1 Introduction

Colleges and universities (herein universities) identify prospective students by purchasing "student lists" from College Board, ACT, and other vendors. A Student list contains the contact information of prospective students who meet the search filter criteria (e.g., test score range, high school GPA, zip codes) specified by the university. Purchased lists are a fundamental input for undergraduate recruiting campaigns, which target individual prospects by mail, email, and on social media (EAB, 2018).

Recent research suggests that student lists are surprisingly important for college access – and degree completion – for millions of students each year. Howell et al. (2021) compared SAT test-takers who opted into the College Board Student Search Service – allowing accredited institutions to "licence" their contact information – to those who opted out.[1] Figure 1 reproduces the main results. After controlling for covariates (e.g., SAT score, parental education, high school), 41.1% of students who participated in Search attended a 4-year college compared to 32.8% of students who opted out, an 8.3 percentage point difference and a 25.3 (`(41.1-32.8)/32.8`) percent change in the relative probability. Participating in Search was associated with larger change in the relative probability of attending a 4-year college for students who identified as Black (24.5%) and Latinx (34.4%) than it was for students who identified as White (21.6%), and this change was larger for students whose parents did not attend college (40.6%) than it was for students whose parents had a BA (18.9%).

A report series from *TICAS* argues that student list products systematically exclude underrepresented students in two ways (Jaquette & Salazar, 2022; Jaquette, Salazar, & Martin, 2022; Salazar, Jaquette, & Han, 2022). The first is which prospective students are in the underlying database. Historically, student list products sold by College Board and ACT exclude non test-takers, but rates of test-taking differ by race and class [CITE]. Second, several search filters (e.g., AP score, geodemographic segment) used to control which prospect profiles are

[1] For a similar analysis of ACT's Educational Opportunity Service see Moore (2017).

purchased facilitate the exclusion of students from communities of color and low-income communities. Prior scholarship on recruiting (e.g., Salazar, Jaquette, & Han, 2021; Stevens, 2007) ignores student list products because this literature assumes that recruiting is something done by individual colleges and universities. However, Jaquette et al. (2022) show that university recruiting behavior is structured by products sold by vendors and consultancies from the edtech/enrollment management industry.

Scholarship from critical data studies unpacks the business models and algorithms of digital platforms (e.g., Coursera). “Rentier capitalism” – as in rent tenants pays a landlord – generates profit by charging monetary rent and data rent (Sadowski, 2019, 2020). Data rent is the “digital traces” created by users interacting with the platform, which often become the basis for new products. Student list data are extracted from user-data created by students laboring on platforms (e.g., taking the SAT, searching for college on Naviance) and these data are packaged to universities for monetary rent. Other scholarship shows how platforms reinforce racial inequality by embedding structural inequality within platform algorithms (e.g., Benjamin, 2019; Noble, 2018). In student list products, several search filters used to target prospects (e.g., zip code, text score) reflect historical inequality in educational opportunity. The nascent “platform studies in education” literature (Nichols & Garcia, 2022) observes that platforms increasingly perform core functions in education and calls for critical scholarship to inform policy regulations about edtech. However, the literature has not investigated how third-party platforms structure educational opportunity along the dimensions of race, class, and geography.

This study investigates the College Board Student Search Service product. We ask, what is the relationship between student list search filters and the characteristics of students who are included versus excluded in student lists purchased from College Board [KS - RQ HERE OR BEFORE ABOVE PARAGRAPH]?

We develop a conceptual framework by drawing from scholarship about algorithmic products from the sociology of race and critial data studies. Student list products are algorithmic selection devices that — similar to Google Ads or Facebook -– allow advertisers to control the prospective customers through the use of search filters. Structural racism is systematic racial bias in which processes viewed as neutral or common-sense systematically advantage dominant groups (Bonilla-Silva, 1997; Ray, 2019). Structurally racist inputs are determinants of a selection device that are correlated with race because non-white people have been historically excluded from the input (Hirschman & Bosk, 2020). We conceptualize several student list search filters as structurally racist inputs. We develop propositions about the relationship between structurally racist search filters and racial exclusion.

We assess propositions using a nationally representative sample of high school students from the High School Longitudinal Study (HSLS:09) and also data about student lists purchased by public universities, which we collected via public records requests. We reconstruct search filters and filter thresholds from the College Board Student Search Service product. We simulate student list purchases using theoretically motivated and commonly observed search filters with the goal of understanding how combinations of search filters and filter thresholds yield racial and socioeconomic inequality included versus excluded prospects.

The manuscript is organized as follows. First, we provide background on student list products, situating them vis-a-vis the process of recruiting students and summarizing recent dynamics in the market for student list data. Second, we review empirical scholarship on recruiting, focusing on scholarship from sociology. Third, we develop a a conceptual framework and propositions. Next, we describe methods and present results. Finally, we discuss implications for policy and scholarship. Several search filters may satisfy all "unfair practices" criteria of the FTC Act. Because of the systematic link between student lists and student loans, student list vendors may be classified as Consumer Reporting Agencies, thereby falling under the regulatory jurisdiction of the Consumer Finance Protection Bureau. Our broader contribution is to policy research on college access. Extant scholarship analyzes students, schools, or universities, often in relation to federal or state policies. We propose a critical, empirical research program on college access that focuses on third-party organizations and products in the edtech sector, which now dwarfs the market for for-profit Title IV institutions. As courts challenge progressive policies like affirmative action, policy research should go on the offensive. Use theories of structural inequality to investigate structural racism by edtech and target this research at regulatory agencies with relevant jurisdiction.

2 Background: Student List Products

2.1 Situating Lists Vis-a-vis Recruiting

Student lists are a match-making intermediary connecting universities to prospective students. The U.S. higher education market can be conceived as a national voucher system, whereby tuition revenue – composed of household savings and grants and loans from federal, state, and private sources – follow students to whichever institution they enroll in. Students want to attend college but do not know all their options, where they would be admitted, and how much it will cost. Universities have a financial incentive to provide access to students. Additionally, universities pursue some mix mix of broad enrollment goals (e.g., academic profile, racial diversity), while also meeting the needs of various campus constituencies (e.g., College of Engineering needs majors, marching band needs players) (Stevens, 2007). Universities cannot realize these goals solely from prospects who contact the university on their own. They must find prospects who can convinced to apply. However, universities don't know who they are, where they are, or how to contact them. Student lists overcome the problem faced by universities, providing the contact information of prospects who satisfy their criteria.

The "enrollment funnel" – depicted in in Figure 2 – is a conceptual model used in the enrollment management industry to describe stages in the process of recruiting students. The funnel begins with a large pool of "prospects" (i.e., prospective students) that the university would like to convert into enrolled students. "Leads" are prospects whose contact information (or "profiles") has been purchased. "Inquiries" are prospects that contact your institution and consist of two types: first, inquiries who respond to an initial solicitation (e.g., email) from the university; and second, "student as first contact" inquiries who reach out to the university on their own (e.g., sending ACT scores, taking a virtual tours that records IP address). Applicants consist of inquiries who apply plus "stealth applicants" who do not contact the university before applying. The funnel narrows at each successive stage in order to convey the assumption of "melt" at each stage (e.g., a subset of "inquiries" will apply). Practically, the enrollment funnel informs interventions that increase the probability of "conversion" from one stage to another (Campbell, 2017). For example, financial aid packages are used to convert admits to enrolled students (e.g., McPherson & Schapiro, 1998).

At the top of the enrollment funnel, universities identify leads by buying student lists. The sum of purchased leads plus student-as-first-contact inquiries constitutes the set of all prospects the university has contact information for who may receive targeted recruiting interventions. Based on data provided by university clients, Ruffalo Noel Levitz (2020) reported that 28% of public universities purchased less than 50,000 names annually, 44% purchased 50,000-100,000 names, 13% purchased 100,000-150,000 names, and 15% purchased more than 150,000 names. Ruffalo Noel Levitz (2018) asked clients to rate different "first contact" interventions (e.g., off-campus recruiting visit) as sources of inquiries and enrolled students. For the median public university, purchased lists accounted for 26% of inquiries, which ranked #1, and accounted for 14% of enrolled students, which ranked fourth after "application as first contact" (19%), campus visit (17%), and off-campus visit (16%).[2]

[2] For private non-profit institutions, 34% of private institutions purchased fewer than 50,000 names, 24% purchased 50,000-100,000 names, 23% purchased 100,000-150,000 names, and 18% purchased more than 150,000 names (Ruffalo Noel Levitz, 2020). Additionally, student list purchases were the highest source of inquiries, accounting for 32% of inquiries and were tied with off-campus recruiting visits as the highest source of enrolled students, accounting for 18% of enrolled students (Ruffalo Noel Levitz, 2018).

2.2 The Market for Student List Data

Historically, the student list business has been dominated by College Board and ACT. In the 21st Century, student lists have been central a surprising level of dynamism in the broader enrollment management industry. Jaquette et al. (2022) describes key dynamics that shaped the contemporary market for student list data. The first dynamic is the centrality of enrollment management consulting firms to the student list business. Although universities are the paying customers of student list products, many universities outsource student list purchases to enrollment management consulting firms. Furthermore, student lists are an essential input to the predictive models and recruiting interventions (e.g., emailing prospects) the consultancies provide.

The second dynamic is competition followed by concentration. In the 2000s, advances in technology yielded new sources of student list data, creating opportunities for new vendors. Start-ups entered the student list market by creating college search engines that asked students to submit information in order to receive recommendations about colleges and scholarships. Another new data source is college planning software (e.g., Naviance) sold to high schools and used by high school students and guidance counselors. In the 2010s, the enrollment management industry experienced a surge in horizontal followed by vertical acquisitions. Horizontal acquisitions occurred when one enrollment management consulting firm acquired a competitor (e.g., e.g., RuffaloCODY acquired Noel-Levitz in 2014) [CITE]. Vertical transformations transformed the student list business. For example, K-12 software provider PowerSchool entered the student list business by acquiring the Naviance college planning and Intersect student recruiting platforms from Hobsons. Enrollment consultant EAB entered the market for student list data through acquisitions (e.g., Cappex college search engine) and by becoming the exclusive reseller of the Intersect. As in the past, however, all student list products derive student list data from the user-data of students laboring on platforms.

