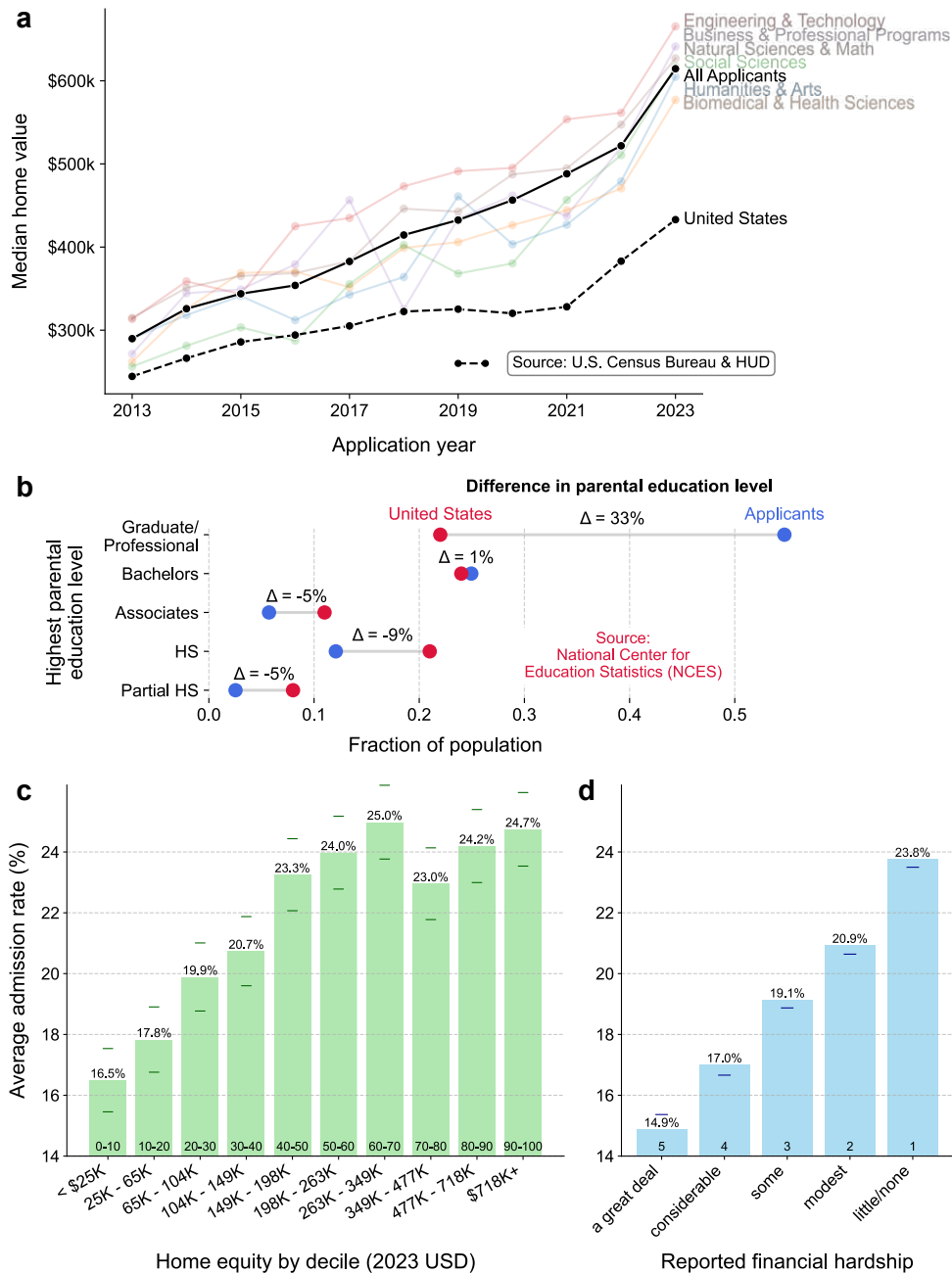


Fig. 1. Socioeconomic composition of doctoral applicants and admissions.

a, Median single-family home values of applicants versus the U.S. population. Home values for sales not recorded during the applicant's application year are adjusted using the *Annual Five-Digit ZIP Code FHFA House Price Index*³³; black lines denote all applicants (solid) and the U.S. (dashed)³⁴, with colors stratifying the applicant pool across common academic fields. Medians are shown annually from 2013-2023. **b**, Highest level of parental education among PhD applicants (blue) with that of parents of children under age 18 in 2022 (red)³⁵; percentage differences between these two groups are displayed above each category. **c**, Average admission rates of U.S. applicants by their permanent residence property values in 2023 USD—binned by deciles, each containing 1,272(3) applicants. Home equity values are adjusted to 2023 USD using the *Consumer Price Index for All Urban Consumers*³⁶. Equity calculations evolve the principal or refinanced mortgage assuming on-time payments under a 30-year fixed rate loan. **d**, Admission rates of U.S. applicants by self-reported financial hardship when funding their undergraduate education; "a great deal" (5,389), "considerable" (12,151), "some" (21,893), "moderate" (19,526), and "little/none" (25,877).

Even though faculty do not have explicit information regarding housing wealth, we find that socioeconomically-advantaged applicants have a significant advantage in doctoral admissions. Admission rates increase across home equity deciles—PhD applicants with the least home equity are admitted at 16.5%, compared to 23.0-25.0% among those with the most home equity (Fig. 1C). The median home equity for admitted applicants was 19.0% higher than rejected applicants ($\Delta = \$36,261$, Mann-Whitney U, $r = 0.09$, $n = 12,727$, $P < .001$).

Each application includes a self-reported measure of financial hardship during undergraduate study, which we find to be negatively correlated with admission (Fig. 1D). Applicants reporting “a great deal” of financial hardship have much lower admission rates compared to those reporting “little/none” hardship (14.9% vs. 23.8%, respectively). Students who have received a Federal Pell Grant, which is reserved for those with exceptional financial need to fund their undergraduate education, also see a penalty in doctoral admission rates ($\Delta = -3.6 \pm 0.3\%$). Applicants elaborate on their experienced financial hardship in a 500 character short response, typically citing the need to work additional jobs or take out loans to fund their undergraduate education—constraints which are likely to limit their time spent on academic preparation^{37,38}. These findings suggest that faculty are systematically admitting applicants from wealthier backgrounds, likely not through direct consideration of socioeconomic status, but by favoring application features that are shaped by greater socioeconomic advantage.

Assessment of factors associated with doctoral program admittance

Having established a wealth gap in doctoral admissions—one that exists prior to and independent of the admissions decision—we next examine how faculty evaluate and admit doctoral applicants. The application package consists of applicant-provided evidence of scholarly achievements, as well as recommendations from applicant-chosen recommenders who vouch for these achievements through written evaluations and quantitative ratings.

