Structuring College Access: The Market Segment Model and College Board Geomarkets

Ozan Jaquette (UCLA)   Karina Salazar (University of Arizona)

ABSTRACT

The Structure of College Choice (Zemsky & Oedel, 1983) created “Geomarkets” and the “Market Segment Model.” Geomarkets carve states and metropolitan areas into smaller geographic units, meant to define local recruiting markets. The Market Segment Model predicts how student demand for a particular college varies by Geomarket, based on the socioeconomic characteristics of households. Geomarkets became an input for two College Board products that help colleges recruit students. First, the Enrollment Planning Service (EPS) software recommends specific Geomarkets and high schools from which colleges should recruit. Second, the Student Search Service sells the contact information of prospective students – referred to as “student lists” – and colleges can filter by Geomarket to determine which prospect profiles they purchase. We draw from scholarship on quantification, particularly the discussions of correlation and homophily by Chun (2021), to conceptualize how recruiting products incorporate Geomarkets. We address two research questions: What is the socioeconomic and racial variation between Geomarkets and how does this variation change over time? How does the socioeconomic and racial composition of included versus excluded prospects vary when student list purchases filter on particular Geomarkets? We answer RQ1 by analyzing Census data from 1980, 2000, and 2020. We answer RQ2 using data on student lists purchased by public universities, which we collected by issuing public records requests. We utilize a quantitative case study design. Metropolitan areas are cases. Analyses consist of descriptive statistics and interactive maps.

1 Introduction

On January 9, 2019 a public research university purchased College Board *Student Search Service* order no. 448922, thereby obtaining the contact information of out-of-state prospective students, who would subsequently receive brochures, postcards, email, and targeted social media. This was a large student list purchase, yielding $122,426$ "names." At \$0.45 per name, the order cost \$$55,091.70$. The order specified three "search filters" to control which names were purchased: first, the 2020 high school graduating class; second, under the heading "College Board Exams," PSAT score range of 1070 - 1180; and third, under the heading "Geography", the order filtered on 67 "Geomarkets," identified by titles like "IL08 - Northwest Suburbs" and "IL09 - North Shore."[1] The Geomarket filter carves states and metroplitan areas into smaller geographic units. Upon closer inspection, this student list purchase selected all Geomarkets in the Chicago-land area except for "IL10 - City of Chicago," which is significantly poorer and has a higher share of Black residents. Similarly, in the DC/Maryland/Virginia metro area, the order selected the affluent "MD02 - Montgomery Metropolitan," "MD03 - Central Maryland (w/o Baltimore)," "VA01 - Arlington & Alexandria," and "VA02 - Fairfax County" Geomarkets, but excluded the "MD05 - Prince Georges Metropolitan," "MD07 - Baltimore (Urban)," and "DC01 - District of Columbia" Geomarkets.

Sociology offers several perspectives to describe the process of sorting students into colleges as a process that tends toward social reproduction (Domina et al., 2017; Stevens et al., 2008). On the student demand side, the status attainment model argues that student postsecondary destination is a function of family socioeconomic background, particularly parental education (e.g., Blau & Duncan, 1967; Fishman, 2020; Karen, 2002; Sewell, 1971; Sewell & Shah, 1967). The cultural capital model argues that bourgeois households bestow the pedigree, social networks, and information necessary for their progeny to claim spots at selective colleges (Bourdieu, 1984, 1988; Huang, 2023) [ADD CITES]. Domina et al. (2017) describes schools as "sorting machines" that create, incorporate, and assign students to categories that determine educational opportunities, often reinforcing reinforcing racial and class inequality. On the supply-side, the credentialism literature recognizes that colleges have a financial incentive for educational credentials to be required to claim

[1] On the same day, the university purchased two additional orders that filtered on PSAT score ranges of $1190-1260$ and $1270-1520$, respectively, but selected the same HS graduating class and Geomarkets.

advantageous social and economic positions (Weber, 1948), and this race for credentials benefits affluent households (Labaree, 1997) [cite others]. Scholarship on enrollment management offers an agentic depiction of organizational behavior, showing that colleges expend considerable resources recruiting desirable prospects (Cottom, 2017; Holland, 2019; Salazar et al., 2021; Stevens, 2007).