Third, incumbents College Board and ACT attempted to retain their competitive advantage amidst the test-optional movement. Both organizations embraced data science by developing new search filters (e.g., ACT's predicted probability of enrolling) based on statistical models. Additionally, organizations leveraged their oligopoly position in the student list business to sell enrollment management consulting, offering clients information about prospects that is not included in purchased lists. However, the test-optional movement poses an existential threat to the college entrance exam. As fewer prospective students take College Board and ACT assessments, their competitive advantage in the coverage of college-going high school students is eroding, and private equity edtech firms such as EAB and PowerSchool are positioned to acquire market share ceded by College Board and ACT. Whereas College Board and ACT historically sold names at a price per-prospect (e.g., $0.50 in 2021), for-profit edtech firms maximize profit by wrapping a large proprietary database of prospects within a software-as-service product (e.g., EAB's Enroll360) that universities must purchase in order to obtain access to these prospects.

3 Scholarship on Recruiting from Sociology

We position our study vis-a-vis scholarship on recruiting. Most scholarship on enrollment management focuses on latter stages of the enrollment funnel, particularly which applicants get admitted and financial aid leveraging to convert admits to enrolled students. Fewer studies investigate the earlier "recruiting" stages of identifying prospects, acquiring leads, and soliciting inquiries and applications. Whereas scholarship from economics tends to analyze the effects of specific recruiting interventions (e.g., Andrews, Imberman, & Lovenheim, 2020; Miller & Skimmyhorn, 2018; Smith, Howell, & Hurwitz, 2021), scholarship from sociology tends to observe how recruiting happens "in the wild" so to speak. Our review focuses on scholarship from sociology. We identify a key blind spot, one that is shared by scholarship from economics and the broader interdisciplinary field of education research.

Scholarship from sociology has analyzed recruiting from the perspective of students, high schools, and postsecondary institutions (e.g., Cottom, 2020; Holland, 2019; Posecznick, 2017; Salazar et al., 2021; Stevens, 2007). This literature primarily utilizes ethnographic or case-study designs, and often analyzes recruiting as part of a broader analysis of college access or enrollment management. Holland's (2019) analysis of pathways from high school to college exemplifies scholarship that engages with recruiting from the perspective of high school students (e.g., McDonough, 1997). Students from groups underrepresented in higher education were drawn to colleges that made them feel wanted because they felt "school counselors had low expectations for them and were too quick to suggest that they attend community college" (Holland, 2019, p. 97). These students were strongly influenced by marketing material and high school recruiting visits, including small-group representative visits and instant decision events. By contrast, affluent students with college-educated parents were less taken by such overtures and more concerned with college prestige.

Several studies analyze connections between colleges and high schools from an organizational perspective. Off-campus recruiting visits are often conceptualized as an indicator of enrollment priorities and/or a network tie indicating the existence of a substantive relationship (Jaquette, Han, & Castaneda, forthcoming). Stevens (2007) provides an ethnography of enrollment management at a selective, tuition dependent private college sensitive about rankings. The college valued recruiting visits to (affluent) high schools as a means of maintaining relationships with guidance counselors at feeder schools. Khan (2011) analyzes the other side of the coin, showing how guidance counselors at an elite private school get under-qualified appicants into top colleges by exploiting colleges' desire for information about which applicants will enroll if admitted. Salazar et al. (2021) analyzed off-campus recruiting visits by 15 public research universities. Most universities made more out-of-state than in-state visits. These out-of-state visits focused on affluent, predominantly white public and private schools. Salazar (2022) analyzed by public research universities to out-of-state metropolitan areas, finding that universities engage in "recruitment redlining – the circuitous avoidance of predominantly

Black and Latinx communities along recruiting visit paths" [p. X]. In contrast to branding about the commitment to racial diversity (Holland & Ford, 2021), these studies find that the recruiting efforts of selective institutions prioritize affluent, predominantly white schools and communities.

A smaller set of studies analyze recruiting at open-access institutions that target working adults (e.g., Cottom, 2017; Posecznick, 2017). Cottom (2017) is simultaneously an ethnography of enrollment management by for-profits and an analysis of political economy. For-profits found a niche in Black and Latinx communities precisely because traditional colleges ignored these communities. They systematically targeted women of color and generated profit by encouraging these students to take on federal and private loans. This business model exemplifies "predatory inclusion," the logic of "including marginalized consumer-citizens into ostensibly democratizing mobility schemes on extractive terms" (Cottom, 2020, p. 443).

Collectively, empirical scholarship on recruiting from sociology – and other disciplines – assumes that recruiting is something done by individual colleges and universities. As a consequence, the recruiting literature ignores the role of third-party products and vendors. This blind spot has two root causes. First, scholarship on recruiting has not considered scholarship from critical data studies, which shows that digital platforms increasingly perform core organizational functions (e.g., Sadowski, 2019, 2020) and that digital platforms utilize algorithms that reinforce racial inequality (e.g., Benjamin, 2019). Second, scholarship on recruiting ignores the broader enrollment management industry that surrounds colleges and universities.

Jaquette et al. (2022) provide a conceptual analysis of the market for student list data in relation to the enrollment management industry. Although universities make choices about which names to purchase, these choices are structured by the algorithmic architecture of student list products — which prospects are included in the product, the targeting behaviors allowed by the product, the targeting behaviors encouraged by the product. Furthermore, many universities are uninformed about which prospective students they target because they outsource student list purchases to enrollment management consultancies (Salazar et al., 2022). These student list purchases substantially determine which prospective students will receive recruiting interventions at subsequent stages of the enrollment funnel. Thus, products and services sold by third-party vendors structure the recruiting behavior of individual universities and, in turn, college access opportunities for students.

Salazar et al. (2022) issued public records request to collect data about student lists purchased by public universities in four states. Their analyses sought to investigate the College Board student list product, rather than the behavior of universities purchasing the product. The primary research question was, what is the relationship between student list filter criteria and the characteristics of prospects included in purchased lists? For example, an analysis of several "women in STEM" purchases – which filtered on a combination of SAT/AP score, GPA, state, and intended major – showed that the racial and socioeconomic composition of targeted prospects differed dramatically from their surrounding metro area. Because of data limitations, however, Salazar et al. (2022) could not determine which filters and filter thresholds were driving this exclusion. This paper advances beyond Salazar et al. (2022) in two ways. First, we develop theoretically motivated propositions about search filters that are likely to yield racial inequality. Second, we test propositions using a nationally representative sample of high school students. These data allow us to examine who is included and excluded as filters and filter thresholds are changed and the analyses yield inferences that generalize to student populations of interest to policymakers.

4 Conceptual Framework

We develop a conceptual framework in two parts. First, we introduce concepts at the nexus of algorithms and structural racism, drawing largely from the sociology of race, but also from critical data studies and critical geography. Second, we apply these concepts to student list products and develop propositions about the relationship between search filters and exclusion.

4.1 Algorithms and Structural Racism

Algorithms and actuarialism. Algorithms are “sets of instructions written as code and run on computers” (Burrell & Fourcade, 2021, p. 215). Sociologists observe that algorithmic products utilize actuarial methods and are based on the logic of actuarialism (Burrell & Fourcade, 2021; Hirschman & Bosk, 2020; Simon, 1988). Pioneered by the insurance industry, actuarial methods proceed in two steps. First, apply statistical techniques to previous cases in order to identify factors positively and negatively associated with an outcome of interest. Second, apply these results to future cases in order to make predictions and to assign levels of risk to each case. Actuarialism is the ideology that equates fairness with risk, as determined by predicted probabilities. Under the logic of actuarialism, individuals or businesses that have characteristics associated with loan default should be charged higher interest rates.

Actuarial methods standardize decision-making by replacing individual judgment with decisions based on a formula.[3] Hirschman & Bosk (2020) observes that actuarial methods can promote racial equity if the primary source of inequality is racial bias of individual decision-makers. For example, Korver-Glenn (2018) shows that homes in predominantly white neighborhoods received higher appraisal values than those in non-white neighborhoods because appraisers have discretion in selecting similar comparision homes (“comps”) for the valuation.

However, scholarship from the sociology of race argues that actuarial methods may not reduce racial inequality due to structural racism. Drawing from Bonilla Silva’s (1997) concept of “racialized social systems,” structural racism is defined as “a form of systematic racial bias embedded in the ‘normal’ functions of laws and social relations” (Tiako, South, & Ray, 2021, p. 1143), whereby processes viewed as neutral or common-sense systematically advantage

[3] Burrell & Fourcade (2021) observes that the adoption of actuarial methods across many industries was buoyed by concerns about racial equity following antidiscrimination legislation in the 1970s.

dominant groups and disadvantage marginalized groups. Hirschman & Bosk (2020, pp. 352–353) states that "actuarialism tends to bake [racial] inequality into the decision-making process, transmuting social disadvantages into seemingly objective measures of individual riskiness." We discuss two mechanisms of structural racism in algorithmic products: (1) structurally racist inputs; and (2) market segmentation and micro-targeting.

Structurally racist inputs Actuarial products predict future outcomes by modeling the determinants of the outcome using historical data. Burrell & Fourcade (2021, p. 224) state that "predicting the future on the basis of the past threatens to reify and reproduce existing inequalities of treatment by institutions." Even when actuarial products do not include race as a determinant, they often include determinants that are highly correlated with race. *Structurally racist inputs* are determinants of an outcome that systematically disadvantage non-white people and are correlated with race because non-white people have been historically excluded from this input (Harcourt, 2015; Hirschman & Bosk, 2020).

Obermeyer, Powers, Vogeli, & Mullainathan (2019) provide an empirical example of structurally racist inputs. They found that an algorithm used by hospital systems to predict patient health care needs under-predicted the needs of Black patients because the algorithm used health care costs as a proxy for health needs, but Black patients tend to receive less care relative to their health needs than other patients. An example by Norris (2021) reconstructed Moody's city government credit rating algorithm, which assigns credits scores to cities based on determinants thought to predict loan default. The algorithm does not include the percent of residents who are Black, but it does include median household income, which is correlated with percent black because of historic wage discrimination. Once household income is included in the model, percent Black is not longer a significant predictor of city credit rating. Thus, household income is a "racialized input" Norris (2021), defined as a seemingly neutral, structurally racist input that masks the structural racism of an algorithm by "explaining away" the relationship between race and the outcome.

Geographic structurally racist inputs emerge from algorithms that make predictions based on where people live (Benjamin, 2019). Because targeting by race can be profitable, platforms often capitalize on residential segregation by using geography as an input to circumvent laws prohibiting race as an input. Thus, Benjamin (2019, p. 147) writes, "racialized zip codes are the output of Jim Crow policies and the input of New Jim Code practices."