Using a multivariate logistic regression, we model the associations of applicant-provided and recommender-provided features with admission. All application features are ranked and normalized relative to each doctoral program and with respect to other applications submitted in the same term, which accounts for variation in admissions standards across programs and application cycles. Figure 2A is a density heatmap of admissions probabilities, as a function of applicant-provided and recommended-provided features, built from 52,888 datapoints. Using the model coefficients in Table 1, each application is transformed into an admissions probability, which determines the color of each square ($0\% < \text{pink} < 5\% < \text{yellow} < 12\% < \text{blue} < 25\% < \text{green}$). Darker squares contain more applications, while lighter squares contain fewer. Recommender-provided features are averaged across the total number of recommendations for a given application (see S4).

Figure 2A shows that both applicant-provided and recommender-provided features contribute similarly to an admissions decision. While most applications exhibit intrinsic qualities and recommendation letters that are well-correlated (both strong or both weak), the data also captures a tradeoff—where strong recommendations can compensate for weaker applicant credentials, and similarly strong grades for weaker letters. We perform a sensitivity analysis and show that these trends are robust, and persist even when we jitter the feature-weighting schemes for both the applicant- and recommender-provided features (Figs. S2-S4).

Having built this multivariate regression model, we next estimate the relative importance of different application features on the admissions decision. We analyze the change of admissions odds across one standard deviation, or one unit change for binary features. All else equal, each additional 0.3 points in undergraduate GPA (e.g., the difference between 3.6 and 3.9) is associated with 1.7× higher odds of admission; having participated in a mentored research experience is associated with 1.6× higher admission odds; and coming from a school ranked 250 places lower (according to the QS World University Rankings) is associated with 0.77× lower admissions odds.

The comments left by faculty reviewing applications further illustrate how admissions committees synthesize and evaluate these application features. In Figure 2B, we show representative excerpts from faculty evaluations given on admitted applications, for each quadrant in Figure 2A. Using natural language processing, excerpts were selected algorithmically from comments containing the most frequent contextual bigrams and trigrams. Notably, comments in quadrant IV—corresponding to stronger recommendations and weaker applicant-provided data—often reference biographical and demographical aspects of the application that indicate overcoming hardships from non-traditional or disadvantaged backgrounds.

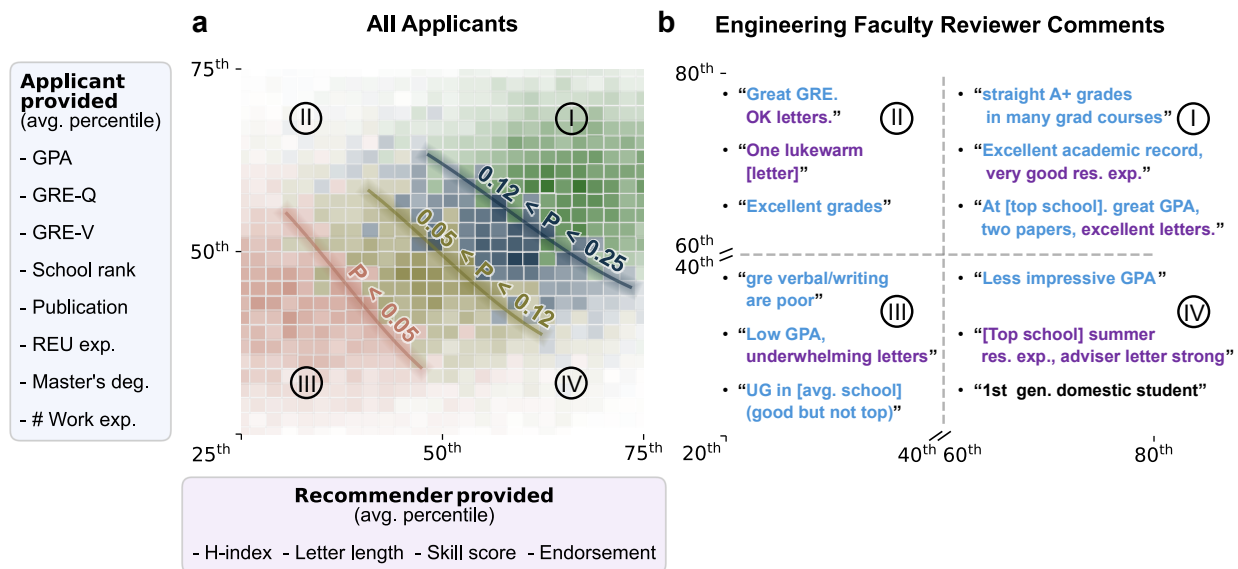


Fig. 2. Tradeoffs between applicant and recommender signals in doctoral admissions.

a, The tradeoff between applicant- and recommender-provided features and their role on the probability of being admitted. Application features are converted to a within group (term + program) percentile and binned. Bins are colored according to the probability quartile of the most common applicant ($0 < \text{pink} < 0.05 < \text{yellow} < 0.12 < \text{blue} < 0.25 < \text{green}$). The darker the bin the greater density of applicants present within that bin. Decision boundaries are fit using a cubic spline. **b**, Comments left by engineering college faculty when reviewing applications. Comments are drawn from admitted applications from the same quadrants in **(a)**, illustrated on the same axes. Quadrant III contains the weakest applicants with reviews commenting on their poor test scores, low GPAs, average ranked schools and underwhelming letters. Quadrant I contains the strongest applicants with excellent academic records, excellent letters, and very good research experience at top schools. Quadrants II and IV contain applicants that are a combination of strong and weak in their applicant-provided and recommender-provided materials. GRE scores are unavailable for applications during (and for some programs, after) the COVID-19 pandemic.

Table 1. Logistic regression coefficients for a model of doctoral admissions.

Recommender provided features are averaged across an applicant's recommendations. Citizenship is a binary variable indicating U.S. Citizens or permanent residents versus non-U.S. Citizens. Citizenship is included because baseline admissions rates differ significantly between domestic and international applicants (20.5% versus 9.1%, respectively).