We argue that these perspectives ignore an important mechansim of social reproduction: third-party vendors create products that sort students on behalf of college. These products often amplify historic inequality in educational opportunity by using past snapshots of student demand to make recommendations about where colleges should prioritize recruiting efforts. In the contemporary era of private equity funded platform capitalism, third-party recruiting products have become increasingly powerful, granular, and ubiquitous. However, we argue that a watershed moment in their develeopment was the creation of Geomarkets by the College Board and UPenn professor, Robert Zemsky.

In 1983, Robert Zemsky and Penny Odell authored *The Structure of College Choice*. College Board underwrote the project – providing data and funding – and published the book as part of their nascent efforts "to help colleges estimate their enrollment potential, efforts which had faltered largely because the smallest geographic unit used in these analyses was the state" (Zemsky & Oedel, 1983, p. x). The broader project was conceptualized as an effort to to "capture and quantify" (p. 11) the knowledge of admissions officers, based on the idea that "a good recruiter knows where to look for prospective applicants, as seen in the students' willingness or reluctance to travel" (p. 11). Zemsky & Oedel (1983) developed the Market Segment Model by analyzing the SAT score-sending behavior of 1980 high school seniors. The thesis of the Market Segment Model is that student demand for your institution is a function of social class and geography. Therefore, colleges should recruit from territories that contain large populations of your target social class. Geomarkets are geographic borders meant to define local recruiting markets, a territory of an admissions recruiter. Figure X, panel A shows Geomarkets in the Chicago-land area. As a planning tool for colleges, the Market Segment Model predicts how demand for a particular college varies by Geomarket, based on characteristics of households in the Geomarket.

Geomarkets have been incorporated into the supply-side structure of college access in three ways. Anecdotally, Geomarkets became an organizing principle for how college admissions offices allocate admissions recruiters to territories. Figure X panel B shows that Chicago-land Geomarkets define recruiting territories for University of Chicago admissions recruiters (Appendix Figures BLAH BLAH). Second, Geomarkets are the basis for the College Board Enrollment Planning Service (EPS), which was in 1984 and remains active today. EPS software recommends which Geomarkets a college should recruit from and which schools/communities they should prioritize within targeted Geomarkets. According to Noel-Levitz (1998), in 1995, 37% of 4-year public institutions and 49% of 4-year private institutions used EPS, while 41% of 4-year publics and 16% of 4-year privates used ACT's market analysis product, which was based on EPS.

Third, Geomarkets were incorporated into College Board's student list product, Student Search Service. Student lists contain the contact information of prospective students and have been the primary source of lead generation in U.S. higher education since 1972, when College Board began selling names (Belkin, 2019; Jaquette et al., 2022). Ruffalo Noel-Levitz (2022b) reported that 87% of private and 86% of public four-year institutions purchase student lists. The Student Search Service database consists of College Board test-takers. Colleges pay a fee for each prospect (e.g., $0.50 in 2021). Colleges control which prospect profiles they purchase by selecting search filters, such as high school graduation year, SAT score, AP score, state, etc. Geomarket filters enable colleges to include/exclude prospects from particular Geomarkets.

Scholarship on enrollment management has focused on college behavior (e.g., Cottom, 2017; Salazar, 2022; Salazar et al., 2021; Stevens, 2007), contributing to the belief that recruiting is a function of individual colleges. Scholarship has failed to investigate the broader enrollment management industry – consisting of third-party vendors, their products and consulting solutions – as a set of mechanisms that structure college access. These mechanisms, which lie upstream of the behavior of individual colleges, are incorporated by individual colleges in ways that funnel certain kinds of students to certain kinds of institutions. This omission is surprising because sociology is concerned with sorting (Domina et al., 2017), particularly sorting into college.

This manuscript analyzes College Board Geomarkets as a case study of quantification. In particular, we draw from the discussions of correlation and homophily from Chun (2021). Zemsky & Oedel (1983) identified the correlates of 1980 SAT score-sending behavior and concluded that student demand for higher education is primarily a function of social class. The Market Segment Model argues that homophily – actors that share characteristics form connections – is the organizing principle of competition and college choice, stating that "the hierarchical structure of collegiate competition largely reflects the stratified social and economic dimensions of the communities from which colleges draw their students" (Zemsky & Oedel, 1983, p. 72). Scholarship on quantification demonstrates that making recommendations based on past correlations amplifies the effects of historic structural inequality (Burrell & Fourcade, 2021). The snapshot of student demand in 1980 – itself a consequence historic structural inequality – was programmed into recruiting products that colleges utilize to identify and target prospective students. The result is a supply-side that amplifies structural inequalities observed on the demand-side.

