From Flagship to Firm: Gatekeeping, Employer Sorting, and the Returns to College

Aleksei Opacic*

Harvard University

October 16, 2025

Abstract

Firms increasingly drive college graduates' labor market earnings, yet the vast literature on the value-added of college has largely overlooked their role. Drawing on human capital theory and institutional theories of credentialism, closure, and organizational matching, I argue that college-to-workplace pipelines are a critical driver of college value-added. To test this, I assemble a novel US employer-employee matched dataset merged with postsecondary and high school academic records. Firm placement explains much of the variation in earnings premiums between colleges. Absent firm sorting, the range of counterfactual earnings differences across colleges would fall by 56%, and over half of the earnings advantage to attending the state flagship comes from access to higher-paying employers. These sorting effects extend broadly across the distribution of high-wage firms, and not merely a handful of elite employers. Crucially, sorting effects do not simply reflect skill-based advantages, implying that college quality derives as much from institutional linkages to the labor market as from human capital development. Policies aimed at broadening recruitment pipelines, rather than solely improving instructional inputs, are therefore essential to reducing inequalities in the economic returns to higher education.

*Department of Sociology, Harvard University, aopacic@g.harvard.edu. Thanks to seminar participants at Harvard University for helpful feedback, as well as to Clem Aeppli, Christina Ciocca Eller, Jessi Streib, Nathan Wilmers, and Xiang Zhou. An earlier version of this paper was presented at the American Sociological Association Annual Conference. The analyses and conclusions presented in this work are solely those of the author and do not reflect the views or opinions of the state's Employment or Education Departments. This research was partly supported by a Russell Sage Foundation Dissertation Grant (RSF Grant#: 2401-47047), grants from the Harvard Institute for Quantitative Social Science, the Harvard Center for Population Studies, as well as a Linda G. Hammett Ory Fellowship.

Introduction

Direct pipelines between specific colleges and employers have long been a staple feature of the American school-to-work transition. In the 1950s, the CIA targeted Ivy League college campuses in order to recruit the "best and the brightest",[1] while from the 1970s, select firms in law and management consulting became increasingly desired career destinations for graduates of top colleges (Granfield, 1992). McKinsey founder Marvin Bower was even so committed to recruiting Harvard Business School graduates that he established a fellowship program in his name at the school (Lemann, 1999), and the company soon had full-time employees tasked with recruiting from just this one university. Still today, significant shares of graduates from top colleges land at a remarkably narrow set of employers operating in a narrow set of sectors (Davis and Binder, 2019; Townsend, 2023).

Yet the vast literature on the economic returns to attending specific colleges - or "college value-added" - largely abstracts away from employers themselves. This work convincingly documents sizable earnings payoffs to measures of college quality: attendance at more selective institutions generally yields higher earnings than attendance at less selective ones (Brand and Halaby, 2006; Hoekstra, 2009; Zimmerman, 2014; Goodman et al., 2017; Witteveen and Attewell, 2017; Smith et al., 2020; Mountjoy, 2024). What remains less understood is why attending different colleges translates into such divergent labor market outcomes.

Two theoretical perspectives dominate. Human capital explanations for educational returns focus on education's productivity-enhancing effects, with workers paid wages corresponding to their marginal contribution (Becker, 1962). By contrast, longstanding sociological accounts emphasize the ways in which education acts as a sorting device into lucrative segments of the labor market (Bowles and Gintis, 1976; Parkin, 1983; Bills, 2003; Domina et al., 2017; Collins, 2019), where network advantages and mechanisms of social closure - such as licensing and credentials - bolster the economic value of degrees and certifications (Weeden, 2002; Weeden and Grusky, 2012). While both frameworks highlight important mechanisms, their empirical application to the case of college value-added has been limited. And crucially, neither approach centers the firm as the organizational site through which human capital, credentials, and networks are converted into differential economic rewards.

In this paper, I argue that the college one attends affects earnings partly via its effects on firm placement. I propose that the sorting of students across employers - specifically, into high- and low-premium firms - is an important yet under-recognized mechanism underpinning the added value of college in the labor market. Through this process of *firm refraction*, colleges transmit and magnify educational advantage.

[1] https://www.muckrock.com/news/archives/2018/jan/26/cia-college-tour-campus-protest/.

Integrating this mechanism into established human capital and sociological explanations clarifies how educational and employment institutions intersect to construct and sustain labor market inequality.

To develop this argument, I bridge the literature on college value-added with a growing body of work highlighting the changing structure of American pay inequality. Recent decades have seen rising earnings inequality among workers performing similar tasks within the same occupation. Working at an employer that pays above or below the market return to individual characteristics like human capital or productivity has become an increasingly important driver of earnings differences between workers (Card et al., 2013; Song et al., 2019; Wilmers and Aeppli, 2021). But while the consequences of firm-level sorting are well documented, the role of the college in sorting individuals into particular firms remains understudied. Prior sociological work emphasizes the role of social closure on the grounds of educational credentials at elite employers (Ho, 2009; Rivera, 2011, 2012), and documents systematic pipelines between colleges and employers (e.g. Davis and Binder, 2019). Yet the full extent of this pattern - its reach, effects on earnings, and contribution to college value-added - is unknown.

I address these issues by performing a unique linkage of administrative employer-employee pay records to postsecondary and high school academic records for the population of recent high school graduates in a US state.[2] I begin by examining the extent of college-based firm sorting, and its role in explaining college value-added. Next, I examine whether this sorting is confined to elite firms, or extends throughout the broader distribution of high-wage employers. Finally, I assess one possible explanation: mechanical sorting based on worker human capital and productivity.

I document three findings. First, firm sorting is a powerful mechanism underpinning college value-added. Absent firm sorting, the range of counterfactual earnings differences across colleges would fall by 56%. Firm sorting accounts for a full half of the total earnings effects of attending the Flagship - the college with the largest earnings returns. Second, firm sorting effects are a pervasive form of labor market stratification, wherein the benefits of attending a more selective college are reflected throughout the upper half of the firm pay structure. Finally, firm sorting advantages overwhelmingly reflect non-human capital resources conferred by colleges, implying that a substantial portion of college premiums derives from credential and network, rather than skill-based, advantages. These findings point to college-to-workplace pipelines as a critical, yet so far under-examined, engine of postsecondary inequality. Efforts to close earnings gaps through higher education must attend to how postsecondary institutions connect their graduates to firms.

[2] I anonymize the focal state throughout this paper to comply with my data use agreement governing my use of the state's administrative data.

College value-added

Higher education is a critical driver of economic outcomes. Much work has documented important earnings payoffs to college quality, such as selectivity or prestige (Bowen and Bok, 1998; Black and Smith, 2004; Brand and Halaby, 2006; Witteveen and Attewell, 2017; Goodman et al., 2017; Smith et al., 2020; though see Dale and Krueger, 2002, 2011). Even the exact college one attends appears to have a large effect on earnings outcomes (Hoekstra, 2009; Zimmerman, 2014; Goodman et al., 2017; Mountjoy and Hickman, 2021; Ciocca-Eller, 2023; Bleemer, 2023; Chetty et al., 2023; Mountjoy, 2024). Understanding why these returns arise is critical not only for explaining educational stratification, but for identifying which institutional features - whether instructional, cultural, or organizational - generate value that could be strengthened elsewhere in the higher education system. This is important because elite institutions with large earnings returns (even net of the characteristics of these students) predominantly serve white, wealthy students, while colleges with lower earnings returns primarily enroll students from underserved backgrounds (Reardon et al., 2012). Yet, surprisingly little work addresses this question.

The comparatively small quantitative literature that does focuses overwhelmingly on human capital-based explanations: attending a particular college may be productivity-enhancing (Becker, 1962, 1964). Colleges differ significantly on characteristics such as expenditure per student, tuition costs, faculty salaries, and ratio of student to faculty, with more selective colleges typically possessing greater resources (e.g. Behrman et al., 1996; Dale and Krueger, 2002; Thomas and Zhang, 2005; Hoekstra, 2009; Zhang, 2012). These factors have been shown to most directly predict completion rates (e.g. Gansemer-Topf and Schuh, 2006), but resource disparities may also shape long-term labor market outcomes by affecting the quality and efficacy of skill development that employers value (Behrman et al., 1996; Dale and Krueger, 2002). In an extreme case, colleges may even differ in the type of knowledge they impart (Biasi and Ma, 2022). Even if schooling does not increase human capital directly, employers may use schooling to screen applicants, faced with imperfect information about potential productivity (Stiglitz, 1975).

In contrast, sociological perspectives emphasize the role of education as a sorting device into particular segments of the labor market (Bills, 2003; Domina et al., 2017; Collins, 2019). Credentialism and network theories propose that education yields wage premiums not because of imparting superior knowledge, but because the highly educated control access to elite positions (Bills, 2003, p. 452). Bowles and Gintis (1976) argued that institutions effectively work to channel students to particular strata of the labor force, while Collins (2019, p.9) described education as “an artificial device for monopolizing access to lucrative occupations”. Licensing (Parkin, 1983), credentialism, and certifications all contribute to closure (Weeden, 2002; Weeden and Grusky, 2012; Abbott, 2014). At the same time, social capital and network resources conferred on students can be activated in job searches via contacts, referrals, and information provision (Stuber, 2006; Armstrong and Hamilton, 2013; Thiele and Gillespie, 2017). These networks are institutionalized through Greek life and other exclusive campus organizations, but also through informal peer cultures and interactions that structure everyday interactions on campus (Stevens et al., 2008; Stuber, 2011; Ridgeway, 2014; Michelman et al., 2022).

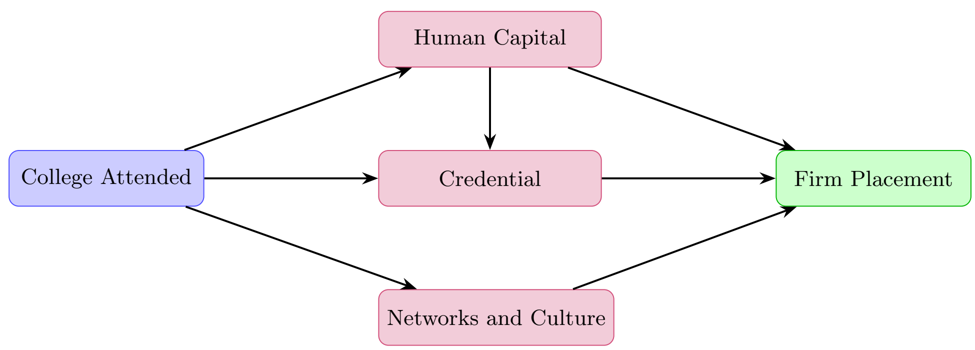
Employers are left implicit in many of these accounts. Human capital models presume a homogeneous labor market in which all firms reward productivity in similar ways. Credentialist accounts recognize education as a closure device, but rarely acknowledge that workplaces themselves are key sites of exclusion (Baron and Bielby, 1980; Tomaskovic-Devey and Avent-Holt, 2019). And research that highlights college-based networks does not fully flesh out the implications of facilitating access to *specific* employers.

Such an omission would be inconsequential in a purely competitive labor market. When an individual's wage is market-determined independent of any individual employer, certain colleges may produce more skilled graduates or provide them with credentials and network resources, but their graduates should earn more regardless of where they work: jobs, not employers, matter for pay. Yet, where one works is a growing determinant of pay and source of pay inequality in the US (Groshen, 1991). Research has found consistent differences in the pay firms offer to workers, even workers with observationally similar skills (Abowd et al., 1999; Card et al., 2013; Song et al., 2019; Wilmers and Aeppli, 2021). Highly productive "superstar" firms in the technology and professional services sectors offer exceptionally high compensation, while subcontractors and firms in retail pay substantially less (Autor et al., 2020; Krippner, 2012). These differences in pay across firms are quantitatively important, accounting for up to one third of overall earnings inequality (Song et al., 2019; Wilmers and Aeppli, 2021).

Education and firm placement

In the following, I build on these insights to reframe a core assumption of much research on the returns to college. Rather than translating directly into higher wages, college resource advantages - skills, credentials and networks, as summarized in Figure 1 - may operate partly through what I term *firm refraction*. Human capital acquired at more selective colleges may increase productivity but, additionally, facilitate access to higher-paying employers. College credential advantages may operate via informal workplace-level closure by opening doors to resource-rich firms, rather than restricting access to certain occupational groups. And campus cultures may embed advantages by socializing students into expectations, pipelines, and networks tied to specific employers. This perspective conceptualizes colleges and employers as intersecting institutions that jointly structure opportunity, converting educational hierarchies into durable inequalities in employment and pay. By centering the college as a source of firm sorting, it offers a fuller account of how colleges generate labor market inequality.

Figure 1: Schematic Diagram Linking College Attended to Firm Placement



Human capital and firm sorting

Higher-paying firms hire higher-skilled workers. Earnings bifurcation across employers has coincided with increasing segregation across different types of workers: highly-skilled workers are increasingly employed by firms that pay more, on average (Song et al., 2019). Much of this pattern reflects increasing occupation-firm alignment (Wilmers and Aeppli, 2021): high paid software engineers work at high-premium firms, while low-paid cooks work at low-premium firms. But could these patterns of sorting partly reflect human capital inputs conferred by colleges?

If higher-paying firms privilege worker productivity in hiring and recruitment decisions, the answer is yes. Pay differences across firms appear, in part, to be driven by organizational differences across firms (Rosenfeld, 2021), from formal structures like firm size and differences in production processes like R&D intensity (Barth et al., 2016), to management practices (Osterman, 2006; Bender et al., 2018; Bloom et al., 2021) and job-specific task allocations (Hackman and Oldham, 1976; Deming and Kahn, 2018; Wilmers, 2020; Oldham and Hackman, 2010; Ranganathan, 2023). These differences - which translate into pay and productivity differences (Kalleberg et al., 1981; Card et al., 2018) - may lead high-paying employers to compete for who they perceive to be the most productive and potentially high-performing workers via intensive skill assessment processes. Meanwhile, assuming that students across colleges have a preference for (and knowledge of) high-paying opportunities, graduates from colleges that bolster their human capital may perform better on employment tests, craft stronger applications, or interview more effectively. These colleges tend to be more selective, as the previous section outlined, though important differences in skill development may remain even among institutions without a clear rank ordering.

Just as important is the type of human capital cultivated across colleges. College major is a key form of specific human capital investment, with significant variation in earnings across students with different majors (Altonji et al., 2012; Andrews et al., 2022; Bleemer and Mehta, 2022). Colleges that specialize in fields directly valued by high-wage firms - such as engineering, finance, or computer science - could reinforce sorting into high-premium firms.

Credentialism and social closure

Human capital is imperfectly observed by employers: in contexts of imperfect information, managers may rely on signals such as college prestige or selectivity as a filtering criterion in the applicant screening process - even when candidates from different institutions are similarly productive (Thurow, 1975; Rivera, 2015; Neely, 2022). High-paying employers may be especially likely to do this, due to the premium they place on productivity: degrees from more selective institutions may be seen as evidence of differential college value-added (Deming et al., 2016), or as a proxy for underlying ability (Rivera, 2012). In particular, Rivera (2011) found that elite employers largely outsource screening of both hard and soft skills to college admission committees. Importantly, as Figure 1 demonstrates, credentialism channels are not mutually exclusive from human capital mechanisms. At an individual level, students who struggle with college-level coursework are less likely to complete a degree (Bound and Turner, 2011); in aggregate, degree completion itself may serve as evidence of acquired skills at college (Spence, 1973; Deming et al., 2016).

Assessments of skill and competence are important, but likely not the only, drivers of employer decisions: hiring decisions at profitable firms may additionally reflect efforts to protect existing advantages via social exclusion (Parkin, 2018; Rivera, 2020). More productive, higher-paying firms may be especially likely to practice such exclusion. When incumbents at resource-rich firms are themselves class privileged, managers may select culturally similar workers to facilitate social cohesion and reinforce rent-extraction processes (Fligstein and Fernandez, 1988; Burawoy and Wright, 1990; Weeden and Grusky, 2014; Tomaskovic-Devey and Avent-Holt, 2019). Over time, these processes generate credential-based segregation across firms, concentrating advantaged workers in resource-rich workplaces while excluding others.

Employer exclusion could also operate more diffusely through selective recruitment across campuses. Higher-paying employers - both at the very top of the firm pay distribution (Rivera, 2015; Davis and Binder, 2019) and more generally (Streib, 2023) - selectively recruit across campuses. Meanwhile, less prestigious employers sometimes avoid the highest-ranked universities altogether, judging such recruitment "not financially sensible" (Streib, 2023). These selective recruitment practices provide students with strong cues in the low-information context of the labor market transition (Deming and Kahn, 2018; Streib, 2023). These practices not only constrain information about available opportunities and access to application pipelines, but could also spill over into supply-side behavior: anticipating exclusion, graduates of less selective colleges may deliberately target mid-tier employers rather than risk rejection from elite firms (cf. Pager and Pedulla, 2015). In turn, these selective recruitment practices provide students with strong cues in the low-information context of the labor market transition (Deming and Kahn, 2018; Streib, 2023). Employer exclusion practices could also spill over into supply-side behavior - the targeting of non-elite firms at non-elite colleges - in anticipation of exclusion (cf. Pager and Pedulla, 2015).

Campus foundations of firm sorting

Although schools may share broad institutional features, individual campuses often develop their own distinctive cultures and social dynamics (Hamilton et al., 2018) and may themselves amplify patterns of firm sorting. Most directly, colleges differ in their informal and formal connections with employers (Rosenbaum et al., 2007; Davis and Binder, 2016). Some of these connections are mechanical, through distinct alumni networks resulting from credentialism and employer exclusions, but others may be more actively curated. Colleges with robust career services may actively procure partnerships with prestigious employers in select industries, such as the Corporate Partnership Programs (CPPs) documented by Davis and Binder (2016). While especially concentrated at large, research-oriented universities, CPPs are also found across a sizable number of less selective state universities (Davis and Binder, 2016). University career services offer powerful job search heuristics for students, who often source job leads through university databases and recruiting events (Manduca, 2024).

In addition to shaping informational environments, colleges influence labor market sorting through more diffuse, cultural channels. Scholars writing in the neo-institutionalist tradition emphasize how organizations shape individual identities and ambitions (Meyer, 1977; Granfield, 1992; Stevens et al., 2008; Stuber, 2011; Binder and Wood, 2012). Educational institutions are no exception: organizational differences across colleges can serve as powerful primers for student identity and career ambitions. For example, Binder et al. (2016) note that college freshmen at elite universities arrive on campus largely unaware of potential career paths, but an almost cult-like campus culture soon cements around the elite firms that advertise at campus events (see also Granfield, 1992). While partly driven by selective employer recruitment across campuses, identity shaping is also likely stoked by what Clark (1972) termed "organizational sagas": colleges vary in the density of their networks, the expectations they project about career success, and the localized status systems that define which jobs are seen as prestigious or appropriate for graduates (Binder et al., 2016). These processes may contribute to a broad prestige hierarchy, funneling students from more selective institutions into higher-premium firms. But, they likely also generate meaningful variation both within tiers of the college hierarchy - insofar as network ties, employer-facing channels, and institutional linkages to high- and low-premium firms develop unevenly across campuses even without clear national prestige. - and across the firm pay distribution - insofar as these processes play out among employers generally.

In light of the foregoing discussion, I propose three hypotheses regarding the role of firms in explaining college value-added.

> **Hypothesis 1:** *Colleges differ meaningfully in the extent to which they sort students into high- and low-premium firms, even net of selection into colleges, and differences in firm placement account for a substantial share of the overall earnings returns to colleges.*
>
> **Hypothesis 2:** *Attendance at a more selective college provides access to high-premium firms beyond those at the very top of the firm pay distribution.*
>
> **Hypothesis 3:** *Human capital can partly, but not fully, explain college effects on firm placement.*

Background and data

I analyze these questions in the context of one US state's public 4-year college sector. Per the terms of my data use agreement, I keep the identity of the state and institutions anonymous. I therefore refer to the setting as Summit State and use generic institutional labels. The set of seven 4-year universities in Summit State comprise a flagship research university, several branch campuses, a large state (R1) university, and multiple regional universities. The flagship is a large, highly selective public institution; College 4 is also highly ranked nationally, while the remaining institutions primarily serve regional markets. By focusing on a system where most students attend non-elite institutions but where meaningful differences in institutional prestige still exist, and therefore offer a clearer perspective on how firm sorting and wage disparities emerge among a set of colleges which serve a large portion of the college-bound population, outside the context of highly selective private colleges (Ho, 2009; Rivera, 2011, 2015; Binder et al., 2016).[3] Notably, these institutions serve distinct student populations,

[3]Table 10 in Appendix F.1 confirms that the public 4-year postsecondary sector in Summit State is a favored destination for postsecondary goers, and especially for students pursuing postsecondary education

reflecting the state's stratified pathways into higher education. Figure S21 shows that the state flagship (the Flagship) and, to a lesser degree, College 4, comprises a more selective, higher-income student body, while regional public universities enroll students from diverse socioeconomic and racial backgrounds. All the 4-year colleges have seen a steady increase in racial-ethnic diversity over time, driven primarily by increasing representation of Hispanic students in the postsecondary sector.

Linking college students to firms and pay presents a methodological challenge, since traditional survey data typically do not contain employer identifiers, and employer-employee datasets on which firm pay premiums can be established tend to lack granular information on educational credentials or pre-college background characteristics. To overcome this challenge, I perform a unique linkage of three administrative datasets covering recent high school graduates, postsecondary enrollees, and the near full population of employees in Summit State between 2005 and 2022.[4] These data contain two critical components: (1) detailed information on educational attainment and academic records for a large share of Summit state college-goers, and (2) longitudinal records on earnings and employer affiliations for a substantial portion of the state's workforce. Together, repeated observations on both individuals and firms enables a decomposition of earnings into worker and firm-specific pay effects and an adjudication of the role of firm sorting in explaining returns to different degrees.

The employer-employee data come from the state's Unemployment Insurance (UI) program, which requires nearly all employers to submit quarterly reports on employee wages and hours. These reports cover most private, local, and state public-sector workers and include consistent longitudinal employer identifiers, along with basic employer characteristics such as industry (NAICS code), location, and employment counts. Several groups - self-employed, federal employees, and active-duty military - are not covered. Education records come from the State Education Research Center (SERC). Postsecondary files include enrollment, course-level GPA, credits, degree attainment, institution attended, and financial aid filings. High school records provide GPA, school attended, and demographic characteristics (e.g., gender, race/ethnicity, disability status, and FRPL eligibility). Linkages across data sources are performed using Social Security Number (SSN) and full name.

I construct a school-year panel of 2006-2012 Summit State public high school graduates,[5] linked to their primary employer. Each worker-school-year observation spans fiscal quarter 3 through quarter 2 of the following year. My main outcome is annualized earnings, measured 10 years after high school graduation (with robustness checks at 15

---

in-state.

[4]For all analyses, I use wages observed up to and including the end of school year 2022, that is, up to financial year 2022:Q2.

[5]I define a graduating cohort in terms of the end of the academic school year - that is, the 2006 cohort refers to students graduating high school in 2006.

years where available), and adjusted to 2022 dollars using the CPI-U. Annual wages are calculated from observed quarterly earnings, and workers with multiple jobs are assigned to the employer where they received the largest share of their compensation, provided they are observed at their primary employer for at least two quarters in any given school year.[6]

I link these employer-employee data to postsecondary records from Summit State's public colleges (2005–2021). Postsecondary files provide information on institution attended, GPA, credits, degree attainment, and financial aid. To ensure that education is complete before analyzing wages, I assign to each individual their highest credential attained in the *prior* school year, prioritizing post-BA, BA, AA, and other degrees. If a student obtains multiple BA credentials in any given year, I attach to that student an in-state public degree if obtained, followed by a public or private 4-year out of state degree. If a student obtains multiple in-state BAs in a given year, I follow a rank ordering (flagship > other public universities > regional colleges) to capture potential institutional signaling effects. Students who pursue graduate education are retained in the sample, but their degree attainment is coded to their highest BA institution; in this way, graduate study is treated as endogenous to BA completion. I also use enrollment records to identify first enrollment institution, credits, and cumulative GPA.[7] Finally, I link individuals to high school transcript records, as well as to enrollment and degree certification records from private in-state colleges, as well as public and private out-of-state colleges via the National Student Clearinghouse.

My core sample includes all high school cohorts beginning their postsecondary studies at a 4-year institution in Summit State for whom I observe positive UI-covered earnings in the relevant period and for whom I am able to estimate an employer fixed effect (see details below). I restrict the sample to 4-year starters for two reasons. First, students who begin at 2-year institutions and later transfer to 4-year colleges differ systematically from 4-year entrants in both demographics and academic preparation (Bowen et al., 2009; Melguizo et al., 2011). Restricting treatment effects to 4-year starters reduces potential unobserved selection given differences across these groups. Second, transfer students typically attend a larger number institutions, complicating the identification of distinct campus effects. Given the salience of distinct campus environments as a potential mechanism for firm sorting, restricting the sample to 4-year

[6] I drop worker-firm observations with missing information on these attributes, and also assign a unique NAICS and location code to each employer per year. I also exclude from my analyses the small number of students working in firms operating in the Mining, Quarrying, and Oil and Gas Extraction industry (NAICS 21).