Dependent variable: <i>Offered admission</i>			
Applicant provided		Recommender provided	
GPA	1.758*** (0.057)	H-index	0.007*** (0.001)
GRE-Q	0.008* (0.003)	Skill score	0.069*** (0.004)
GRE-V	0.036*** (0.003)	Letter word length	0.002*** (0.000)
Undergraduate institution rank	-0.001*** (0.000)	Endorsement	1.318*** (0.098)
Publication	0.119*** (0.028)	non-US Citizenship	-0.859*** (0.036)
REU experience	0.443*** (0.034)	Intercept	-26.700*** (0.685)
Master's degree	0.001 (0.033)	Observations	52,888
Number of work experiences	0.007 (0.005)	Pseudo R ²	0.219

Note: *p<0.05; **p<0.01; ***p<0.001

The importance of an influential recommender

Participating in mentored research experiences and receiving a strong recommendation letter is critical to expressing future research potential on a doctoral application. However, we find that not only is it important to receive a strong recommendation, it is also important to obtain the recommendation from an influential academic. For recommenders that are professors, researchers, or lecturers, we linked their name and institution to bibliometric records to obtain their *h*-index and publishing history. Although *h*-index²² is an imperfect statistic^{39–42}, it is a quantitative proxy for both visibility and prestige, as well as time in academia. In our dataset, 71.7% of applications have at least one recommender with published work. Using each recommender's unique author identifier, we further estimate letter-writing experience by counting the number of total recommendations they have submitted within our 11-year dataset of application files.

To determine the importance of recommender prestige, we next aim to understand: of those students who receive similarly positive endorsements, how much influence does the reputation and experience of the letter writer have on admission? Each recommender writes a letter and scores applicants on six criteria—field knowledge, motivation/perseverance, independence, communication, teaching potential, and research ability—rating the applicant relative to others they have known at similar career stages. Ratings range from “Exceptional-upper 5%” to the “Bottom 50%”, with a “No basis for judgement” category which we exclude. We map these categories to integers from 5 to 1 and sum them to create a total recommendation score (0–30). Recommenders also provide their overall endorsement of the applicant, most often selecting “Highly recommended” (87%). Using these summed ratings, we model a logistic regression with a generalized estimating equations approach, specifying application-level clusters to account for the fact that one single application, with one admissions decision, receives multiple (typically three) recommendations.

Figure 3A shows that higher *h*-index recommenders are associated with higher odds of admissions, even when holding recommendation scores constant (0.016, $P = .002$). For example, letters from recommenders with an *h*-index of 40 versus 20 are associated with a 23.8% increase in admission on average ($\Delta P_{\text{admit}} = 4.4\%$, from 18.5% to 22.9%). We notice a similar trend using a field-weighted *h*-index, which accounts for citation differences between specific fields (see S5). Since more productive faculty tend to teach at more prestigious universities⁴³, access to these recommenders may be limited for students from lower socioeconomic backgrounds, who are more likely to attend lower ranked undergraduate institutions⁴⁴.

More experienced letter writers—measured as the number of letters written across the dataset—are more successful in having their students admitted (Fig. 3B). Experienced letter writers are associated with higher admission at every level of recommendation score (0.088, $P < .001$) and the established positive effect of applicant skill diminishes when the recommender is someone who writes many letters (-0.002, $P = .03$). In other words, the benefit in admissions rate with experienced letter writers is greatest for applicants who receive lower recommendation scores; these recommenders may provide substantial narrative details in their letter that could help contextualize or distinguish candidates beyond their tabulated scores.

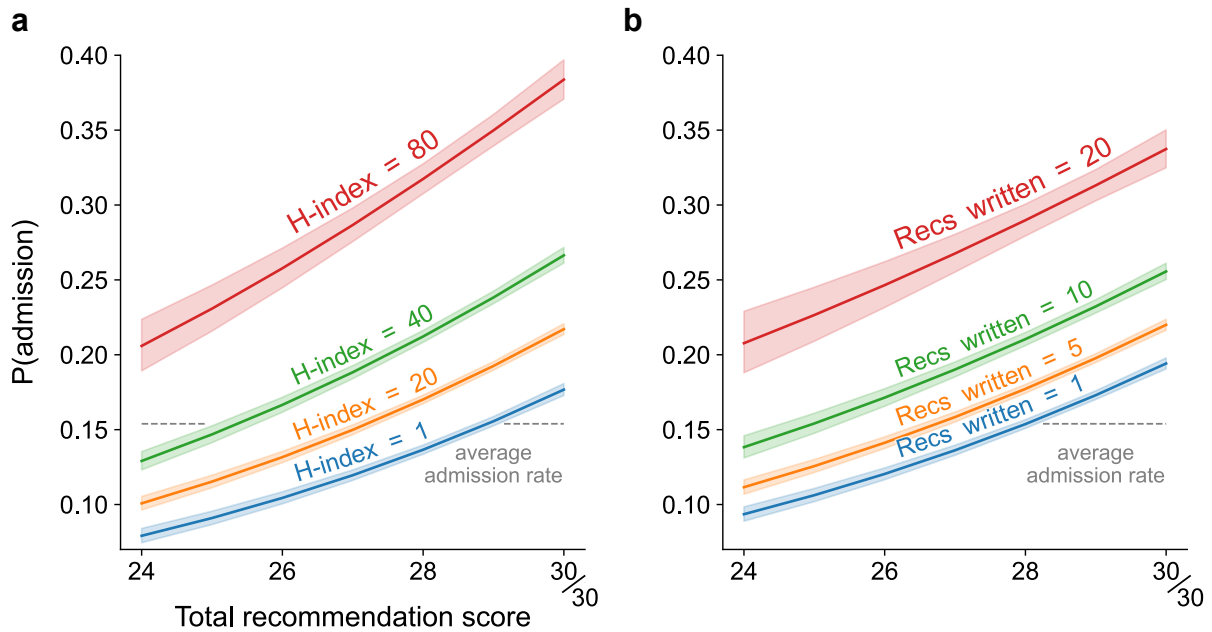


Fig. 3. Interaction of recommendation strength and recommender prominence in doctoral admissions.

a,b, Interaction plots for the modeled probability of admission for levels of total recommendation score, with the second factor, **(a)** recommender *h*-index, **(b)** frequency of letters written by a single recommender, represented by different colored lines. Interactions model the success of the recommender's applicant on being admitted. Recommendations with ratings in the top 2 (of 5) categories are considered for similar comparison. The gray dashed line indicates the average admission rate for all PhD applicants. The cumulative number of recommenders with *h*-indices greater than or equal to the given thresholds is 139,632 (≥ 1), 50,094 (≥ 20), 16,383 (≥ 40), and 2,313 (≥ 80). For repeat recommenders, the corresponding counts are 142,286 (≥ 1), 36,504 (≥ 5), 12,092 (≥ 10), and 2,404 (≥ 20).

Application features as mediators of socioeconomic status in admission

Thus far we have examined socioeconomic status, application features, and recommendations as separate predictors of admission; without yet resolving how they relate to one another. Application features may transmit (*i.e.*, mediate) some of the association between socioeconomic status and admission seen in Figures 1C-D. We therefore apply mediation analysis⁴⁵, which decomposes the total association between socioeconomic status and admission probability into indirect effects transmitted through application features and a remaining direct effect. Our mediation analysis results are represented by triangles in Figure 4D, which denote statistically significant (colored) or insignificant (gray) pathways. Triangle size reflects the strength of mediation, with smaller, semi-transparent symbols indicating effects below 0.005, equivalent to less than a 0.5 percentage point difference in admission probability. The sign of the association is indicated by triangle orientation (upward for positive, downward for negative). Examples of the path coefficients from these mediation analyses are shown in Figures 4A-C. See S6 for more details.