The concepts "space" and "place" from critical geography (Agnew, 2011; Bell, 2007) provide insight about geographic structurally racist inputs. Place denotes a holistic understanding of a geographic location that incorporates the "history, peoples, and purposes within the political, social, and economic landscape" (Bell, 2007, p. 317). By contrast, space simply refers to a physical location which can be described in terms of quantifiable spatial features (e.g., distance, demographics, economic activity). Geospatial analyses typically adopt this view of space "as a location on a surface where things 'just happen' " (Agnew, 2011, p. 318).

Market research conceives of geography as space and exploits racial segregation as a means of identifying customers (Benjamin, 2019; Noble, 2018). For example, geodemography is a branch of market research that estimates the behavior of consumers based on where they live. College Board (2011, p. 1), which develops geodemographic segment search filters, states that "the basic tenet of geodemography is that people with similar cultural backgrounds, means, and perspectives naturally gravitate toward one another or form relatively homogeneous communities; in other words, birds of a feather flock together." By contrast, scholarship on racial segregation conceives of geography as place. Racial segregation is a function of historic and contemporary structurally racist laws, policies, and practices promoting residential segregation (Harris, 1993; J. Howell & Korver-Glenn, 2018; Korver-Glenn, 2018; e.g., Rothstein, 2017). Algorithms that categorize people based on geographic location (space) without considering historic and contemporary structures (place) that produce residential segregation are to reproduce historical race-based inequality in opportunity.

Market segments and micro-targeting. A second source of structural racism in products utilizing actuarial methods are the related processes of market segmentation and micro-targeting. Market segmentation categorizes customers into groups (e.g., “married sophisticates,” “rural everlasting”) that are useful for advertisers (Federal Trade Commission, 2014). Micro-targeting is the process of using data to precisely identify and target granular segments of society (Cotter, 2022).

In sociology, Fourcade & Healy (2013) define “classification situations” – which encompasses market segmentation – as the use of actuarial techniques to categorize consumers into different groups. Historically, classifications were binary; consumers with “good” credit were offered loans and those with bad credit were not. Advances in data analytics (e.g., individual credit scores) enabled finer classifications; customers are classified into many groups, or along a continuum, alongside the emergence of tiered products targeting different consumer groups with different benefit levels.For example, “payday loans” charge high interest rates to consumer groups that were previously denied credit altogether. Thus, at one end of the continuum, these classifications are similar to Cottom’s (2017, 2020) concept “predatory inclusion,” while at the other end of the continuum marginalized populations are excluded from attractive product offerings.

Scholarship from critical data studies shows that racial exclusion is a predictable consequence market segmentation (Benjamin, 2019; Cotter, Medeiros, Pak, & Thorson, 2021; Noble, 2018). The process of developing a classification system requires developers to make a series of inherently subjective decisions (e.g., who is in the dataset, which measures to utilize, which categories to identify), creating opportunities for individual biases of developers and structurally racist inputs to enter the algorithm (Noble, 2018). Furthermore, classification systems are developed to optimize profit. In turn, “audiences are treated as a commodity to be bought and sold. When audience segments are under-valued in the market, demand among advertisers for the ability to reach them will be relatively low, which decreases the likelihood that a corresponding segment will be produced” (Cotter et al., 2021, p. 3).

Critical data studies has also investigated micro-targeting (e.g., Benjamin, 2019; Cotter, 2022; Cotter et al., 2021). For example, Cotter et al. (2021, p. 1) state that micro-targeting by Facebook "is driven not by a goal of making all users available to advertisers, but of making the 'right' individuals available. Facebook advises that advertisers 'Implement a targeting strategy that focuses on reach and precision and eliminates waste.'" A theme from scholarship on micro-targeting in politics is that these technologies *could* be used to to increase outreach to marginalized groups, but in practice they are not. Instead, "campaigns routinely 'redline' the electorate, ignoring individuals they model as unlikely to vote, such as unegistered, uneducated, and poor voters (Kreiss, 2012, p. 74-75).[4]

4.2 Mechanisms of Exclusion in Student List Products

Conceptualizing student-list products. Student list products allow universities to choose prospective students from within some database of prospects by selecting on search filters. Salazar et al. (2022) categorize the filters available in the College Board Student Search Service product into the four buckets of geographic, academic, demographic, and student preferences (e.g., desired campus size, intended major), as shown in Table X.

Student list products have similarities and differences to algorithmic products that have been analyzed within sociology and critical data studies. Sociologists often study algorithmic products that assign scores based on the value of input determinants. For example, Moody's algorithm assigns credit scores to cities based on inputs correlated with default in previous cases (Norris, 2021). Similar to these products, student lists utilize search filters tha can be conceptualized as structurally racist inputs. For example, the College Board product enables universities to filter prospects based on 5-digit zip code, but zip codes are racially segregated.

Algorithmic products that make decisions purely based on scores (e.g., most credit offers) remove racial inequality caused by explicit or implicit individual decision-making bias

[4] WE PULLED THIS QUOTE FROM COTTER 2021

(Hirschman & Bosk, 2020). By contrast, student list products are similar to purchasing ads from Google or Facebook in that advertisers (universities) choose prospective customers by selecting on search filters. Universities purchasing lists may be thoughtful about avoiding structurally racist search filters. However, this individual discretion raises the possibility of racial disparities due to individual bias or lack of knowledge about the products.

Market segmentation and micro-targeting are at the core of student list products. Universities are encouraged to execute multiple student list purchase, each targeting different market segments (Waxman, 2019). Purchases target granular populations by simultaneously filtering on several filters. As in political advertising (Cotter, 2022), micro-targeting has become a branding strategy for student list products. For example, College Board Student Search promises to "create a real pipeline of best-fit prospects" College Board (n.d.) while ACT Encoura uses the tag-line "find and engage your best-fit students" Encoura (n.d.). Recently, student list products have developed new search filters that utilize market segmentation and micro-targeting approaches that have been analyzed within critical data studies (Cotter et al., 2021; Noble, 2018). For example, College Board (2011) utilized market segmentation methodologies to create "geodemographic segment" filters, which classify each high school and each neighborhood to a group based on the past college-going behaviors of students.[5] Other new filters are designed to increase precision, for example College Board's "interest in my peers" filter and ACT's predicted probability of enrollment filter. To the extent that universities want to target students from affluent schools and communities (Salazar et al., 2021), contemporary student list products facilitate this goal with great efficiency.

4.2.1 Predicting Exclusion

Our analyses focus exclusion due to structural inequality embedded in the underlying student list product as opposed to exclusion that emerges from the individual bias of people purchasing lists. We posit two broad sources of structural inequality in student list products: (1) who

[5] In 2021, College Board added three new geodemographic filters (College Board, 2021).

is included in the underlying database; and (2) utilizing structurally racist inputs as search filters for selecting prospects from the underlying database. This section draws from the theoretical discussion above to develop propositions about the relationship between search filters and exclusion, focusing on academic and geographic search filters.

Academic filters. College Board academic filters include high school graduating class, SAT score, PSAT score, AP score by subject, high school GPA, and high school class rank.

The first source of structural inequality in student list products is which prospective students are included in the underlying database. Historically, College Board and ACT student list products exclude students who do not take at least one of their assessments (e.g,. SAT, AP, PSAT).[6]. Prior research shows that rates of SAT/ACT test-taking differ by race and class (e.g., Bastedo & Jaquette, 2011). Similarly, the percentage of students who take AP exams vary across race, particularly for STEM exams (Kolluri, 2018), and Black students are more likely than white students to attend a high school with few AP course offerings (Rodriguez & McGuire, 2019). These findings motivate the following proposition, which we analyze separately by assessment (SAT, PSAT, AP) and for taking any assessment.

P1: The condition of taking standardized assessments (X) is associated with racial disparities in who is included versus excluded (Y)

Second, search filters that condition on test scores thresholds are a source of exclusion that builds on differences in test-taking. Average standardized test scores differ by race and by class [CITE]. Prior research finds that access to test preparation varies by race and class (Park & Becks, 2015). Furthermore, prior research finds that SAT question items are racially and socioeconomically biased (Freedle, 2003; e.g., Santelices & Wilson, 2010). Therefore, we conceptualize test score filters as a structurally racist input in student list products. By contrast, prior research suggests that high school GPA is a less biased measure of performance (Alon & Tienda, 2007; Bastedo & Jaquette, 2011; Posselt, Jaquette, Bielby, & Bastedo, 2012)

[6] footnote: recently added college search engine blah blah

and that GPA is a strong predictor of postsecondary student success (Allensworth & Clark, 2020; Niu & Tienda, 2010). We test the following proposition separately by assessment (SAT, PSAT, AP) and for GPA:

P2: As test score threshold increases, the proportion of underrepresented minority students included in student lists declines relative to the proportion who are excluded

Geographic filters. Geographic search filters enable universities to target prospects based on where they live. College Board Student Search Service offers two kinds of geographic search filters: filters based on known geographic borders (e.g., state, CBSA, county, zip code); and filters based on new geographic borders that College Board creates using historic data about test-takers (e.g, geomarket, geodemographic segment). At present we do not have access to data on borders created by College Board.

We conceptualize geographic search filters as structurally racist inputs because these search filters are built on the back of historic and contemporary policies and practices promoting racial segregation in American schools and communities. Drawing from critical geography (e.g., Agnew, 2011), targeting prospective students based on geographic location (space) without consideration to the historic and contemporary structures (place) that produce residential segregation is likely to reinforce historical race-based inequality in educational opportunity.

We expect that utilizing finer geographic filters (e.g., zip code rather than county) is associated with greater racial and socioeconomic dispairities in student list purchases because American residential segregation occurs at fine-grained geographic levels (Korver-Glenn, 2022).

P3. Filtering on smaller geographic localities is associated with greater racial disparities in included vs. excluded than filtering on larger geographic localities.

Prior research on recruiting consistently finds that selective private and public research universities disproportionately target affluent schools and communities (Jaquette et al., forthcoming; Salazar, 2022; Salazar et al., 2021; Stevens, 2007). These findings suggest that universities may filter on affluent zip codes when purchasing student lists. We expect that filtering for affluent neighborhoods is positively associated with racial exclusion because of historical and contemporary practices that exclude people of color from living in many affluent neighborhoods.

