

Which Colleges Equalize Earnings Between Low-Income and High-Income Graduates?

A Model-Selection Analysis of the College Scorecard

Abstract

Research on higher education and earnings has focused largely on whether selective institutions produce higher *average* earnings for their graduates, with a long-running debate over whether any premium reflects institutional treatment or selection. This study reframes the question: rather than asking which institutions produce high earners, it asks which institutions *equalize* earnings between their low-income and high-income graduates ten years after enrollment. Using institution-level data from the U.S. Department of Education's College Scorecard merged with the American Council on Education's 2025 Institutional Classifications ($n = 2,354$ institutions; $n = 1,348$ for regression), I fit a linear regression with BIC model selection and a nested-model mediation analysis. Out of more than fifteen candidate predictors, the selected model retains only Pell share and four major-composition variables. Traditional markers of institutional quality, including admission rate, cost, control type, and region, are all dropped. A nested-model comparison further shows that roughly half of the apparent relationship between selectivity and equalization runs through differences in major composition. The results suggest that within-school earnings equalization is shaped more by who enrolls and what they study than by institutional prestige.

1 Introduction

College graduates earn considerably more than workers with only a high school diploma, and that premium has grown over time (Oreopoulos & Petronijevic, 2013). Within the population of graduates, however, outcomes vary enormously across institutions, and a central question in the economics of education is how much of this variation reflects what colleges *do* for their students versus *whom* they admit. Chetty and colleagues, using tax records linked to enrollment data, found that certain elite institutions were associated with dramatically higher upward mobility for low-income students. Dale and Krueger, comparing students admitted to similar sets of colleges, found that the selectivity premium largely disappeared once unobserved student ability was accounted for, implying that much of the apparent benefit of elite schools is selection rather than treatment.

Most of this work asks whether selective institutions produce higher earnings on average. The present study takes a different angle. It asks which institutions *equalize* earnings, that is, which colleges produce similar outcomes for low-income and high-income students who attend the same school. This shifts the focus from access to treatment on the admitted. Two institutions can produce identical average earnings while differing sharply in whether their low-income graduates close the gap with higher-income peers, and the latter question is arguably more central to whether higher education functions as a mobility mechanism.

Research question. Which institutional and structural characteristics predict the size of the earnings gap between low-income and high-income graduates ten years after enrollment, and does major composition account for any apparent effect of selectivity? I hypothesize that student composition (Pell share) and major composition will dominate institutional-characteristic predictors, consistent with the Dale–Krueger view that observable institutional features contribute less than selection into them.

2 Data and Methods

The primary data source is the U.S. Department of Education’s College Scorecard, a public institution-level dataset compiled from IPEDS, the National Student Loan Data System, and de-identified Treasury tax records. I merged it on the IPEDS UNITID identifier with the American Council on Education (ACE) 2025 Institutional Classifications, the successor to the Carnegie Classification framework. The analytical sample is restricted to institutions that report median ten-year earnings, enroll at least 500 undergraduates, and report both the low- and high-income-tier earnings variables. After filtering and merging, the dataset contains 2,354 institutions; the regression analyses run on a complete-case subset of 1,348, the reduction driven mainly by missingness on admission rate.

Outcome. The outcome is the ratio of low-income-tier median earnings (families earning \$0–\$30K at first federal-aid receipt) to high-income-tier median earnings (families earning \$75K+), measured ten years after enrollment. A ratio of 1.0 indicates that low- and high-income graduates of the same institution earn the same amount; values below 1.0 indicate that low-income graduates earn less. This captures *within-institution* equalization rather than between-institution differences in average earnings. The distribution is approximately symmetric (skewness ≈ -0.04), with a median of 0.80 and an interquartile range of 0.72 to 0.87, so the raw ratio is modeled directly. The median institution’s low-income graduates earn roughly 80 cents on the dollar relative to higher-income peers.

Predictors. Predictors fall into three groups. *Student composition* is the share of students receiving Pell Grants, the standard proxy for the economic background of a student body. *Institutional characteristics* include admission rate, average net cost, undergraduate enrollment, a control factor (public, private nonprofit, for-profit), and Census-region fixed effects. *Major composition* is built

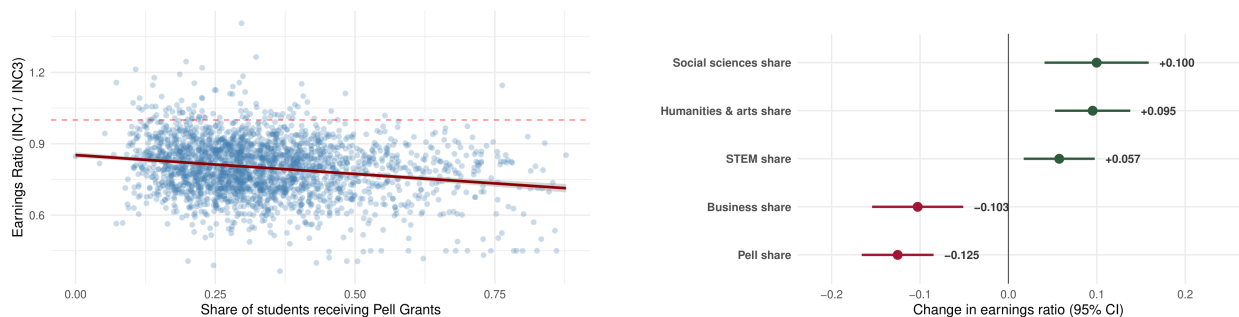
from the Scorecard’s CIP fields, aggregated into six buckets: STEM, Health, Business, Social sciences, Humanities and arts, and a residual Applied/Vocational bucket used as the omitted reference. Because the six shares sum to one at every institution, each major coefficient is read as the effect of moving a percentage point of degrees from applied/vocational toward the named bucket.