Our analyses address the two research questions, which speak to how Geomarkets are utilized within EPS and within the Student Search Service student list product. First, what is the socioeconomic and racial variation between geomarkets in metropolitan areas and how does this variation change over time? We address this question by spatially joining the geomarket shapefile to Census data about socioeconomic and racial characteristics. Second, how does the socioeconomic and racial composition of included versus included prospects vary when student list purchases filter on particular geomarkets? This research question is motivated by Zemsky & Oedel (1983), which recommends that selective colleges target affluent Geomarkets. Additionally, prior research on enrollment management behavior finds that selective colleges focus recruiting visits on high schools in affluent, predominantly white communities (Salazar et al., 2021; Stevens, 2007). We analyze this research question using data from actual student list purchases, which were collected using public records requests. We analyze student list purchases that include all Geomarkets in a metropolitan area in only to assess which prospects would be included/excluded by had the student list purchase filtered on particular Geomarkets.

The following section introduces background and scholarship about enrollment management and introduces salient concepts from scholarship on quantification, particularly correlation and homophily. Second, we introduce the case, describing how Zemsky & Oedel (1983) developed Geomarkets and the Market Segment Model and how these concepts were incorporated into College Board recruiting products. Third, we describe data and methods. Fourth, we present results. [We find that…?PARAGRAPH ON RESULTS?] Finally, we discuss implications for scholarship and for policy. The sociology of education should move beyond the fixation on schools and policies as the primary sorting machines. DiMaggio & Powell (1983, p. 148) remind us that the organizational field includes the totality of relevant actors." Third-party vendors have structured college access for decades and their influence is growing. The contemporary enrollment management industry is characterized by private equity owned firms selling software-as-service platforms that utilize the same logic as Zemsky & Oedel (1983), for example EAB's "pipeline analytics" machine learning product [CITE]. By investigating high-leverage actors and products in the organizational field, sociology can develop important insights that shape future federal and state regulatory policy. [REVISE THIS PARAGRAPH]

2 Enrollment Management and Quantification

Enrollment Management

Enrollment management is simultaneously a profession, an administrative structure, and an industry. As a profession, enrollment management (EM) integrates techniques from marketing and economics in order to "influence the characteristics and the size of enrolled student bodies" (Hossler & Bean, 1990, p. xiv). As an administrative structure, the EM office typically controls the activities of admissions, financial aid, and recruiting (Kraatz et al., 2010).

Figure 1 depicts the "enrollment funnel," which modifies the traditional "marketing funnel" to depict broad stages in the process of recruiting students (EAB, 2019; Litten et al., 1983). The funnel begins with a large pool of "prospects" (i.e., prospective students) that the university would like to enroll. "Leads" are prospects whose contact information has been obtained. "Inquiries" are prospects that contact the institution, including those that respond to an initial solicitation (e.g., email) and those that reach out on their own (e.g., sending SAT scores). The purpose of the enrollment funnel is to inform recruiting interventions that target one or more stages. These interventions seek to increase the probability of "conversion" across stages (Campbell, 2017). At the top of the enrollment funnel, purchasing student lists is the primary means of converting prospects to leads (Jaquette et al., 2022). Purchased leads are served emails, brochures, and targeted social media designed to solicit inquiries and applications (Ruffalo Noel-Levitz, 2022b).

Scholarship at the nexus of enrollment management and college access can be categorized by which part(s) of the enrollment funnel it speaks to. The majority of scholarship focuses on the admissions stage, analyzing which admissions criteria are utilized and/or which applicants are admitted (e.g., Hirschman et al., 2016; Killgore, 2009; O. Y. A. Poon & Bastedo, 2022; Posselt, 2016; Taylor et al., 2024). Scholarship from the economics of education often investigate financial aid leveraging, which seeks to convert admits to enrolled students (Hurwitz, 2012; e.g., Leeds & DesJardins, 2015).