[7] I define an individual's first enrollment institution to be that at which they first enrolled in college following the semester in which they graduated high school. Following Mountjoy and Hickman (2021), I omit enrollments occurring over the summer. For students enrolled in more than one institution in their first contact with the postsecondary system, I assign their institution to be the institution at which they attempted most credit hours.

starters offers a more focused analysis.[8]

Table 1 presents descriptive statistics for this core sample ($N = 48,769$). Among 4-year starters, there is meaningful heterogeneity by institution attended. For example, freshman admits at the State Flagship have the highest final high school GPA (3.69), while those at Colleges 7 and 6 have the lowest (3.21 and 3.17, respectively). The proportion of freshman admits who were ever on free and reduced price lunches also differs considerably across colleges, with the lowest rates at Colleges 4 and 5, as well as the flagship, and the highest at College 7. Among 4-year college entrants, there is notable variation in labor market outcomes by institution. Graduates of the flagship are the most likely to work in large firms (60%), in the Primary Metropolitan Area (62%), and earn the highest wages (log wage of 3.70). In contrast, students from regional colleges such as College 7 and College 6 are more likely to work in smaller firms located outside the Primary Metropolitan Area, and have lower average earnings and wage rates. There are also important differences in the likelihood that earnings are observed in administrative data: the probability of observed earnings is highest for Colleges 2, 3 and 6, and is lowest for the state flagship and College 4. Because I only observe earnings for individuals covered by the Summit State UI system (that is, for individuals who are employed in-state), treatment effects should be interpreted as local to the population of in-state, UI-covered workers. If certain colleges (e.g., selective 4-year institutions) are more likely to place graduates into jobs out of state, observed earnings may understate the full returns to those institutions. It is also difficult to ascertain whether out-of-state employment reflects employment at better-premium employers, or in better jobs at similar firms.

[8] In order to ensure comparability across analyses that require non-missing industry, firm size, and firm location information, I define an individual to have observed earnings only if the individual has non-missing information on these employer attributes. In practice, this reduces the sample trivially, by just over 200 individuals.

Table 1: Mean Values of Individual Characteristics, Academic Metrics, and Labor Market Outcomes by College Attainment Status by First College Attendance

			College of First Attendance						
	Non 4-Year	4-Year	College 7	College 6	College 5	College 4	College 3	College 2	Flagship
Female	0.49	0.55	0.59	0.52	0.59	0.51	0.49	0.57	0.55
Black	0.05	0.03	0.07	0.03	0.02	0.03	0.05	0.06	0.03
Hispanic	0.13	0.07	0.11	0.09	0.04	0.07	0.09	0.08	0.06
White	0.70	0.72	0.76	0.81	0.82	0.82	0.47	0.55	0.57
English Speaking	0.86	0.88	0.90	0.93	0.94	0.93	0.74	0.81	0.79
Final HS GPA	2.72	3.43	3.21	3.17	3.46	3.37	3.27	3.25	3.69
Ever Homeless	0.02	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00
Ever FRPL	0.41	0.22	0.34	0.23	0.13	0.18	0.30	0.32	0.23
Ever SPED	0.11	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01
Ever ELL	0.05	0.02	0.02	0.02	0.01	0.01	0.06	0.04	0.03
Firm in PMA	0.37	0.49	0.25	0.43	0.51	0.45	0.60	0.57	0.62
Firm is Multi-Est.	0.33	0.30	0.32	0.28	0.28	0.29	0.33	0.37	0.31
Firm Size > 1000	0.47	0.55	0.54	0.52	0.53	0.54	0.58	0.58	0.60
Firm Size > 10,000	0.16	0.21	0.16	0.15	0.18	0.19	0.24	0.24	0.30
Log Earnings	10.62	10.96	10.74	10.82	10.87	10.98	10.99	10.87	11.15
Log Wage	3.26	3.54	3.36	3.42	3.48	3.54	3.54	3.46	3.70
Pr(Earnings >0)	0.51	0.58	0.61	0.62	0.59	0.58	0.62	0.62	0.56
N (Individuals)	177,394	48,769	5,416	5,864	9,037	11,838	1,244	936	14,434
N (Firms)	30,877	12,224	2,676	3,063	3,913	4,937	778	640	4,559
Pr(Estimated FE)	0.91	0.91	0.92	0.92	0.91	0.90	0.90	0.91	0.90

Notes: “PMA” refers to the state’s Primary Metropolitan Area. 4-Year refers to high school graduates who began their postsecondary education at an in-state public, 4 year college. Non 4-Year refers to everyone else. "Pr (Estimated FE)" reflects the proportion of firms that appear in the largest connected component of the worker-firm network used in the employer fixed effects estimation, and comprise the firms for which I am able to estimated a firm effect.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Methodology

My analysis proceeds in two steps. The first step exploits the longitudinal nature of my linked employer-employee earnings data - specifically, fact that I observe workers as they move across different employers over time. I leverage these longitudinal earnings data and tools from labor economics [an AKM model (Abowd et al., 1999)] to decompose individuals' total earnings into a fixed component ("person effects"), as well as a component deriving from firm of employment ("firm effects"). The estimated firm effects give me a consistent and continuous measure of firms' pay premiums, which I can then use an outcome variable to tease out the firm component of college earnings returns. I treat person effects primarily as controls to help estimate the firm effects, though I additionally use them as a proxy for human capital in some analyses that follow.

Earnings decomposition

In the first step, I assume that the (log) earnings paid to worker i in year t is generated by the following model:

$$\ln y_{it} = \alpha_i + Z_{it}^{\top}\beta + \psi_{F(i,t)} + \epsilon_{it}, \tag{1}$$

where α_i is an individual fixed effect, Z_{it} is a vector of time-varying controls (including year fixed effects), $F(i,t)$ is an index function that maps worker i to their employer in year t, $\psi_{F(i,t)}$ is the firm-specific wage premium, and ϵ_{it} is an error term that captures idiosyncratic shocks. The $\psi_{F(i,t)}$ terms are my focal parameters. Firm effects can be interpreted as capturing the extent to which employer j pays workers a wage premium to their employees, in excess of the wage we would expect on the basis of their own characteristics alone.[9] Importantly, the firm effects in Equation (1) are only identifiable within "connected sets" of workers that are linked by worker mobility across firms (Abowd et al., 2002). A connected set is a network of firms linked together by observed worker mobility: two firms belong to the same set if there is a chain of workers moving between them over time. Because fixed effects are only identified relative to one another, estimates from separate sets cannot be compared on the same scale. I therefore restrict estimation of Equation (1) to the "largest connected set", which yields a single, internally consistent scale of firm effects. Appendix D provides further details.

I then apply these estimated firm effects to my sample of 2006-2012 Summit State public high school graduates. Specifically, for each individual, I attach the estimated

[9]Note that in all analyses that follow, I normalize firm fixed effects with respect to the average firm effect in the retail sector.

firm effect corresponding to the firm at which that individual was employed 10 years following high school graduation. Table 1 shows that just over 90% of 4-year college attendees are matched to an employer fixed effect (that is, 90% of these attendees work at firms in the largest connected set) and that this match rate is consistent across attendees at different colleges. I use this sample for my main empirical analyses estimate causal effects of college attendance via and net of firms

I use this model to estimate firm effects rather than an approach which jointly estimates firm and college effects (such as with traditional mediation analysis - see Table 24 in Appendix E), because it adjusts for patterns of selection into firms that are unobserved by the researcher. If workers differ in time invariant ways in their productivity in ways not captured by pre-college control variables, the person effect (α_i) parameters absorb this variation. The model does, however, assume away any time-varying unobserved confounders that may predict both firm selection and wages.

Estimating the role of firm sorting

In the second step, I use the firm effect attached to each individual to parse out the role of employer sorting in educational returns. Let $\mathcal{J} = \{\text{College 7}, \ldots, \text{College 2}, \text{Flagship}\}$ denote the set of 4-year public colleges in Summit State; Y denote an individual's earnings 10 years following high school graduation, and let M denote an individual's firm, also at 10 years following high school graduation. I use potential outcomes notation to refer to counterfactual quantities under receipt of a BA from different colleges. Specifically (suppressing subscripts i), I use $Y(j)$ to denote an individual's potential earnings were they to attend college $j \in \mathcal{J}$ (possibly contrary to fact). I assess the value-added of a attendance at a particular institution relative to attendance at College 7, the college with the lowest raw earnings return in the state. The average treatment effect of attending college j relative to attending College 7 can be written as:

$$\text{ATE}^Y(j) = \mathbb{E}[Y(j) - Y(\text{College 7})]. \tag{2}$$

To assess the role of firm sorting in mediating the earnings return to a credential from school j (Hypothesis 1), I employ a causal mediation framework. To do so, I make use of the "nested" counterfactual $Y(\text{College 7}, M(j))$. Intuitively, this quantity captures an individual's potential earnings, were they to attend College 7, but worked at the firm that attendance at school j would have directed them to. To assess mediation via firms, we can decompose the ATE of attending school j relative to College 7 as

$$\text{ATE}^Y(j) = \underbrace{\mathbb{E}[Y\big(j, M(j)\big) - Y\big(\text{College 7}, M(j)\big)]}_{\text{Explained by earnings gains net of firm sorting}}$$

$$+ \underbrace{\mathbb{E}[Y\big(\text{College 7}, M(j)\big) - Y\big(\text{College 7}, M(\text{College 7})\big)]}_{\text{Explained by between-firm sorting}}.$$

To identify the quantity $\mathbb{E}[Y\big(\text{College 7}, M(j)\big)$, I exploit the fact that the earnings decomposition in Equation (1) implies that the value-added to working at a given firm is constant across employees. Under this structural model for earnings, the portion of the total effect of attending college j versus College 7 driven by firm sorting is equal to

$$\text{ATE}^{\psi}(j) = \mathbb{E}[\psi(j) - \psi(\text{College 7})], \tag{3}$$

that is, the average treatment effect of attending college j on a worker's firm premium (see Appendix A). Here, $\psi(j)$ denotes the potential firm premium of student i if they attended college $j \in \mathcal{J}$, corresponding to an individual's employer 10 years after high school graduation.[10]Importantly, the quantity in Equation (3) also captures the the extent to which college j shapes access to high-premium firms, directly addressing Hypothesis 1). Under the assumption of no unobserved confounding for the effect of college attendance on within- and between-firm earnings, Equations (2) and (3) can be identified via the standard identification formula for the average treatment effect (see Appendix B). To make this assumption more plausible, I leverage the administrative registries to compile a large battery of pre-college and college-level covariates. These include basic demographic variables (gender, race, age, high school test scores), pre-college characteristics such as FRPL, English language ability, and disability status, as well as high school fixed effects. In some analyses, I estimate cumulative effects of college attendance and completion on earnings and firm premiums; for these analyses, I use information on first term credit attainment and credit attempts, first term college GPA, and cumulative institution GPA per student at the time of graduation or, if failing to graduate, of dropout. Due to the high-dimensional nature of my control variables, I estimate all quantities using ridge regressions and a double machine learning approach for valid inference. All quantities are estimated for the population of high school graduates who are begin their postsecondary studies at a 4-year college (high school graduating classes 2006-2012). Appendix B provides further details.

[10] I assume that the employer fixed effects estimated via Equation 1 are measured with negligible error, given the full-population nature of the UI data. Uncertainty estimates then come from uncertainty in estimation of average treatment effects. Kline (2024) further shows that downward bias resulting from failing to account for correlation between estimated firm effects can outweigh the upward bias attributable to treating firm effects as random draws from a super population.

Results

I first document observed gaps in both earnings (teal bars) and firm premiums (orange bars) across 4-year college enrollees. Figure 2, top panel, plots mean log earnings gaps between students who first enrolled in each 4-year public college, relative to students first enrolling at College 7, 10 years out from high school graduation. Point estimates are shown in Table 16 in Appendix H. There is substantial earnings variation by institution attended. Students who enroll at College 7 have the lowest average earnings (10.738 log points) and lowest mean firm premiums. Students who enroll at the flagship stand out, earning on average .411 log points more than College 7 attendees, and obtaining a firm with a pay premium that is .131 points higher than that attained by College 7 attendees. This is followed by Colleges 2 and 4, with earnings premiums of .258 and .246, and firm premiums of .125 and .073, respectively, and then by Colleges 5 and 6. These raw log earnings gaps are sizable, especially at such an early stage following labor market entry, and in fact exceed the .34 log earnings gap between 4-year starters and the non-4-year starting population (Table 1).

Do firms mediate the college premium?

Raw differences in earnings and firm premiums across different college attendees may be the result of selection into colleges and the effects of college attendance itself. To illustrate the strength of these differences even among observationally similar students, Figure 3 presents a descriptive view of firm placement by pre-college achievement. Even among students with comparable high school GPAs, institution attended is a strong predictor of the type of firm at which a person works. Flagship and College 2 attendees are employed at markedly higher-premium firms than peers from other colleges, especially those who attended Colleges 5, 6 and 7. Institutional differences remain pronounced across the entire GPA distribution: the hierarchy in firm placement is largely preserved even among high-achieving students. For example, students attending the Flagship rather than College 7 with a high school GPA between 2.8 and 3.0 work at firms paying an average premium .11 log points higher (corresponding to a 12% earnings increase). Among students with GPAs between 3.8 and 4.0, this gap rises to .15 log points (implying roughly a 16% earnings increase).

This simple descriptive comparison underscores how sharply firm sorting tracks institutional affiliation, rather than pre-college academic preparation. To more formally adjust for selection on a broader range of pre-college characteristics, Figure 2, bottom panel, plots the treatment effects of attending different Summit State colleges relative to College 7 attendance (teal bars), using the large set of control variables described previously. Table 17 in Appendix H presents corresponding point estimates. Raw earnings gaps depreciate significantly after adjusting for selection into college (teal bars). Relative to a baseline of 10.91 log earnings - which captures a person's earnings if they counterfactually attended College 7 - attending the state's flagship campus causes an average increase in earnings by .134 log points (14%).[11] Beyond the flagship, attending College 2 or College 4 corresponds to smaller but still sizable earnings gains, at .06 and .085 log points, respectively. Adjusting for selection into colleges further alters the rank ordering of earnings returns across colleges. Colleges 5 and 6 have raw earnings premiums relative to College 7, but this does not reflect a causal relationship: College 6 attendance provides almost no earnings return, relative to College 7, while attendance at College 5 provides a sizable earnings loss of approximately 5%. These is driven by differential selection across colleges. College 7 attendees have significantly lower high school GPAs, and comprise a large Black and Hispanic student population. By contrast, College 5 serves a predominantly White and academically strong student body; the average high school GPA of freshman enrollees stands only behind the Flagship (Table 1). These findings echo those in Ciocca-Eller (2023), who found that colleges' rankings in terms of observed BA completion rates do not necessarily map onto the causal effects of each college on BA completion.

The orange bars in Figure 2 show the effect of attending college j relative to College 7 on expected firm premiums (Equation 2). Attending the state flagship yields the largest firm premium, amounting to a 7% increase in pay relative to College 7 attendance. This is followed closely by College 2 attendance, and to a lesser extent, Colleges 3 and 4. By contrast, College 5 and 6 attendees work at firms with lower estimated pay premia than those who attended College 7. While college attendance effects on firm premiums are smaller than unadjusted raw gaps in firm premiums across colleges, the degree of attenuation is far less pronounced than the attenuation of total overall earnings gaps across colleges. For example, while the flagship log earnings gap (relative to College 7) falls by .277 log points after adjusting for selection, the flagship firm premium gap (relative to College 7) falls by only .73 log points. This implies that characteristics that predict college attendance are more strongly associated with within-firm earnings gains, rather than with earnings gains which result from cross-firm sorting. These patterns suggest that colleges meaningfully affect firm assignment, supporting Hypothesis 1. While Flagship or College 2 attendance sorts students into the highest-paying employers, there is still meaningful variation in firm placement among the regional colleges, with College 7 in fact conferring a positive firm advantage relative

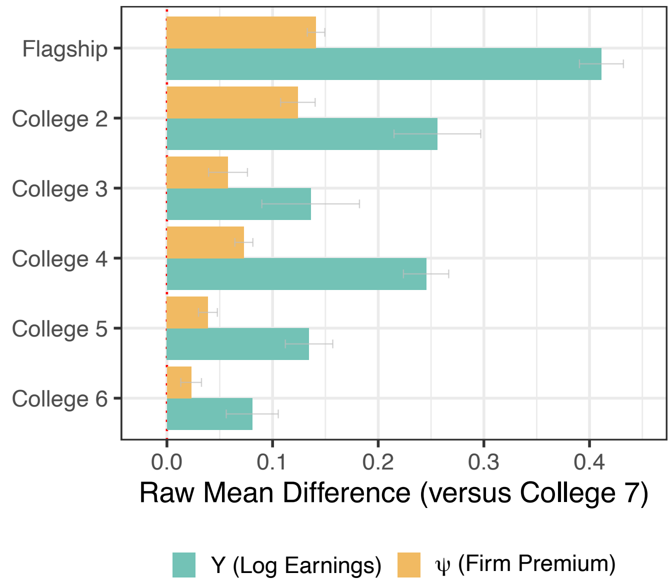
[11] Qualitatively, my estimate of flagship attendance on earnings is broadly similar to, albeit somewhat smaller than, to those presented by Hoekstra (2009), who uses a regression discontinuity design to estimate the economic return to attending the flagship relative to students' fallback options as between 19% and 24%. That my estimates are smaller could reflect either cross-college variation in institutional impact, or larger treatment effects among the set of compliers whose admission or enrollment decisions were determined by the admission cutoff studied

to College 5 and College 6.

How significant are these firm sorting patterns for college returns overall? Under the structural model for earnings described in Equation 1, the firm premium returns (orange bars) in Figure 2, bottom panel, represent the fraction of college premiums explained by firms (Equation 3). Firms account for a full half of the 14% earnings return to Flagship attendance (relative to College 7), and almost entirely account for the earnings advantages of attending Colleges 2 and 3, and negative effects of College 6. By contrast, within-firm differences appear to matter more for Colleges 4 and 5: firm sorting accounts for approximately one-third and one-fifth, respectively, of these colleges' total earnings return relative to College 7. Importantly, the range of counterfactual earnings differences under attendance at different colleges would fall dramatically if firm sorting channels did not exist. The range of total potential earnings between the Flagship and College 5 (the college with the lowest counterfactual earnings returns) is .18 (.134+.046). Absent firm sorting, this would abate to 0.101 (.066+.035), implying a reduction of 56%. Taken together, these results support Hypothesis 1 - that firm placement contributes meaningfully to overall earnings returns to different colleges.

Figure 2: Raw Gaps in and Treatment Effects on Log Earnings and Firm Premiums Among Summit State 4-Year Starters

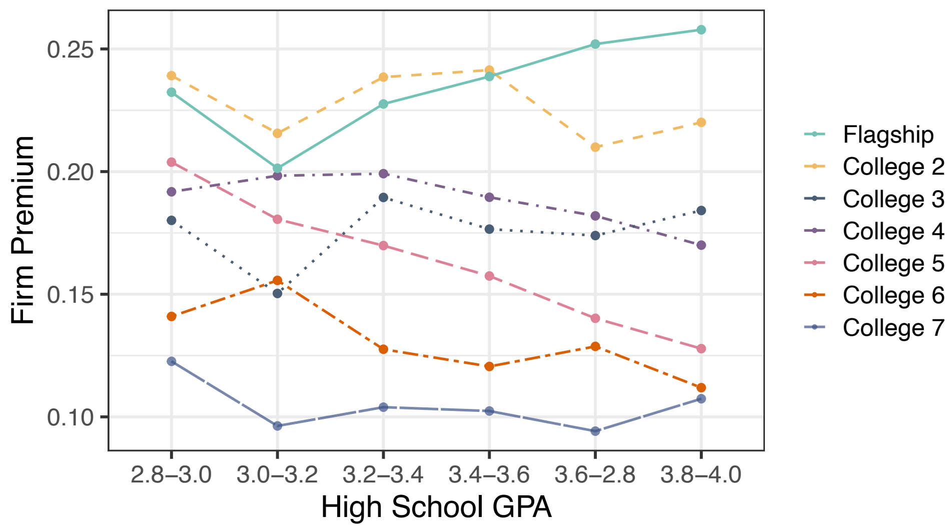
(a) Raw gaps in log earnings (teal bars) and firm premiums (orange bars)



(b) Treatment effects on total log earnings (teal bars) and via firm sorting (orange bars)

Source: Summit State administrative UI earnings, postsecondary and high school records.

Figure 3: College Effects on Firm Premium by High School GPA Buckets



Source: Summit State administrative UI earnings, postsecondary and high school records.

How pervasive is firm sorting?

Attending the state flagship sorts students into firms that, on average, pay more. But to what extent is driven by access to the highest-paying, elite firms?

To answer this question, I re-estimate treatment effects on firm attainment defined in terms of its percentile rank in the overall firm premium distribution. Specifically, I define a set of binary outcomes indicating whether a student works at a firm at or above the 60th, 70th, 80th, 90th, or 95th percentiles of the firm fixed effect distribution. I re–estimate Equation (3), replacing the ψ terms with a set of indicator variables, equal to one if an individual works at a firm at or above percentile p, for $p \in 60, 70, 80, 90, 95$.[12] This percentile-based approach provides insight into whether college effects are concentrated at the top of the firm hierarchy, or whether more selective colleges facilitate access to better-premium firms more broadly throughout the distribution.

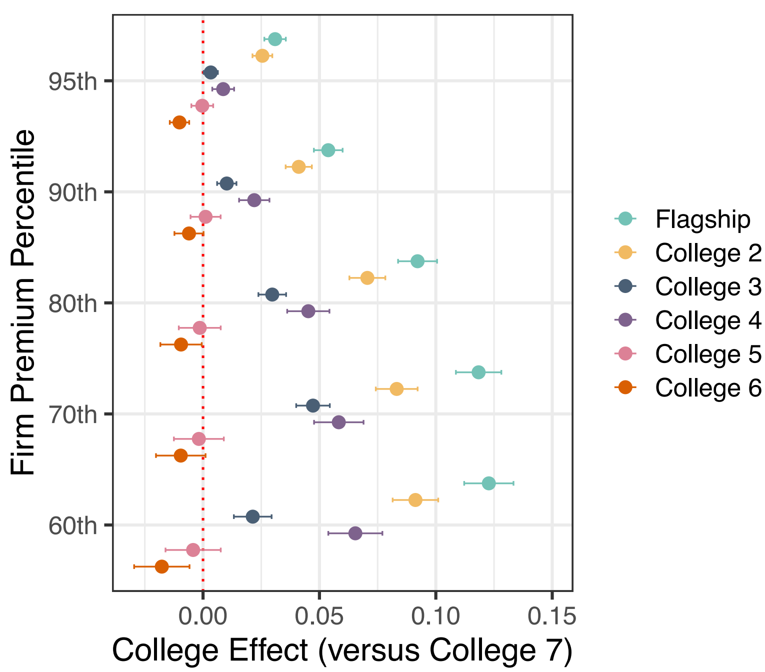
Figure 4 shows these estimates, with each point capturing the effect of attending a given college (versus College 7) on the probability of working at a firm at or above

[12] I define percentiles of pay premiums with respect to the population of 4-year college attendees, rather the full sample of high school graduates, to capture sorting effects for the types of firms college graduates are likely to work at. There is considerable employer segregation between college goers and graduates and those without postsecondary experience.

the indicated percentile. The full set of point estimates is provided in Table 18 in Appendix H. Attending the Flagship increases the likelihood of working at a firm in the top 95th percentile of the pay distribution by 3.1 percentage points, but increases the probability of working at a firm at or above the 90th percentile by 5.4 percentage points, the 80th percentile by 9.2 percentage points, and the 70th and 60th percentiles by 11.8 and 12.3 percentage points, respectively, relative to College 7. Similar patterns hold for College 2 and, to a lesser extent, Colleges 3 and 4. In contrast, Colleges 5 and 6 show no meaningful firm sorting advantages across the distribution, with College 6 estimates even negative at lower percentiles.

Firm sorting, then, is not limited to access to only the most elite firms (Hypothesis 2). Rather, more selective colleges such as the Flagship and College 4 appear to systematically expand access to higher-premium employers across the broader distribution of firms. This points to a more diffuse but pervasive form of labor market stratification, where the benefits of attending a more selective college are reflected throughout the upper half of the firm pay structure.