Models. The full candidate regression specifies the earnings ratio as a linear function of Pell share, admission rate, average cost, enrollment, control, region, and the five non-reference major shares, with an i.i.d. normal error. I apply stepwise selection on the Bayesian Information Criterion (BIC, penalty $k = \log n \approx 7.2$), preferred over AIC because the outcome has a low signal-to-noise ratio and the goal is interpretive: its preference for parsimony yields cleaner coefficients. To test whether major composition mediates selectivity, I compare two nested models on the same complete-case sample: a “before” model with admission rate (the selectivity proxy) and the controls but no majors, and an “after” model adding the five non-reference major shares, reporting the change in the admission-rate coefficient, the proportion mediated, and a nested F -test. Diagnostics confirm linearity and constant variance are approximately satisfied; with $n = 1,348$ inference relies on the Central Limit Theorem, and all retained predictors have variance inflation factors below 1.3.

3 Results

3.1 Descriptive baseline

Before modeling, the bivariate relationship between the strongest single predictor, Pell share, and the outcome is negative and highly significant (Pearson’s $r = -0.199$, 95% CI $[-0.238, -0.160]$, $p < 0.001$; Figure 1a): institutions enrolling more Pell-eligible students tend to show *larger* within-school earnings gaps, a counterintuitive pattern, since one might expect schools serving more low-income students to equalize outcomes for them. Among the major buckets, social sciences ($r = 0.30$) and STEM ($r = 0.21$) show the strongest positive bivariate correlations with the ratio.



(a) The puzzle: higher Pell share, wider gaps.

(b) The resolution: composition, not prestige.

Figure 1: (a) The earnings ratio falls as Pell share rises ($r = -0.199$, $p < 0.001$); the dashed line marks perfect equalization. (b) Forest plot for the BIC-selected model with 95% confidence intervals: green coefficients narrow the gap, garnet widen it. Once Pell share and field mix are accounted for, no traditional prestige variable survives selection.

3.2 Which characteristics predict equalization?

The full candidate model achieves $R^2 = 0.131$. BIC selection drops admission rate, average net cost, undergraduate enrollment, institutional control, region, and health share, retaining five predictors: Pell share and four of six major-composition variables (Table 1, Figure 1b). The final model achieves $R^2 = 0.090$. This is modest in absolute terms, but expected given the aggressive BIC penalty and the low-signal nature of institution-level earnings-gap outcomes. Every retained coefficient is significant at $p < 0.01$.

The surviving coefficients split into two clean groups. Pell share ($\beta = -0.125$) and Business

Table 1: Final BIC-selected model for the earnings equalization ratio ($n = 1,348$).

Variable	Estimate	Std. Error	t	p
(Intercept)	0.8646	0.0128	67.58	< 0.001
Pell share	-0.1253	0.0207	-6.05	< 0.001
STEM share	0.0575	0.0205	2.81	0.005
Business share	-0.1027	0.0263	-3.91	< 0.001
Social sciences share	0.0998	0.0300	3.32	< 0.001
Humanities & arts share	0.0953	0.0217	4.39	< 0.001

share ($\beta = -0.103$) widen the within-school gap, with both confidence intervals entirely below zero, while STEM (+0.057), Humanities and arts (+0.095), and Social sciences (+0.100) shares all narrow it. Equalization is pulled down by Pell composition and pre-professional (business) field mix, and pulled up by liberal-arts and STEM field mix.

The variables that did *not* survive selection are the traditional workhorses of the institutional-quality literature: admission rate, average net cost, control type, and region all failed the BIC penalty (admission rate’s full-model coefficient was -0.016 , $p = 0.29$). The absence of these prestige and selectivity variables is itself the finding: once Pell composition and field mix are accounted for, conventional measures of institutional quality add no detectable explanatory power to within-school equalization.

3.3 Does major composition mediate selectivity?

This motivates the mediation test. Although admission rate does not survive joint selection, it is a significant predictor on its own, so the question is whether its apparent effect runs through major composition. In the “before” model the admission-rate coefficient is -0.030 ($p = 0.038$); after adding the five major shares it shrinks to -0.016 ($p = 0.290$), a 47.9% reduction and a transition from significant to non-significant, while a nested F -test confirms the major-composition block adds significant explanatory power ($F(5, 1327) = 9.03$, $p < 0.001$). Nearly half of the apparent selectivity-equalization relationship runs through what students study: selective institutions equalize more partly because they concentrate enrollments in fields with more uniform outcomes, not because of anything intrinsic to selectivity.