A growing literature analyzes the earlier "recruiting" stages of identifying leads, soliciting inquiries, and soliciting applications. Salazar et al. (2021) conceptualize recruiting behavior as an indicator of college enrollment priorities. Ethnographies by Stevens (2007) and Khan (2011) identify connections between private school guidance counselors and college admissions officers as a mechanism for social reproduction. In turn, recruiting visits to high schools – and subsequent phone calls – are a means of maintaining ties with guidance counselors at feeder schools and establishing relationships with prospective students (Ruffalo Noel-Levitz, 2022b; Stevens, 2007). Quantitative case-studies of off-campus recruiting visits by public research universities and by selective private universities reveal a preference for visiting private schools and affluent, predominantly white public schools (Jaquette et al., 2024; Salazar, 2022; Salazar et al., 2021). From the student perspective, Holland (2019) finds that underrepresented students were drawn to colleges that made them feel wanted, often attending institutions with lower graduation rates and requiring larger loans than other college options. Cottom (2017) shows that for-profit colleges found a niche in Black and Latinx communities because traditional colleges ignored these communities.

By focusing on the behavior of colleges, scholarship on EM implicitly states that recruiting is something done by individual colleges. The broader EM industry includes relevant stakeholders in the organizational field (DiMaggio & Powell, 1983), including professional associations (e.g., National Association for College Admission Counseling) and third-party servicers (e.g., College Board, EAB) that supply products and consulting solutions to colleges. We argue college access is structured by third-party servicers and products that interact with direct-providers (colleges). Although sociologists have hinted at the ways enrollment management contributes to inequality in college access (e.g., Kraatz et al., 2010), scholarship has failed to make third-party servicers and products the object of empirical analysis.

Drawing from scholarship on organizational theory, enrollment management processes require colleges to make many "make or buy" (Coase, 1937; Scott & Davis, 2007) decisions about whether to perform a given task in-house or outsource it to a third-party vendor (Jaquette et al., 2022). Unfortunately, few empirical studies make EM vendors or products the object of empirical analysis. The most widely-known class of EM vendor is EM consulting firms, which provide advice and implementation in the areas of marketing, recruiting, pricing and financial aid, and student success. As "creating a class" has becomes complicated and high-stakes (Stevens, 2007), many colleges hire EM consulting firms to develop and/or implement recruiting campaigns. The two largest firms – Ruffalo Noel Levitz and EAB – claim to serve more than 3,000 colleges and universities collectively (EAB, n.d.; Ruffalo Noel Levitz, 2023).[2]

While the contemporary EM industry is characterized by software-as-service products sold to colleges by private equity backed firms (Jaquette et al., 2022), College Board products played a pivotal role in transforming recruiting from an in-house process to a process structured by third-party products. Every year, colleges must figure out which prospects should receive promotional material and which high schools to visit. The Student Search Service – created in 1972 and copied by ACT's Educational Opportunity Service – became the ubiquitous means of converting prospects into leads that could be served targeted recruiting interventions. Second, the Enrollment Planning Service (EPS), launched in 1984, is an early software-as-service product that provides recommendations about which Geomarkets colleges should recruit from and which high schools they should visit within targeted Geomarkets. Both products are applications of quantification. Student Search Service includes or excludes prospects from a student list purchase based on value of selected variables while EPS encourages colleges to target Geomarkets and high schools based on the characteristics of households. The multi-disciplinary literature on quantification helps conceptualize how third-party products utilize Geomarkets and the potential consequences of this utilization on equality of opportunity.

Quantification

Espeland & Stephens (2008, p. 402) define quantification as "the production and communication of

[2] In our data collection, attempting to obtain data about student list purchases from all public universities in four states, at least half of these universities outsourced student list purchases to an EM consultancy (Jaquette et al., 2022; Salazar et al., 2022).

numbers" (p. 402). Reacivity, discipline, and authority are three interrelated themes that describe the effects of quantification. *Reactivity* is the idea that salient quantitative measures cause people and organizations to change their behavior. Quantification *disciplines* actors to react in particular ways. Quantification changes decision-making power *authority*, often weakening the discretion of local decision-makers. Scholarship about U.S. News & World Report (USNWR) Law School rankings demonstrate the effects of quantification (Espeland & Sauder, 2007, 2016; Sauder, 2008; Sauder & Espeland, 2009). Law school rankings were developed as a means of informing prospective students about the relative quality of different programs. Rankings affect (reactivity) where students apply and enroll (Espeland & Sauder, 2016). Rankings also affected hiring decisions of firms because clients evaluated firms based on the prestige of law schools attended by firm lawyers. Once law schools realized that "important groups of constituents — students, faculty, trustees, employers, other media — were using rankings to make decisions that had large consequences for schools …[then] schools felt pressured to take them seriously" (Espeland & Stephens, 2008, p. 415). Rankings *disciplined* the behavior of law schools. For example, characteristics valued by the rankings system (e.g., LSAT scores) became more important for decisions about admissions and merit aid (Espeland & Sauder, 2016). Rankings also weakened the *authority* of admissions personnel to make independent decisions class size and which applicants to admit (Espeland & Sauder, 2016).