Figure 4: College Attendance Effects on Probability of Reaching a Firm with a Premium $\geq$ pth Percentile



Source: Summit State administrative UI earnings, postsecondary and high school records.

Human capital mechanisms

The last sections demonstrated that firms are an important conduit of educational premiums, but did not unpack why. Here, I assess one possible explanation: that human capital gains at certain colleges facilitate access to higher-paying firms.

I measure human capital in three ways. First, using college GPA at time of graduation or at dropout, for non-completers. Second, using estimated "person effects" in Equation 1, following recent approaches to measuring skill-based matching between workers and firms (Gerard et al., 2021). Third, using field of study. Each of these approaches has its drawbacks. In particular, GPA is not standardized across colleges; person effects can only be estimated among the subset of workers who switch firms and reflect a composite of human capital and a variety of other factors, and neither capture how time-varying human capital may unfold for graduates of different colleges. Still, in aggregate this exercise helps assess the utility of human capital as a mechanism for the firm sorting patterns I document.

To parse out non-human capital channels, I estimate the amount of treatment effect that operates net of human capital channels. Letting HC be the measures of human capital listed above, and ψ be the estimated firm effects from Equation 1, I decompose the firm premium to attending College j relative to College 7 ($\text{ATE}^{\psi}(j)$) as

$$\begin{aligned}\text{ATE}^{\psi}(j) = & \underbrace{\frac{1}{2}\sum_{r\in\{j,\text{College 7}\}} \mathbb{E}[Y\big(j, HC(r)\big) - Y\big(\text{College 7}, HC(r)\big)]}_{\text{Explained by non-HC factors}} \\ & + \underbrace{\frac{1}{2}\sum_{r\in\{j,\text{College 7}\}} \mathbb{E}[Y\big(r, HC(j)\big) - Y\big(r, HC(\text{College 7})\big)]}_{\text{Explained by HC-based sorting}}. \end{aligned} \tag{4}$$

I estimate these quantities using the mediation estimator described in Appendix B. Figure 5 displays estimates of college firm premium effects, juxtaposed against the portion that persists net of each of the three human capital measures. For all but one college (College 5), neither person effects nor college GPA explain their firm sorting advantages relative to College 7. The reason for this is explored in Figure 6, which shows the partial effect of these two human capital measures on firm premiums (top two rows) and of college GPA on earnings (bottom row).[13] College GPA and person effects have only very small effects on firm premiums. An increase of a person effect by 1 log point corresponds to an firm premium gain of .05. To make sense of this effect

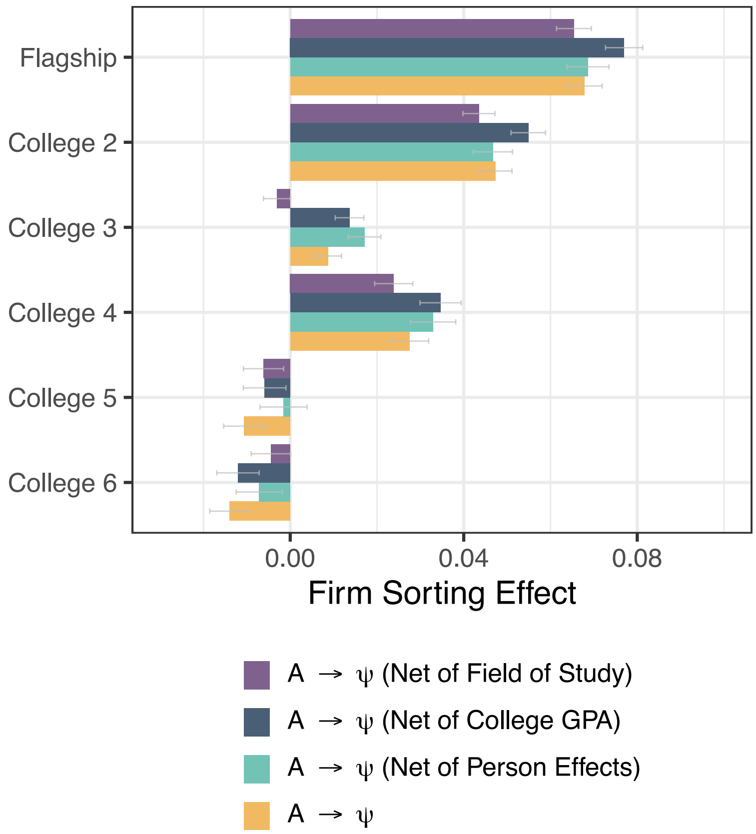
[13]These are all estimated net of the same covariates used for the main analyses, in addition to first college attended.

size, an increase of 1 log point of person effects would be equivalent to shifting from the 10th to 93rd percentile in the person effects distribution. The effect of college GPA on firm sorting advantages is similarly small: increasing college GPA by a full point would yield a firm premium advantage of .03. To put this effect size into perspective, the bottom row of Figure 5 shows that increasing college GPA by one point would yield an effect on overall earnings that is 7 times as large the effect on firm premium.

Nor do field-of-study differences across colleges explain firm sorting advantages. Summit State's colleges do vary markedly in their curricular profiles, with the Flagship school emphasizing STEM and engineering, leading to professions that tend to be clustered in high-premium firms (Wilmers and Aeppli, 2021), and the regional colleges emphasizing humanities, education and professional fields (see Table 14). These curricular differences shape students' fields of study, which in turn could structure the set of firms they might enter, accounting for some of the observed firm premiums. That they do not reflects the fact that colleges convey firm sorting advantages within every field of study (see Figure S12). The exceptions are for Colleges 3 and 6, where field of study is an important contributor to firm sorting advantages relative to College 7. While this tentatively suggests that, at less selective colleges, sorting across firms stems from curricular differences rather than the interface between employer and college, some caution should be attached to this interpretation, because a student's chosen field of study could be, in part, endogenous to employer behaviors such as campus recruitment.

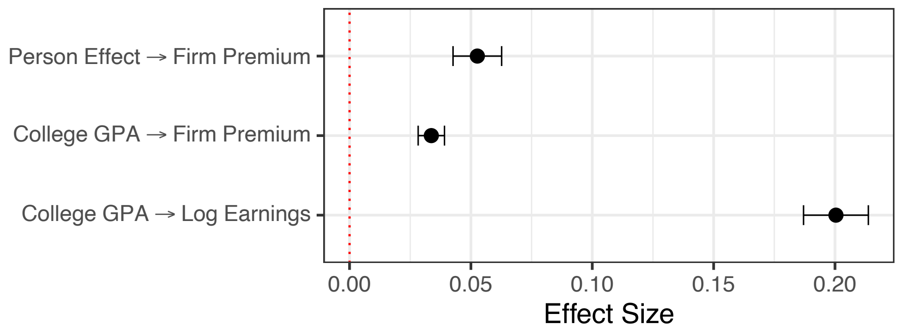
Collectively, these analyses suggest that human capital is not a primary - or even substantial driver - of the firm-sorting advantages across colleges that I have documented, partly supporting Hypothesis 3. This is not to say that human capital is not an important driver of college returns overall. In particular, of the 50% of the Flagship premium unexplained by firm sorting, human capital is likely a critical component. Rather, it suggests that the other half on the Flagship premium is driven primarily by non-skill-based resource advantages conferred by the school.

Figure 5: Firm Sorting Effects, Isolated from Human Capital Components



Source: Summit State administrative UI earnings, postsecondary and high school records.

Figure 6: Partial Effects of Human Capital Measures on Firm Premium and Earnings



Source: Summit State administrative UI earnings, postsecondary and high school records.

Supplementary analyses

I test the robustness of my findings through several additional analyses. All results are shown in Appendix E.

First, I analyze whether firm sorting effects vary across race and class background. Absent a measure of family income for all students, I measure class background in two ways: whether a student received free or reduced price lunches in high school, and the family income of students who received financial aid. Across colleges, there is little evidence of heterogeneity in overall effects or in the firm sorting advantages conveyed by the college. Similarly, the results for race suggest that college effects on firm placement are comparable across racial groups.

Second, I test whether geographic and industry sorting can account for the observed college-firm premium differences. Figure S2 in Appendix D displays average firm premiums across Summit State counties and the locations of each college. Notably, the state flagship is in a county (which I call the "Primary Metropolitan Area") home to some of the highest paying firms in the state. Yet, Figure S8 shows that firm placement advantages across colleges do not simply reflect geographic sorting. Flagship attendees obtain access to substantially higher-paying employers, even compared with other college attendees who find work in the same area. Nor do firm-sorting advantages simply stem from industry-based channeling (Figure S10).

Third, I examine the cumulative or joint effects of attendance *and* completion at a given college. While my main analyses follow prior definitions of college attendance as a treatment (e.g. Hoekstra, 2009; Mountjoy and Hickman, 2021), this approach conflates attendance effects the effects of completing a degree - possibly at other campuses.[14]

[14]Table 14 in Appendix F.2 shows that out-transfer is indeed a common experience: of the 90% of the

The results from this exercise (shown in Figure S6) are, nevertheless, consistent with my main results.

Finally, I examine college effects on earnings, and their mediation by firms, fifteen years after high school for the subset of cohorts for which I can observe this horizon. Table 14 shows that firms remain a central component of college wage premiums, even as the precise pattern of earnings premiums changes (though some effects are imprecisely estimated). Notably, the explanatory power of firms for Flagship attendance abates by 20 percent - a decline not driven by erosion of firm sorting advantages, but by the within-firm earnings advantages conferred by Flagship attendance. This shift could reflect several processes, including higher rates of promotion and opportunities for on-the-job skill training among Flagship attendees (cf. Tomaskovic-Devey et al., 2005), or continual human capital investment via graduate education.[15] Graduate education has large labor market returns (Altonji et al., 2016; Altonji and Zhong, 2021), but may be associated more with within-firm earnings advancement than from between-firm sorting.

Discussion

Colleges do not simply teach; they also act as powerful sorting devices (Kerckhoff, 1976; Domina et al., 2017). While it has long been recognized that the relationship between education and earnings may vary across "occupations, communities...and labor markets" (Bills, 2003, p.457) this variation is often abstracted from employers themselves. In this paper, I have argued that a consequential, yet under-appreciated, dimension of this sorting is across workplaces.

Drawing on a novel linked employer-employee dataset for high school graduating classes in a US State, I find that firm sorting explains a substantial share of earnings returns across institutions. Half of the earnings premium associated with attending Summit State's flagship public university, relative to a lower-ranked regional university, is explained by differences in access to higher-premium firms. Crucially, these patterns persist even after accounting for detailed controls for students' backgrounds and prior academic performance. Firm sorting is also not confined to access to a handful of elite employers. Rather, it is pervasive across the distribution of firms: selective colleges channel graduates into better-premium firms throughout the wage structure, not just

Flagship starters earning a BA degree 10 years following high school graduation, 92% completed a degree at the Flagship; by contrast, of the 62% of College 7 starters earning a BA degree 10 years following high school graduation, only 79% completed a degree at College 7.

[15]Table 1 shows that, ten years after high school, 17 percent of the Flagship attendees had completed a graduate degree compared with 9 percent of College 7 attendees; fifteen years out, these shares had risen to 22 and 13 percent, respectively.

into the highest resourced organizations. Further, human capital does not explain the firm sorting advantages conferred by colleges.

These findings carry important theoretical and policy implications. Theoretically, they suggest that models of education and inequality must place greater emphasis on employer behavior and organizational sorting. Existing work on school-to-work transitions stresses the role of individual-level skill formation and utilization (Witte and Kalleberg, 1995; Robst, 2007; DiPrete et al., 2017; Bol et al., 2019; Quadlin et al., 2021; Cassidy and Gaulke, 2024), credentialism and social networks. My results show that these processes translate into higher earnings largely when they secure access to higher-paying firms. Moving beyond an exclusive focus on individual-level skills, or on models of social exclusion that abstract away from workplaces, requires recognizing the ways in which educational institutions and employer practices are jointly implicated in the production of inequality.

From a policy perspective, the results suggest that efforts to reduce postsecondary wage gaps are likely to be unsuccessful via a sole focus on increasing degree completion or improving instructional quality. I show that a significant portion of the returns to colleges appears to stem from factors that are distinctly non-skill-based. While ensuring that all colleges impart students with skills for the workforce is surely an important goal, such an approach will not fully address earnings disparities if graduates of different institutions continue to face unequal access to higher-paying firms. Colleges thus have a role to play beyond the classroom: strengthening career services, alumni networks and employer partnerships can help ensure that skills students acquire translate into better labor market outcomes. Expanding connections between regional colleges and high-wage employers, investing in career services infrastructure that targets a broader range of firms may materially change the labor market trajectories of their graduates. At the same time, structural constraints - including employers' entrenched preferences for certain credentials and firm-based practices of social closure - limit what colleges can achieve without broader labor market reforms. Narrowing wage gaps will require not only strengthening college inputs, but also encouraging high-paying employers to broaden their hiring pipelines.

There are several important limitations of my analyses. While the data used provide rare granularity linking students to their educational and employment histories, they are limited to one state's public college system and early career outcomes. Long-term effects may differ, especially as firm-specific networks and career ladders unfold over time. To the extent that between- and within-earnings trajectories evolve differently for graduates of different colleges, my analyses could mischaracterize the long-run effects of college and the relative importance of firm sorting.[16] Nevertheless, early outcomes

[16]Preliminary analyses in this direction suggest some waning of the importance of firm sorting 15 years after high school graduation (Figure (S11), Appendix (E)), driven by an increased within-firm returns to flagship

may be important: being hired at a higher-premium firm as a labor-market entrant is an important predictor of long-term earnings (Arellano-Bover, 2024). Establishing how common upward firm mobility is across different groups, and its importance for the evolution of earnings inequality across the life-course, remain important and open questions for future research

My estimated earnings returns also pertain to the intensive margin of earnings, and exclude a large number individuals who obtain out-of-state employment. To the extent that out-of-state employment is concentrated in better-premium opportunities, and that such employment is more common among more selective college graduates, these patterns could downwardly bias my estimates of firm-based mediation. However, it is difficult to definitively ascertain the size, or direction, of the bias. Moreover, while I used an extensive set of controls to adjust for selection into colleges, unobserved confounders, such as family social capital or informal job search networks, may still bias estimates to some extent. Finally, my data do not enable me to parse the role of credential advantages versus campus environments in facilitating access to high-paying firms. Distinguishing precisely what matters most for firm sorting - from broader cultural logics, to more direct network effects such as referrals and resources by campus career centers - is an important next step.

Still, the broader implications are clear: firm sorting is a major, under-recognized contributor to the stratified college wage premium. As labor markets grow more unequal and as returns to firm affiliation continue to rise, the stakes of credential-based sorting are only likely to deepen. In the short run, firm assignment shapes wage gaps. In the long run, it likely shapes career trajectories, wealth accumulation, geographic mobility, and even health outcomes. Early access to higher-premium firms compounds advantages over time, while exclusion from such firms constrains upward mobility for graduates of less selective colleges. If policymakers and educators are serious about using higher education as a tool for reducing inequality, they must attend not only to the supply side of skills as taught in colleges, nor solely to the demand side of employer hiring - but to the organizational interface between schools and firms. It is in the everyday logics, practices, and pipelines that link campuses to employers where much of inequality in opportunity is forged. Recognizing this interface - and acting to reshape it - may be a powerful lever to reduce inequalities in the college-to-work transition.

attendance over time, rather than by waning firm sorting patterns over time. Parsing out to what extent this pattern results within-firm progression on the basis of educational credentials per se, as opposed to graduate education returns, is an important direction for future work.

A Proof of Mediation Under Additive Separability

In this section, I formally show that, under an additive separability assumption, the mediation of the average treatment effect (ATE) of college on earnings is equivalent to the ATE of college on firm pay premiums. As in the main text, let $A_i = j$ indicate that individual i attended college j; let X_i denote pre-college covariates; let M_i denote the firm at which individual i is employed ten years after high school graduation; and let Y_i denote log earnings. Assume the following nonparametric structural equation model (NPSEM) with independent errors governs the joint distribution of (X_i, A_i, L_i, M_i):

$$\begin{aligned}
X_i &= f_X(\epsilon_{X,i}) \\
A_i &= f_A(X_i, \epsilon_{A,i}) \\
L_i &= f_A(X_i, A_i, \epsilon_{Z,i}) \\
M_i &= f_M(X_i, A_i, L_i, \epsilon_{M,i})
\end{aligned}$$

Additionally, assume that earnings for worker i in year t are generated according to the following model:

$$\ln y_{it} = \alpha_i + Z_{it}^{\top}\beta + \psi_{F(i,t)} + \epsilon_{it}, \tag{5}$$

where α_i is an individual fixed effect, Z_{it} is a vector of time-varying controls (including year fixed effects), $F(i,t)$ is an index function that maps worker i to their employer in year t, $\psi_{F(i,t)}$ is the firm-specific wage premium, and ϵ_{it} captures idiosyncratic shocks. To link this panel data model to the NPSEM, define Y_i to be the logged earnings for individual i ten years after high school graduation as $Y_i \equiv \ln y_{i,10 \text{ years}}$, and let $M_i \equiv F(i, 10 \text{ years})$. Equation (5) then implies a semi-parametric model for Y_i of the form:

$$Y_i = f_Y(X_i, A_i, L_i, \epsilon_{Y,i}) + \beta^{\top} Z_{i,10 \text{ years}} + g(M_i), \tag{6}$$

where $g(M_i) = \psi_{F(i,10 \text{ years})}$ is the firm-specific wage premium faced by individual i, by construction. That is, additive separability in Equation (5) implies that Y_i is additively separable in $(X_i, A_i, Z_i, \epsilon_{Y,i})$ and the firm M_i.

Under ignorability for A_i on Y_i, the average treatment effect of attending college j relative to attending College 7 is:

$$\text{ATE}_{A\to Y} = \int_x \left[\mathbb{E}[Y_i \mid X_i = x, A_i = j] - \mathbb{E}[Y_i \mid X_i = x, A_i = \text{College 7}]\right] dP(x).$$

I aim to decompose this total effect into direct and indirect (mediated) components. Let $Y_i(a, m)$ denote the potential outcome under treatment a and mediator value m. Then the ATE can be written as:

$$\text{ATE}_{A\to Y} = \underbrace{\mathbb{E}[Y_i(j, M_i(j)) - Y_i(\text{College 7}, M_i(j))]}_{\text{Direct effect (net of firms)}}$$

$$+\underbrace{\mathbb{E}[Y_i(\text{College } 7, M_i(j)) - Y_i(\text{College } 7, M_i(\text{College } 7))]}_{\text{Indirect effect (via firms)}}.$$

Under the above assumptions, we can write

$$\begin{aligned}
&\mathbb{E}[Y_i(\text{College } 7, M_i(j)) - Y_i(\text{College } 7, M_i(\text{College } 7))] \\
&= \mathbb{E}[f_Y(X_i, A_i(\text{College } 7), L_i, \epsilon_{Y,i}) + \beta^\top Z_{i,10 \text{ years}} + g(M_i(j))] \\
&-\mathbb{E}[f_Y(X_i, A_i(\text{College } 7), L_i, \epsilon_{Y,i}) + \beta^\top Z_{i,10 \text{ years}} + g(M_i(\text{College } 7))] \\
&= \mathbb{E}[g(M_i(j))] - \mathbb{E}[g(M_i(\text{College } 7))] \\
&= \mathbb{E}_X\mathbb{E}[g(M_i(j))|A_i = j, X] - \mathbb{E}_X\mathbb{E}[g(M_i(\text{College } 7))|A_i = \text{College } 7, X] \\
&= \mathbb{E}_X\mathbb{E}[g(M_i)|A_i = j, X] - \mathbb{E}_X\mathbb{E}[g(M_i)|A_i = \text{College } 7, X] \\
&= \mathbb{E}_X\mathbb{E}[\psi_{F(i,10 \text{ years})}|A_i = j, X] - \mathbb{E}_X\mathbb{E}[\psi_{F(i,10 \text{ years})}|A_i = \text{College } 7, X].
\end{aligned}$$

Therefore, under additive separability, the mediated effect of college on earnings through firm pay premiums is equivalent to the ATE of college on firm pay premiums. Note further that because of firm effect additivity, this decomposition is unique (that is, the reference category does not matter). In other words, $\mathbb{E}[Y_i(\text{College } 7, M_i(j)) - Y_i(\text{College } 7, M_i(\text{College } 7))] = \mathbb{E}[Y_i(j, M_i(j)) - Y_i(j, M_i(\text{College } 7))]$.

B Further Details on Estimation Procedures

B.1 Average Treatment Effect of College Attendance (Figure 2, and Figures S16, S17 and S18)

I estimate the average treatment effect (ATE) of attending college j relative to College 7 using efficient influence functions (EIFs). Let the observed data consist of units indexed by $i = 1, \dots, n$, with variables (Y_i, A_i, X_i), where $Y_i \in \{\text{top10}, \text{FE}, \text{lnearn}\}$ denotes the outcome of interest, $A_i \in \{j, \text{College 7}\}$ is the treatment assignment indicator, and X_i is a vector of observed pre-college covariates.

I define the conditional mean outcome under treatment j as $\mu_j(X) = \mathbb{E}[Y \mid A = j, X]$, and similarly for the control group $\mu_{\text{College 7}}(X) = \mathbb{E}[Y \mid A = \text{College 7}, X]$. The population-level treatment effect is identified (under ignorability) as:

$$\text{ATE}_j = \mu_j - \mu_{\text{College 7}}, \quad \text{where} \quad \mu_j = \mathbb{E}_X[\mu_j(X)], \quad \mu_{\text{College 7}} = \mathbb{E}_X[\mu_{\text{College 7}}(X)].$$

To estimate these target quantities, I use the efficient influence function (EIF) for the mean outcome under each treatment group. Let $\hat{\mu}_j(X)$ denote an estimate of $\mu_j(X)$ obtained from an outcome regression model, and let $\hat{\pi}_j(X) = \hat{\Pr}(A = j \mid X)$ denote the estimated propensity score. Then the EIF-based estimator for μ_j is:

$$\hat{\mu}_j^{\text{EIF}} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{\mu}_j(X_i) + \frac{\mathbb{I}(A_i = j)}{\hat{\pi}_j(X_i)} \left(Y_i - \hat{\mu}_j(X_i)\right) \right\}.$$

The estimator $\hat{\mu}_{\text{College 7}}^{\text{EIF}}$ is defined analogously. The EIF-based estimator for the average treatment effect of attending college j relative to College 7 is then:

$$\widehat{\text{ATE}}_j^{\text{EIF}} = \hat{\mu}_j^{\text{EIF}} - \hat{\mu}_{\text{College 7}}^{\text{EIF}}.$$

In practice, I estimate $\hat{\mu}_j(X)$ and $\hat{\pi}_j(X)$ (the "nuisance functions") using linear, binomial or multinomial ridge regressions (depending on whether the outcome variable is continuous, binary or categorical, respectively), and implement K-fold cross-fitting. In this procedure, the data is partitioned into $K = 5$ folds. Cross-fitting preserves the asymptotic efficiency of the EIF estimator even when using a nonparametric method like ridge regression. Specifically, for each fold $k = 1, \dots 5$, the nuisance functions are estimated on the training folds (excluding the current fold), and predictions are made on the held-out fold. This process ensures that each observation's EIF term is computed using nuisance estimates that are fit on independent data. The final ATE estimate is computed by aggregating EIF contributions across all folds. Under this procedure, standard errors for $\widehat{\text{ATE}}_j^{\text{EIF}}$ can be obtained using the empirical variance of the estimated EIF. Asymptotically valid confidence intervals can then be constructed using the normal approximation.

Additionally, in Figures S16, S17 and S18 in Appendix I.2, I display counterfactual probabilities of working in different locations (or in a multi-establishment firm), industries, and firms of different sizes, respectively. These estimates are obtained via $\hat{\mu}_j^{\text{EIF}}$, where Y_i is replaced by a categorical variable denoting the location, industry, or size of an employee's firm.

B.2 Mediation Quantities

To assess how firm-based sorting overlaps with other dimensions of stratification (such as industry, location and firm size), I also estimate several mediation-related quantities.

Direct Effects (Figures S8 and S10)

The first of these is the controlled direct effect (CDE) of attending college j relative to College 7 on the firm premium of an individual's firm (FE), net of a mediator m. The CDE isolates the portion of the treatment effect that operates *not through* the mediator, by fixing the mediator at a specific level m. Figures X and Y, for example, display the CDE of attending college j relative to College 7 for industry and location, or firm size.