P4. As purchases filter on higher levels of zip-code affluence, the proportion of underrepresented minority students included in student lists declines relative to the proportion who are excluded

Filtering on academics and geography. Actual student list purchases filter on several criteria rather than one. Salazar et al. (2022) analyzed data on 830 student lists purchased by 14 public universities. The average purchase specified X.X criteria and XX% of purchases simultaneously specified at least one academic and one geographic filter. Filtering on multiple search criteria facilitates micro-targeting of desired prospects. The flipside of micro-targeting is exclusion. We suggest that filtering on multiple structurally racist inputs has a compounding effect on racial inequality in which prospects are included versus excluded. To assess this claim, we draw on the Salazar et al. (2022) sample of orders placed by public research universities and select several orders that utilized common, but potentially problematic filter criteria. We analyze the racial composition of students included versus excluded from these purchases. Next, we simulate marginal changes to order criteria to gain insight about how structurally racist inputs drive exclusion.

5 Methods

5.1 Data

Analyses utlize two data sources. First, the primary data source is the High School Longitudinal Study of 2009 (HSLS09) conducted by the National Center for Education Statistics

(NCES). HSLS09 is a nationally representative survey that follows a cohort of more than 23,000 students from 944 schools entering the ninth grade in Fall 2009. Follow-up surveys were administered to students in Spring 2012 (when most were in 11th grade), in 2013, in 2016, and NCES collected high school transcripts in 2013-14. HSLS provides the extensive student-level demographic, geographic, and academic variables needed to create academic and geographic filters used within student list purchases.

Our analysis sample includes students who meet all of the following conditions: completed Spring 2012 follow-up 1 survey; completed 2013 update survey; and obtained high school transcript data. Of the 23,503 respondents included in HSLS09, our unweighted analysis sample consists of the 16,525 students who meet all conditions. The survey weight variable `W3W2STUTR` is designed for respondents who meet these conditions. After weighting, these 16,525 students represent the population of approximately 4.187 million U.S. 9th graders in 2009 [STATE 95% CI FOR THIS POINT ESTIMATE?]. The analysis sample is smaller for analyses that utilize variables that have missing values for some respondents. [? SAY MIN ANALYSIS SAMPLE SIZE IS THIS?].

The second data source consists of the "order summaries" of student lists that public universities purchased from College Board. These data are used to inform hypothetical student list purchases in the final set of analyses. As described in Salazar et al. (2022), we collected these data by issuing public records requests to all public universities in five states (CA, IL, TX, MN, and one anonymous state) about student lists purchased from 2016-2020. Salazar et al. (2022) analyzed 804 College Board orders, which yielded XXXX prospect profiles. These orders were placed by 14 public universities. Salazar et al. (2022) shows that some universities placed many more orders and/or purchased many more prospect profiles than others. Figure 3 shows the filters utilized in these orders, separately for research universities and ma/doctoral universities.

5.2 Variables

Our research question is, what is the relationship between student list search filters and the characteristics of students who are included versus excluded from College Board student list purchases? In turn, our dependent variable measures student demographic characteristics and our independent variables are measures of student list filters, which can be utilized in isolation or in combination with one another. Descriptive statistics for analysis variables are shown in Table X [KARINA NEEDS TO CREATE TABLE].

Dependent variable. Following our conceptual framework, this manuscript is primarily concerned with the racial composition of prospects who are included in student list purchases compared to the racial composition of prospects who are excluded. Our primary dependent variable is the student race/ethnicity composite variable `X2RACE`, which includes the following seven categories: American Indian/Alaska Native, non-Hispanic; Asian, non-Hispanic; Black/African-American, non-Hispanic; Hispanic; More than one race, non-Hispanic; Native Hawaiian/Pacific Islander, non-Hispanic; and White, non-Hispanic.[7] We also conducted analyses that utilized parental education (`X2PAREDU`) and family income (`X2FAMINCOME`) as dependent variables, but we exclude these analyses because of manuscript space limits.

Independent variables. Independent variables are measures of student list filters. Choices about independent variables were based on our conceptual framework and the set of student list filters observed in our public records request data collection, shown in Figure 3. Our conceptual framework restricts analytic focus to academic filters and geographic filters, leaving demographic filters and student preferences filters for a future analysis [?EXCEPT INTENDED MAJOR FOR WOMEN IN STEM?].

Propositions **P1** and **P2** focus on academic filters. **P1** is concerned about which students take standardized assessments, which determines inclusion in the underlying College Board student list database. **P2** is concerned with test score thresholds utilized to filter prospects.

[7] We collapse the categories "Hispanic, no race specified" and "Hispanic, race specified" into a single category.

We create. For **P1**, we create dichotomous measures for each of the following assessments (input variables in parentheses) based on test score variables from the high school transcript file: PSAT/PreACT (`X3TXPSATCOMP`); SAT/ACT (`X3TXSATCOMP`); any AP exam (variables with names that start with `X3TXAP`); and any STEM AP exam. For **P2**, we use these same input variables to create test score measures for PSAT/PreACT; SAT/ACT; highest AP exam score; and highest AP STEM exam score. We also create a measure of high school GPA in academic courses (`X3TGPAACAD`), which is a question asked in the pre-test questionnaire of College Board assessments. Consistent with how College Board filters work, **P2** variables are analyzed as categorical rather than continuous variable. To select thresholds for **P2** variables – for example, an SAT score thresholds of 1000+, 1100+, 1200+, etc. – we considered what the product allows, what we observed in orders collected via public records requests, and the goal of parsimony.

A limitation of measures created for **P1** and **P2** is that HSLS does have separate measures for SAT and ACT. Instead, SAT and ACT test scores are converted to the same scale, but we do not know which students took which assessments. The same is true for PSAT and PreACT assessment.

Propositions **P3** and **P4** focus on geographic filters filters. Drawing from Figure 3, we create measures for student home state (`X2GSTATE`), county (`X2GCNTY`), zip code (`X2GZIPCD`), and CBSA (based on crosswalk with home zip code). Next, we attach income data to localities by merging in data from the American Community Survey (ACS) YYYY 5-year estimates. We do not create independent variables for geomarket filter or geodemograhic segment filter because these filters utilize geographic borders that College Board created using proprietary, historical data about test-takers.

5.3 Analyses

Analyses utilize simple descriptive statistics, with appropriate statistical tests. All analyses compare the racial composition of included versus excluded prospects when particular filters and/or filter thresholds are utilized to purchase prospect profiles.

Consider a hypothetical purchase that that all prospects took an AP STEM exam. Conceptually, two types of comparisons are possible. Comparison type one compares the proportion of students from a particular race/ethnicity group (e.g. Black) who are included to the proportion of students not from that particular race who are included. For example, using unweighted sample sizes, `91/1,655=5.5%` of Black students took an AP STEM exam and `1,712/14,870=11.5%` of non-Black students took an AP STEM exam. The test for difference in proportions compares the 5.5% of Black students who are included to the 11.5% of non-Black students who are included, and this test is run separately for each race/ethnicity group.

Comparison type two compares the racial composition of the included group to the racial composition of the excluded group. For example, Black students comprise 5.05% (`=91/1803`) of AP STEM test-takers and Black students comprise 10.6% (`=1564/14722`) of students who do not take an AP STEM exam. The test for difference in proportions compares whether the proportion of included prospects who identify are Black differs from the proportion of excluded prospects who identify as Black, and this test is run separately for each race/ethnicity group.

While the significance tests from comparison type one and comparison type two are mathematically equivalent, the two comparisons differ conceptually. The first comparison analyzes the probability of being targeted from the student perspective; that is, do students who identify as Black have a higher/lower probability of being included than students who do not identify as Black? The second comparison focuses on the racial composition targeting from the university perspective; that is, what is the racial composition of prospects who are targeted by a particular set of filters versus the racial composition of prospects who are excluded by these filters? This manuscript focuses on comparison type two because we are interested in how student list products structure the racial composition of university recruiting efforts.

Analyses for propositions P1 through P4 examine purchases that utilize individual filters in isolation. The final set of analyses examine purchases that utilize academic and geographic filters in combination with one another, with choice of filters informed by commonly observed combinations from the public request data and also by theoretical considerations.

5.4 Limitations

This manuscript uses HSLS09 to recreate the College Board Student Search Service. One limitation is that HSLS variables for SAT test-taking and test scores also include ACT test-takers, with ACT scores converted to the SAT scale. The same is true for the PSAT and PreACT. The Student Search service includes students who take at least one College Board assessment, but we cannot differentiate between College Board and ACT test-takers, so our analyses incorrectly treat ACT test-takers as College Board test-takers. We considered restricting the analysis sample to states where the majority of students take the SAT rather than the ACT. We chose not to take this approach because the ACT "Educational Opportunity Service" student list product – now, named Encoura – includes academic and geographic filters that are nearly identical to the College Board filters that are the focus of this manuscript (Schmidt, 2022). Thus, analyses can be interpreted as who would be included/excluded by both College Board and ACT student list products.

Second, test-takers have the opportunity to opt-out of the College Board Student Search Service and the ACT Educational Opportunity Service but HSLS09 has no reasonable proxy for whether students opt-in or opt-out. Moore (2017) finds that 86% of ACT test-takers opt-in, but does not investigate the student characteristics associated with opting in. Third, the HSL09 cohort pre-dates the increase in test-optional admissions policies and decline in test-takers which occurred since the onset of Covid. This undermines the external validity of our findings with respect to current cohorts of high school students. That said, several for-profit vendors have developed student list products (e.g., Intersect by PowerSchool) poised to acquire market share ceded by College Board and ACT, and these products use filters that are similar to College Board and ACT products (Feathers, 2022). Our analysis of structurally racist inputs and exclusion yields insights across student list products. Fourth, we could not make measures for high school class rank, an academic filter, or for geomarket and geodemographic filters, which utilize proprietary College Board data.

6 Findings

6.1 Academic Filters

We begin by describing the racial characteristics of prospects who completed standardized assessments in comparison to those who did not, which determines inclusion in the underlying College Board student list database. Figure 1 presents the racial/ethnic composition of prospects included (i.e., completed assessment) and excluded (i.e., did not complete assessment) across SAT, PSAT, and AP exams. For example, the top left graph shows that more than 1.8 million prospects completed the SAT and would have presumably been included in the College Board student list database. In comparison, more than 2.3 million prospects did not complete the SAT and would be excluded from the database. White students make up 53% of included students who completed the SAT and 51% of excluded students who did not. Table X reports statistical tests for proportions between included and excluded students by race/ethnicity. Differences in White student proportions across included and excluded prospects are statistically significant ($p<0.000$). While Asian and Multiracial students make up nearly equal proportions in both included and excluded prospects, Latinx students make up only 17% of included prospects relative to 26% of excluded prospects ($p<0.000$). Black students similarly make up 12% of included prospects but 15% of excluded prospects (p<0.000).