The analysis of UK school "league tables" by McArthur & Reeves (2022) shows how quantification can be a mechanism of social reproduction. In 1992, the UK government began publishing school league tables, which ranked schools based on student performance on national exams taken at age 16. League tables facilitate making evaluative comparisons between schools regardless of geographic proximity. They also discipline households and schools to conceive of school quality in terms of test scores, which are substantially a function of the class composition of schools. Using Census data measured at the local authority level (similar to a U.S. county), McArthur & Reeves (2022) find that localities with higher performing schools experienced growth in the share of managerial/professional households following the adoption of league tables and a decline in the share of working-class households. Using longitudinal survey data, managerial/professional households were more likely to move to localities with higher ranked schools following the introduction of school league tables. The introduction of league tables contributed to social reproduction because professional/managerial households were more aware of these consumer-facing metrics and had resources to respond by moving to more expensive neighbhorhoods, near higher performing schools.

Correlation. The interdisciplinary literature on quantification (Mennicken & Espeland, 2019) includes important contributions from the field of critical data studies (e.g., Noble, 2018; O'Neil, 2016). In particular, the discussions of correlation (chapter 1) and homophily (chapter 2) by Chun (2021) introduce ideas salient to the analysis the Market Segment Model and Geomarkets. Correlation measures the extent to which two or more variables move together. Predictive analytics are based on correlation and are developed in two steps. First, apply statistical techniques to previous cases (training data) in order to identify factors positively and negatively associated with an outcome of interest. Second, apply these results (e.g., regression coefficients) to future cases in order to make predictions and/or to assign levels of risk to each case. Chun (2021) provides the example of Kosinski et al. (2013), who develop a method to predict sensitive personal attributes (e.g., gender, political party) based on Facebook Likes. These models predict outcomes based on correlations without requiring knowledge about underlying causal relationships. Chun (2021, p. 50) writes that

> correlation grounds big data's so-called revolutionary potential. As Wired editor Chris Anderson infamously declared …, big data proved that 'correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.'

Due to data limitations, predictive analytics often utilize "proxy" variables (Chun, 2021; O'Neil, 2016), which are variables that are correlated with a variable of interest. For example, "e-scores" utilize proxy variables to identify "people like you" and then predict your buying behavior based on the past buying behavior of people like you. O'Neil (2016, p. 146) states that "the modelers for e-scores have to make do with trying to answer the question 'How have people like you behaved in the past?' when ideally they would ask, 'How have you behaved in the past?' " We observe similar behavior in market research about college access, when enrollment managers use 'which colleges did students near you consider' as a proxy for, 'which colleges would you consider?'

Many studies show that predictions based on correlations reproduce structural inequality (for a review see Burrell & Fourcade, 2021). The correlations observed during the training data stage are a snapshot of relationships between variables at a particular point in time. Observed correlations may be a function of enduring structural inequality, but underlying causes are not considered by applications of predictive models. Reviewing scholarship about algorithms, Burrell & Fourcade (2021, p. 224) state that "predicting the future on the basis of the past threatens to reify and reproduce existing inequalities." Disproportionately targeted/excluded populations are predicted to have a higher risk of an outcome, which amplifies subsequent targeting/exclusion. This phenomenon has been termed the "ratchet effect" (Harcourt, 2015) and "pernicious feedback loops" (O'Neil, 2016).[3] As we discuss below, Zemsky & Oedel (1983) analyzed 1980 score-sending behavior (training data) and inferred that demand for college is a function of class. They then recommended that selective colleges recruit from localities that have a high share of affluent, college-educated households.

Homophily. Homophily is the idea actors who share common characteristics are likely to form connections with one another, or "birds of a feather flock together" (McPherson et al., 2001). Homophily is a core concept of network science, in which actors (nodes) are connected to one another directly and indirectly via network ties (edges). In *The company We Keep: Colleges and Their Competition*, Zemsky & Oedel (1983, Chapter 4) conduct a network analysis whereby two colleges are defined as competitors if a large number of students submit SAT scores to both colleges. Chun (2021) argues that because network science models often draw from rational choice theory, they assume that homophily is the result of voluntary action by individuals.