We find that socioeconomic status is most mediated by test scores and grades, and least mediated by research related experiences—such as publications, work experiences and participation in mentored research experiences (Fig. 4D). This suggests that the socioeconomic inequalities in PhD admissions arise primarily through standardized test scores rather than research engagement, highlighting research as a pathway toward more socioeconomically blind admissions.

Using the admissions model visualized in Fig. 2A, we introduce each SES variable individually into the model to examine if any remaining effect exists once all application components are controlled for. As summarized in Table 2, several SES variables reverse sign—parental education, financial hardship, and Federal Pell Grant status each shift from positive (or negative) associations in the direct model to associations in the opposite direction under the conditioned model. These reversals suggest that once all application components are controlled for, admissions committees may have in the past positively weighted socioeconomic disadvantage—consistent with holistic review processes that recognize applicants’ achievements within the context of their socioeconomic background⁴⁶.

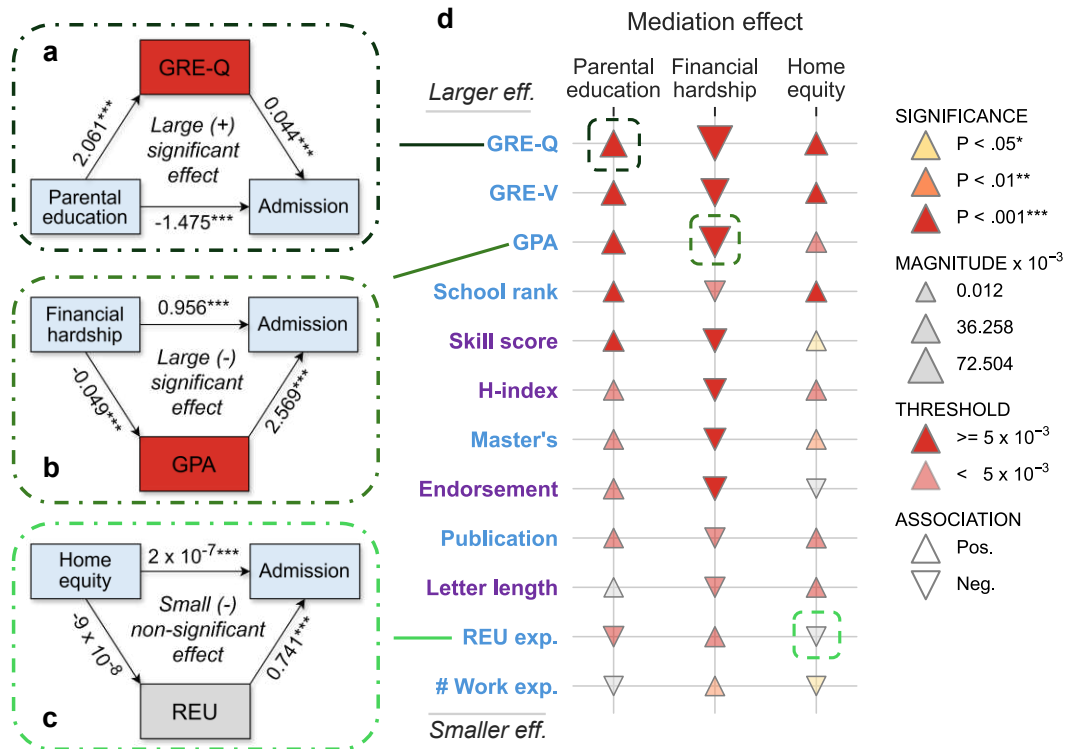


Fig. 4. Comparison of mediation pathways linking socioeconomic factors to doctoral admissions.

a,b,c, Mediation analysis of socio-economic factors on admission through applicant provided (blue) and recommender provided (purple) application features. Panels **a-c** examine the path coefficients on the two-step regression of, for example, parent education on GRE-V and admission on parent education and GRE-V. Mediation effects are color coded according to their significance in **(d)** with red being the most significant ($P < .001^{***}$) and gray being not significant ($P > .05$). **d,** Mediation effect sizes (magnitude) and significance (color). Effects < 0.005 (half a percentage point change in admission probability) are semi-transparent; triangle direction indicates effect sign (up, positive; down, negative). Features are ordered by overall significance and magnitude across socioeconomic factors. Treatment-control contrasts: parental education (bachelor's vs. graduate/professional degree), financial hardship (little/none vs. some), and home equity (25th vs. 75th percentile). School rank is inverted so positive effects indicate better (lower) ranks. Analyses are restricted to U.S. citizens/permanent residents residing in the U.S., using the maximum available sample with non-missing socioeconomic and application features.

Table 2. Socioeconomic status regression coefficients in models with and without application features.

Logistic regression coefficients for socioeconomic status (SES) variables under (1) SES as the sole regressor and (2) SES and all application components as regressors. Each model is estimated on the same sample of applicants for whom that linked SES variable is observed and all application features. The sample sizes in order of the table are 3,717, 16,559, 24,690, 24,690. Note that home equity still shows a significant association with admission in the larger sample (1.97×10^{-7} , $P < .001$, $N = 12,727$), despite not showing significance in the smaller sample below.

SES variable	(1)	(2)
DV: Admission	Direct model	Conditioned model
Home equity	9.5×10^{-8} (8.8×10^{-8})	5.7×10^{-8} (1.2×10^{-7})
Parental education	0.108^{***} (0.018)	-0.082^{***} (0.023)
Financial hardship	-0.139^{***} (0.012)	0.034^* (0.016)
Pell Grant status	-0.268^{***} (0.034)	0.172^{***} (0.043)

Note: * $P < .05$; ** $P < .01$; *** $P < .001$

1 Discussion

2 The last few decades have seen rising social tensions and erosion of trust in academic
3 institutions, in part due to a perceived *Matthew Effect*⁴⁷—where accumulated advantage begets further
4 advantage. Here, we quantitatively characterized a wealth gap in PhD admissions, where applicants
5 from wealthier backgrounds are not only more likely to apply to doctoral programs, but are also more
6 likely to be admitted. From analyzing the faculty evaluations for applicants, we found that faculty
7 do not overtly prefer students from wealthier backgrounds, but rather, students with wealthier
8 backgrounds more readily acquire the academic credentials that are most influential for doctoral
9 admissions. The responsibility for broadening access to higher education therefore lies not only in the
10 final admissions decision, but also in earlier intervention to elevate promising future scientists who are
11 coming from lower socioeconomic status backgrounds.