Other standardized assessments resulted in similar included prospects that were on average made up of larger proportions of White and Asian students and smaller proportions of Latinx, Black, and American Indian/Alaska Native students than excluded groups, lending support for our first proposition. For example, Figure 1 also shows the composition of included versus excluded prospects by AP exam completion in any subject on the top right panel. Similar to SAT, White (54% versus 51% White) and Asian students (8% versus 3%) make up statistically significant (p<0.000) larger proportions of included prospects. While an equal proportion of included and excluded students are Hispanic (22%), Black students make up a smaller share (p<0.000) of included prospects (8%) than excluded prospects (15%). When inclusion versus exclusion is determined by completing an AP exam in a STEM subject, total included students declines to nearly half (383,669) of those included via completing any AP exam. Moreover, Black (7% versus 14%), Hispanic (19% versus 22%), Multiracial (7% versus 14%), and American Indian/Alaska Native (0.2% versus 0.7%) students make up smaller statistically significant proportions of the included prospects relative to excluded prospects based on completion of an AP STEM exam

Proposition 2 suggests the proportion of underrepresented minority students included in student lists decline relative to the proportion who are excluded as assessment score thresholds increase. In order to test this proposition, we analyze the racial composition of included versus excluded students at minimum score thresholds commonly used across student list purchase orders for SAT, PSAT, and AP exams. For example, Figure 2 presents these results for SAT (left panel) and PSAT assessments (right panel). For the top left panel, each bar represents the racial composition of included prospects who completed the SAT exam and scored at the minimum threshold indicated. On the bottom left panel of Figure 2, each bar represents the racial composition of excluded prospects who did not complete the SAT exam in addition to students who did complete the exam but did not meet the minimum score threshold indicated. Statistical tests for differences in proportion for Figure 2 are reported in online Appendix A for space considerations.

As SAT score thresholds increase from less than 1000 to greater than 1400 in Figure 2, proportions of included White and Asian students increase while proportions of included Hispanic and Black students decrease. For example, White students make up a statistically significant ($p<0.000$) smaller share of included (47%) than excluded (53%) prospects scoring less than 1000 on the SAT, which results in an equal share of Hispanic students (22%) and a greater share of included Black students (19% versus 12%) relative to excluded prospects at this score threshold. However, Hispanic student proportions in included versus excluded prospects decrease to 12% versus 25% at scores greater than 1000, 9% versus 23% at scores greater than 1200, and down to 5% versus 22% at scores greater than 1400. Similarly, Black student proportions in included versus excluded prospects decrease to 6% versus 16% at scores greater than 1000, 4% versus 14% at scores greater than 1200, 2% versus 14% at scores greater than 1300, and down to making up 0% of included prospects at scores greater than 1400. These proportional differences across score thresholds are statistically significant ($p<0.05$) for both Hispanic and Black students (online Appendix A).

While making up relatively small proportions of the overall sample, declines in proportions of American Indian/Alaska Native students and Native Hawaiian/Pacific Islander students within included versus excluded prospect groups are statistically significant as score thresholds increase (online Appendix A). In order to more equitably capture these differences, we report the number of students rather than their overall representational proportion within included versus excluded groups. For instance, more than 7,600 American Indian/Alaska Native students and nearly 2,500 Native Hawaiian/Pacific Islander students are represented in the included prospects relative to the more than 20,800 and 16,300 represented in the excluded prospects at SAT scores less than 1000, respectively. However, American Indian/Alaska Native students decline to zero and Hawaiian/Pacific Islander students decline to 435 students in the included prospects group by the 1300 or greater SAT score threshold

PSAT results are also shown in Figure 2 for composite scores that range from 60 to 240. [8] Similar to SAT, as PSAT composite score thresholds increase from less than 120 to greater than 220, proportions of included White and Asian students increase while proportions of included Hispanic and Black students decrease relative to excluded prospects. Online Appendix B shows all comparisons between included and excluded students across PSAT score thresholds are statistically significant at the p<0.000 level, with the exception of multiracial students at the 220 or greater minimum score threshold.

We find similar racial disparities in included versus excluded prospects across AP exam score thresholds, providing strong support for Proposition 2. Figure 3 shows similar results as Figure 2 for AP exams. As AP score thresholds for any subject exam (left panel) increase from one to five, proportions of included White and Asian students increase while proportions of included Hispanic, Black, Multiracial, American Indian/Alaska Native, and Native Hawaiian/Pacific Islander decrease relative to excluded prospects. For example, the 110,360 included prospects (relative to excluded prospects) who had a score of one on any subject AP exam were on average 38% White (52% excluded), 6% Asian (4% excluded), 27% Hispanic (22% excluded), and 21% Black (14% excluded). By an AP score threshold of four or greater, included prospect proportions shift (relative to excluded) to 60% White (51% excluded), 10% Asian (3% excluded), 20% Hispanic (22% excluded), and 4% Black (4% excluded).[9]. Similar patterns are evident for AP STEM exam completion (right panel).

Given differences in completion rates for standardized assessments by race, our conceptual framework outlines an interest in whether using GPA filters leads to greater racial parity between included versus excluded students relative to standardized assessments as thresholds

[8]PSAT exams taken 2014 or before receive composite scores that range from from 60 to 240. PSAT exams taken 2015 or later are scored via a range from 320 to 1520. Our lower bound PSAT composite score thresholds of 120, 170, 200, and 220 for HSLS students who completed the exam prior to 2014 equate to minimum score thresholds of 890, 1220, 1410, and 1510 on the 2015 or later PSAT scale, respectively. CITE

[9]Proportional differences for these specific racial/ethnic categories at reported score thresholds are statistically significant at the p<0.000 level and reported in online Appendix B

increase. We therefore analyze the racial composition of included versus excluded students at minimum thresholds commonly used across student list purchase orders for high school GPA. Figure 4 shows the racial composition of included (top panel) and excluded (bottom panel) students across less than 2.0, 2.0 or greater, 3.0 or greater, and 3.5 or greater thresholds of GPA. Similar to standardized assessments, Figure 4 suggests proportions of included prospects increase for White students and Asian students (although modestly) while proportions of included prospects decrease for Hispanic and Black students as GPA thresholds increase. For instance, White student proportions increase from 37% (relative to 57% excluded, $p<0.000$) at GPA less than 2.0 to 71% (relative to 57% excluded, $p<0.000$) at GPA 3.5 or greater. On the other hand, Hispanic and Black included student proportions (relative to excluded proportions) decrease from 30% (19% excluded, $p<0.000$) and 22% (11% excluded, $p<0.000$) at GPA less than 2.0 to 11% (24% excluded, $p<0.000$) and 4% (15% excluded, $p<0.000$) at GPA 4.0 or greater, respectively. The 12,591 American Indian/Alaska Native students included at GPA less than 2.0 (15,928 excluded, $p<0.000$) also decline to 1,304 (27,216 excluded, $p<0.000$) at GPA 3.5 or greater score threshold.

However, Figure 4 shows that GPA filters at "middle" thresholds (2.0 to 3.0) lead to smaller declines in proportions of included Hispanic and Black students relative to middle thresholds of SAT and PSAT filters. In increasing GPA from 2.0 or greater to 3.0 or greater, the proportions of Hispanic and Black included students decrease by 3 percentage points or less. In comparison, increasing PSAT from scores 120 or greater to 200 or greater results in an up to nine percentage point decrease in the number of included Black students. This pattern of lesser relative declines in the proportion of underrepresented minority students included at "middle" thresholds is also evident by AP filters (see Figure 3). However, given the disparities in AP course availability and exam completion rates, a considerable smaller number of overall included students are captured by AP filters than GPA.

6.2 Geographic Filters

Proposition 3 and Proposition 4 conceptualize how the use of geographic filters may result in greater racial disparities in proportions of included prospects relative to excluded prospects. For instance, Proposition 3 suggests as purchases filter on higher levels of zip-code affluence, the proportion of underrepresented minority students included in student lists will decline relative to the proportion who are excluded. In order to test this proposition, we analyze the racial composition of included versus excluded students when filtering by zip code median household income. In order to deal with median household incomes varying widely across the U.S., we categorized all zip codes into percentiles based on levels of median household income within their respective Core Based Statistical Areas (CBSA). For example, median household income percentiles based on the 378 zip codes within the Los Angeles metropolitan area are \$55,256 at the 20th percentile, \$70,804 at the 40th percentile, \$89,709 at the 60th percentile, and \$108,316 at the 80th percentile (in 2022 CPI). So the Los Angeles zip code 92649, which captures parts of the Huntington Beach area, with a median household income of \$109,159 (in 2022 CPI) would be categorized as zip code in the 80th percentile of affluence within CBSA. This approach also aligns with common ways in which student list orders purchase prospect's contact information by filtering on zipcodes within specific CBSAs.

Figure 5 presents the racial composition of zipcodes that included (top panel) versus (excluded panel) when filtering based on percentile of affluence within CBSA. The figure suggests that as zip code affluence increases, included prospects have larger proportions of White students and smaller proportions of Latinx and Black students relative to excluded prospects. For example, Hispanic and Black students make up 30% and 27% of included prospects and 20% and 11% of excluded prospects at zip codes below the 20th percentile of affluence, respectively. The proportions of Hispanic and Black students within included prospects decline as zip code affluence increases up though the 89th percentile. For zipcodes in 90th percentile or higher of affluence within CBSA, the proportions of Hispanic students within included prospects declines to 11% relative to making up 23% of excluded prospects (p<0.000). Similarly, Black students make up 9% of included prospects relative to making up 14% of excluded prospects (p<0.000) within the most affluent zip codes. [10]

To contextualize these findings, Figure 6 presents similar results as Figure 5 for two specific CBSAs: Los Angeles and New York. The left panel of Figure 6 shows the racial composition of included (top) and excluded (bottom) students across percentiles of zip code affluence for Los Angeles. Similar to results across all CBSAs in Figure 5, proportions of included prospects relative to excluded prospects increase for White students while proportions decline for Hispanic and Black students as zip codes become more affluent. For instance, White student proportions increase from 5% (relative to 17% excluded, p<0.000) at the 20th percentile, to 28% (relative to 12% excluded, p<0.000) at 50th-79th percentiles, and up to making up 76% of included prospects (relative to 12% excluded, p<0.000) for zip codes at the 90th percentile or higher of affluence. On the other hand, Hispanic and Black included student proportions (relative to excluded proportions) decrease from 73% (54% excluded, p<0.000) and 14% (7% excluded, p<0.000) at zip codes in the lower 20th percentiles of affluence to 4% (60% excluded, p<0.000) and 3% (9% excluded, p<0.000) at the 90th percentile or higher of affluence, respectively. While New York provides a different racial composition of students than Los Angeles, similar and statistically significant patterns persist.