To formally define this quantity, let M denote the mediator variable (e.g., industry, location, or size), and let $\text{FE}(j, m)$ denote the potential outcome that would be observed if treatment were set to college j and the mediator were fixed to $M = m$. The *controlled direct effect* (CDE) of treatment j relative to College 7 at mediator level m is defined as:

$$\text{CDE}_j(m) = \mathbb{E}[\text{FE}(j, m)] - \mathbb{E}[\text{FE}(\text{College } 7, m)].$$

To express the CDE as a function of observed data, I invoke an additional no unmeasured confounding for the mediator: $(\text{FE}(j, m) \perp\!\!\!\perp A \mid X)$ and $(\text{FE}(j, m) \perp\!\!\!\perp M \mid A = j, X)$. This assumption rules out any unobserved selection into the mediator (such as location) net of pre-college observed characteristics. As discussed in the main text, this is unlikely to hold. In particular, regionally-based students, for example, who attend College 7, who obtain post-college employment in the Primary Metropolitan Area are likely positively selected, relative to their peers who work elsewhere. In practice, such patterns of selection are likely to downwardly bias my estimates of the CDE. Under this assumption, we can identify $\mathbb{E}[Y(j, m)]$ as:

$$\mathbb{E}[Y(j, m)] = \mathbb{E}_X \left[\mathbb{E}[Y \mid A = j, M = m, X]\right] = \mathbb{E}_X \left[\mu_j(m, X)\right],$$

where $\mu_j(m, X) = \mathbb{E}[Y \mid A = j, M = m, X]$ is the conditional mean outcome under treatment j and mediator level m, given covariates X. Similarly, $\mathbb{E}[Y(\text{College } 7, m)]$ is identified as:

$$\mathbb{E}[Y(\text{College } 7, m)] = \mathbb{E}_X\left[\mu_{\text{College } 7}(m, X)\right].$$

Thus, the controlled direct effect is identified as the difference in these conditional expectations averaged over the distribution of X :

$$\text{CDE}_j(m) = \mathbb{E}_X\left[\mu_j(m, X) - \mu_{\text{College } 7}(m, X)\right].$$

This identification result enables estimation using either plug-in estimators (i.e., regression followed by averaging) or doubly robust estimators such as those based on the efficient influence function (EIF). In practice, I estimate $\mathbb{E}[Y(j, m)]$ using the efficient influence function (EIF) estimator:

$$\hat{\mu}_j^{\text{EIF}}(m) = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}_j(m, X_i) + \frac{\mathbb{I}(A_i = j, M_i = m)}{\hat{\pi}_j(m, X_i)\hat{\pi}_j(X_i)}\left(\text{FE}_i - \hat{\mu}_j(m, X_i)\right)\right\},$$

where $\hat{\mu}_j(m, X_i)$ is an estimate of $\mathbb{E}[Y \mid A = j, M = m, X_i]$ (obtained from a ridge regression model), $\hat{\pi}_j(m \mid X_i) = \Pr(M = m \mid X_i, A = j)$ is the estimated joint propensity of treatment and mediator, and $\mathbb{I}(A_i = j, M_i = m)$ is the indicator function for the joint treatment-mediator status. The estimated controlled direct effect can then be estimated as:

$$\widehat{\text{CDE}}_j^{\text{EIF}}(m) = \hat{\mu}_j^{\text{EIF}}(m) - \hat{\mu}_{\text{College } 7}^{\text{EIF}}(m).$$

As in the ATE estimation procedure, cross-fitting is used to mitigate overfitting when estimating nuisance functions the $\mu_j(m, X)$, $\pi_j(X)$, and $\pi_j(m \mid X)$. The data is split into K folds; nuisance models are trained on $K - 1$ folds, and predictions are made on the held-out fold. This ensures that each observation's EIF contribution is based on out-of-fold estimates, preserving the efficiency and asymptotic normality of the final estimator even when using flexible, high-dimensional learners. Standard errors for $\widehat{\text{CDE}}_j^{\text{EIF}}(m)$ are computed using the empirical variance of the estimated EIF terms, analogous to the ATE case.

Formal Mediation Analyses (Figure 5, and Figures S8 and S10)

Overlap between firm-sorting across colleges and geographic-, industry- and size-based sorting can be summarized by an additional mediation quantity, which captures the portion of the treatment effect that is not mediated by M (summarizing across levels of M). Specifically, I estimate the natural direct effect (NDE) of attending college j relative to attending College 7, where the mediator M (e.g., industry, location, or firm size) is fixed at the value it would naturally take under treatment j.

To formally describe this procedure, let $\text{FE}(a, m)$ denote the potential outcome under treatment $A = a$ and mediator fixed to $M = m$, and let $M(j)$ denote the potential value of the mediator under treatment j. Then, the natural direct effect is defined as:

$$\text{NDE}_j = \mathbb{E}[\text{FE}(j, M(j))] - \mathbb{E}[\text{FE}(\text{College 7}, M(j))].$$

When estimating direct effects, I use school j as the reference group. For example, when comparing the effect of attending college j on firm premium net of industry, I ask: to what extent does attending school j relative to College 7 yield access to a higher-premium firm, if College 7 students worked in the same industries as the Flagship students. s

Under the assumptions mentioned previously, the NDE is identified via:

$$\begin{aligned} \text{NDE}_j &= \mathbb{E}_X\left[\mathbb{E}_{M|X,A=j}\left[\mu(j, M, X) - \mu(\text{College 7}, M, X)\right]\right] \\ &= \mathbb{E}_X\left[\mu_j(X) - \mathbb{E}_{M|X,A=j}\mu(\text{College 7}, M, X)\right], \end{aligned}$$

where $\mu(a, m, X) = \mathbb{E}[\text{FE} \mid A = a, M = m, X]$. Note that the first term, μ_j, can be estimated using the EIF-based estimator as described above. To estimate the second term $\mathbb{E}[\text{FE}(\text{College 7}, M(j))]$, I use its efficient influence function (EIF), which enables doubly robust estimation. The EIF for this quantity is:

$$\begin{aligned} \phi_i = \; & \mathbb{E}_{M|X_i,A=j}\left[\mu(\text{College 7}, M, X_i)\right] \\ & + \frac{\mathbb{I}(A_i = j)}{\pi_j(X_i)}\left(\mu(\text{College 7}, M_i, X_i) - \mathbb{E}_{M|X_i,A=j}\left[\mu(\text{College 7}, M, X_i)\right]\right) \\ & + \frac{\mathbb{I}(A_i = \text{College 7})}{\pi_{\text{College 7}}(X_i)} \cdot \frac{\pi_j(X_i, M_i)}{\pi_{\text{College 7}}(X_i, M_i)} \cdot \left(\text{FE}_i - \mu(\text{College 7}, M_i, X_i)\right), \end{aligned}$$

where $\pi_j(X) = \Pr(A = j \mid X)$ and $\pi_{\text{College 7}}(X) = \Pr(A = \text{College 7} \mid X)$ are estimated propensity scores, $\pi_j(X, M) = \Pr(A = j \mid X, M)$ and $\pi_{\text{College 7}}(X, M) = \Pr(A = \text{College 7} \mid X, M)$ are treatment probabilities conditional on X and M.

The EIF-based estimator of $\mathbb{E}[\text{FE}(\text{College 7}, M(j))]$ is then obtained by averaging $\hat{\phi}_i$ over the sample, where $\hat{\phi}_i$ is estimated by plugging in estimates of nuisance functions in Equation **??**. The final estimator of the NDE is therefore:

$$\widehat{\text{NDE}}_j = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\psi}_i - \hat{\phi}_i\right\},$$

where $\hat{\psi} = \hat{\mu}_{ij}(X_i) + \frac{\mathbb{I}(A_i=j)}{\hat{\pi}_j(X_i)}\left(Y_i - \hat{\mu}_j(X_i)\right)$.

Cumulative Treatment Effects (Figures S19 and S6)

Figure S19 in the main text and Figure S6 in Appendix E rely on cumulative treatment effects of attendance at college j and BA attainment at college k, relative attendance and completion at College 7. Now, let the observed data consist of units indexed by $i = 1, \ldots, n$, with variables $(Y_i, A_i, X_i, Z_i, D_i)$, where as before, X_i denotes pre-college covariates, A_i denotes college attendance, $Y_i \in \{\text{FE}, \text{lnearn}\}$ and, additionally, D_i denotes degree completion and Z_i denotes a vector of intermediate (post-treatment) covariates. In what follows, Z_i include first-term GPA, GPA at graduation or (for non-completers), at stop-out, number of credits attempted and earned in the first semester, and their squares.

I define the cumulative treatment as the joint event $(A_i = j, D_i = k)$, that is, of attending college j and completing a degree at college k. To estimate the cumulative treatment effect of attending college j and completing a BA at college k, I consider the potential outcome $Y_i(j,k)$, which represents the firm premium that individual i would face if they both attended and graduated from college k. The target quantity is:

$$\mathbb{E}[Y_i(j,k)] - \mathbb{E}[Y_i(\text{College 7}, \text{College 7})],$$

that is, the average outcome under the path of attending college j, completing a degree from college k, relative to attending and completing a degree from College 7. In Figure S19, $j = \text{Flagship}$ and $k = \mathcal{J} \setminus \text{Flagship}$,that is, all colleges bar the Flagship. In Figure S6 in Appendix E , $j = k$.

Identification of this quantity relies on a sequential ignorability assumption, namely that there is no unobserved confounding for college attendance or for college completion. While in practice this is unlikely to be fully met, my large set of pre-college and post-college covariates makes this assumption more plausible. Under these assumptions, I can estimate $\mathbb{E}[Y(j,k)]$ using the efficient influence function (EIF) estimator:

$$\hat{\mu}_{j,k}^{\text{EIF}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}_j(X_i, Z_i) + \frac{\mathbb{I}(A_i = j)}{\hat{\pi}_j(X_i)}\left(\hat{\mu}_j(X_i, Z_i) - \hat{\mu}_j(X_i)\right) + \frac{\mathbb{I}(A_i = j, D_i = k)}{\hat{\pi}_j(X_i)\cdot\hat{\gamma}_j(X_i, Z_i)}\left(Y_i - \hat{\mu}_j(X_i, Z_i)\right)\right\},$$

where

- $\hat{\mu}_j(X_i, Z_i) = \mathbb{E}[Y \mid A_i = j, D_i = 1, X_i, Z_i]$ is the outcome regression,
- $\hat{\mu}_j(X_i) = \mathbb{E}_{Z|X_i, A_i = j}[\hat{\mu}_j(X_i, Z)]$ is the marginal regression over Z_i,
- $\hat{\pi}_j(X_i) = \Pr(A_i = j \mid X_i)$ is the treatment (attendance) propensity score, and
- $\hat{\gamma}_j(X_i, Z_i) = \Pr(D_i = k \mid A_i = j, X_i, Z_i)$ is the probability of BA completion at college k given attendance and post-treatment covariates.

As in previous sections, I use K-fold cross-fitting to estimate the nuisance functions $\hat{\mu}_j(X, Z)$, $\hat{\pi}_j(X)$, and $\hat{\gamma}_j(X, Z)$ to mitigate overfitting. Each observation's EIF contribution is computed using out-of-fold predictions. Standard errors are obtained from the empirical variance of the EIF terms.

C Further Details on Sample Construction

Data

Linked employer-employee data come from the state's employment department, obtained from employer wage reports filed through the state's Unemployment Insurance (UI) program. Specifically, I obtain quarterly earnings and hours worked from fiscal years 2005:Q1 to 2022:Q2 for nearly all private-sector, nonfederal public, and local government workers covered under the state's Unemployment Insurance (UI) system, representing a substantial share of the state's civilian labor force.[17] Employers submit quarterly wage reports to the employment department on each employees' wages and hours worked, as well as a (quarterly) summary file on employer characteristics including industry (North American Industry Classification System - NAICS - code), location of firm, and total counts of employees for each month. Employers can be either firms or establishments, depending on whether they operate exclusively in Summit State or across multiple states.[18] The data are organized at the employee-employer-quarter level, meaning an individual may appear multiple times within a quarter if employed by more than one firm. I discuss the process of attaching a single employer to each person year in detail below. A crucial component of these data is that they contain longitudinal employer identifiers that are consistent across years.[19]

Data Construction

In this section, I provide further details about the construction of my earnings panels. First, for all panels, I begin by imputing missing hours based on an imputed wage rate for the contiguously employed, and drop all observations with negative wage values in a given year-quarter. If the worker has missing, zero, or negative income in a given quarter, I assume the worker is unemployed or working out of state, If the respondent

[17]Nearly all employers with employees are required to participate in the UI Program if they pay wages to employees, regardless of the dollar amount. Participation includes registering with the state and submitting quarterly reports, and is not contingent on a worker filing a UI claim. It is these quarterly reports that I use. Several classes of worker are not reported in state UI records. These include self-employed workers and business owners, church employees, as well as federal employees, such as U.S. Postal Service (USPS), federal civilian employees, and active duty and retired military.

[18]I use the terms "firm" and "employer" interchangeably to refer to a person's place of employment. Technically, within the sample, the employer is the set of establishments operating in Summit State under a given owner. In the event that the company only operates inside Summit State, the employer is the firm; for companies that have registered addresses outside the state, the employer is a an establishment, or set of establishments. For companies operating solely in Summit State and with only one address, the firm is the establishment.

[19]Mergers, acquisitions, liquidation, changes in ownership and other types of restructuring will result in a new employer identification code.

has zero reported hours but positive income, I use that individual in the sample (that is, I assume that the employer failed to report the hours of that worker, or that these were wages earned in one quarter but paid in another). If a worker has no reported hours in a given quarter, I classify hours - and therefore the worker's wage rate - as missing, but still use the observation to construct annualized earnings. For each calendar quarter, I then identify each worker's *primary employer* as the firm from which they earned the largest share of their total earnings, and annualize wages from their primary employers, so long as the worker is observed at the firm for at least two consecutive quarters in a given year. These earnings and wage variables, and the corresponding employer, are the basis of my primary analysis. I then annualize the remaining quarterly earnings and hours within each calendar year, conditional on the worker having at least two consecutive quarters of employment with the same primary employer in that year. Annual earnings, hours, and hourly wage rates are computed using this cleaned panel. The resulting unit of observation is the *worker-year*, defined by the worker's relationship with their primary employer in a given school year.

I link these earnings data to individuals' estimated firm premiums, which I obtain from a separate linked employer–employee panel. An employment spell is defined as a sequence of at least five consecutive quarters during which a worker remains employed by the same primary employer. To avoid measurement error from job transitions or partial employment, I drop the first quarter of each spell (which may reflect mid-quarter job starts) and the last two quarters (which may capture separation-related wage fluctuations). Together with the fact that I keep only an individual's primary employer in a given wage quarter (i.e. that with the largest income share) a record will be dropped if an employer is an individual's primary source of income in one quarter, but if the individual is still observed at that employer in the following quarter but that employer is no longer the individual's primary income source (or if the person skills a quarter at an employer before returning). Essentially, this means that an individual must have begun an employment spell before the second quarter (inclusive) in order for that spell to count as an observation in the year. Similarly, when a given employment spell ends, leaving the employer must occur in the fourth quarter, otherwise that spell will not be counted as an observation for a given wage year.[20]

I next impose a series of sample restrictions on this panel. First, I exclude workers who appear with more than nine employers in a calendar year. I also drop workers with annual earnings below $4,270 or with an hourly wage rate less than or equal to $3 (both in 2022 dollars). To eliminate outliers in hours worked, I restrict the sample to

[20]For instance, if a person is employed at employer j for 20161, 2, 3,4, 20171, 2, 3. They have over 4 quarters for that spell so part of the spell is kept: 20162, 20163, 20164, 20171. If this is the primary income source in all years, I end up assigning the employer and the total or annualized wages from the employer in 2016, but the person will have no observed wages in 2017 because there is only one quarter observed at the employer (which does not meet the inclusion threshold of two wage quarters in a given year at the primary employe

workers with between 400 and 4,800 annual hours, to workers with strictly fewer than 10 employers over the period 2005-2022, and drop employers with fewer than 5 high school graduates observed over the full period.

Second, I filter wage observations according to the following exclusion criteria, to determine when an individual completes their education. First, for high school graduates never observed in the postsecondary enrollment or awards files, I take all years following high school graduation (that is, in the school year following high school graduation onwards). If an individual is observed to complete a postsecondary degree (in or out of state), I take all person earnings years following the last credential attained sub BA award (inclusive) that an individual attained, ignoring graduate awards (again, taking wage observations only beginning the school year following award attainment). Finally, for college stop outs (i.e., those that at some point enrolled in college but did not complete it, I define an individual to be stopped out if they are observed as a college enrollee in year 5, and no longer observed as a college enrollee in any subsequent years $t, t+1, ..., T$. For individuals who stop out and re-enroll, I only take earnings for that individual following their final enrollment spell. Note that while high school graduates and NSC completers will not feature in my estimation sample for college returns, they can be used to increase precision for estimating the employer effects. Together, these restrictions ensure that I make firm inferences from individuals' earnings (i) only once an individual has completed their formal schooling, and (ii) based on either a partial quarter of employment or in the quarters either before or of job loss.

Assignment of Employers to Employer Characteristics

Assigning a consistent location and industry to each firm is essential for analyzing firm-level variation in wages and outcomes. However, the underlying administrative data present several challenges. Summit State does not require multi-industry and multi-site employers to break down their employment and wage data by site and industry. Even when employers do provide these breakouts, it is not clear which employees are working in which locations or industries. While only a very small number of firms in Summit State report data for multiple locations, according to the state's employment department, these firms account for approximately 30% of total employment.. Additionally, location and industry reporting may be inconsistent across workers within the same firm and year.[21]

To address these issues, I implement a rule-based procedure to assign each firm a single county and a single NAICS industry code per year (or, when this is not feasible, to explicitly flag the firm as multi-establishment). The procedure begins by determining

[21]Note in addition that the UI wage data reports only the location of the employer, and not where an individual employee is located.

the modal county and NAICS code for each firm-year. Then, at the firm level, I assign a unique location and industry as follows. If the firm appears in only one unique county (excluding "Multiple Counties"), I assign that county as the firm's location. If a firm is flagged as multi-establishment but consistently reports a single location, I retain that location. If a firm appears in multiple counties or reports both a specific location and "Multiple Counties," I assign it the label "Multiple Counties." I assign a single NAICS code for each firm-year.

To preserve as many observations as possible, I disregard missing values when more specific information is available. For example, if a firm appears in Jane County and as missing, I assign the firm to Jane County; similarly, if a firm appears with both a valid NAICS code and a missing code, I retain the non-missing value. Firms with no usable location or industry information across all associated workers are excluded from the analysis.

D AKM Model Details

D.1 Model Results

The first column of Table 3 presents summary statistics for the annualized linked employer-employee panel which I use to fit the AKM model (Equation 1 in the main text) among workers in the years following education completion - that is, after processing the quarterly records imposing the sample restrictions described in Appendix C. The left column shows summary statistics for the full annualized panel, while the right column shows summary statistics for the fully connected set. The employer effects are identified only within the "connected set" of employers that are linked by worker transitions between those employers, so the AKM estimation is necessarily restricted to the largest connected set of employers. This consists of 43 percent of employers in the full annualized panel, 70 percent of workers in the panel, and 81 percent of worker-year observations in the panel. Because identification of employer fixed effects comes from workers moving between primary employers, it is important to know how much mobility there is in the sample. The table shows that the largest connected set includes over half a million unique workers, and approximately 43 percent of those workers changed primary employer at least once during the panel 2005-2022.

I estimate the AKM model (Equation 1 in the main text) using the linked employer-employee panel. Table 5 shows the resulting model fit (rightmost column) and implied earnings variance decomposition. Logged earnings is decomposed into five components: worker effects, firm effects, year effects, the covariance between worker and employer effects (sorting of workers and employers), and a residual. I omit the worker-year and firm-year covariances, which capture, collectively, approximately 3% of the total variance.

The estimated variance of earnings for workers within 10 years of high school graduation is .36. 47% of Worker fixed effects explain the largest share of earnings variation, at 47%. Firm effects explain a smaller but still sizable share: approximately 20% of the variation in earnings. These shares are very similar to those estimated by prior authors using UI wage records in Summit State (redacted to keep the state anonymous).

Table 3: Summary Statistics for the Full Sample of Workers and the Largest Connected Set (AKM Sample)

Statistic	**Full Sample**	**Connected Set**
Person-year observations	2,913,597	2,355,953
Unique individuals (SSNs)	744,364	521,605
Unique firms	127,534	54,811
Unique movers (2+ firms)	262,674	224,080
Mover observations	361,497	300,470

Source: Summit State administrative UI earnings, postsecondary and high school records.

Table 5: Variance Decomposition of Log Earnings

Variance Component	**Raw Variance**	**Share of Total**	**Adj.** R^2
Person Effect (Var(α_i))	0.17	0.47	.84
Firm Effect (Var(ψ_j))	0.07	0.20	
Person-Firm Sorting ($2 \cdot \text{Cov}(\alpha_i, \psi_j)$)	0.05	0.12	
Residual (Var(ε_{it}))	0.04	0.12	
Total Variance of Log Earnings	**0.36**	**1.00**	

Notes: The decompositions also include covariances between worker and employer fixed effects and year fixed effects. Because these covariances explain only about 3 percent of the variation, they are omitted from the table.
Source: Summit State administrative UI earnings, postsecondary and high school records.

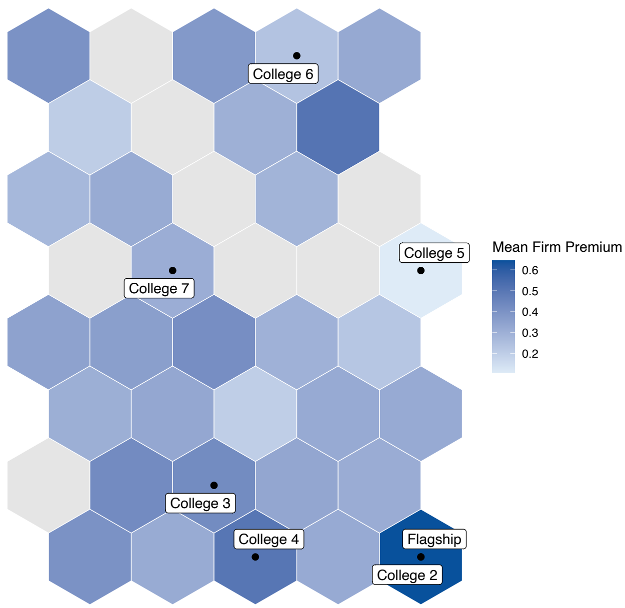
D.2 Estimated Firm Premiums by Industry

Figure S1: Mean and Variance of Estimated Firm Premiums, by 2-digit Industries

Notes: Average and variance of firm pay premiums across industries, normalized relative to the Accommodation and Food Services sector (NAICS 72), which is the lowest-paying sector in terms of both estimated firm effects and average worker characteristics. Bars represent the average estimated firm fixed effect within each 2-digit NAICS sector, where firm effects are estimated using the AKM decomposition described in the main text (Equation 1). Industry-level averages and variances are weighted by the number of person-quarter observations at each establishment.
Source: Summit State administrative UI earnings, postsecondary and high school records.

D.3 Estimated Firm Premiums across Counties

Figure S2: Mean of Estimated Firm Premiums, by County



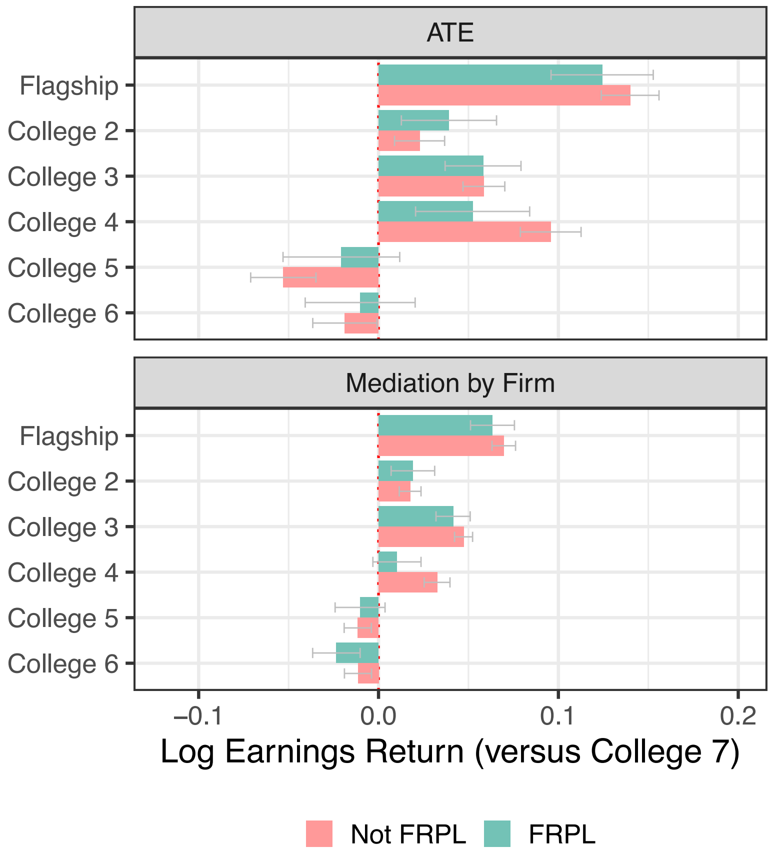
Notes: Average estimated firm wage premium for each Summit State county, based on the location of firms between 2018 and 2022. The array is stylized in order to keep the state and its institutions anonymous. Firm locations are assigned using administrative records from this period, and averages are weighted by the average number of person-year observations at each firm across this same period. Several low-population counties are excluded (white colored) due the low number of firms (under 50) employing high school graduates. Multi-establishment firms are excluded from county-specific averages; their average firm premium, across the full sample, is 0.46 (not shown).
Source: Summit State administrative UI earnings, postsecondary and high school records.