Chun (2021) problematizes the idea that homophily is a naturally occurring phenomenon. Because network science models – like correlational models above – typically do not observe how historical structures affect relationships, they "erase historical contingencies, institutional discrimination, and economic realities" (Chun, 2021, p. 95) that cause behavior consistent with homophily. Second, in commercial social networks, homopphily is more than an assumption; rather, it is programmed into algorithms that create connections between users. Thus, "social networks create and spawn the reality they imagine; they become self-fulfilling prophecies" (Chun, 2021, p. YY). We observe

[3] An often cited example is the LSI-R recidivism model which predicts a prisoner's chances of re-arrest and is used by 24 states (O'Neil, 2016). Because the algorithm uses zip code as an input, people who live in highly policed neighborhoods have a higher predicted probability of being arrested, which leads to more policing in those neighborhoods, which perpetuates racialized inequality in arrests. Note that predictive models such as the LSI-R model are not merely used for social science purposes. Rather, they reproduce structural inequality because they direct the allocation of future resources based on a snapshot of historical inequality.

similar phenomena in our case. Based on a snapshot of 1980 SAT score-sending, Zemsky & Oedel (1983) concludes that like-colleges compete for like-students (i.e., students of similar social origin). In turn, Zemsky & Oedel (1983) reasons that colleges should target the Geomarkets and high schools targeted by peer colleges. This logic is subsequently programmed into EPS software that recommends which Geomarkets and high schools to recruit from. Thus, homophily observed on the demand side – which is a function of historic structural inequality — is programmed into the supply-side.

Homophily is central to market research products that categorize customers. Geodemography emerged in the 1970s as a branch of market research that estimates the behavior of consumers based on where they live (Burrows & Gane, 2006). Market segments are subgroups within a larger market that have similar consumer demand. Early geodemographic classifications of consumers (e.g., PRIZM by Claritas Corporation) were derived from publicly available Census data, which disaggregated data to the zip code level. The Claritas Corporation had a financial incentive to argue that people living near one another share similar consumer preferences because geographic localities could then be categorized into market segments that would be useful for direct mail marketing campaigns (McKelvey, 2022). Later, the development of individual credit scores (e.g., FICO score) enabled merchants to classify consumers into many, fine-grained groups (M. Poon, 2007). Fourcade & Healy (2013) introduce the concept "classification situations" to describe the expansion of actuarial techniques to categorize customers into many, ordinally ranked groups. Merchants and lenders began tying these classifications to tiered products that targets different consumer groups with different levels of benefits and costs (Fourcade & Healy, 2024).[4] Classification situations engender markets where a vertical hierarchy of products are matched to a vertical hierarchy of consumers. We observe similar processes in our case study. Zemsky & Oedel (1983) categorizes students into four market segments – local, in-state, regional, and national – based on their score sending behavior and then evaluates the attractiveness of each Geomarket based on how many "regional" and "national" students live there.

We make targeted contributions to scholarship on quantification in the sociology of education. First, in the case of college rankings, extant research shows how consumer-facing quantification disciplines

[4] For example, "payday loans" charge high interest rates to consumer groups that were previously denied credit altogether.

the application/enrollment behavior of prospective students, the hiring behavior of employers, and the behavior of colleges that are disciplined to pursue customers with characteristics valued by rankings (Espeland & Sauder, 2007, 2016; Sauder & Espeland, 2009). As Stevens (2007) demonstrates in the chapter *Numbers*, EM is fundamentally concerned with quantification. However, sociological analyses of enrollment management focus on the behavior of colleges and their agents, yielding the implicit belief that inequalities created by EM are a function of individual colleges reacting to rankings and other macro structures in their environment. To our knowledge, the sociology of education has not considered producer-facing products that quantify students on behalf of producers. Market research and investigative journalism indicates that the vast majority of colleges purchase third-party EM products to identify prospects and to decide which schools and communities to target [Marcus (2024);]. Drawing from scholarship from critical data studies (e.g., Benjamin, 2019; Chun, 2021; O'Neil, 2016), recruiting products grounded in the logic of predictive analytics take a snapshot of student demand – without considering historical structures that produce inequality – and then recommend that colleges divert recruiting resources to localities with strong student demand. Based on a snapshot of existing social stratification, market research matches verticially categorized consumers to vertically categorized producers, thereby amplifying the effect of initial stratification on subsequent stratification. Third-party EM products have been sold since the 1980s and have become more abundant with the ubiquity of software-as-service platforms, but have remained invisible to sociologists. We make third-party EM products the object of empirical analysis, though we cannot claim to assess how the population of colleges utilize these products.