These findings suggest that purchases filtering on higher levels of zip-code affluence lead to smaller proportions of underrepresented minority students included in student lists relative to the proportion who are excluded, providing support for Proposition 3. However, we acknowledge that categorizing zip codes within CBSA limits the number of rural zip codes captured within the included prospect groups. By categorizing zip code affluence within CBSA, only rural zip codes within micropolitan statistical areas (i.e., areas that have at least one urban cluster of at least 10,000 people with commuting ties to adjacent metropolitan

[10] Online Appendix C reports statistical tests for proportions between included and excluded students by race/ethnicity for zip code affluence.

areas that have higher degrees of social and economic activity) will be captured via CBSA.

To analyze whether filtering on smaller geographic localities is associated with greater racial disparities in included prospects relative to excluded prospects (Proposition 4), we compare the racial characteristics of prospects based on zip code filters versus county filters. We categorize all zip codes and counties based on levels of median household income within their respective Core Based Statistical Areas (CBSA), given results for Proposition 3, to analyze whether relationships between racial composition and geographic level change across levels of affluence. Therefore, zip codes and counties are categorized as low, moderate, and high income based on their median household incomes falling below 30th percentile, within 30th-70th percentiles, or greater than 70th percentile of affluence within their respective CBSAs.

Figure 7 presents the racial composition of prospects when using zip code filters (left panel) in comparison to county filters (right panel). There are modest differences between included versus excluded groups when comparing zipcode filters to county filters at low and moderate levels of affluence. For example, Hispanic students are slightly overrepresented within included relative to excluded prospects (23% versus 22%, $p<0.000$) by zip code at moderate levels of affluence but underrepresented (20% versus 22%, $p<0.000$) at moderate affluence levels for county. Similarly, Black students are slightly underrepresented within included relative to excluded prospects (12% versus 15%, $p<0.000$) by zip code at moderate levels of affluence but overrepresented (20% versus 13%, $p<0.000$) at moderate affluence levels for county.

Differences between the use of zip code filters and county filters are most evident in high levels of affluence. For instance, Hispanic and Black students make up 16% and 8% of included prospects and 24% and 16% of excluded prospects for the most affluent zip codes, respectively. However, in the most affluent counties, Hispanic students make up an equal share of included and excluded prospects (22%). Similarly, filtering on the most affluent counties leads to smaller differences in proportions of Black students within included and excluded prospects

(11% versus 16%, p<0.000). These results suggest filtering for smaller geographic localities (i.e., zip codes) is associated with greater racial disparities in included prospects relative to excluded prospects in comparison to larger geographic localities (i.e., counties) at higher levels of affluence, which results for Proposition 3 suggest are the thresholds with the greatest lack of racial parity between included versus excluded students.

6.3 Academic and Geographic Filters

Our last set of analyses focus on assessing whether filtering on multiple criteria compound the effect of racial disparities in which prospects are included versus excluded. We draw on Figure 3 to select common filters used across orders. We begin by combining the two most common academic filters: GPA and SAT. Figure X (top panel) presents the racial composition of prospects included when filtering on GPA greater than or equal to 3.0 while simulating increases to minimum SAT thresholds at increments of 50 beginning at scores just above the sample median of 1010. For space considerations, we only present the included prospect group The figure suggests that even at the the lowest SAT score, White students make up much larger proportions while Black and Hispanic students make up significantly smaller proportions of included prospects when filtering for both GPA and SAT. For example, White students make up 72% of included prospects when filtering for GPAs greater than or equal to 3.0 in combination with SAT scores greater than 1050, whereas Hispanic and Black students make up 10% and 3%, respectively. Racial disparities only grow as SAT thresholds increases. Moreover, these racial disparities are greater than when filtering for similar thresholds for GPA (Figure X) and SAT score (Figure X) individually. The bottom panel of Figure X suggest similar results are evident when combining a GPA filter greater than or equal to 3.0 and a PSAT filter. While Hispanic and Black students make up larger proportions at lower thresholds of PSAT in comparison to SAT when combined with GPA, the racial disparity for Black students is still greater in the combination of filters than when filtering for similar thresholds of PSAT score (Figure X) individually.

In order to assess the effects of combining academic and geographic filters, Figure X adds a zip code filter to the GPA and SAT/PSAT order simulations presented above. We again deal with median household incomes varying widely across the U.S. by categorizing all zip codes into percentiles based on levels of median household income within CBSAs. The top panel of Figure X presents the racial composition of included students when filtering for GPAs greater than or equal to 3.0, SAT scores greater than or equal to 1050, and zip codes at various levels of affluence. In comparison to racial disparities in included versus excluded prospects driven by just zip code affluence in Figure 5, the combination of zip code with GPA and SAT filters leads much greater disparities even at lower levels of affluence. For example, Figure X shows White students make up 72% of included prospects when filtering for GPAs greater than or equal to 3.0 in combination with SAT scores greater than 1050 within the lowest income zipcodes (<20th percentile), whereas Hispanic and Black students make up 9% and 7%, respectively. The proportions of Hispanic and Black included prospects resulting from the combination of filters are considerably lower than the 30% of Hispanic and 27% of Black included prospects resulting from only filtering by zip code (Figure 5). Greater racial disparities result from the the combination of filters across all levels of zip code affluence in comparison to only filtering by zip code, although proportional differences are modest at higher incomes. Similar patterns are evident when combining similar zip code and GPA filters with a PSAT filter for composite scores greater than or equal to 150.

7 Discussion

Prior scholarship on recruiting assumes that recruiting is something done by individual colleges and universities. Universities identify prospective students by purchasing student lists. College Board began selling student lists in 1972 (Belkin, 2019), but prior research has not investigated how student list products structure the connection between universities and prospective students. We ask, what is the relationship between student list search filters and the characteristics of students who are included versus excluded in student lists purchased from College Board? We develop a conceptual framework about structural racism in algorithmic products by drawing from sociology and critical data studies. Structurally racist inputs are determinants of a selection device that are correlated with race because historically dominated racial groups have been historically excluded from the input (Hirschman & Bosk, 2020). We propose that several academic filters and geographic filters utilized by the College Board Student Search Service product are structurally racist inputs. We assess propositions about the relationship between search filters and racial exclusion using a nationally representative sample of 9th graders from 2009.

Results for proposition **P1** .... [KARINA WRITE HIGH-LEVEL SUMMARY OF RESULTS; ONE OR TWO PARAGRAPHS]

These results have policy implications for federal regulatory agencies concerned with consumer protection and equality of opportunity. Consider zip code filters. Given the history of racial segregation, there is no equality of opportunity rationale for products that enable universities to target one zip code and exclude students from the neighboring zip code. Over the last decade, the Federal Trade Commission (FTC) has become increasingly concerned about algorithmic products that "categorize consumers in ways that can result in exclusion of certain populations" (Federal Trade Commission, 2016, p. 9). The FTC enforces the FTC Act, which applies to all organizations engaged in interstate commerce. Section 5 of the FTC Act prohibits "unfair" practices. A practice is defined as unfair if it meets the three criteria of (1) causes substantial harm to consumers, (2) harm cannot be reasonably avoided, and (3) harm not outweighed by benefits to other consumers and to competition (FDIC, 2018). Zip code filters may cause substantial harm to consumers (criterion #1) because students who live in nearby non-targeted zip codes are excluded from college access opportunities. Consumers cannot reasonably avoid the injury (criterion #2) because they cannot easily move to a different zip code. The benefit to targeted consumers may not outweigh the harm to excluded consumers (criterion #3). However, other student list filters may satisfy the unfair practices criteria more unequivocally.

This manuscript is the first word on student list products, not the last word. Future research should examine filters that utilize predictive analytics, which make predictions about future cases from models of past cases. In ACT's "Enrollment Predictor" filter, "every student in the Encoura®Data Cloud is scored on their likelihood to enroll at your institution" (Schmidt, 2022). College Board has developed geographic filters that create geographic borders based on historic, proprietary data on college enrollment. The "geomarket" filter carves metropolitan areas into distinct markets. "Geodemographic Segment" filters utilize cluster analysis to allocate individual high schools and individual census tracts into distinct clusters based on historic college-going behavior. The analysis of Moody's city government credit rating algorithm by Norris (2021) shows that these filters can be recreated – or closely approximated – and then deconstructed using publicly available data sources.

Another topic for future research, "demographic search filters" allow universities to target prospects by race, ethnicity, gender, and first-generation college students. College Board argues that these filters facilitate the recruitment of underrepresented student populations. However, analyses by Salazar et al. (2022) observed that "women in STEM" purchases yielded profound racial and socioeconomic inequality. Additionally, purchases that filtered for underrepresented minority students often disproportionately targeted students from affluent, predominantly white schools and communities.

In addition to empirical analyses, legal scholarship informs how regulatory agencies interpret the law (e.g., Russell, Reidenberg, Martin, & Norton, 2019). For example, Lawler & Dold (2021, p. 4) argue that universities that "give advice and assistance to students seeking loans to pay for tuition" – the activities of financial aid office – are "covered persons" under the Consumer Financial Protection Act and may be regulated by the Consumer Financial Protection Bureau. We recommend that future legal scholarship analyze whether and how student list vendors can be regulated as "consumer reporting agencies." A consumer reporting agency is an entity that sells information about prospective consumers that leads to the extension of credit (i.e., loans) (15 U.S.C §1681a). We suggest that student list vendors are consumer reporting agencies because there is a systematic link between student lists and student loans. In particular, the first stage of the enrollment funnel is to identify "leads" by purchasing student lists and the last stage is convert admits to enrolled students by offering financial aid packages. Consumer reporting agencies are regulated under the Fair Credit Reporting Act and the Consumer Finance Protection Act. Legal scholarship can inform how FTC and CFPB interpret their regulatory authority over student list vendors and products.

The broader contribution of this manuscript is to motivate critical education policy research that focuses on third-party products and vendors. The majority of policy research in education analyzes students, schools, or universities, often in relation to federal, state, or local policies. Scholarship from critical data studies shows that structural racism in digital platforms is "a feature, not a bug" (Benjamin, 2019; Noble, 2018) because racial exploitation is the defining feature of capitalism (Robinson, 2000) and the defining feature of platform capitalism (Cottom, 2020). Drawing from critical data studies, the nascent "platform studies in education" literature urges scholars "to go beyond pedagogical and technical questions toward social, political, and economic critiques" (Napier & Orrick, 2022, p. 207). However, this literature has not investigated how platforms structure educatonal opportunity along racial, class, and geographic dimensions. We propose a critical, empirical literature that bridges education policy and platform studies by utilizing structural theories of inequality to investigate third-party products and vendors in education.