E Robustness Checks Cited in Main Text

E.1 Effect Heterogeneity by Socio-Economic Origin

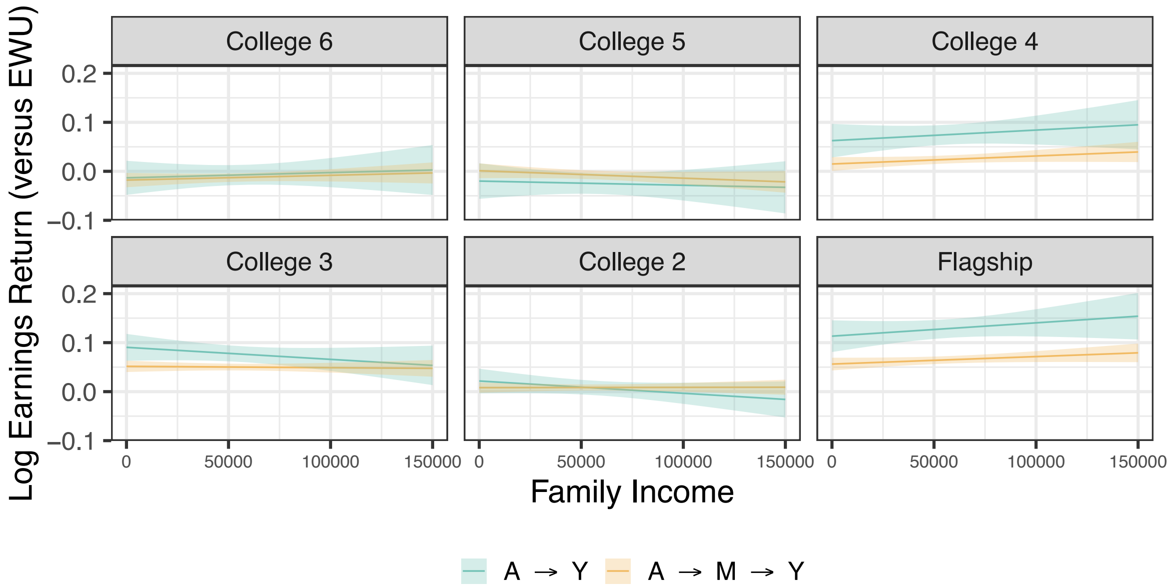
Figure S3 compares the ATE of attendance at each college (top panel) and its mediation by firm sorting (bottom panel), comparing students ever eligible for free/reduced-price lunch (FRPL, teal bars) to those who were not (red bars). At the flagship, both FRPL and non-FRPL students earn substantial earnings premiums relative to College 7, with nearly identical magnitudes and similar shares mediated by firm placement. Across Colleges 2–3 and 5–6, the returns for FRPL and non-FRPL students are also closely aligned, and differences are not statistically significant. The only meaningful divergence appears at College 4, where non-FRPL students capture somewhat larger gains than their FRPL peers. Taken together, these results suggest that firm sorting contributes comparably to the college earnings premium across socioeconomic groups, with little evidence that class background systematically alters the payoff to selective college attendance.

Figure S3: College Attendance Effects and Mediation via Firms, by Free and Reduced Price Lunch (FRPL) Status



Notes: The top panel shows estimated average treatment effects (ATEs) of attending each college (versus College 7) on log earnings ten years after high school graduation, by FRPL status (colors). The bottom panel shows the portion of the total effect mediated through differences in employer assignment, by racial-ethnic group (colors). The colleges included are Summit State's Flagship school, and Colleges 2 to 6. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting, and modified to address effect heterogeneity. Specifically, I use the EIF-based estimator $\hat{\mu}_j^{\mathrm{EIF}}(r) = \frac{1}{n_r} \sum_{i:R=r} \left\{ \hat{\mu}_j(X_i) + \frac{\mathbb{I}(A_i=j)}{\hat{\pi}_j(X_i)} \left(Y_i - \hat{\mu}_j(X_i)\right) \right\}$, where $R \in \{\text{Not FRPL}, \text{FRPL}\}$ denotes FRPL status, and n_r denotes the total number of people in the sample belonging to each $R = r$ group. Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. Grey whiskers indicate 95% confidence intervals. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Figure S4: College Attendance Effects and Mediation via Firms, by Family Income, Among Receivers of Financial Aid

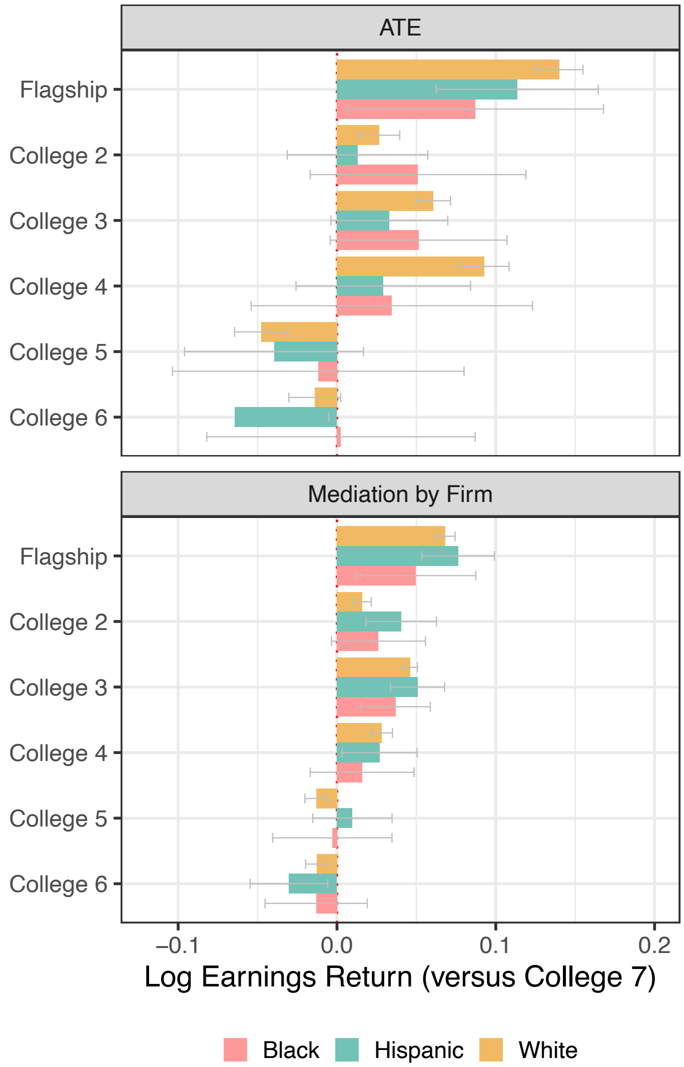


Notes: The top panel shows estimated average treatment effects (ATEs) of attending each college (versus College 7) on log earnings ten years after high school graduation, by family income. The bottom panel shows the portion of the total effect mediated through differences in employer assignment, by racial-ethnic group (colors). The colleges included are Summit State's Flagship school, and Colleges 2 to 6. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting, and modified to address effect heterogeneity. Ribbons denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). High school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

E.2 Effect Heterogeneity by Race-Ethnicity

I estimate treatment effects across three racial-ethnic groups: White, Hispanic and Black. These models correspond to the conditional average treatment effects (CATEs) of college attendance at college j relative to College 7 across racial-gender groups, as well as conditional mediation via firms, and provide insight into whether credentialing and college effects apply uniformly across individuals. Figure S5 shows that, for White students, college attendance is associated with substantial earnings returns, particularly attendance at the flagship. A sizable portion of these returns is mediated through employer placement. For Hispanic students, flagship attendance also provides a strong, through slightly attenuated effect, for which mediation via firms represents a larger component. Among Black students, estimated returns are markedly smaller, though substantially less precise, reflecting the very small number of Black students attending different 4-year colleges, but still indicate meaningful firm-mediated effects for several colleges, particularly the Flagship and College 2. Overall, while the magnitude and precision of college effects vary across groups, the results suggest that firm sorting plays a substantial role in mediating the college earnings premium across all racial/ethnic groups studied, with especially strong contributions for Hispanic and White students.

Figure S5: College Attendance Effects and Mediation via Firms, by Racial-Ethnic Groups



Notes: The top panel shows estimated average treatment effects (ATEs) of attending each college (versus College 7) on log earnings ten years after high school graduation, by racial-ethnic group (colors). The bottom panel shows the portion of the total effect mediated through differences in employer assignment, by racial-ethnic group (colors). The colleges included are Summit State's Flagship school, and Colleges 2 to 6. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting, and modified to address effect heterogeneity. Specifically, I use the EIF-based estimator $\hat{\mu}_j^{\text{EIF}}(r) = \frac{1}{n_r} \sum_{i:R=r} \left\{ \hat{\mu}_j(X_i) + \frac{\mathbb{I}(A_i=j)}{\hat{\pi}_j(X_i)} (Y_i - \hat{\mu}_j(X_i)) \right\}$, where $R \in \{\text{White}, \text{Black}, \text{Hispanic}\}$ denotes racial-ethnic group, and n_r denotes the total number of people in the sample belonging to each $R = r$ group. Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. Grey whiskers indicate 95% confidence intervals. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

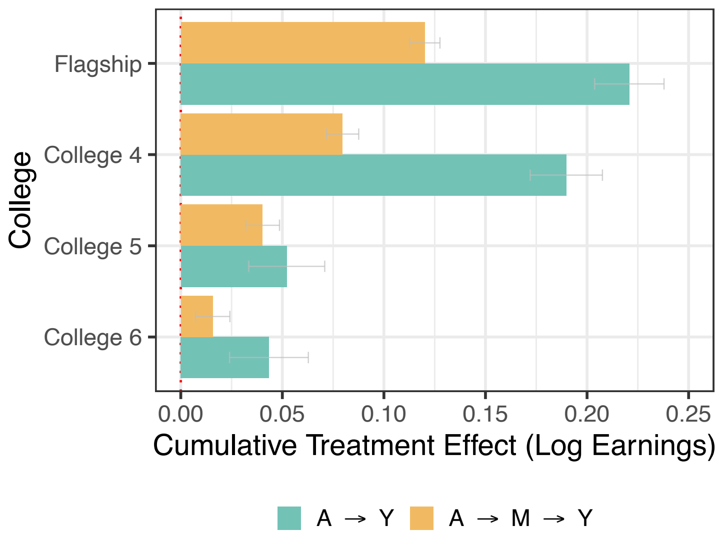
E.3 Joint Effect of Attendance and Completion, and Mediation via Firms

In this section, I examine the cumulative or joint effects of attendance *and* completion at a given college. My main analyses follow prior definitions of college-as-treatment (e.g. Hoekstra, 2009; Mountjoy and Hickman, 2021) by examining the earnings returns to attending different 4-year colleges on earnings, and mediation via firm placement. Nevertheless, such an approach elides an important margin: BA completion. Because the effect of attending a particular college partly reflects differential likelihoods of degree completion (both any college, and at different colleges), estimated effects conflate attendance effects (that is, exposure to any given campus) with enrollment and completion effects at other campuses.[22] I therefore additionally estimate the cumulative treatment effect (CTE) of attendance and completion each of the state's 4-year public colleges, using the DML procedure described in Appendix B. The results from this exercise (shown in Figure S6) are consistent with my main results.

There are two important differences compared with my main results. The first is that estimated CTEs are much larger than the estimated ATEs of college attendance as shown in Figure 2. The reason for this discrepancy is that counterfactual rates of BA completion are almost identical, at approximately 77% across all colleges (Figure S15 in Appendix I.1). Given the positive effect of BA completion on earnings, non-completion attenuates ATE estimates of college attendance, relative to CTE estimates of college attendance *and* completion. Second, the estimated treatment effects of College 5 and College 6 attendance and completion are positive, more aligned with the raw earnings gaps displayed in the top panel of Figure 2 than with the negative estimated treatment effects presented in the figure's bottom panel. The reason for this is shown in Figure S14 in Appendix I.1: out-transfer (mechanically, to a more selective institution) is a far more common experience among College 7 than among College 5 and 6 attendees (with College 4 a particularly common destination for College 7 attendees). For this reason, CTE estimates of College 7 attendance and completion attenuate the ATE of College 7 relative to College 6 or College 5 attendance, given that College 7 attendance is more likely to lead to a credential from a more selective institution.

[22]Table 14 in Appendix F.2 shows that out-transfer is indeed a common experience: of the 90% of the Flagship starters earning a BA degree 10 years following high school graduation, 92% completed a degree at the Flagship; by contrast, of the 62% of College 7 starters earning a BA degree 10 years following high school graduation, only 79% completed a degree at College 7.

Figure S6: Cumulative Treatment Effects (CTEs) of Attendance and Completion on Firm Outcomes



Notes: Point estimates and standard errors displayed in Figure S19 in the main text. Teal bars show estimated cumulative treatment effects of attending (as a freshman admit) and completing a degree at different colleges on log earnings 10 years following high school graduation, relative to attending (as a freshman admit) and completing a BA at College 7. Orange bars show mediation via employer placement. Estimates are obtained using the double machine learning (DML) approach described in Appendix B, with ridge regression used for all nuisance models. Covariates include those from previous models (demographics, academic background, and high school characteristics), along with additional controls for degree completion selection: GPA at graduation or stop-out, first-term credits attempted and earned, and their squares.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Table 7: Cumulative Treatment Effects (CTEs) of Attendance and Completion on Firm Outcomes

College	ATE	Mediation via Firm
College 7	10.872 (0.0066)*	0.102 (0.003)*
College 6	0.043 (0.0099)	0.016 (0.0043)
College 5	0.052 (0.0095)	0.04 (0.0042)
College 4	0.19 (0.0091)	0.08 (0.004)
College 1	0.221 (0.0087)	0.12 (0.0038)

Notes: Point estimates and standard errors displayed in Figure S6. Estimated cumulative treatment effects of attending (as a freshman admit) and completing a degree at different colleges on log earnings 10 years following high school graduation, relative to attending (as a freshman admit) and completing a BA at College 7, and mediation via employer placement. Estimates are obtained using the double machine learning (DML) approach described in Appendix B, with ridge regression used for all nuisance models. Covariates include those from previous models (demographics, academic background, and high school characteristics), along with additional controls for degree completion selection: GPA at graduation or stop-out, first-term credits attempted and earned, and their squares.
Source: Summit State administrative UI earnings, postsecondary and high school records.

E.4 Mediation by Firm Location and Industry

College-based firm sorting may still emerge even in the absence of employer social exclusion or distinct campus cultures directing students to high-premium firms. Here, I examine two alternative explanations for the observed patterns: geographic-based sorting and industry segregation, Indeed, Figures S16 and S17 in Appendix I.2 show that there is substantial sorting into different locations and industries across colleges.

E.4.1 Location

First, I examine whether the firm premium advantage enjoyed by students attending the Flagship might be driven more mechanically by acting as a gateway to its local (high-wage) labor market. Figure S2 in Appendix D displays average firm premiums across Summit State counties and the locations of each college. Notably, the state flagship is in a county (which I call the "Primary Metropolitan Area") with some of the highest firm premiums in the state; Table 1 also showed that Flagship attendees are much more likely than College 7 attendees to work in a firm in the nearby Primary Metropolitan Area.

The natural question then arises of whether the Flagship obtains its firm sorting advantages primarily by anchoring students in the nearby high-wage metropolitan area.[23] To formally assess the role of geographic sorting in explaining the observed firm premium differences across colleges, I estimate mediation of college effects on firm premia net of location (see Appendix B for methodological details). The results, displayed in Figure S7, show that only a small portion of firm sorting advantages can be accounted for by location.[24]

The reasons for this are explored in Figure S8, which presents estimates of controlled direct effects (CDEs) of college attendance on firm fixed effects, conditioning on firm location. Figure S8 presents these estimates, where each panel holds firm location constant - either (1) a single-establishment firm located in the Primary Metropolitan Area (left panel), (2) a multi-establishment firm (middle panel), or (3) a single-establishment

[23]In my analyses, I control for high school fixed effects, mitigating the related concern that the Flagship students, many of whom may originate from the state's Primary Metropolitan Area or its surrounding areas, are more inclined to remain in the region after graduation even without any college effects per se.

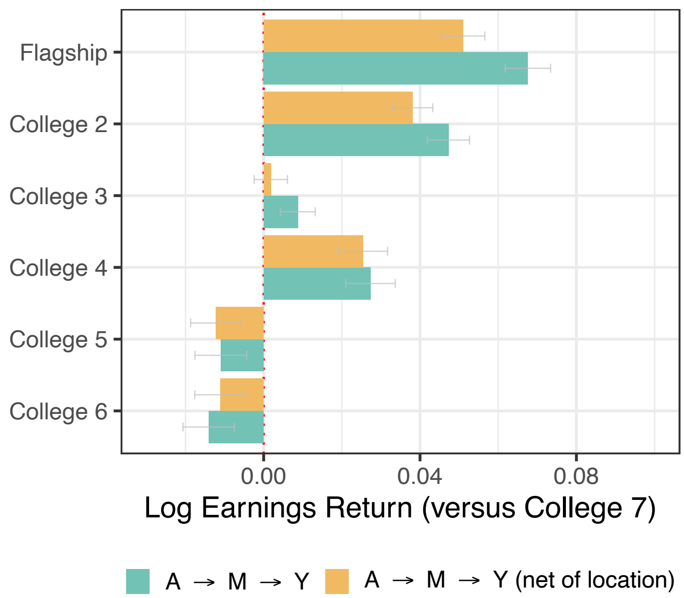
[24]Some care should be taken over the causal interpretation of this quantity for two reasons. First, for one's first set of jobs after college, firm and location are in many ways simultaneously determined; "Primary Metropolitan Area" is not a strict mediator for firm as an outcome. Despite not having a strictly causal interpretation, it is nevertheless useful for summarizing the gap in firm premia among workers based in the Primary Metropolitan Area, net of pre-college characteristics. Second, given that College 7 is distant from the Primary Metropolitan Area, College 7 students who do work in the Primary Metropolitan Area may be positively selected on bridging funds and motivation for certain types of work, among other characteristics. I expect that this pattern of selection may downwardly bias my estimates of the direct effect, producing a conservative estimate

firm located outside the state's Primary Metropolitan Area (right panel).[25] By conditioning on firm location, these estimates capture the portion of the college effect on firm wage premiums not attributable to sorting into geographic placement based on person attributes. Intuitively, these quantities capture a person's firm pay premium if they attended college j, relative to College 7, if we fixed their post-college employment location to either the Primary Metropolitan Area, or another area.

The figure shows that firm placement advantages across colleges do not simply reflect geographic sorting. Attending the Flagship and obtaining employment in the Primary Metropolitan Area, relative to attending College 7 and obtaining employment in the Primary Metropolitan Area, yields access to significantly (.06 log point) higher-premium employers - a figure almost identical to the overall firm premium to the Flagship attendance (Figure 2). Notably, College 2 also shows a substantial advantage within the Primary Metropolitan Area, while College 3, despite being located in the Primary Metropolitan Area, does not. The pattern is similar among workers at multi-establishment firms: the Flagship and College 2 attendees are placed at higher-premium employers than comparable students from College 7, largely mirroring the overall firm premia to these schools. By contrast, given employment outside of the Primary Metropolitan Area at a single-establishment employer, firm premium differences returns across colleges are far smaller - and in the case of College 4, College 5, and College 6, even negative relative to College 7. Figure S7 in Appendix E.4 summarizes the contribution of geographic sorting to the Flagship's firm premium: a full three quarters of its firm-sorting effect comes from non-geographic sorting. Note that the importance of geography may work in subtler ways beyond the physical location of graduates to different places. For example, the Flagship may have stronger career center connections to the nearby Primary Metropolitan Area (e.g. Manduca, 2024), facilitating high-premium firm attainment compared with College 7 graduates.

[25] As discussed in detail in Appendix (C), I am unable to discern the location of employment for employers with multiple establishments. I therefore categorize single establishment firms into whether they are located in the "Primary Metropolitan Area" or another county ("Other Area"), or whether the employer has multiple establishments.

Figure S7: College Effects on Firm Premia, Marginally and Net of Location



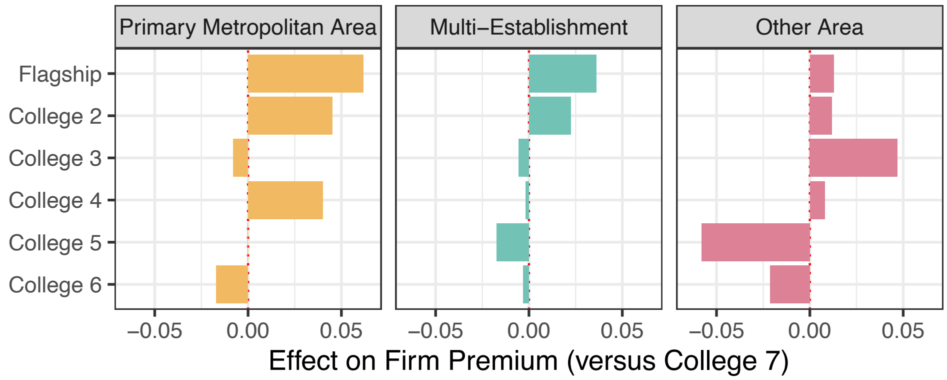
Notes: Teal bars show estimated average treatment effects (ATEs) of attending each college (versus College 7) on firm pay premiums, ten years after high school graduation. Orange bars show mediation of this effect via firm location, defined as a categorical variable denoting whether a firm is a multi-establishment enterprise, located in the state's Primary Metropolitan Area, or located in another county. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Table 8: College Effects on Firm Premia, Marginally and Net of Location

College	Mediation via Firm	Mediation via Firm, Net of Place
College 6	-0.014 (0.0033)	-0.011 (0.0033)
College 5	-0.011 (0.0034)	-0.012 (0.0033)
College 4	0.027 (0.0032)	0.025 (0.0032)
College 2	0.047 (0.0028)	0.038 (0.0026)
College 3	0.009 (0.0023)	0.002 (0.0022)
Flagship	0.068 (0.003)	0.051 (0.0028)

Notes: Point estimates and standard errors shown in Figure S16. See Notes for Figure S16.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Figure S8: Direct Effects of College Attendance on Firm Premium, Net of Employer Location



Notes: Bars show estimated controlled direct effects (CDEs) of attending each college (versus College 7) on firm pay premiums. "Primary Metropolitan Area" and "Other Area" denotes working for single-establishment employers in the Primary Metropolitan Area or a non-Primary Metropolitan county, respectively. "Multi-Establishment" denotes working for a firm with multiple establishments; in the administrative UI records, I am unable to locate the specific establishment of employment of workers employed by multi-establishment firms. Each panel holds firm location fixed - to the Primary Metropolitan Area (left panel) multi-establishment firms (middle panel) or other areas (right panel) - and estimates the effect of college attendance on the fixed effect of a student's employer within that location category, ten years after high school graduation. Estimates reflect the effect of college attendance on the average fixed effect of a student's employer ten years after high school graduation, adjusting for firm location. By conditioning on firm location, these estimates isolate the portion of the college effect that operates net of sorting into geographic or organizational location, and thus adjust for selection into different types of firms. The colleges included are Summit State's Flagship school, and Colleges 2 to 6. Estimates are obtained using a double machine learning (DML) approach described in Appendix B, with ridge regression used to estimate both propensity scores and outcome models. Control variables include student race/ethnicity, sex, high school graduation cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

E.4.2 Industry

Second, I examine whether firm sorting stems from industry-based channeling. Industries differ systematically in their average wage premiums (Yi et al. 2024). Variation in the types of industries into which different colleges channel their graduates (such as through variation in industry-specific campus targeting) could account for firm premium effects. Using the same analytic strategy as for location, I find (in Figure S9) that industry does not explain the observed sorting patterns.