12 Our results suggest that mentored research experiences may be able to overcome this wealth
13 gap. Unlike standardized tests and GPAs, which strongly mediate socioeconomic status, participation
14 in mentored research experiences is less correlated with socioeconomic status, while remaining strong
15 predictors of admission. These tangible experiences are discussed in the applicant's statement of
16 purpose or recommendation letters and are critical to assessing research potential¹⁷. By prioritizing
17 demonstrated research ability over common academic metrics, admissions committees can identify
18 talent that traditional signals might miss.

19 In order to maximize the talent and research potential of academic institutions, there is
20 therefore an onus to expand federally-funded academic programs such as The National Science
21 Foundation's Research Experiences for Undergraduates (REU). Programs like these draw students
22 from under-resourced universities to more research-intensive institutions, where they are given the
23 opportunity to conduct summer research with their research faculty. As shown here, recommendations
24 from these faculty can play a significant role in improving admissions odds. Under their tutelage and
25 research mentorship, students can acquire the necessary research skills and endorsements to become a
26 more competitive doctoral applicant. Moreover, faculty endorsement of an applicant's research
27 capabilities provides essential external validation to reveal these 'hidden gem' researchers to
28 admissions committees, regardless of their socioeconomic origins.

1 **Methods**

2 **Ethics review**

3 This study was reviewed and approved by the University of Michigan Institutional Review Board
4 (IRB study HUM00225492). Consistent with the IRB's approval of this study as expedited, minimal-
5 risk research on pre-existing records, study-specific informed consent was not required; applicant
6 materials were used under the institution's existing submission acknowledgement, and faculty
7 evaluations were used as pre-existing institutional records governed by an approved Memorandum of
8 Understanding (MOU). This MOU specifies the institutional requirements for data security,
9 confidentiality, subject identifiability, and disposal. The acquisition and linkage of third-party data
10 (school rankings, property records, census data, and bibliometric records) to the application records
11 were outlined and approved in the study protocol.

12 To protect privacy and confidentiality, direct identifiers were used solely for record linkage within
13 the institution's data warehouse and were removed before analysis. Data were stored and analyzed
14 exclusively on university-managed resources—the research group's compute cluster, a university-
15 managed desktop workstation with an encrypted local drive, and an approved university Dropbox
16 folder (for initial sharing)—all restricted to IRB-approved study personnel authenticating with secure
17 university credentials. All reported results are aggregate statistics from which no individual
18 applicant, recommender, or evaluator can be identified, and representative faculty-comment excerpts
19 were anonymized (see Fig. 2b methods). Data retention and disposal followed the approved data-
20 management and security plan and the source institution's MOU.

21 **Study overview**

22 We analyzed 175,702 PhD applications submitted across 154 doctoral programs to a single large U.S.
23 R1 public research university between the 2013 and 2023 application cycles. Each application
24 package contains a generic application form, an academic statement of purpose, a personal statement,
25 a curriculum vitae, three or more recommendation forms with accompanying letters, transcripts, and
26 (for engineering applicants) one or more faculty evaluations. We linked these applications to four
27 external data sources: the QS World University Rankings²¹ (for applicant undergraduate institutions
28 and recommender affiliations), Elsevier's Scopus⁴⁸ and SciVal⁴⁹ databases (for recommender
29 bibliometrics), the CoreLogic Property Basic⁵⁰, Owner Transfer⁵¹, and Mortgage⁵² databases (for
30 applicant home real-estate records), and Freddie Mac⁵³, Federal Housing Finance Agency (FHFA)³³,
31 Bureau of Labor Statistics (BLS) and Census Bureau public data³⁶ (for inflation, house-price
32 indexing, mortgage rates, and parental-education benchmarks).

33 The final analytic samples used in the main text are: 52,888 applications with complete applicant-
34 and recommender-provided features (Fig. 2, Table 1); 12,727 U.S.-resident applications with
35 validated home-equity values (Figs. 1A,C and 4); 514,572 recommendations linkable to bibliometric
36 records (Fig. 3); and subsets ranging from 3,717 to 24,690 U.S.-resident applications with the
37 relevant socioeconomic-status (SES) covariate observed (Fig. 4, Table 2). The 11-year analysis
38 window was determined by the earliest year in which applications were digitized at the host
39 institution; no statistical sample-size calculations were performed because the analysis uses the full
40 available administrative record.

1 **Construction of the application database**

2 **Tabular records from the university data warehouse.** GPA (on the reported scale), GRE-
3 Quantitative and GRE-Verbal scaled scores, prior-degree history, prior work-experience entries,
4 citizenship status, prior Federal Pell Grant receipt, application program code, application term, and
5 the final admission decision were extracted directly from the host institution's data warehouse for all
6 Fall-matriculating PhD applications between 2013 and 2023. A binary master's-degree flag was
7 derived from the prior-degree history. Engineering-college faculty evaluations—containing up to six
8 rubric scores, an overall numeric grade, and free-text comments—were available in tabular form and
9 linked to applications by application ID, department identifier, department name, and term (3.9%
10 unlinked due to multi-program applications or department-naming inconsistencies).

11 **PDF-derived features.** Several application features were not warehoused and were text-mined from
12 the original compiled PDF using PyMuPDF⁵⁴. We first segmented each PDF deterministically into its
13 constituent sections (generic form, academic statement of purpose, personal statement, curriculum
14 vitae, and one block per recommendation) by exploiting the consistent page headers and bookmark
15 structure of the application template; segmentation logic and edge cases are detailed in **S1.2.1**. Of the
16 175,702 Fall PhD applications, 6.8% ($n = 11,873$) could not be segmented and were processed as
17 ordered page-text lists. Two applications failed extraction outright (one image-only PDF, one read
18 error from an oversized embedded image).

19 Within segmented sections, form-field values (recommender first name, last name, title, institution;
20 applicant permanent address; financial-hardship Likert response; maximum parental education) were
21 extracted from predefined bounding boxes whose coordinates were fixed for each application cycle's
22 template. Checkbox selections (used to identify participation in named mentored research programs:
23 McNair, RISE, PREP, MARC, MICHHERS, Mellon Mays, LSAMP, GEM, Bridges to the
24 Baccalaureate, and Big Ten SROP) were detected by computing the black-pixel ratio inside each
25 candidate checkbox against a grayscale template; ratios above 0.1 were coded as selected.