4 Discussion

Two findings emerge. Within-school earnings equalization is predicted primarily by who enrolls (Pell share) and what they study (major composition), with institutional prestige, cost, control, and region contributing nothing detectable once composition is accounted for; and about half of the apparent selectivity effect is itself mediated by major composition. Both align with the Dale-Krueger reading of the earnings-premium literature: what looks like an institutional treatment effect is, on closer inspection, partly a composition story.

Three caveats bound these conclusions. The data are a single cross-sectional snapshot, so the mediation result describes associational structure, not a causal pathway; the Scorecard’s earnings tiers cover only federally aided students, making the ratio a within-institution measure among aid recipients; and the regression sample over-represents institutions with formal admissions. The most valuable extension would link these findings to student-level longitudinal data, decomposing the field-of-study effect into its selection and treatment components while relaxing the federally-aided-only restriction.

References

- Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2017). *Mobility Report Cards: The Role of Colleges in Intergenerational Mobility* (NBER Working Paper 23618). National Bureau of Economic Research.
- Dale, S. B., & Krueger, A. B. (2002). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. *The Quarterly Journal of Economics*, 117(4), 1491–1527.
- Dale, S. B., & Krueger, A. B. (2014). Estimating the Return to College Selectivity over the Career Using Administrative Earnings Data. *Journal of Human Resources*, 49(2), 323–358.
- Oreopoulos, P., & Petronijevic, U. (2013). *Making College Worth It: A Review of Research on the Returns to Higher Education* (NBER Working Paper 19053). National Bureau of Economic Research.
- Witteveen, D., & Attewell, P. (2017). The Earnings Payoff from Attending a Selective College. *Research in Social Stratification and Mobility*, 48, 48–56.
- American Council on Education. (2025). *2025 Institutional Classifications*.
- U.S. Department of Education. (2026). *College Scorecard Data*.

Generative AI Statement

No generative AI was used to conduct the research, perform the statistical analysis, create the figures, or write the paper.

Appendix: Institutional Mission and Equalization (Secondary Analysis)

As a secondary analysis, I treated institutional mission as categorical rather than collapsing it into continuous proxies, using a two-way ANOVA with ACE Research Activity (research doctoral, research colleges, teaching-focused) and Size (small, medium, large) as factors, fit on the full 2,354-institution sample. Because the 3×3 design is unbalanced, I used Type III sums of squares.

The Research Activity main effect is highly significant ($F(2, 2345) = 22.5, p < 0.001$). Estimated marginal mean ratios are 0.899 for research doctoral, 0.880 for research colleges, and 0.765 for teaching-focused institutions, a gap of roughly thirteen percentage points between research-intensive and teaching-focused schools. The Size main effect is not significant ($p = 0.116$), and the Research Activity $\times$ Size interaction is significant at the 0.05 level ($F(4, 2345) = 3.04, p = 0.017$), but the qualitative ordering holds across all three size groups (Figure 2).

This result complicates a strict composition story: even after the regression accounts for Pell share and field mix, research-intensive institutions still equalize substantially more than teaching-focused ones, suggesting an institutional-mission component, perhaps related to faculty research engagement, career-services infrastructure, or resource intensity, that is not fully reducible to student and field composition. The homogeneity-of-variance assumption is violated (Levene's $p < 0.001$); the two-way F -test is approximately robust at large samples, and the result is interpreted with appropriate caution given the smallest cell holds only seven observations.

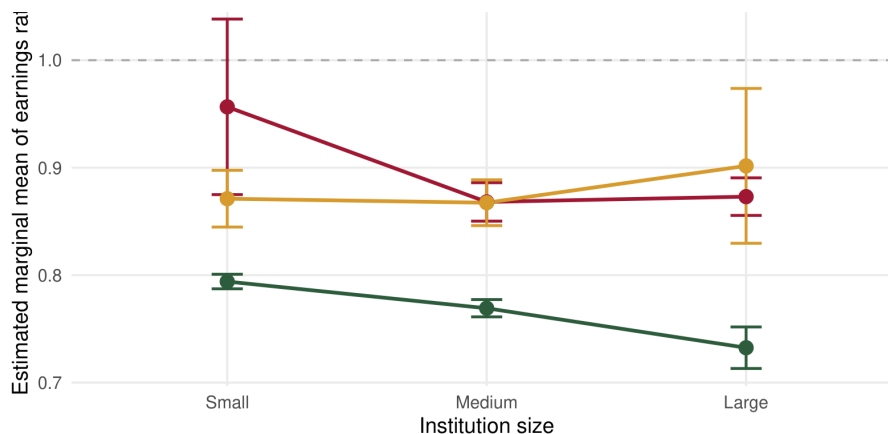


Figure 2: Estimated marginal means of the earnings ratio across the 3×3 Research Activity $\times$ Size design, with 95% confidence intervals. Garnet denotes research doctoral, gold research colleges, and green teaching-focused institutions. The two research-intensive series sit above the teaching-focused series at every size level; the dashed line marks perfect equalization (ratio = 1.0).