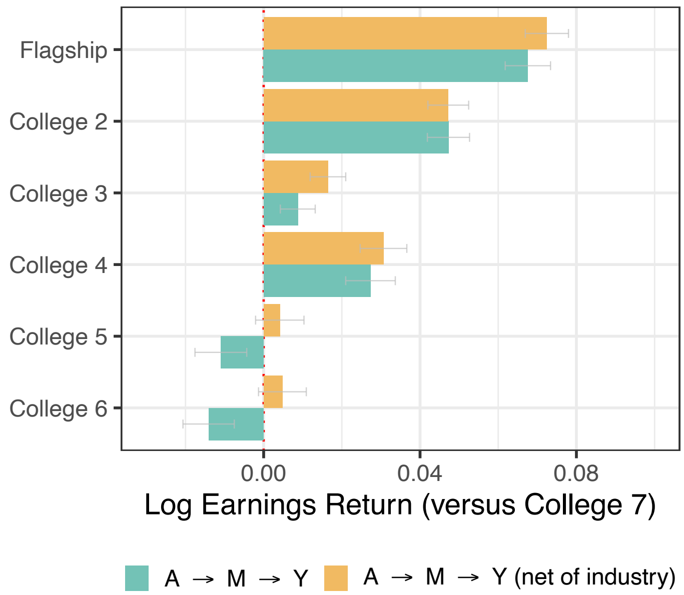
To show how firm sorting advantages persist even within industry groups, I estimate controlled direct effects (CDEs) of college attendance on firm premiums, now controlling for the industry of the employer, and for students' field of study. Figure S10 presents these estimates across four major sectors of employment (Finance and Insurance, Health Care and Social Assistance, Information, and Professional, Scientific, and Technical (PST) Services). I fix industry and estimate the average firm premium associated with each college, net of sorting into particular industry sectors (fields of study). Neither industry nor field of study sorting can explain firm premium payoffs to college attendance. Notably, within the Information and PST Services sectors, the Flagship and College 2 graduates enjoy huge firm premiums; within Information, the Flagship attendance sorts workers into firms that pay roughly 0.20 log points more College 7 attendance. Despite a positive payoff, the Flagship firm premium effects are significantly smaller in the Finance and Insurance and Health Care Industries.[26]

In sum, geographic and industry-based based sorting only partly overlaps with the firm sorting mechanism that I document here.[27]

[26]This is partly a mechanical artifact of the type of firms within these industries. Figure S1 in Appendix (C) shows that, despite having one of the highest (and lowest) average pay premiums, the variance of pay premiums in the Finance and Insurance (and Health Care) industry is significantly lower than in other sectors. This means that cross-firm sorting effects are attenuated relative to other industries.

[27]In Appendix E.4, I conduct a formal mediation analysis for the effects of college attendance on firm premiums net of location, industry and, additionally size.. I find that these characteristics explain only a small amount of the total flagship firm premium.

Figure S9: College Effects on Firm Premia, Marginally and Net of Industry



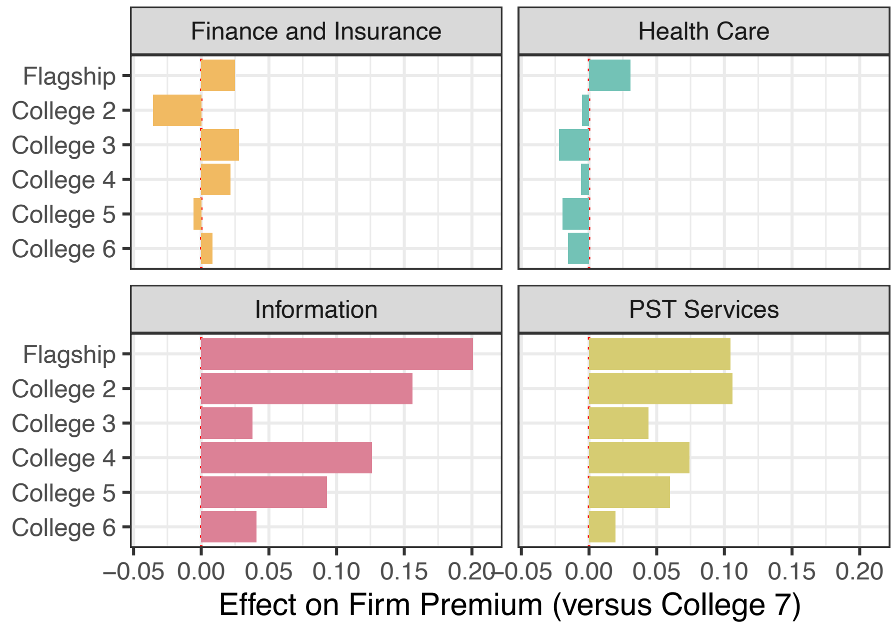
Notes: Teal bars show estimated average treatment effects (ATEs) of attending each college (versus College 7) on firm pay premiums, ten years after high school graduation. Orange bars show mediation of this effect via the industry in which the firm operates, captured via 2-digits NAICS codes. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Table 9: College Effects on Firm Premia, Marginally and Net of Industry

College	Mediation via Firm	Mediation via Firm, Net of Industry
College 6	-0.014 (0.0033)	0.005 (0.0031)
College 5	-0.011 (0.0034)	0.004 (0.0032)
College 4	0.027 (0.0032)	0.031 (0.003)
College 2	0.047 (0.0028)	0.047 (0.0027)
College 3	0.009 (0.0023)	0.016 (0.0023)
Flagship	0.068 (0.003)	0.072 (0.0028)

Notes: Point estimates and standard errors shown in Figure S9. See Notes for Figure S9.
Source: Summit State administrative UI earnings, postsecondary and high school records.

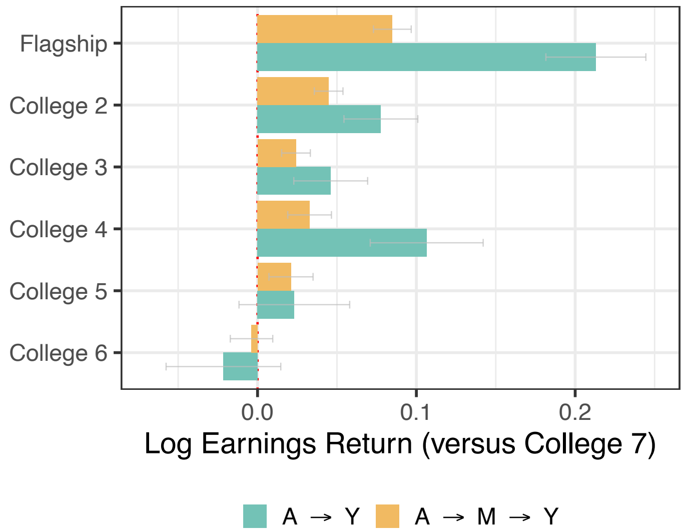
Figure S10: Direct Effects of College Attendance on Firm Premium, Net of Employer Industry



Notes: Bars show estimated controlled direct effects (CDEs) of attending each college (versus College 7) on firm pay premiums, adjusting for industry of employment. Each panel holds industry fixed and estimates the effect of college attendance on the fixed effect of an individual's employer within that industry, measured ten years after high school graduation. Estimates reflect the effect of college attendance on the average fixed effect of a student's employer ten years after high school graduation, adjusting for industry sorting. By conditioning on firm industry, these estimates isolate the portion of the college effect that operates net of sorting into particular industry sectors, and thus adjust for selection into different types of firms. The colleges included are Summit State's Flagship school, and Colleges 2 to 6. Estimates are obtained using a double machine learning (DML) approach described in Appendix B, with ridge regression used to estimate both propensity scores and outcome models. Control variables include student race/ethnicity, sex, high school graduation cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

E.5 College Attendance Effects, 15 Years Following High School Graduation

Figure S11: College Attendance Effects and Mediation via Firms: 15 Years Following High School Graduation



Notes: Teal bars show estimated average treatment effects (ATEs) of attending each college (versus College 7) on log earnings ten years after high school graduation. Orange bars show estimated ATEs on firm pay premiums, capturing the portion of the total effect mediated through differences in employer assignment. The colleges included are Summit State's Flagship school, and Colleges 2 to 6. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. Grey whiskers indicate 95% confidence intervals. $N = 23,289$; high school graduating classes of 2006-2007.
Source: Summit State administrative UI earnings, postsecondary and high school records.

F Descriptive Statistics on Educational Outcomes

F.1 Engagement with non-Summit State Public Postsecondary System

This section shows that the public 4-year postsecondary sector in Summit State is a favored destination for postsecondary goers, and especially for students pursuing postsecondary education in-state. Among 4-year college starters (that is, freshman admits) among the 2006-2012 high school graduating cohorts, 59% (24/(24+4+5+8)) began at a 4-year Summit State public college (corresponding to 24% of all high school graduates). The in-state 2-year public college sector enrolls the largest share of Summit State high school graduates, at 54%. By contrast, only 10% (4/(24+4+5+8)) of 4-year starters began at a 4-year private in-state college. Students enrolling in an in-state public 4-year college are broadly similar to those attending a 4-year private or out-of-state college (though are more likely to be white, and have slightly lower high school GPAs) (Table 12 in Appendix F.1) .

Table 10: High School Graduates' Engagement with the Postsecondary Sector

	Proportion of HS Graduates (N = 448,795)
Ever Attended 4 Year College	0.59
BA Within 10 Years of HS Completion	0.42
First College: In-State 4-Year Public	0.24
First College: In-State 2-Year Public	0.54
First College: In-State 4-Year Private	0.04
First College: In-State 2-Year Private	0.01
First College: OOS 4-Year Public	0.05
First College: OOS 4-Year Private	0.08
First College: OOS 2-Year Private	0.00

Notes: "Ever Attended 4 Year College" denotes whether an individual ever attended a 4-year college with 10 years following high school graduation. High school graduating classes 2006-2012 shown.
Source: Summit State administrative UI earnings, postsecondary and high school records, and National Student Clearinghouse.

Table 12: Characteristics of High School Graduates by First College Type

	First College Attended	
	4 Year Summit State Public	4 Year Private or OOS
Female	0.54	0.56
Black	0.03	0.04
Hispanic	0.06	0.05
White	0.72	0.79
English Speaking	0.88	0.93
Final HS GPA	3.44	3.34
Ever Homeless	0.01	0.01
Ever FRPL	0.21	0.20
Ever SPED	0.01	0.03
Ever ELL	0.02	0.01
Probability of Non-Zero Earnings	0.59	0.38
N (Individuals)	85,313	63,161

Notes: Descriptive statistics for students whose first college attended was either a Summit State public four-year institution (or "4 Year Summit State Public") a private or out-of-state four-year institution ("4 Year Summit State or OOS"). High school graduating classes 2006-2012, who were were first-time college enrollees.
Source: Summit State administrative UI earnings, postsecondary and high school records, and National Student Clearinghouse.

F.2 Educational Outcomes (by First College Attended)

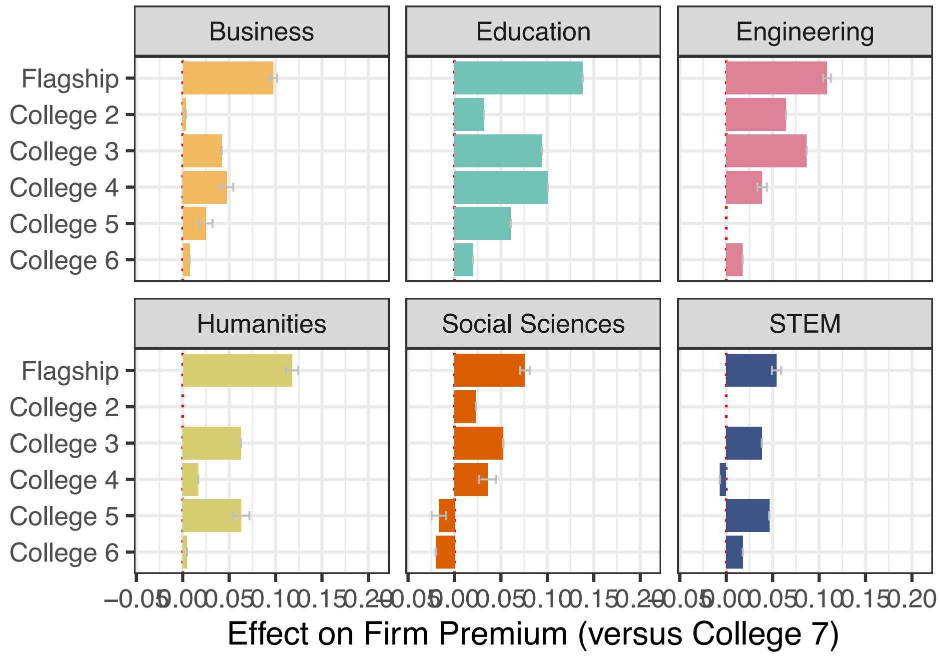
Table 14: Educational Outcomes by College of First Enrollment

	College Attended						
	College 7	College 6	College 5	College 4	College 3	College 2	Flagship
First Term GPA	2.85	2.92	2.76	2.86	3.13	2.89	3.21
First Term Credits	11.02	10.75	11.67	11.40	13.57	12.44	12.17
AA Within 10 Years of HS Grad.	0.13	0.13	0.13	0.12	0.12	0.21	0.08
BA Within 10 Years of HS Grad.	0.62	0.67	0.82	0.77	0.77	0.71	0.90
BA from First College	0.49	0.56	0.72	0.66	0.62	0.56	0.83
GPA at Graduation	3.31	3.17	3.17	3.23	3.20	3.16	3.31
Major: Humanities	0.12	0.14	0.24	0.09	0.20	0.12	0.12
Major: STEM	0.10	0.09	0.09	0.09	0.10	0.02	0.19
Major: Social Science	0.19	0.17	0.22	0.18	0.16	0.30	0.26
Major: Health	0.08	0.03	0.03	0.02	0.03	0.02	0.03
Major: Business	0.16	0.13	0.15	0.18	0.23	0.15	0.12
Major: Engineering	0.08	0.08	0.06	0.13	0.14	0.12	0.16
Major: Education	0.10	0.12	0.05	0.04	0.01	0.00	0.02
Major: Other Professional	0.14	0.12	0.16	0.19	0.12	0.21	0.08
Ever Transferred to 2-Year In-State Public	0.35	0.33	0.29	0.35	0.34	0.37	0.22
Ever Transferred to 4-Year In-State Public	0.19	0.16	0.14	0.15	0.31	0.24	0.08
Ever Transferred to NSC	0.11	0.09	0.06	0.07	0.06	0.09	0.03
Has Post-Graduate Degree	0.09	0.07	0.10	0.11	0.06	0.10	0.17

Notes: Educational outcomes for students whose first college attended was a Summit State public four-year institution, among high school graduating classes 2006-2012. All statistics refer to outcomes within 10 years of high school graduation.
Source: Summit State administrative UI earnings, postsecondary and high school records, and National Student Clearinghouse.

G Supplemental Human Capital Results

Figure S12: Direct Effects of College Attendance on Firm Premium, Net of Field of Study



Notes: Bars show estimated controlled direct effects (CDEs) of attending each college (versus College 7) on firm pay premiums, adjusting for field of study. Each panel holds field of study (at time of graduation) fixed and estimates the effect of college attendance on the fixed effect of an individual's employer ten years after high school graduation. Estimates reflect the effect of college attendance on the average fixed effect of a student's employer ten years after high school graduation, adjusting for field of study sorting. I exclude individuals who did not graduate with a BA, because these students often have an undeclared field of study. The colleges included are Summit State's Flagship school, and Colleges 2 to 6. Estimates are obtained using a double machine learning (DML) approach described in Appendix B, with ridge regression used to estimate both propensity scores and outcome models. Control variables include student race/ethnicity, sex, high school graduation cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

H Table Outputs: Main Results

H.1 Raw Earnings Gaps

Table 16: Raw Log Earnings Gaps Among Summit State 4-Year Starters

College	Raw Earnings Premium	Raw Firm Premium
College 7	10.738 (0.0091)	0.089 (0.0036)
College 6	0.081 (0.0126)	0.023 (0.005)
College 5	0.135 (0.0114)	0.039 (0.0046)
College 4	0.245 (0.0109)	0.073 (0.0043)
College 2	0.256 (0.021)	0.124 (0.0083)
College 3	0.136 (0.0236)	0.058 (0.0094)
Flagship	0.411 (0.0106)	0.141 (0.0042)

Notes: Point estimates and standard errors displayed in Figure 2 in the main text. Raw log earnings returns, relative to College 7, for students who attended (as a freshman admit) six public 4-year institutions in Summit State. The colleges included are Summit State's Flagship school, and Colleges 2 to 6. Earnings returns are measured ten years after high school graduation, while firm returns correspond to the estimated fixed effect of an individual's employer, also ten years after high school graduation. Standard errors are in parentheses. Numbers with asterisks denote means, rather than differences in means. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

H.2 College Attendance Effects

Table 17: College Attendance Effects and Mediation via Firms

College	ATE	Mediation via Firm
College 7	10.91 (0.0052)	NA
College 6	-0.017 (0.0079)	-0.014 (0.0033)
College 5	-0.046 (0.0081)	-0.011 (0.0034)
College 4	0.085 (0.0076)	0.027 (0.0032)
College 2	0.06 (0.0063)	0.047 (0.0028)
College 3	0.015 (0.0052)	0.009 (0.0023)
Flagship	0.134 (0.0071)	0.068 (0.003)

Notes: Point estimates and standard errors displayed in Figure 2 in the main text. Estimated average treatment effects (ATEs) of attending each college (versus College 7) on log earnings ten years after high school graduation, and mediation via employer assignment. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). Standard errors are in parentheses. Numbers with asterisks denote counterfactual means, rather than treatment effects. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

H.3 Quantile Treatment Effects

Table 18: College Attendance Effects on Probability of Reaching a Firm with a Premium $\geq$ pth Percentile

	Percentile (p) of Firm Premium				
College	95th	90th	80th	70th	60th
College 6	-0.01 (0.0021)	-0.006 (0.0032)	-0.009 (0.0045)	-0.01 (0.0054)	-0.018 (0.0061)
College 5	0 (0.0024)	0.001 (0.0033)	-0.001 (0.0046)	-0.002 (0.0055)	-0.004 (0.0061)
College 4	0.009 (0.0024)	0.022 (0.0033)	0.045 (0.0046)	0.058 (0.0054)	0.065 (0.0059)
College 2	0.025 (0.0022)	0.041 (0.0029)	0.071 (0.0039)	0.083 (0.0046)	0.091 (0.005)
College 3	0.003 (0.0016)	0.01 (0.0021)	0.03 (0.003)	0.047 (0.0037)	0.021 (0.0041)
Flagship	0.031 (0.0024)	0.054 (0.0032)	0.092 (0.0043)	0.118 (0.005)	0.123 (0.0054)

Notes: Point estimates and standard errors displayed in Figure 4 in the main text. Estimated average treatment effect (ATE) of attending a given college (versus College 7) on the probability of working at a firm in the indicated percentile or higher of the firm fixed effect (firm premium) distribution, measured ten years after high school graduation. Estimates are computed separately for the 60th, 70th, 80th, 90th, and 95th percentiles of the firm premium distribution, defined within the set of 4-year freshman enrollees. Effects are estimated using a double machine learning (DML) approach described in Appendix B, with ridge regression used for both propensity score and outcome models. Control variables include student race/ethnicity, sex, high school graduation cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). Standard errors are in parentheses. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

H.4 Human Capital Sorting

Table 19: College Attendance Effects on Firm Premium, Isolated from Human Capital

College	Firm Sorting Effect	Net of PE	Net of GPA	Net of FOS
College 6	-0.014 (0.0023)	-0.007 (0.0027)	-0.012 (0.0025)	-0.004 (0.0023)
College 5	-0.011 (0.0024)	-0.002 (0.0028)	-0.006 (0.0025)	-0.006 (0.0024)
College 4	0.027 (0.0023)	0.033 (0.0027)	0.035 (0.0024)	0.024 (0.0022)
College 2	0.047 (0.0019)	0.047 (0.0023)	0.055 (0.002)	0.044 (0.0019)
College 3	0.009 (0.0016)	0.017 (0.0019)	0.014 (0.0017)	-0.003 (0.0016)
Flagship	0.068 (0.0021)	0.069 (0.0025)	0.077 (0.0022)	0.065 (0.0021)

Notes: Point estimates and standard errors displayed in Figure 4 in the main text. Estimated average treatment effect (ATE) of attending a given college (versus College 7) on the probability of working at a firm in the indicated percentile or higher of the firm fixed effect (firm premium) distribution, measured ten years after high school graduation. Estimates are computed separately for the 60th, 70th, 80th, 90th, and 95th percentiles of the firm premium distribution, defined within the set of 4-year freshman enrollees. Effects are estimated using a double machine learning (DML) approach described in Appendix B, with ridge regression used for both propensity score and outcome models. Control variables include student race/ethnicity, sex, high school graduation cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). Standard errors are in parentheses. $N = 48,769$ (mediation); $N = 32,976$ (person effects); $N = 43,231$ (college GPA), and $N = 47,881$ (field of study); high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

I Additional Results

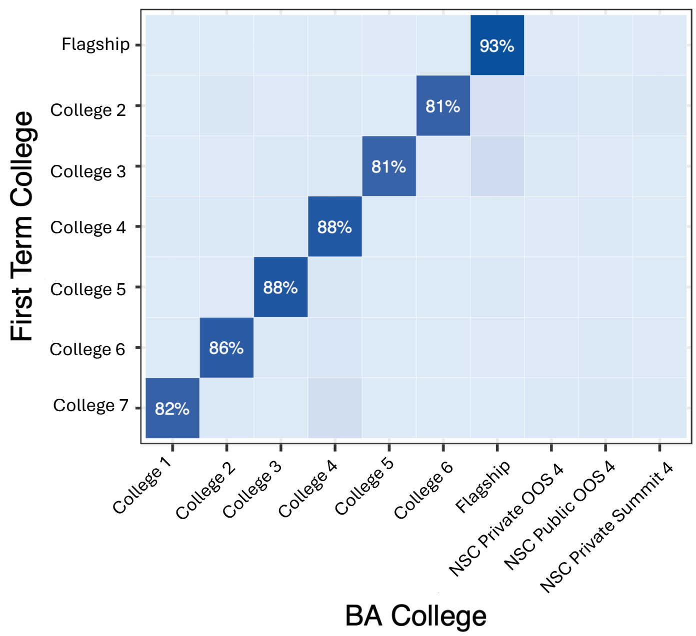
I.1 BA Completion and Transfer Students

Table 20: Predictors of Transferring out of the state's Flagship

	Transfer	
	(1)	(2)
Constant	0.1810***	
	(0.0193)	
Female	-0.0094***	-0.0104***
	(0.0028)	(0.0028)
Black	-0.0290***	-0.0329***
	(0.0088)	(0.0064)
Hispanic	0.0063	-0.0097
	(0.0068)	(0.0076)
white	0.0147***	0.0086**
	(0.0031)	(0.0035)
finalgpa	0.0057	0.0014
	(0.0051)	(0.0046)
Ever Homeless	0.0212	0.0130
	(0.0218)	(0.0265)
Ever FRPL	0.0034	-0.0038
	(0.0037)	(0.0042)
Ever SPED	0.0021	0.0047
	(0.0202)	(0.0218)
Ever ELL	0.0219**	0.0269***
	(0.0085)	(0.0099)
first_term_gpa	-0.0435***	-0.0407***
	(0.0028)	(0.0041)
first_term_cred	-0.0016***	-0.0016***
	(0.0005)	(0.0005)
Standard-Errors	IID	HS
Observations	19,720	19,720
R^2	0.01765	0.04986
Within R^2		0.01574
HS fixed effects		✓

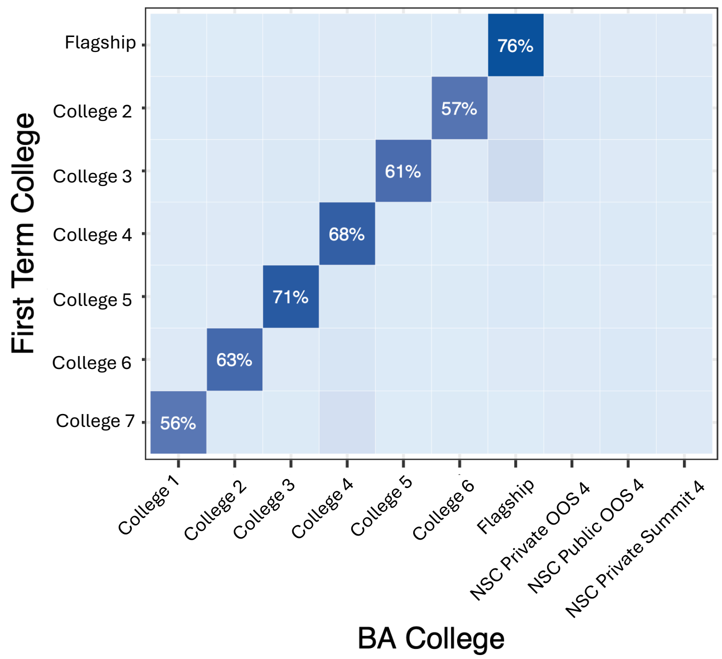
Notes: Linear probability model estimates of the likelihood that a student transfers from the state's Flagship, given that they began their postsecondary studies at the Flagship. The outcome variable equals one if the student completes a BA at a different institution than where they initially enrolled. Column (1) includes student demographic and academic covariates only; column (2) adds high school fixed effects to account for unobserved differences across sending schools. Key predictors include first-term GPA and first-term credits, both of which are strongly negatively associated with transfer likelihood. Female, Black, and English Language Learner (ELL) students show statistically significant differences in transfer rates. Standard errors are clustered at the individual level in column (1) and at the high school level in column (2).
Source: Summit State administrative UI earnings, postsecondary and high school records, as well as National Student Clearinghouse.