26 **Text-derived features.** Two binary applicant features—publication and REU—were derived by
27 regular-expression matching over the concatenation of an applicant's statement of purpose, personal
28 statement, curriculum vitae, and recommendation letters. The publication regex matched
29 publication(s), publish(ing), or published. The mentored-research regex matched the literal acronyms
30 REU, SRIP, SURE, SROP, UROP, SURP, and SURF as well as their expanded forms (the full
31 pattern list is reproduced in S1.3). We validated both flags by manually annotating 200 randomly
32 sampled applications stratified 50:50 by predicted label, obtaining F1 scores of 0.95 (REU) and 0.87
33 (publication). The final truth value of REU is computed as the logical OR of this text-derived feature
34 and the named mentored research programs in the section above.

35 **Recommender scores and letter length.** Each recommendation form solicits ratings on six skill
36 dimensions—field knowledge, motivation/perseverance, independence, communication, teaching
37 potential, and research ability—on a five-point scale from “Bottom 50%” (1) to “Exceptional – upper
38 5%” (5), plus a “No basis for judgement” option that we excluded from aggregation. Across an
39 applicant's recommendations we took the median per dimension and averaged across dimensions to
40 produce a single skill score per application. For the visualizations in Fig. 2 we instead mapped

categories to the midpoint of their percentile range (97.5, 87.5, 72.5, 57.5, 25). An overall endorsement (4 = “Highly recommended” to 1 = “Not recommended”) was averaged across recommenders. Letter word counts were computed by joining hyphen-broken lines, replacing other line breaks with spaces, tokenizing on whitespace, and averaging across an application’s letters.

External data linkage

Undergraduate institutional rank. Applicant undergraduate-institution strings were linked to the year-specific QS World University Rankings²¹ (2013–2023) by computing the Levenshtein distance between each of the 1,270 QS institution names and every distinct applicant-supplied institution name. For each QS entry the five closest application strings, together with country, state, and city, were presented to an annotator on the research team for verification. Year-normalized QS rank was matched to each application’s submission cycle. In total, 65.6% of applications (n = 115,321) received a linked QS rank.

Recommender bibliometrics. Recommenders whose self-reported title contained “professor”, “research”, “researcher”, or “lecturer” (514,572 recommendations) were queried against the Elsevier Author Search endpoint using first name, last name, and affiliation, yielding 123,127 unique author profiles. For each profile we retrieved h-index, by-year h-index, and subject-area publication counts from SciVal (snapshot October 2024; configuration parameters in S5). Letter-writing experience was operationalized as the count of recommendations a unique authorID submitted across the 11-year corpus.

For Fig. 3 we used the raw h-index; for the robustness check in Fig. S5 we computed a field-weighted h-index percentile $FWH_i = \mathbb{E}_{f \sim w_i} [F_f(h_i)]$, where $F_f(h_i)$ is the cumulative distribution

function of h -index within field f and $w_{if} = \frac{n_{if}}{n_i}$, is author i ’s share of publications in field f .

Home real estate. We treated each U.S.-resident applicant’s self-reported permanent address as their primary familial residence and linked it to CoreLogic Property Basic⁵⁰, Owner Transfer⁵¹, and Mortgage⁵² records. Free-form address fields were standardized to United States Postal Service (USPS) conventions (punctuation stripped, case normalized, zip codes truncated to five digits, street suffixes and cardinal/unit designators substituted in both forward and reverse to generate every permissible spelling permutation; full grammar in S2.1). A composite key made up of address information was matched against CoreLogic records constructed from the SITUS street, city, state, and zip code fields. To increase yield, addresses were also geocoded using a local Nominatim⁵⁵ with OpenStreetMap⁵⁶ instance and matched to CoreLogic parcels by latitude/longitude with secondary verification of house number and zip code.

A matched parcel was retained only if (i) the buyer or owner surname on the deed transfer matched the applicant’s surname; (ii) the parcel was flagged owner-occupied (CoreLogic owner-occupancy code M, O, or S); and (iii) the borrower surname on the mortgage matched the surname of the applicant. Multiple matches were resolved by selecting the deed transfer or mortgage closest in time to a December 1st application deadline of the year preceding the matriculation term. Refinanced principal mortgages dated before the application were used in place of the original purchase

mortgage. We restricted parcels to single-family residences, condominiums, duplexes, and apartments transacted at arm's length on or after 1 January 2000, and excluded transactions of \$1,000 or less to remove obvious non-arms-length transfers and quitclaim deeds⁵⁷. Parcels with no recorded mortgage were assumed unencumbered.

Home-equity calculation. Home equity at the time of application was computed as home equity = estimated market value – estimated debt. Market value was the most recent sale price, adjusted to the application date with the *Annual Five-Digit ZIP Code FHFA House Price Index*^{®33} to account for local appreciation. Estimated debt was the amortized balance of the principal or refinanced mortgage under a 30-year fixed loan,

$$B = L \times \frac{(1+r)^n - (1+r)^k}{(1+r)^n - 1}$$

where L is the loan amount, r is the monthly rate (national Freddie Mac average at the mortgage origination month), n = 360, and k is the number of months between origination and the application date. All home-equity values were converted to 2023 USD with the BLS Consumer Price Index for All Urban Consumers³⁶.

Modeling admissions decisions

Multivariate logistic regression. The relationship between the binary admission outcome and the 13 applicant-, recommender-, and biographical features listed in **S4** was modelled with a multivariate logistic regression. To absorb between-program and between-cohort variation in admissions standards, the model included a full program × term interaction, with Computer Science and Engineering PhD and Fall 2013 as the reference groups. To avoid spurious null effects for the mentored-research flag, programs with fewer than 20 applicants or a computed REU rate below 2.5% (i.e., programs whose disciplinary culture is unlikely to involve REU-style training) were excluded from the REU coefficient. Coefficients and standard errors calculated from the estimated covariance matrix of the fitted model are reported in Table 1. Two-sided 95% confidence intervals were computed using an asymptotic normal (z-based) distribution to estimate p-values. Model performance was evaluated with 5-fold cross-validation, yielding an F1 of 0.50 ± 0.01; a representative held-out fold contained 1,062 true positives, 1,179 false positives, 7,453 true negatives, and 883 false negatives, reflecting the substantial residual noise inherent to admissions prediction.

Visualizing the applicant–recommender tradeoff. For Fig. 2a, each applicant-provided and recommender-provided feature was converted into a within-program, within-term percentile. The two feature sets were averaged (equal weight within each set) to place each application in a 2D plane, which was discretized into 0.02 × 0.02 bins. Each bin was colored according to which of four admission-probability quartiles (0–0.05, 0.05–0.12, 0.12–0.25, 0.25–1.0) contained the most applications in that bin; bin opacity was scaled to the count of applications relative to the most populous bin. The three decision boundaries between adjacent quartiles were extracted by applying a Gaussian filter to smear the discrete transitions, locating the median vertical crossing of the 0.5 contour at each horizontal position by linear interpolation, and fitting a degree-3 univariate spline

(smoothing factor $0.3 \times$ number of boundary points). Splines were truncated where the underlying bin density fell below 25% of the maximum. We confirmed the qualitative robustness of these decision boundaries by re-running the procedure under three independent $\pm 10\%$ random perturbations of the equal-weight scheme (Figs. S2–S4).