Student list products are a model topic for this critical, empirical literature because they substantially structure college access and are a source of profit for third-party vendors. Sadowski (2019) develops the concept "data as capital," drawing from Marx (1978) to describe how platforms monetize user-data, often by becoming the basis for a new product.

The formula $M - C - M'$ represents economic capital, whereby money M is invested to produce commodity C, which is sold for a larger amount of money M' (Marx, 1978). Student list data are capital derived from the user-data of students laboring on a platform, whether that be taking a standardized assessment or searching for scholarships on a 'free' college search engine. Processes that profit from student list data follow the general formula for economic capital. College Board uses the cycle $M - C - M' - C - M''$, investing money (M) to create tests (C), which are sold to households for M' and also yield student list data (C), which are sold to universities (M'') looking for students. New entrants to the market for student list data (e.g., PowerSchool, EAB) add another link to the cycle. Instead of selling names at a price-per-prospect (e.g., $0.50) like College Board, they wrap proprietary databases of prospects within software-as-service products that recruit these prospects (e.g., Intersect, Enroll360), which are then sold to universities for an annual subscription.

Critical policy research should examine how college access is structured by vendors and consultancies in the broader enrollment management industry. Many universities depend on enrollment management consulting firms to develop and implement recruiting campaigns. In our data collection, roughly half the public universities outsourced student list buys to consulting firms (Salazar et al., 2022), and these universities tended to be uninformed about who they were recruiting. Several enrollment management consultancies sell algorithmic products designed to make recommendations about list buys (Fire Engine RED, 2022; Ruffalo Noel Levitz, 2021). For example, Ruffalo Noel Levitz offers an algorithm that tells universities how many names to buy from each zip code [CITE]. Beyond name buys, consultancies develop and implement strategy about digital advertising, direct mail, which high schools to visit, and tuition pricing and financial aid. To the extent that unviersities outsource enrollment management to consultancies, these consultancies substantially structure college access, but scholarship on college access assumes that universities perform these functions in-house.

The enrollment management industry also structures "student success" in higher education and k-12. For example, EAB's Starfish student success software that incorporates university administrative data, issues notifications when students get "off-track," promising to "make student intervention easy and integrated" (EAB, 2022). Scholars have modeled the determinants of student success and are assessing the fairness of alternative predictive models of student success CITE STUFF LIKE THIS. However, scholarship has not investigated the fairness of third-party predictive models sold to universities. Do these models use race/ethicity as an input? Do commercial student success platforms achieve higher graduation rates by pushing students out of certain majors?

START HERE NEXT: OBFUSCATION AND HOW TO EMPIRICALLY INVESTIGATE THIRD-PARTY PRODUCTS/VENDORS

8 References

Agnew, J. A. (2011). Space and place. In J. A. Agnew & D. N. Livingstone (Eds.), *The SAGE handbook of geographical knowledge* (pp. 316–330). Los Angeles: SAGE.

Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, *49*(3), 198–211. https://doi.org/10.3102/0013189X20902110

Alon, S., & Tienda, M. (2007). Diversity, opportunity, and the shifting meritocracy in higher education. *American Sociological Review*, *72*(4), 487–511. Journal Article. Retrieved from <Go to ISI>://000248696500001

Andrews, R. J., Imberman, S. A., & Lovenheim, M. F. (2020). Recruiting and supporting low-income, high-achieving students at flagship universities. *Economics of Education Review*, *74*, 101923. https://doi.org/https://doi.org/10.1016/j.econedurev.2019.101923

Bastedo, M. N., & Jaquette, O. (2011). Running in place: Low income students and the dynamics of higher education stratification. *Educational Evaluation and Policy Analysis*, *33*(3), 318–339.

Belkin, D. (2019). For sale: SAT-Takers' names. Colleges buy student data and boost exclusivity. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/for-sale-sat-takers-names-colleges-buy-student-data-and-boost-exclusivity-11572976621

Bell, C. A. (2007). Space and place: Urban parents' geographical preferences for schools. *The Urban Review*, *39*(4), 375–404. https://doi.org/10.1007/s11256-007-0059-5

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new jim code*. Medford, MA: Polity.

Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American Sociological Review*, *62*(3), 465–480. https://doi.org/10.2307/2657316

Burrell, J., & Fourcade, M. (2021). The society of algorithms. *Annual Review of Sociology*, *47*, 213–237. Journal Article. https://doi.org/10.1146/annurev-soc-090820-020800

Campbell, A. (2017). *Higher education marketing: How to master your admissions funnel.*

Hop Online. Retrieved from https://hop-online.com/blog/higher-education-marketing-admissions-process/

College Board. (n.d.). College board search solutions. The College Board. Retrieved from https://cbsearch.collegeboard.org/solutions

College Board. (2011). *Segment Analysis Service: An educationally relevant geodemographic tagging service*. College Board. Retrieved from https://secure-media.collegeboard.org/mSSS/media/pdf/segment-analysis-service-overview.pdf

College Board. (2021). Introducing environmental attributes. *YouTube*. Retrieved from https://www.youtube.com/watch?v=VmTU9sb4ZiY

Cotter, K. (2022). Selling political data: How political ad tech firms' discourses legitimate microtargeting. In *17th international conference on information for a better world - shaping the global future (iConference)* (Vol. 13192, pp. 195–208). CHAM: Springer International Publishing Ag. https://doi.org/10.1007/978-3-030-96957-8_18

Cotter, K., Medeiros, M., Pak, C., & Thorson, K. (2021). "Reach the right people": The politics of "interests" in facebook's classification system for ad targeting. *Big Data & Society*, *8*(1), 16. https://doi.org/10.1177/2053951721996046

Cottom, T. M. (2017). *Lower ed: The troubling rise of for-profit colleges in the new economy*. New Press, The.

Cottom, T. M. (2020). Where platform capitalism and racial capitalism meet: The sociology of race and racism in the digital society. *Sociology of Race and Ethnicity*, *6*(4), 441–449. https://doi.org/10.1177/2332649220949473

EAB. (2018). *Making your digital ads count: 15 lessons on new and emerging techniques in undergraduate recruitment marketing*. EAB.

EAB. (2022). Starfish: Scale your student success efforts. Retrieved from https://eab.com/products/starfish/

Encoura. (n.d.). Encoura. ACT. Retrieved from https://encoura.org/

FDIC. (2018). VII. Unfair and deceptive practices. In *FDIC consumer compliance examination* *manual* (pp. 1–8).

Feathers, T. (2022). College prep software naviance is selling advertising access to millions of students. *The Markup*. Retrieved from https://graphics.reuters.com/USA-ELECTION/DATA-VISUAL/yxmvjjgojvr/https://themarkup.org/machine-learning/2022/01/13/college-prep-software-naviance-is-selling-advertising-access-to-millions-of-students

Federal Trade Commission. (2014). *Data brokers: A call for transparency and accountability*. Federal Trade Commission. Retrieved from https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Federal Trade Commission. (2016). *Big data: A tool for inclusion or exclusion? Understanding the issues*. Federal Trade Commission. Retrieved from https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

Fire Engine RED. (2022). *Data services: Search modeling*. Fire Engine RED. Retrieved from https://www.fire-engine-red.com/data-services/

Fourcade, M., & Healy, K. (2013). Classification situations: Life-chances in the neoliberal era. *Accounting Organizations and Society*, *38*(8), 559–572. Journal Article. https://doi.org/10.1016/j.aos.2013.11.002

Freedle, R. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review*, *73*(1), 1–43.

Harcourt, B. E. (2015). Risk as a proxy for race: The dangers of risk assessment. *Federal Sentencing Reporter*, *27*(4), 237–243. https://doi.org/10.1525/fsr.2015.27.4.237

Harris, C. I. (1993). Whiteness as property. *Harvard Law Review*, *106*(8), 1707–1791. https://doi.org/10.2307/1341787

Hirschman, D., & Bosk, E. A. (2020). Standardizing biases: Selection devices and the quantification of race. *Sociology of Race and Ethnicity*, *6*(3), 348–364. https://doi.org/10.1177/2332649219844797

Holland, M. M. (2019). *Divergent paths to college: Race, class, and inequality in high schools*. Rutgers University Press. https://doi.org/10.36019/9780813590288

Holland, M. M., & Ford, K. S. (2021). Legitimating prestige through diversity: How higher education institutions represent ethno-racial diversity across levels of selectivity. *Journal of Higher Education*, *92*(1), 1–30. https://doi.org/10.1080/00221546.2020.1740532

Howell, Jessica, Hurwitz, M. H., Mabel, Z., & Smith, J. (2021). *Participation in student search service is associated with higher college enrollment and completion*. College Board. Retrieved from https://cbsearch.collegeboard.org/pdf/college-outreach-and-student-outcomes.pdf

Howell, J., & Korver-Glenn, E. (2018). Neighborhoods, race, and the twenty-first-century housing appraisal industry. *Sociology of Race and Ethnicity*, *4*(4), 473–490. https://doi.org/10.1177/2332649218755178

Jaquette, O., Han, C., & Castaneda, I. (forthcoming). The private school network: Recruiting visits to private high schools by public and private universities. In S. Burd (Ed.), *Lifting the veil on enrollment management: How a powerful industry is limiting social mobility in american higher education*. Book Section, Cambridge, MA: Harvard Education Press.

Jaquette, O., & Salazar, K. G. (2022). *Student list policy: Problems, regulations, and a solution*. Washington, DC: TICAS. Retrieved from #

Jaquette, O., Salazar, K. G., & Martin, P. (2022). *The student list business: Primer and market dynamics*. washington, DC: TICAS. Retrieved from https://ticas.org/wp-content/uploads/2022/09/The-Student-List-Business_-Primer-and-Market-Dynamics.pdf

Khan, S. R. (2011). *Privilege: The making of an adolescent elite at st. Paul's school* (p. 232). Princeton, N.J.: Princeton University Press.

Kolluri, S. (2018). Advanced placement: The dual challenge of equal access and effectiveness. *Review of Educational Research*, *88*(5), 671–711. https://doi.org/10.3102/0034654318787268

Korver-Glenn, E. (2018). Compounding inequalities: How racial stereotypes and discrimination accumulate across the stages of housing exchange. *American Sociological Review*, *83*(4), 627–656. https://doi.org/10.1177/0003122418781774

Korver-Glenn, E. (2022). *Race brokers: Housing markets and racial segregation in 21st century urban america* (pp. pages cm). New York, NY: Oxford University Press.