Figure S13: Observed Conditional Probabilities of Degree Completion Across Colleges, Given First College and Given BA Attained



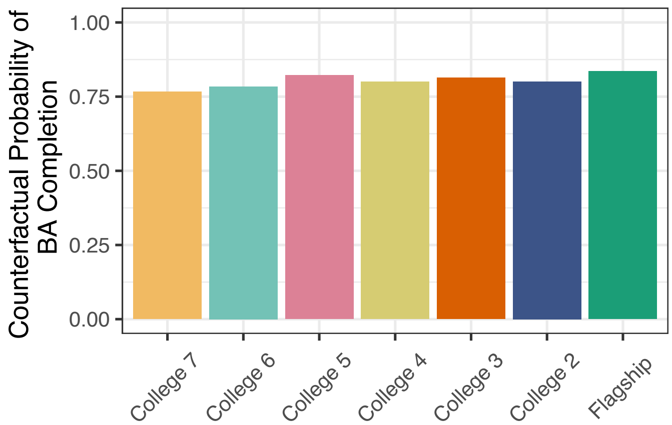
Notes: Observed probabilities of BA completion at a given college, conditional on where a student first enrolled as a freshman admit (following high school graduation). Each cell is colored according to the share of students who began at the institution on (y-axis) and ultimately completed a BA at the institution (x-axis), with darker colors indicating a greater share of students. Diagonal entries indicate the share of students who completed a degree at their first-term college. Off-diagonal entries (now shown, due to low sample sizes in some cells) represent cross-institutional completion, including completions at private or out-of-state four-year institutions. Cell entries condition on BA completion; entries will sum row-wise to the probability of BA completion, given that an individual begins their postsecondary education at a given college. $N = 85,313$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records, as well as National Student Clearinghouse.

Figure S14: Counterfactual Conditional Probabilities of Degree Completion Across Colleges, Given First College and Given BA Attained



Notes: Counterfactual probabilities of BA completion at a given college, conditional on where a student first enrolled as a freshman admit (following high school graduation). Each cell is colored according to the counterfactual share of students who began at the institution on (y-axis) and ultimately completed a BA at the institution (x-axis), with darker colors indicating a greater share of students. Diagonal entries indicate the counterfactual share of students who completed a degree at their first-term college. Off-diagonal entries (now shown, due to low sample sizes in some cells) represent cross-institutional counterfactual completion, including completions at private or out-of-state four-year institutions. Cell entries condition on BA completion; entries will sum row-wise to the counterfactual probability of BA completion, given that an individual begins (possibly contrary to fact) their postsecondary education at a given college. High school graduating classes of 2006-2012. Counterfactual probabilities are obtained by reweighing the full sample of 4-year starters by $\frac{\mathbb{I}(A=j)}{\hat{\Pr}(A=j|X)}$, that is, a weight equal to 1 divided by the probability of attending the college that they did in fact attend. $\hat{\Pr}(A = j \mid X)$ is estimated via ridge regression with the following covariates: student race/ethnicity, sex, high school graduation cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), and ever classified as an English language learner (ELL). $N = 83,519$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records, as well as National Student Clearinghouse.

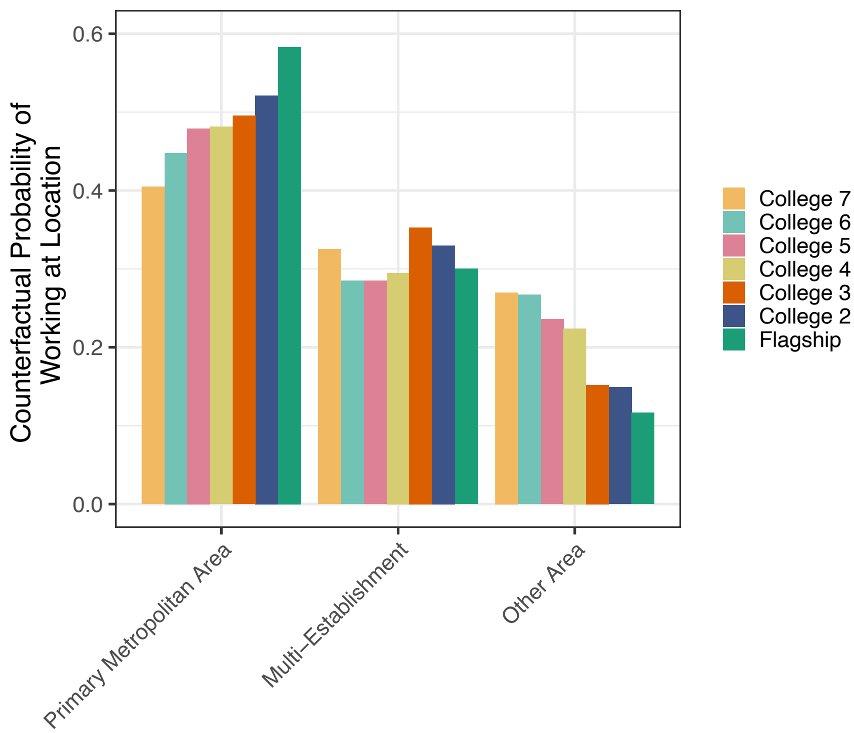
Figure S15: Counterfactual Probabilities of Any BA Attainment by College Attended



Notes: Bars show counterfactual probabilities of attaining a BA (at any college) within 10 years following high school graduation. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

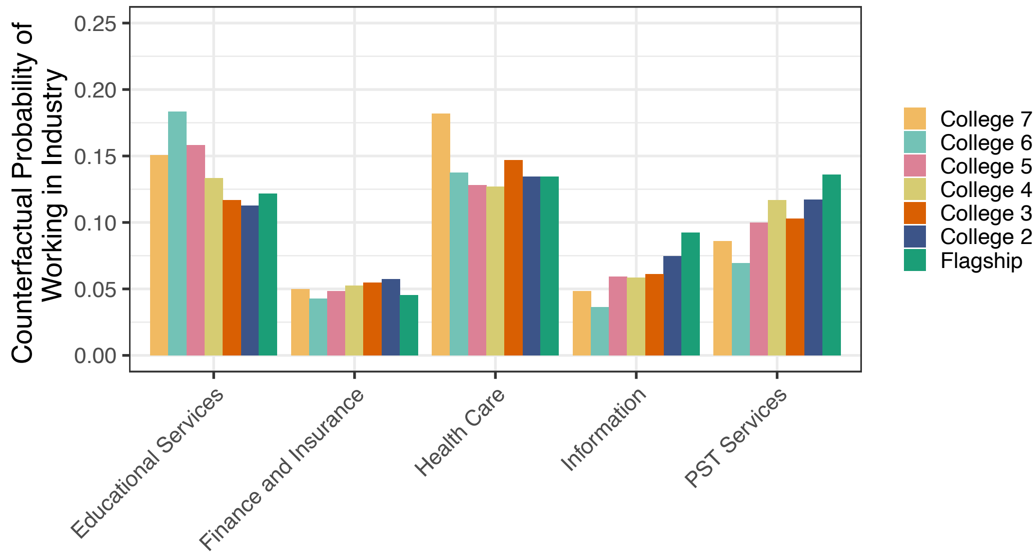
I.2 Counterfactual Firm Location, Industry and Size

Figure S16: Counterfactual Probabilities of Working in the Primary Metropolitan Area, a Multi-Establishment Firm, or Outside the Primary Metropolitan Area, by College Attended



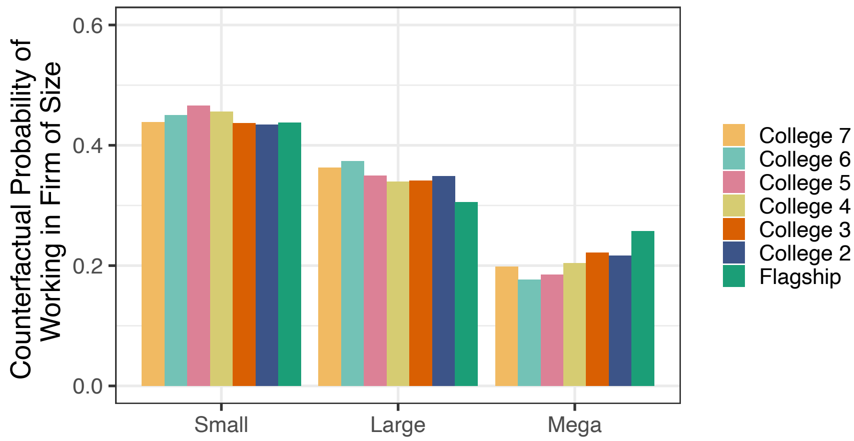
Notes: Bars show counterfactual probabilities of working in a firm located in either the state's Primary Metropolitan Area, a multi-establishment firm, or in a non-Primary Metropolitan Area county. "Primary Metropolitan Area" and "Other Area" denotes working for single-establishment employers in the state's Primary Metropolitan Area or a non-Primary Metropolitan Area, respectively. S "Multi-Establishment" denotes working for a firm with multiple establishments; in the administrative UI records, I am unable to locate the specific establishment of employment of workers employed by multi-establishment firms. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Figure S17: Counterfactual Probabilities of Working in Select Industries, by College Attended



Notes: Bars show counterfactual probabilities of working in firm that operates in a particular industry. "PST Services" refers to the Professional, Scientific, and Technical Services sector. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Figure S18: Counterfactual Probabilities of Working in Firm of Given Size, by College Attended



Notes: Bars show counterfactual probabilities of working in firm of a given size. "Small" refers to employers with fewer than 1,000 employees; "Large" refers to employers with 1,000-10,000 employees; "Mega" refers to firms with over 10,000 employees. Size is measured as the average (across months) number of employees working at a firm in the relevant year. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

I.3 Screening as a Mechanism

My theoretical framework highlights several ways in which credentials may sort graduates into different types of firm. In principle, explicit credentialing (that is, employers' screening behaviors - on account of either human capital screening or social closure) could be empirically distinguished from informational network effects (which capture job seekers' search behaviors) by observing differences in application behavior to firms by credential type, and then in offer likelihood given application). Unfortunately, in the administrative data, job application information is not observed.

Nevertheless, to gain some traction on this issue, I exploit the fact that a non-trivial proportion of individuals who begin their studies at one college attain a BA degree at another. Figure S13 in Appendix (I.1) shows that, conditional on attaining a BA degree, some 9% of students who initially attended the Flagship as a freshman admit end up completing their degree elsewhere.[28] I use this variation to examine the effect of attaining a BA degree from the Flagship, given that a person starts their postsecondary education at the Flagship, relative to starting college at the Flagship and completing elsewhere (mechanically, at a lower-ranked school). The core intuition here is that a student may still benefit from non-human capital inputs from the Flagship even if they transfer prior to degree completion at the Flagship. By contrast, any purely credentialing advantages are likely reduced, for the simple reason that the Flagship starters but non-completers cannot list a degree from the Flagship on their resumes. In the administrative data, the Flagship starters who finish a degree elsewhere roughly split their time between initial and final college, spending on average 6.6 semesters at the Flagship and 7.7 semesters at the school to which they transfer.

Exploiting this variation requires overcoming some important selection patterns. In Table 20 in Appendix I.1, I regress an indicator of out-transfer (from the Flagship) on pre-college as well as college-level covariates such as first-term GPA and credits completed. Column (1) includes student demographic and academic covariates only; column (2) adds high school fixed effects to account for unobserved differences across sending schools. Interestingly, while Female, Black, and English Language Learner (ELL) students show statistically significant differences in transfer rates, the strongest (negative) predictors of out-transfer are college-level covariates; high school GPA effects are statistically insignificant. The importance of college-level covariates suggests that more complex estimation strategy is required, in order to properly account for both pre-college, and college-level, covariates - the latter of which may themselves be affected by college attended (see e.g. Wodtke et al., 2011).

Appendix B details the strategy I employ; in short, I estimate cumulative treatment effects (CTEs) of (i) the Flagship attendance and the Flagship completion, and (ii) the

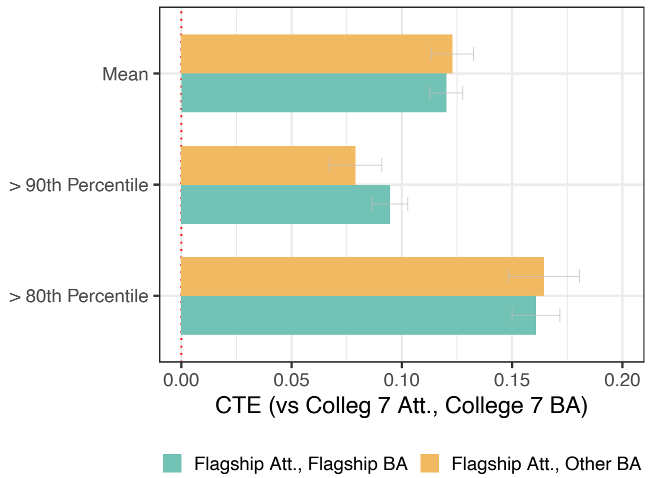
[28] Of the 7.5% of the Flagship starters who transfer out of the Flagship, 56% transfer to an in-state college. Of those, just under half transfer to College 3 or College 2.

Flagship attendance and other school completion, both relative to College 7 attendance and completion. I examine CTEs on three outcomes: (1) the average firm fixed effect, (2) the probability of working at a firm in the top 90th percentile of the firm premium distribution, and (3) the probability of working at a firm in the top 80th percentile. These quantities are all estimated using the set of controls described previously, as well as using college-level controls. In particular, high school fixed effects adjust for the fact that students may be moving closer to their home.

Figure S19 presents results (point estimates are shown in Table 21, Appendix H). The teal bars show the effect of attending and graduating from the Flagship, while the orange bars represent the effect of attending the Flagship but completing a degree elsewhere.[29] Both the Flagship starters-the Flagship completers and the Flagship-starters-other completers experience substantial firm premium advantages relative to College 7 graduates. On average, attending and graduating from the Flagship yields a roughly .12 log point firm premium advantage; by contrast, attending the Flagship and completing a degree elsewhere yields no significant loss or gain in firm premium (-0.003, 95% confidence interval $[-.006, .01]$) . This lack of gain is reflected in yielding no advantage in gaining access to firms at or above the 80th percentile of firm pay premiums. By contrast, there is a small and significant firm advantage to the Flagship completion when it comes to access to the highest premium firms. Specifically, the Flagship degree completion corresponds to a 2 percentage point increase in the probability of gaining access to a firm at or above the 90th percentile of firm pay (95% confidence interval $[.004, .28]$), representing a 20% increase on the baseline of 7.9 percentage points. These differences, while modest, suggest that holding the the Flagship credential confers an added labor market advantage - above and beyond the benefits of exposure to the the Flagship campus or peer environment - for access to the best paying employers. At the same time, a large portion of the total sorting advantage comes not from the credential itself, but from exposure to a distinct campus environment.

[29] Note that the Flagship attendance and completion yields a significantly higher firm premium, relative to the Flagship attendance effects shown in Figure (**??**). The reason for this discrepancy is that counterfactual rates of BA completion are almost identical, at approximately 77% across all colleges (Figure (S15) in Appendix (I.1)). Given the positive effect of BA completion on earnings, non-completion attenuates ATE estimates of college attendance, relative to CTE estimates of college attendance *and* completion.

Figure S19: Cumulative Treatment Effects (CTEs) of Attendance and Completion on Firm Outcomes



Notes: Estimated cumulative treatment effects of attending the University of the Flagship on firm wage outcomes, relative to attending (as a freshman admit) and completing a BA at College 7. Cumulative treatment effects are defined as Flagship attendance and Flagship completion (teal bars), and as Flagship attendance and BA completion at any other in-state, public college (orange bars). Outcomes include: (1) the average firm fixed effect, (2) the probability of working at a firm in the top 90th percentile of the firm premium distribution, and (3) the probability of working at a firm in the top 80th percentile. Percentiles are defined based on firm assignments among all public four-year college starters in Summit State. Estimates are obtained using the double machine learning (DML) approach described in Appendix B, with ridge regression used for all nuisance models. Covariates include those from previous models (demographics, academic background, and high school characteristics), along with additional controls for degree completion selection: GPA at graduation or stop-out, first-term credits attempted and earned, and their squares. Grey whiskers represent 95% confidence intervals. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

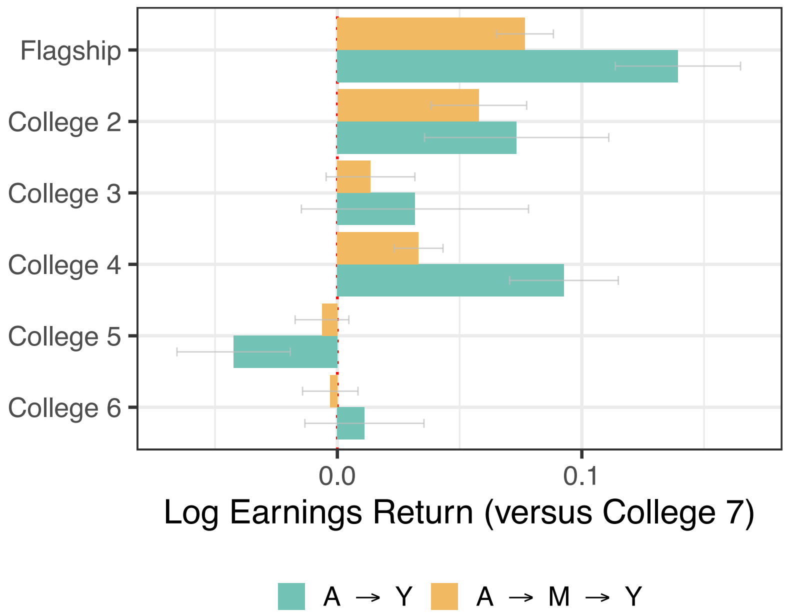
Table 21: Cumulative Treatment Effects (CTEs) of Attendance and Completion on Firm Outcomes

	Flagship Att., Flagship BA	Flagship Att., Other BA	Difference
Mean Firm Premium	0.12 (0.0038)	0.123 (0.0049)	-0.003 (0.0044)
> 90th Percentile of Firm Premium	0.095 (0.0041)	0.079 (0.0061)	0.016 (0.006)
> 80th Percentile of Firm Premium	0.161 (0.0055)	0.164 (0.0082)	-0.004 (0.0079)

Notes: Point estimates and standard errors displayed in Figure S19 in the main text. Estimated cumulative treatment effects of attending (as a freshman admit) and completing a degree at different colleges, relative to attending (as a freshman admit) and completing a BA at College 7. Cumulative treatment effects are defined as the Flagship attendance and the Flagship completion (the Flagship Att., the Flagship BA), and as the Flagship attendance and BA completion at any other in-state, public college (the Flagship Att., Other BA). Outcomes include: (1) the average firm fixed effect, (2) the probability of working at a firm in the top 90th percentile of the firm premium distribution, and (3) the probability of working at a firm in the top 80th percentile. Percentiles are defined based on firm assignments among all public four-year college starters in Summit State. Estimates are obtained using the double machine learning (DML) approach described in Appendix B, with ridge regression used for all nuisance models. Covariates include those from previous models (demographics, academic background, and high school characteristics), along with additional controls for degree completion selection: GPA at graduation or stop-out, first-term credits attempted and earned, and their squares. Standard errors are in parentheses. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records, as well as National Student Clearinghouse.

I.4 Linear-Based Estimates of ATE

Figure S20: College Attendance Effects and Mediation via Firms: Linear-Based Estimator



Notes: Teal bars show estimated average treatment effects (ATEs) of attending each college (versus College 7) on log earnings ten years after high school graduation. Orange bars show estimated ATEs on firm pay premiums, capturing the portion of the total effect mediated through differences in employer assignment. Estimates are obtained using linear models, based on a linear regression of log earnings on indicators for attendance at each college, and control variables (teal bars), and on a linear regression of estimated firm premia on indicators for attendance at each college, and control variables (orange bars). Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). The red dotted line marks the baseline (College 7) comparison group. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

Table 22: College Attendance Effects and Mediation via Firms: Linear-Based Estimator

	Log Earnings	Firm Premium
	(1)	(2)
College 6	0.0111	-0.0029
	(0.0124)	(0.0058)
College 5	-0.0424***	-0.0063
	(0.0118)	(0.0056)
College 4	0.0927***	0.0332***
	(0.0113)	(0.0051)
College 3	0.0733***	0.0579***
	(0.0192)	(0.0100)
College 2	0.0318	0.0136
	(0.0237)	(0.0093)
Flagship	0.1393***	0.0767***
	(0.0131)	(0.0059)
Female	-0.1798***	-0.0731***
	(0.0067)	(0.0028)
Black	-0.0658***	0.0089
	(0.0174)	(0.0063)
Hispanic	-0.0059	-0.0089
	(0.0156)	(0.0055)
Asian	0.0189*	0.0132***
	(0.0105)	(0.0038)
Has GED	0.2026*	0.0332
	(0.1214)	(0.0631)
HS Graduation Year	0.0208***	0.0065***
	(0.0018)	(0.0007)
Ever Homeless	-0.0485	-0.0435***
	(0.0341)	(0.0165)
Ever FRPL	-0.0482***	-0.0060*
	(0.0095)	(0.0035)
Ever SPED	-0.2262***	-0.0539***
	(0.0285)	(0.0100)
Ever ELL	-0.0081	0.0031
	(0.0244)	(0.0089)
HS GPA	-0.3849***	-0.0672***
	(0.0348)	(0.0187)
HS GPA, Squared	0.1034***	0.0148***
	(0.0058)	(0.0030)
Observations	48,808	48,808
R^2	0.11097	0.08110
Within R^2	0.07267	0.04355
HS fixed effects	✓	✓

Notes: Point estimates and standard errors displayed in Figure 22. Column 1 shows estimated average treatment effects (ATEs) of attending each college (versus College 7) on log earnings ten years after high school graduation, obtained from a linear regression of log earnings on indicators for attendance at each college, and control variables. Column 2 displays estimates of mediation via firms, based on a linear regression of estimated firm premia on indicators for attendance at each college, and control variables. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). $N = 48,769$; high school graduating classes of 2006-2012.
Estimated average treatment effects (ATEs) of attending each college (versus College 7) on log earnings ten years after high school graduation, and mediation via employer assignment. Estimates are obtained using a linear model-based, doubly robust estimator. This doubly robust estimator is equivalent to the EIF-based estimator $\hat{\mu}_j^{\text{EIF}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}_j(X_i) + \frac{\mathbb{I}(A_i=j)}{\hat{\pi}_j(X_i)}\left(Y_i - \hat{\mu}_j(X_i)\right)\right\}$ as shown in Appendix B, but where $\hat{\mu}_j(X_i)$ and $\hat{\pi}_j(X_i)$ are estimated using linear models without cross-fitting, rather than using machine-learning and cross-fitting, Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

I.5 Total Effect of any 4-Year College Attendance and Mediation via Firms

In this section, I compare my analyses for different college effects relative to the effect of attending a 4-year college rather than no college at all. Table 23 shows results from an analysis of the effect of any 4-year college attendance on earnings, relative to no postsecondary exposure. Interestingly, nearly three quarters of the 11% return to any 4-year attendance appears to be driven by within-, rather than-between firm sorting. This suggests that, relative to no postsecondary experience, going to college matters not because of firm placement but because of within-firm productivity and job gains. But, conditional on the decision to attend a 4-year college, further earnings gains derive primarily from the firm at which an individual gains their post-college employment.