Faculty-comment selection (Fig. 2b). Within each Fig. 2a quadrant, we restricted to admitted engineering applications, extracted the most frequent contextual bigrams and trigrams from the faculty comment field, and sampled representative comments containing those n-grams. Comments were truncated to the phrase most relevant to the n-gram theme; identifying tokens were anonymized in square brackets.

Recommender-prominence models

To assess the marginal contribution of recommender prestige and experience while holding letter content constant, we modelled admission as a function of the summed six-dimension recommendation score (range 6–30) with one of two interaction covariates: the recommender’s *h*-index, or the recommender’s total letter-writing count in the corpus. Because every application receives multiple recommendations but only one admission decision, we estimated a logistic generalized estimating equations (GEE) model clustered at the application level. The interaction plots in Fig. 3 were generated by evaluating the fitted model at fixed recommender-prominence values; shaded bands are 95% confidence intervals using an asymptotic z-test. The analysis was restricted to recommendations rated in the top two of five overall endorsement categories, to compare similarly positive letters; the field-weighted *h*-index version is reported in Fig. S5.

Mediation analysis

For each pairing of a socioeconomic status (SES) exposure (parental education, financial hardship, home equity, prior Pell Grant receipt) with an application feature serving as mediator, we estimated the average causal mediation effect (ACME) using the *mediation* R package⁴⁵. Continuous mediators were modelled with linear regression, $P(M = T | X) = \alpha_0 + \alpha_1 X$; binary mediators with logistic regression, $P(M = T | X) = \text{logit}^{-1}(\alpha_0 + \alpha_1 X)$. The outcome model was always $P(Y = 1 | X, M) = \text{logit}^{-1}(\beta_0 + \beta_1 X + \beta_2 M + \beta_3 (X \times M))$. Standard errors and significance were obtained from 5,000 bootstrapped resamples.

SES variables were coded as follows: parental education on a 1–5 integer scale (partial HS through graduate/professional degree); financial hardship on a 1–5 integer scale; home equity in 2023 USD; Pell Grant receipt as a 0/1 indicator. Treatment–control contrasts were the 75th versus 25th percentile of home equity (approximately \$370,000 vs \$70,000), bachelor’s versus graduate/professional parental education, the lowest versus second-lowest hardship category, and Pell-recipient versus non-recipient. For each (SES, mediator) pair, the analysis was restricted to the largest sample of U.S.-resident applicants with both variables observed; sample sizes for every pair are tabulated in Table S10. The visual encoding in Fig. 4D and the conditioned models in Table 2 follow the conventions described in S6.1.

Statistics

1 Reported admissions odds and probability differences in the main text come from the logistic
2 regression in Table 1, evaluated at a one-standard-deviation change in continuous features or a one-
3 unit change in binary features, holding all other covariates at their sample means. Two-sided tests
4 were used throughout; significance thresholds are flagged in figure and table notes as $*P < 0.05$, $**P$
5 < 0.01 , $***P < 0.001$. Mann–Whitney U effect sizes (r) are reported alongside P values for the
6 admitted-vs-rejected home-equity comparison. P values are not adjusted for multiple comparisons.

References

1. Chetty, R., Deming, D. J. & Friedman, J. N. Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges. Working Paper at <https://doi.org/10.3386/w31492> (2023).
2. Jerrim, J., Chmielewski, A. K. & Parker, P. Socioeconomic inequality in access to high-status colleges: A cross-country comparison. *Research in Social Stratification and Mobility* **42**, 20–32 (2015).
3. Jerrim, J. *Family Background and Access to 'High Status' Universities*. (2013).
4. Morgan, A. C. *et al.* Socioeconomic roots of academic faculty. *Nat Hum Behav* **6**, 1625–1633 (2022).
5. Abramitzky, R. *et al.* Climbing the Ivory Tower: How Socio-Economic Background Shapes Academia. Working Paper at <https://doi.org/10.3386/w33289> (2024).
6. Hurwitz, M. The impact of legacy status on undergraduate admissions at elite colleges and universities. *Economics of Education Review* **30**, 480–492 (2011).
7. Espenshade, T. J., Chung, C. Y. & Walling, J. L. Admission Preferences for Minority Students, Athletes, and Legacies at Elite Universities. *Social Science Quarterly* **85**, 1422–1446 (2004).
8. Tompsett, J. & Knoester, C. Family socioeconomic status and college attendance: A consideration of individual-level and school-level pathways. *PLOS ONE* **18**, e0284188 (2023).
9. Park, J. J. *et al.* Inequality Beyond Standardized Tests: Trends in Extracurricular Activity Reporting in College Applications Across Race and Class. *American Educational Research Journal* **62**, 336–377 (2025).
10. *Regents of the University of California v. Bakke*, 438 U.S. 265 (1978). vol. 438 265 (1978).
11. *Grutter v. Bollinger*, 539 U.S. 306 (2003). vol. 539 306 (2003).
12. *Students for Fair Admissions v. Harvard*, 600 U.S. 181 (2023). vol. 600 181 (2023).
13. *Students for Fair Admissions v. University of North Carolina*. (2023).

14. Kent, J. D. & McCarthy, M. T. *Holistic Review In Graduate Admissions: A Report from the Council of Graduate Schools*. (2016).
15. Wilson, M. A., Odem, M. A., Walters, T., DePass, A. L. & Bean, A. J. A Model for Holistic Review in Graduate Admissions That Decouples the GRE from Race, Ethnicity, and Gender. *LSE* **18**, ar7 (2019).
16. Young, N. T. Rubric-based holistic review: A promising route to equitable graduate admissions in physics. *Phys. Rev. Phys. Educ. Res.* **18**, (2022).
17. Posselt, J. R. *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping*. (Harvard University press, Cambridge, Massachusetts, 2016).
18. Sackett, P. R., Borneman, M. J. & Connelly, B. S. High stakes testing in higher education and employment: appraising the evidence for validity and fairness. *Am Psychol* **63**, 215–227 (2008).
19. Hoxby, C. M. & Avery, C. The Missing “One-Offs”: The Hidden Supply of High-Achieving, Low Income Students. Working Paper at <https://doi.org/10.3386/w18586> (2012).
20. Miller, C. & Stassun, K. A test that fails. *Nature* **510**, 303–304 (2014).
21. QS World University Rankings: Methodology. *Top Universities* <https://www.topuniversities.com/world-university-rankings/methodology> (2026).
22. Hirsch, J. E. An index to quantify an individual’s scientific research output. *Proc Natl Acad Sci U S A* **102**, 16569–16572 (2005).
23. Jones, A., Schuhmann, P., Soques, D. & Witman, A. So you want to go to graduate school? Factors that influence admissions to economics PhD programs. *The Journal of Economic Education* **51**, 177–190 (2020).
24. Chetty, R., Hendren, N., Kline, P. & Saez, E. Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *Q J Econ* **129**, 1553–1623 (2014).
25. Pfeffer, F. T. & Killewald, A. Generations of Advantage. Multigenerational Correlations in Family Wealth. *Soc Forces* **96**, 1411–1442 (2018).