Lawler, M., & Dold, M. (2021). For-profit schools as covered persons under the CFPA. *Berkeley Law, Center for Consumer Law & Economic Justice*. Retrieved from https://protectborrowers.org/wp-content/uploads/2021/03/SBPC_UCB_For-Profits.pdf

Marx, K. (1978). Capital. In R. C. Tucker (Ed.), *The marx-engels reader* (second). New York: W.W. Norton; Company.

McDonough, P. M. (1997). *Choosing colleges: How social class and schools structure opportunity* (pp. xi, 174 p.). Albany: State University of New York Press.

McPherson, M. S., & Schapiro, M. O. (1998). *The student aid game*. Princeton, NJ: Princeton University Press.

Miller, B. J., & Skimmyhorn, W. L. (2018). I want you! Expanding college access through targeted recruiting efforts. *Education Finance and Policy*, *13*(3), 395–418. https://doi.org/10.1162/edfp_a_00232

Moore, J. (2017). *Do students who opt into ACT's educational opportunity service (EOS) enroll in college at higher rates?* ACT, Inc. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/R1652-benefits-of-act-eos-opt-in-2017-08.pdf

Napier, A., & Orrick, A. (2022). The economic, social, and political dimensions of platform studies in education. *Harvard Educational Review*, *92*(2), 206–208. https://doi.org/10.17763/1943-5045-92.2.206

Nichols, T. P., & Garcia, A. (2022). Platform studies in education. *Harvard Educational Review*, *92*(2), 209–230. https://doi.org/10.17763/1943-5045-92.2.209

Niu, S. X., & Tienda, M. (2010). Minority student academic performance under the uniform admission law: Evidence from the university of texas at austin. *Educational Evaluation and Policy Analysis*, *32*(1), 44–69. https://doi.org/10.3102/0162373709360063

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: New York University Press.

Norris, D. (2021). Embedding racism: City government credit ratings and the institutionalization of race in markets. *Social Problems*. https://doi.org/10.1093/socpro/spab066

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Park, J. J., & Becks, A. H. (2015). Who benefits from SAT prep?: An examination of high school context and race/ethnicity. *Review of Higher Education*, *39*(1), 1–23. Retrieved from <Go to ISI>://WOS:000360777300001

Posecznick, A. (2017). *Selling hope and college merit, markets, and recruitment in an unranked school*. Ithaca ; London: Cornell University Press. Retrieved from https://muse.jhu.edu/book/52616

Posselt, J. R., Jaquette, O., Bielby, R., & Bastedo, M. N. (2012). Access without equity: Longitudinal analyses of institutional stratification by race and ethnicity, 1972-2004. *American Educational Research Journal*, *49*(6), 1074–1111.

Ray, V. (2019). A theory of racialized organizations. *American Sociological Review*, *84*(1), 26–53. https://doi.org/10.1177/0003122418822335

Robinson, C. J. (2000). *Black marxism: The making of the black radical tradition* (pp. xxxiii, 436). Chapel Hill, N.C.: University of North Carolina Press. Retrieved from Table of Contents http://lcweb.loc.gov/catdir/toc/99030995.html

Rodriguez, A., & McGuire, K. M. (2019). More casses, mre acess? Understanding the effects of course offerings on back-white gaps in advanced placement course-taking. *Review of Higher Education*, *42*(2), 641–679. Retrieved from <Go to ISI>://WOS:000454537100016

Rothstein, R. (2017). *The color of law: A forgotten history of how our government segregated America*. Liveright Publishing.

Ruffalo Noel Levitz. (2018). *2018 marketing and student recruitment report of effective* *practices*. Ruffalo Noel Levitz. Retrieved from http://learn.ruffalonl.com/rs/395-EOG-977/images/RNL_2018_Student_Recruitment_Marketing_Report_EM-19.pdf

Ruffalo Noel Levitz. (2020). *2020 marketing and recruitment practices for undergraduate students report*. Ruffalo Noel Levitz. Retrieved from https://learn.ruffalonl.com/rs/395-EOG-977/images/2020_Marketing_Recruitment%20Practices_Undergraduate_Students.pdf

Ruffalo Noel Levitz. (2021). RNL student search and engagement. Retrieved from https://www.ruffalonl.com/enrollment-management-solutions/building-demand/student-search-and-engagement/

Russell, N. C., Reidenberg, J. R., Martin, E., & Norton, T. B. (2019). Transparency and the marketplace for student data. *Virginia Journal of Law and Technology*, *22*(3), 107–157.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 12. https://doi.org/10.1177/2053951718820549

Sadowski, J. (2020). The internet of landlords: Digital platforms and new mechanisms of rentier capitalism. *Antipode*, *52*(2), 19. https://doi.org/10.1111/anti.12595

Salazar, K. G. (2022). Recruitment redlining by public research universities in the los angeles and dallas metropolitan areas. *The Journal of Higher Education*, 1–37. https://doi.org/10.1080/00221546.2021.2004811

Salazar, K. G., Jaquette, O., & Han, C. (2021). Coming soon to a neighborhood near you? Off-campus recruiting by public research universities. *American Educational Research Journal*, *58*(6), 1270–1314. https://doi.org/10.3102/00028312211001810

Salazar, K. G., Jaquette, O., & Han, C. (2022). *Geodemographics of student list purchases: A first look*. Washington, DC: TICAS. Retrieved from https://ticas.org/wp-content/uploads/2022/09/Geodemographics-of-Student-List-Purchases_A-First-Look.pdf

Santelices, M. V., & Wilson, M. (2010). Unfair treatment? The case of freedle, the SAT, and the standardization approach to differential item functioning. *Harvard Educational Review*, *80*(1), 106–133.

Schmidt, D. (2022). *Prospect search filters*. Encoura. Retrieved from https://helpcenter.encoura.org/hc/en-us/articles/360035260452-Prospect-Search-Filters-

Simon, J. (1988). The ideological effects of actuarial practices. *Law & Society Review*, *22*(4), 771–800. Journal Article. https://doi.org/10.2307/3053709

Smith, J., Howell, J., & Hurwitz, M. (2021). The impact of college outreach on high schoolers' college choices: Results from over one thousand natural experiments. *Education Finance and Policy*, 1–25. https://doi.org/10.1162/edfp_a_00334

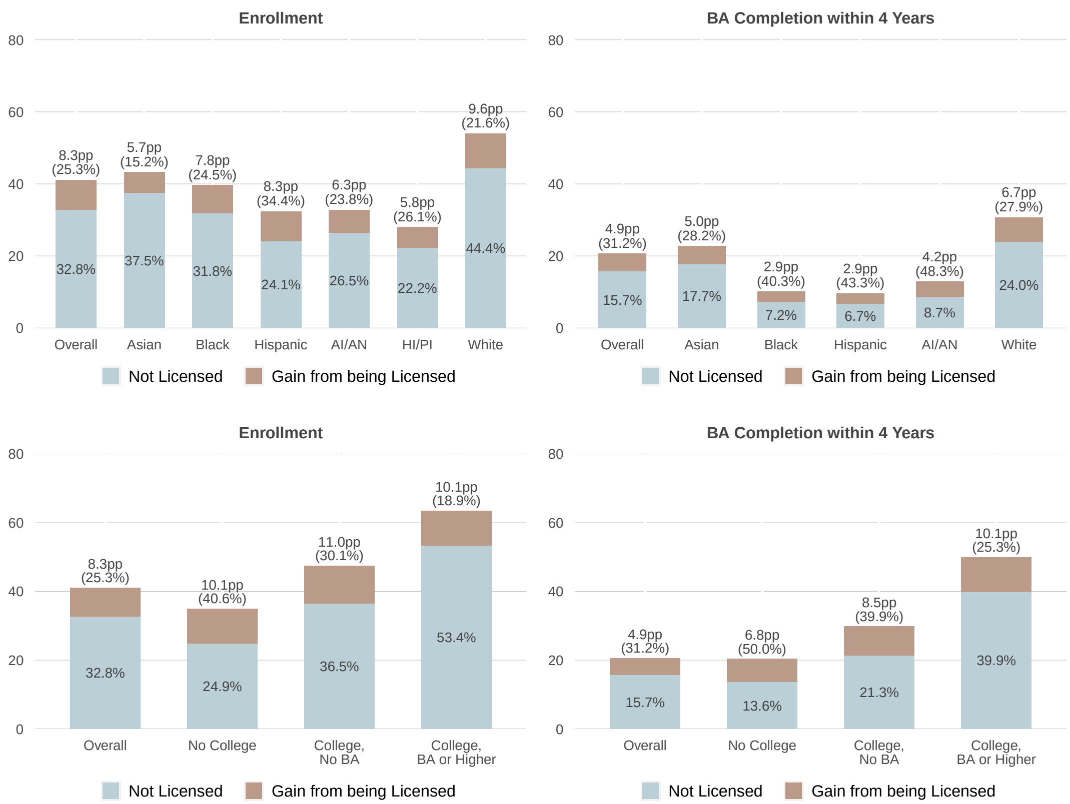
Stevens, M. L. (2007). *Creating a class: College admissions and the education of elites* (p. 308). Cambridge, MA: Harvard University Press.

Tiako, M. J. N., South, E., & Ray, V. (2021). Medical schools as racialized organizations: A primer. *Annals of Internal Medicine*, *174*(8), 1143–1144. https://doi.org/10.7326/m21-0369 %m 34058105

Waxman, B. (2019). The power of list segmentation part i: Are you already doing this? *intead, Global & Local Academic Branding*. Retrieved from https://services.intead.com/blog/the-power-of-list-segmentation-part-i-are-you-already-doing-this

9 Figures

Figure 1: Student Search Service and four-year college enrollment/completion



Notes: AI/AN = American Indian or Alaska Native. HI/PI = Hawaiian or Pacific Islander. Sample for enrollment outcomes is all SAT takers in the 2015–2018 high school graduation cohorts. Sample for completion outcomes is students in the 2015–2016 cohorts. Results are estimated from regressions that include student-level controls for: sex, race/ethnicity, SAT score, parental education level, last Student Search Service opt-in status, graduation cohort, and high school fixed effects. All differences between licensed versus non-licensed students are statistically significant at the 1% level.

Figure 2: The enrollment funnel

Figure 3: Filters used in College Board orders purchased by 14 public universities