Table 23: Effect of Any 4-year Exposure, and Mediation via Firms

College	ATE	Mediation via Firm
High School Only	10.622 (0.003)*	0.093 (0.0011)*
4-Year Attendance	0.108 (0.0037)	0.028 (0.0014)

Notes: Estimated average treatment effects (ATEs) of attending a 4-year college (either as a direct admit or as a transfer students) on log earnings ten years after high school graduation, relative to graduating high school but obtaining no postsecondary exposure, and mediation via employer assignment. Estimates are obtained using a double machine learning (DML) approach, as described in Appendix B, with ridge regression used for both the propensity score and outcome models, estimated with five-fold cross fitting. Grey whiskers denote 95% confidence intervals. Control variables include student race/ethnicity, sex, high school cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). Standard errors are in parentheses. Numbers with stars denote counterfactual means, rather than treatment effects. $N = 227,114$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

I.6 Linear Mediation Analysis

I also estimate a more conventional wage regression that includes firm fixed effects directly. This approach follows the empirical strategy of Engbom and Moser (2017) and allows for a simple decomposition of credential effects into within-firm and between-firm components. Specifically, I estimate the following model:

$$Y_i = \sum_j \beta_j \mathbb{I}(A_i = j) + \psi_{F(i)} X_i^\top \gamma + \gamma_t + \epsilon_i, \tag{7}$$

where y_{it} denotes the earnings of individual i 10 years following high school graduation; $\mathbb{I}(C_{it} = j)$ is an indicator for whether an individual attended college j; and β_j captures the credential-specific earnings premium, relative to a reference group (College 7 attendance). X_i denotes control variables which include student race/ethnicity, sex, high school graduation cohort, high school GPA and GPA squared, high school fixed effects, and indicators for whether the student was ever homeless, ever eligible for free or reduced-price lunch (FRPL), ever received special education services (SPED), or ever classified as an English language learner (ELL). Finally, $\psi_{F(i)}$ is a firm fixed effect for an individual's employer 10 years after high school graduation.

I estimate this model both with and without the firm fixed effects $\psi_{F(i)}$ to assess the extent to which credential premia reflect sorting into higher-premium firms. If attendance at college j increases earnings regardless of an individual's firm placement, the estimated β_j should remain stable with and without firm controls. Conversely, if college attendance primarily serve to facilitate access to higher-premium firms, then controlling for firm fixed effects should substantially attenuate the estimated β_j coefficients.

Table 24 presents results. Column (1) reports estimates from a model without firm fixed effects, while column (2) includes firm fixed effects to account for variation in wages attributable to employer sorting. In the model without firm fixed effects (column 1), I find sizable and statistically significant earnings premiums associated with attending certain four-year public institutions in Summit State. For example, attending the Flagship is associated with a log earnings premium of 0.139, or approximately 15%, relative to College 7 (the omitted category). Similar positive and significant returns are observed for Summit State University (9.3%) and College 2 (7.5%). However, once firm fixed effects are included in column (2), these college-specific earnings premiums decline substantially. The estimated premium for Flagship attendance drops to 1.1% and is no longer statistically significant, while the coefficient for College 2 becomes negative. The premium for Summit State University also falls by more than half, from 9.3% to 3.3%. These results suggest that a substantial portion of the observed earnings advantages for graduates of selective institutions reflects sorting into higher-premium firms, rather than differences in pay within firms. It is important to emphasize that this approach does not account for unobserved, time-invariant worker heterogeneity. If individuals with higher unobserved ability are more likely to attend more selective colleges and sort into higher-premium firms, then the firm fixed effects model may overstate the extent of mediation through firms. That is, controlling for firm fixed effects may "over-control" for endogenous sorting, leading to downward-biased estimates of the credential premium and overstated mediation effects. When calibrated against my main results, which are based on models that do adjust for selection into firms, this indeed appears to be the case.

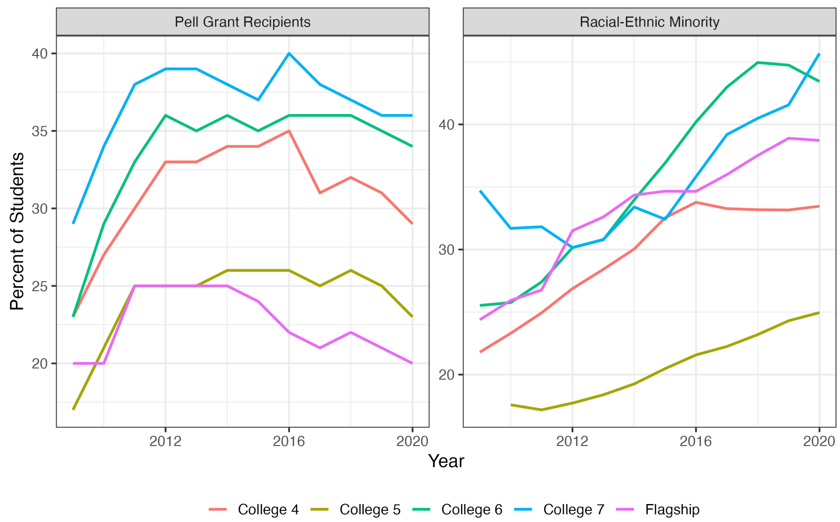
Table 24: Linear Mediation Analysis

	Log Earnings	
	M1: No Firm Effects	M2: With Firm Effects
	(1)	(2)
College 6	0.0111	0.0148
	(0.0124)	(0.0143)
College 5	-0.0424***	-0.0407***
	(0.0118)	(0.0143)
College 4	0.0927***	0.0318**
	(0.0113)	(0.0135)
College 3	0.0733***	-0.0368*
	(0.0192)	(0.0216)
College 2	0.0318	-0.0160
	(0.0237)	(0.0275)
Flagship	0.1393***	0.0103
	(0.0131)	(0.0148)
Female	-0.1798***	-0.0729***
	(0.0067)	(0.0070)
Black	-0.0658***	-0.0722***
	(0.0174)	(0.0198)
Hispanic	-0.0059	0.0014
	(0.0156)	(0.0150)
Asian	0.0189*	-0.0099
	(0.0105)	(0.0098)
Has GED	0.2026*	0.2254***
	(0.1214)	(0.0721)
HS Graduation Year	0.0208***	0.0122***
	(0.0018)	(0.0017)
Ever Homeless	-0.0485	0.0135
	(0.0341)	(0.0390)
Ever FRPL	-0.0482***	-0.0219***
	(0.0095)	(0.0079)
Ever SPED	-0.2262***	-0.1416***
	(0.0285)	(0.0320)
Ever ELL	-0.0081	-0.0398
	(0.0244)	(0.0253)
HS GPA	-0.3849***	-0.2134***
	(0.0348)	(0.0473)
HS GPA, Squared	0.1034***	0.0647***
	(0.0058)	(0.0075)
Observations	48,808	48,808
R^2	0.11097	0.59102
Within R^2	0.07267	0.03244
HS fixed effects	✓	✓
Firm fixed effects		✓

Notes: Results from linear regressions of log earnings ten years after high school graduation on college attended and student characteristics, estimated with and without firm fixed effects. Column (1) includes controls for high school fixed effects, student demographics, and academic background, but omits firm controls. Column (2) adds firm fixed effects to account for wage variation attributable to employer assignment. The coefficients on college indicators represent earnings differences relative to College 7, the omitted category. Comparing columns shows how much of the college earnings premium is accounted for by differences in the types of firms students work for. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $N = 48,769$; high school graduating classes of 2006-2012.
Source: Summit State administrative UI earnings, postsecondary and high school records.

I.7 Trends in Pell Grant and Minority Students at Summit State Public College

Figure S21: Trends in Pell Grant and Minority Students at Summit State Public Colleges, 2009–2020



Notes: Trends in the percentage of Pell Grant recipients (left panel) and racial-ethnic minority students (right panel) at Summit State's public four-year colleges from 2009 to 2020. Each line represents a different institution, highlighting variation in student demographics over time.
Source: Integrated Postsecondary Education Data System.

References

Andrew Abbott. *The system of professions: An essay on the division of expert labor*. University of Chicago press, 2014. ISBN 022618966X.

John M Abowd, Francis Kramarz, and David N Margolis. High wage workers and high wage firms. *Econometrica*, 67:251–333, 1999. ISSN 0012-9682.

John M Abowd, Robert H Creecy, and Francis Kramarz. Computing person and firm effects using linked longitudinal employer-employee data. Technical report, 2002.

Joseph G Altonji and Ling Zhong. The labor market returns to advanced degrees. *Journal of Labor Economics*, 39:303–360, 2021. ISSN 0734-306X.

Joseph G Altonji, Erica Blom, and Costas Meghir. Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu. Rev. Econ.*, 4:185–223, 2012. ISSN 1941-1383.

Joseph G Altonji, Peter Arcidiacono, and Arnaud Maurel. *The analysis of field choice in college and graduate school: Determinants and wage effects*, volume 5, pages 305–396. Elsevier, 2016. ISBN 1574-0692.

Rodney J Andrews, Scott A Imberman, Michael F Lovenheim, and Kevin M Stange. The returns to college major choice: Average and distributional effects, career trajectories, and earnings variability. Technical report, 2022.

Jaime Arellano-Bover. Career consequences of firm heterogeneity for young workers: First job and firm size. *Journal of Labor Economics*, 42:549–589, 2024. ISSN 0734-306X.

Elizabeth A Armstrong and Laura T Hamilton. *Paying for the Party*. Harvard University Press, 2013. ISBN 0674073517.

David Autor, Claudia Goldin, and Lawrence F Katz. Extending the race between education and technology. In *AEA papers and proceedings*, volume 110, pages 347–351. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2020. ISBN 2574-0768.

James N Baron and William T Bielby. Bringing the firms back in: Stratification, segmentation, and the organization of work. *American sociological review*, pages 737–765, 1980. ISSN 0003-1224.

Erling Barth, Alex Bryson, James C Davis, and Richard Freeman. It's where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics*, 34:S67–S97, 2016. ISSN 0734-306X.

Gary Becker. S. human capital: A theoretical and empirical analysis, with special reference to education. *Chicago: University of Press*, 19:4, 1964.

Gary S Becker. Investment in human capital: A theoretical analysis. *Journal of political economy*, 70:9–49, 1962. ISSN 0022-3808.

Jere R Behrman, Mark R Rosenzweig, and Paul Taubman. College choice and wages: Estimates using data on female twins. *The Review of Economics and Statistics*, pages 672–685, 1996. ISSN 0034-6535.

Stefan Bender, Nicholas Bloom, David Card, John Van Reenen, and Stefanie Wolter. Management practices, workforce selection, and productivity. *Journal of Labor Economics*, 36:S371–S409, 2018. ISSN 0734-306X.

Barbara Biasi and Song Ma. The education-innovation gap. Technical report, 2022.

David B Bills. Credentials, signals, and screens: Explaining the relationship between schooling and job assignment. *Review of educational research*, 73:441–469, 2003. ISSN 0034-6543.

Amy J Binder and Kate Wood. Becoming right: How campuses shape young conservatives. 2012.

Amy J Binder, Daniel B Davis, and Nick Bloom. Career funneling: How elite students learn to define and desire "prestigious"jobs. *Sociology of Education*, 89:20–39, 2016. ISSN 0038-0407.

Dan A Black and Jeffrey A Smith. How robust is the evidence on the effects of college quality? evidence from matching. *Journal of econometrics*, 121:99–124, 2004. ISSN 0304-4076.

Zachary Bleemer. Affirmative action and its race-neutral alternatives. *Journal of Public Economics*, 220:104839, 2023. ISSN 00472727. doi: 10.1016/j.jpubeco.2023.104839. URL https://doi.org/10.1016/j.jpubeco.2023.104839.

Zachary Bleemer and Aashish Mehta. Will studying economics make you rich? a regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics*, 14:1–22, 2022. ISSN 19457790. doi: 10.1257/APP.20200447.

Nicholas Bloom, Scott W Ohlmacher, Cristina J Tello-Trillo, and Melanie Wallskog. Pay, productivity and management. Technical report, 2021.

Thijs Bol, Christina Ciocca Eller, Herman G Van De Werfhorst, and Thomas A DiPrete. School-to-work linkages, educational mismatches, and labor market outcomes. *American Sociological Review*, 84:275–307, 2019. ISSN 0003-1224.

John Bound and Sarah Turner. *Dropouts and diplomas: The divergence in collegiate outcomes*, volume 4, pages 573–613. Elsevier, 2011. ISBN 1574-0692.

William G Bowen and Derek Bok. *The Shape of the River. Long-Term Consequences of Considering Race in College and University Admissions.* ERIC, 1998. ISBN 0691002746.

William G Bowen, Matthew M Chingos, and Michael S McPherson. *Crossing the finish line*. Princeton University Press, 2009. ISBN 1400831466.

Samuel Bowles and Herbert Gintis. *Schooling in capitalist America*, volume 57. New York: Basic Books, 1976.

Jennie E Brand and Charles N Halaby. Regression and matching estimates of the effects of elite college attendance on educational and career achievement. *Social Science Research*, 35:749–770, 2006. ISSN 0049-089X.

Michael Burawoy and Erik Olin Wright. Coercion and consent in contested exchange. *Politics & Society*, 18:251–266, 1990. ISSN 0032-3292.

David Card, Jörg Heining, and Patrick Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128:967–1015, 2013. ISSN 0033-5533.

David Card, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36: S13–S70, 2018. ISSN 0734-306X.

Hugh Cassidy and Amanda Gaulke. The increasing penalty to occupation-education mismatch. *Economic Inquiry*, 62:607–632, 2024. ISSN 0095-2583.

Raj Chetty, David J Deming, and John N Friedman. Diversifying society's leaders? the causal effects of admission to highly selective private colleges. Technical report, 2023.

Christina Ciocca-Eller. What makes a quality college? re-examining the equalizing potential of higher education in the united states. *American Journal of Sociology*, 9 2023. ISSN 0002-9602. doi: 10.1086/727891. URL https://doi.org/10.1086/727891. doi: 10.1086/727891.

Burton R Clark. The organizational saga in higher education. *Administrative science quarterly*, pages 178–184, 1972. ISSN 0001-8392.

Randall Collins. *The credential society: An historical sociology of education and stratification*. Columbia University Press, 2019. ISBN 0231549784.

Stacy Dale and Alan B Krueger. Estimating the return to college selectivity over the career using administrative earnings data. Technical report, 2011.

Stacy Berg Dale and Alan B Krueger. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, 117:1491–1527, 2002. ISSN 1531-4650.

Daniel Davis and Amy Binder. *Selling students: The rise of corporate partnership programs in university career centers*, volume 46, pages 395–422. Emerald Group Publishing Limited, 2016. ISBN 1785608312.

Daniel Davis and Amy Binder. Industry, firm, job title: The layered nature of early-career advantage for graduates of elite private universities. *Socius*, 5: 2378023119859711, 2019. ISSN 2378-0231.

David Deming and Lisa B Kahn. Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36:S337–S369, 2018. ISSN 0734-306X.

David J Deming, Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F Katz. The value of postsecondary credentials in the labor market: An experimental study. *American Economic Review*, 106:778–806, 2016. ISSN 0002-8282.

Thomas A DiPrete, Christina Ciocca Eller, Thijs Bol, and Herman G Van de Werfhorst. School-to-work linkages in the united states, germany, and france. *American journal of sociology*, 122:1869–1938, 2017. ISSN 0002-9602.

Thurston Domina, Andrew Penner, and Emily Penner. Categorical inequality: Schools as sorting machines. *Annual review of sociology*, 43:311–330, 2017. ISSN 0360-0572.

Neil Fligstein and Roberto M Fernandez. Worker power, firm power, and the structure of labor markets. *Sociological Quarterly*, 29:5–28, 1988. ISSN 0038-0253.

Ann M Gansemer-Topf and John H Schuh. Institutional selectivity and institutional expenditures: Examining organizational factors that contribute to retention and graduation. *Research in higher education*, 47:613–642, 2006. ISSN 0361-0365.

François Gerard, Lorenzo Lagos, Edson Severnini, and David Card. Assortative matching or exclusionary hiring? the impact of employment and pay policies on racial wage differences in brazil. *American Economic Review*, 111:3418–3457, 2021. ISSN 0002-8282.

Joshua Goodman, Michael Hurwitz, and Jonathan Smith. Access to 4-year public colleges and degree completion. *Journal of Labor Economics*, 35:829–867, 2017. ISSN 0734-306X.

Robert Granfield. Making elite lawyers: Visions of law at harvard and beyond, 1992.

Erica L Groshen. The structure of the female/male wage differential: Is it who you are, what you do, or where you work? *Journal of Human Resources*, pages 457–472, 1991. ISSN 0022-166X.

J Richard Hackman and Greg R Oldham. Motivation through the design of work: Test of a theory. *Organizational behavior and human performance*, 16:250–279, 1976. ISSN 0030-5073.

Laura Hamilton, Josipa Roksa, and Kelly Nielsen. Providing a "leg up": Parental involvement and opportunity hoarding in college. *Sociology of Education*, 91:111–131, 2018. ISSN 0038-0407.

Karen Ho. *Liquidated: An Ethnography of Wall Street*. Duke University Press., 2009.

Mark Hoekstra. The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics*, 91:717–724, 2009. ISSN 0034-6535.

Arne L Kalleberg, Michael Wallace, and Robert P Althauser. Economic segmentation, worker power, and income inequality. *American journal of sociology*, 87:651–683, 1981. ISSN 0002-9602.

Alan C Kerckhoff. The status attainment process: socialization or allocation? *Social forces*, 55:368–381, 1976. ISSN 1534-7605.

Patrick Kline. Firm wage effects. *Handbook of Labor Economics*, 5:115–181, 2024. ISSN 1573-4463.

Greta R Krippner. *Capitalizing on crisis: The political origins of the rise of finance.* Harvard University Press, 2012. ISBN 0674735315.

Nicholas Lemann. The kids in the conference room, 10 1999.

Robert Manduca. "i was open to anywhere, it's just this was easier:" social structure, location preferences, and the geographic concentration of elite college graduates. *Qualitative Sociology*, 47:153–185, 2024. ISSN 0162-0436.

Tatiana Melguizo, Gregory S Kienzl, and Mariana Alfonso. Comparing the educational attainment of community college transfer students and four-year college rising juniors using propensity score matching methods. *The Journal of Higher Education*, 82:265–291, 2011. ISSN 0022-1546.

John W Meyer. The effects of education as an institution. *American journal of Sociology*, 83:55–77, 1977. ISSN 0002-9602.

Valerie Michelman, Joseph Price, and Seth D Zimmerman. Old boys' clubs and upward mobility among the educational elite. *The Quarterly Journal of Economics*, 137:845–909, 2022. ISSN 0033-5533.

Jack Mountjoy. Marginal returns to public universities. Technical report, 2024.

Jack Mountjoy and Brent R Hickman. The returns to college (s): Relative value-added and match effects in higher education. Technical report, 2021.

Megan Tobias Neely. *Hedged out: Inequality and insecurity on Wall Street.* Univ of California Press, 2022. ISBN 0520307704.

Greg R Oldham and J Richard Hackman. Not what it was and not what it will be: The future of job design research. *Journal of organizational behavior*, 31:463–479, 2010. ISSN 0894-3796.

Paul Osterman. The wage effects of high performance work organization in manufacturing. *Ilr Review*, 59:187–204, 2006. ISSN 0019-7939.

Devah Pager and David S Pedulla. Race, self-selection, and the job search process. *American Journal of Sociology*, 120:1005–1054, 2015. ISSN 0002-9602.

Frank Parkin. Marxism and class theory: A bourgeois critique. 1983.

Frank Parkin. *Strategies of social closure in class formation*, pages 1–18. Routledge, 2018.

Natasha Quadlin, Emma D Cohen, and Tom VanHeuvelen. Same major, same economic returns? college selectivity and earnings inequality in young adulthood. *Research in Social Stratification and Mobility*, 75:100647, 2021. ISSN 0276-5624.

Aruna Ranganathan. When the tasks line up: How the nature of supplementary tasks affects worker productivity. *Ilr Review*, 76:556–585, 2023. ISSN 0019-7939.

Sean F Reardon, Rachel B Baker, and Daniel Klasik. Race, income, and enrollment patterns in highly selective colleges, 1982-2004. 2012.

Cecilia L Ridgeway. Why status matters for inequality. *American sociological review*, 79:1–16, 2014. ISSN 0003-1224.

Lauren Rivera. Pedigree: Elite reproduction in hiring, 2015.

Lauren A Rivera. Ivies, extracurriculars, and exclusion: Elite employers' use of educational credentials. *Research in Social Stratification and Mobility*, 29:71–90, 2011. ISSN 0276-5624.

Lauren A Rivera. Hiring as cultural matching: The case of elite professional service firms. *American sociological review*, 77:999–1022, 2012. ISSN 0003-1224.

Lauren A Rivera. Employer decision making. *Annual review of sociology*, 46:215–232, 2020. ISSN 0360-0572.

John Robst. Education and job match: The relatedness of college major and work. *Economics of Education review*, 26:397–407, 2007. ISSN 0272-7757.

James E Rosenbaum, Regina Deil-Amen, and Ann E Person. *After admission: From college access to college success.* Russell Sage Foundation, 2007. ISBN 1610444787.

Jake Rosenfeld. *You're paid what you're worth: And other myths of the modern economy.* Harvard University Press, 2021. ISBN 0674250850.

Jonathan Smith, Joshua Goodman, and Michael Hurwitz. The economic impact of access to public four-year colleges. Technical report, 2020.

Jae Song, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. Firming up inequality. *The Quarterly journal of economics*, 134:1–50, 2019. ISSN 0033-5533.

Michael Spence. Job market signaling. *The Quarterly Journal of Economics,*, 87: 355–374, 1973.

Mitchell L Stevens, Elizabeth A Armstrong, and Richard Arum. Sieve, incubator, temple, hub: Empirical and theoretical advances in the sociology of higher education. *Annual Review of Sociology*, 34:127–151, 7 2008. ISSN 0360-0572. doi: 10.1146/annurev.soc.34.040507.134737. URL https://doi.org/10.1146/annurev.soc.34.040507.134737. doi: 10.1146/annurev.soc.34.040507.134737.

Joseph E Stiglitz. The theory of" screening," education, and the distribution of income. *The American economic review*, 65:283–300, 1975. ISSN 0002-8282.

Jessi Streib. *The Accidental Equalizer: How Luck Determines Pay after College*. University of Chicago Press, 2023. ISBN 0226829685.

Jenny M Stuber. Talk of class: The discursive repertoires of white working-and upper-middle-class college students. *Journal of Contemporary Ethnography*, 35:285–318, 2006.

Jenny M Stuber. *Inside the college gates: How class and culture matter in higher education*. Lexington Books, 2011. ISBN 0739149008.

Megan Thiele and Brian Joseph Gillespie. Social stratification at the top rung: Classed reports of students' social experiences on a selective university campus. *Sociological Perspectives*, 60:113–131, 2017. ISSN 0731-1214.

Scott L Thomas and Liang Zhang. Post-baccalaureate wage growth within four years of graduation: The effects of college quality and college major. *Research in Higher Education*, 46:437–459, 2005. ISSN 0361-0365.

Lester C Thurow. *Generating inequality*. Basic Books, 1975.

Donald Tomaskovic-Devey and Dustin Robert Avent-Holt. *Relational inequalities: An organizational approach*. Oxford University Press, 2019. ISBN 0190624426.

Donald Tomaskovic-Devey, Melvin Thomas, and Kecia Johnson. Race and the accumulation of human capital across the career: A theoretical model and fixed-effects application. *American Journal of Sociology*, 111:58–89, 2005. ISSN 0002-9602.

Bill Townsend. Which company employs the most college graduates in your state?, 8 2023.

Kim A Weeden. Why do some occupations pay more than others? social closure and earnings inequality in the united states. *American journal of sociology*, 108:55–101, 2002. ISSN 0002-9602.

Kim A Weeden and David B Grusky. The three worlds of inequality. *American Journal of Sociology*, 117:1723–1785, 2012. ISSN 0002-9602.

Kim A Weeden and David B Grusky. Inequality and market failure. *American Behavioral Scientist*, 58:473–491, 2014. ISSN 0002-7642.

Nathan Wilmers. Job turf or variety: Task structure as a source of organizational inequality. *Administrative Science Quarterly*, 65:1018–1057, 2020. ISSN 0001-8392.

Nathan Wilmers and Clem Aeppli. Consolidated advantage: new organizational dynamics of wage inequality. *American Sociological Review*, 86:1100–1130, 2021. ISSN 0003-1224.

James C Witte and Arne L Kalleberg. Matching training and jobs: The fit between vocational education and employment in the german labour market. *European Sociological Review*, 11:293–317, 1995. ISSN 1468-2672.

Dirk Witteveen and Paul Attewell. The earnings payoff from attending a selective college. *Social Science Research*, 66:154–169, 2017. ISSN 0049-089X.

Geoffrey T Wodtke, David J Harding, and Felix Elwert. Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *American sociological review*, 76:713–736, 2011. ISSN 0003-1224.

Liang Zhang. *Does quality pay?: Benefits of attending a high-cost, prestigious college*. Routledge, 2012. ISBN 1135923175.

Seth D Zimmerman. The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32:711–754, 2014. ISSN 0734-306X.