26. Saez, E. & Zucman, G. Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data *. *Q J Econ* **131**, 519–578 (2016).
27. Hout, M. & DiPrete, T. A. What we have learned: RC28’s contributions to knowledge about social stratification. *Research in Social Stratification and Mobility* **24**, 1–20 (2006).
28. Torche, F. Is a College Degree Still the Great Equalizer? Intergenerational Mobility across Levels of Schooling in the United States. *American Journal of Sociology* **117**, 763–807 (2011).
29. Dynarski, S. M. Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review* **93**, 279–288 (2003).
30. Understanding the Relationship Between Parental Education and STEM Course Taking Through Identity-Based and Expectancy-Value Theories of Motivation - Ryan C. Svoboda, Christopher S. Rozek, Janet S. Hyde, Judith M. Harackiewicz, Mesmin Destin, 2016.
<https://journals.sagepub.com/doi/10.1177/2332858416664875>.
31. Maltese, A. V. & Tai, R. H. Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education* **95**, 877–907 (2011).
32. Nguyen, M. *et al.* Socioeconomic Diversity in Admissions to MD-PhD Programs, 2014-2019. *JAMA Netw Open* **7**, e241951 (2024).
33. FHFA House Price Index® Datasets | FHFA. <https://www.fhfa.gov/data/hpi/datasets>.
34. Median Sales Price of Houses Sold for the United States.
<https://fred.stlouisfed.org/series/MSPUS> (2025).
35. COE - Characteristics of Children’s Families.
<https://nces.ed.gov/programs/coe/indicator/cce/family-characteristics>.
36. U.S. Bureau of Labor Statistics. All items in U.S. city average, all urban consumers, not seasonally adjusted.
37. Perna, L. W. Understanding the Working College Student. *Academe* **96**, 30–33 (2010).

38. Darolia, R. Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review* **38**, 38–50 (2014).
39. Costas, R. & Bordons, M. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics* **1**, 193–203 (2007).
40. Waltman, L. & van Eck, N. J. The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology* **63**, 406–415 (2012).
41. Adler, R., Ewing, J. & Taylor, P. Citation Statistics. *Statist. Sci.* **24**, (2009).
42. Larivière, V., Ni, C., Gingras, Y., Cronin, B. & Sugimoto, C. R. Bibliometrics: Global gender disparities in science. *Nature* **504**, 211–213 (2013).
43. Zhang, S., Wapman, K. H., Larremore, D. B. & Clauset, A. Labor advantages drive the greater productivity of faculty at elite universities. *Sci Adv* **8**, eabq7056.
44. Chetty, R., Friedman, J. N., Saez, E., Turner, N. & Yagan, D. Mobility Report Cards: The Role of Colleges in Intergenerational Mobility. Working Paper at <https://doi.org/10.3386/w23618> (2017).
45. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software* **59**, 1–38 (2014).
46. Bastedo, M. N., Bowman, N. A., Glasener, K. M. & Kelly, J. L. What are We Talking About When We Talk About Holistic Review? Selective College Admissions and its Effects on Low-SES Students. *The Journal of Higher Education* **89**, 782–805 (2018).
47. Merton, R. K. The Matthew Effect in Science: The reward and communication systems of science are considered. *Science* **159**, 56–63 (1968).
48. Elsevier. Scopus. (<https://www.elsevier.com/products/scopus>)
49. Elsevier. Scival. (<https://www.elsevier.com/products/scival>)
50. CoreLogic, Inc. Property Basic. (2020).

- 1 51. CoreLogic, Inc. Onwer Transfer. (2021).
- 2 52. CoreLogic, Inc. Mortgage. (2022).
- 3 53. Freddie Mac. Primary Mortgage Market Survey. (2025).
- 4 54. PyMuPDF. Artifex (2024). (<https://pymupdf.readthedocs.io/en/latest/>)
- 5 55. Nominatim. (<https://nominatim.org/>)
- 6 56. OpenStreetMap contributors. Planet dump retrieved from <https://planet.osm.org>. (2025).
- 7 57. Lambie-Hanson, L., Li, W. & Slonkosky, M. Real estate investors and the U.S. housing
- 8 recovery. *Real Estate Economics* **50**, 1425–1461 (2022).
- 9
- 10

Acknowledgements

The work by ND, DC, and WS, was supported by the Michigan Institute for Data and AI in Society (MIDAS) Propelling Original Data Science (PODS) Grants. N.D. acknowledges support from the National Science Foundation Graduate Research Fellowship (NSF GRFP) under Grant No. DGE 2241144.

Author contributions

W.S. and N.D. conceptualized the study. N.D. curated the dataset and W.S., D.C. and N.D., were responsible for acquiring the funding. N.D., W.S., M.T., D.C., S.X. and D.R. developed the methodology. N.D. and W.S. created the visualizations. The original manuscript was written by N.D. and W.S. with substantive feedback from M.T., D.C. and D.R.

Competing interests

The authors declare no competing interests.

Data availability

The raw PhD-application records analyzed in this study contain personally identifying and protected educational information and cannot be shared publicly, in accordance with the host institution's data-use agreement and the U.S. Family Educational Rights and Privacy Act (FERPA). De-identified aggregate statistics underlying the published figures and tables are available from the corresponding author on reasonable request, subject to host-institution approval. The CoreLogic Property Basic, Owner Transfer, and Mortgage datasets are commercially available from CoreLogic (now Cotality; <https://www.corelogic.com/contact/>). Elsevier Scopus and SciVal bibliometric data are commercially available through Elsevier (<https://dev.elsevier.com/>). QS World University Rankings data were obtained from <https://www.universityrankings.ch/en/results/QS>. Publicly available reference series used for inflation adjustment and house-price indexing: the FHFA House Price Index (<https://www.fhfa.gov/data/hpi/datasets>), the BLS CPI-U (<https://www.bls.gov/cpi/data.htm>), the Freddie Mac 30-Year Fixed Rate Mortgage Average (FRED series MORTGAGE30US), the U.S. Census Bureau / HUD median sales price series (FRED MSPUS), and NCES family-characteristics tables (<https://nces.ed.gov/programs/coe/indicator/cce/family-characteristics>).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [gradfinalnaturenrdauidSIONLY.pdf](#)