This manuscript analyzes Market Segment Model (Zemsky & Oedel, 1983), which categorized high school students into vertical market segments – local, in-state, regional, national – and simultaneously created local Geomarkets that could be evaluated based on their composition of student market segments. The Market Segment Model and Geomarkets became the basis for the College Board *Enrollment Planning Service* (EPS), which advised colleges which Geomarkets to target. Later, Geomarkets were incorporated into the College Board student list product named Student Search Service. Unlike the analysis of UK school league tables, McArthur & Reeves (2022) we cannot show the effect of quantification on social reproduction. However, by showing which students are included/excluded in student list purchases that filter on particular Geomarkets, our analyses provide novel insight into the underlying mechanism by which quantification can reproduce historical class-based and race-based inequality in educational opportunity.

3 The Market Segment Model and College Board Geomarkets

Creating Geomarkets and the Market Segment Model.

In 1978, set against the backdrop of impending college-age demographic decline, University of Pennsylvania professor Robert Zemsky, was tasked by the President to figure out, " 'Who thinks about Penn?' " and "'What other institutions do they think about when they think about us?" (Zemsky & Oedel, 1983, p. x). To answer these questions, Zemsky began working with the Market Research Committee of the Consortium on Financing Higher Education (COFHE), a consortium of 30 selective private universities founded in the mid-1970s. The project goal became the creation of the Market Segment Model, which had the lofty ambition of predicting student demand for any college from any locality. However, "to gain a truly comprehensive view of the collegiate enrollment market, we needed a database that described most institutions and most students" (Zemsky & Oedel, 1983, p. x). The group approached College Board. "Coincidentally, the Board was reviewing its own efforts to help colleges estimate their enrollment potential, efforts which had faltered largely because the smallest deomgraphic unit used in these analyses was the state" (Zemsky & Oedel, 1983, p. x).

Although the analyses of Zemsky & Oedel (1983) are based on the score-sending behavior of SAT test-takers, the authors had great respect for admissions officers and conceived of the project as an effort to "capture and quantify" (p. 11) their knowledge:

> Admissions officers invariably are tellers of stories – about the colleges they represent, about the colleges they attended, about each other, and about the often vagabond life of college recruiting (p.9)...We have begun with this celebration of sotrytelling...[because] we believe that the intuitios of admissions officers actually comprise a remarkably systematic body of knowledge about the college selection process...Our research this is based opn listening carefully to what admissions officers have to say (pp. 9-10).

The "initial task" was to define geographic boundaries "in a manner consistent with admissions officers' intuitive understanding of student pools" (Zemsky & Oedel, 1983, p. 4). Quoting an admissions officer, Zemsky & Oedel (1983, p. 11) write, " 'There are only three kinds of college-bound students: those who want to live at home, those who want to live on campus but bring their laundry home, and those who want to go far enough from home that Mom and Dad can't visit without calling first.' " As such, Zemsky & Oedel (1983, p. 11) created "three types of boundaries – region, state, and community. The initial regions in the project were New England, Middle States, and the South. Next, they"divided each state into as few as two and as many as thirty community-based enrollment markets or pools, for a total of 143 separate markets" (p. 11). These enrollment markets, later called Geomarkets, were intended to be consistent with the conception of a catchment market from the perspective of admissions counselors. Zemsky & Oedel (1983, pp. 11–12) described the creation of Geomarket borders only briefly:

> In many cases, the market boundaries match formal political and educational divisions, reflecting natural channels of communication. Each major metropolitan area is composed of several markets, usually corresponding to the inner city, a first ring of suburbs, and an outer ring of suburbs. In more sparsely populated areas, communities are sometimes combined in order to make the analysis meaningful."

Having defined geography, the Market Segment Model sought to predict how student demand for a particular college varies by Geomarket, based on the characteristics of households in the Geomarket:

> For our analysis, we sought not a complex mathematical model, but a straightforward classification system that would track the pattern of SAT-score submissions to create a map of student choice. The Market Segment model that we developed was nothing more than a set of simple rules for disaggregating high school seniors into similar groups. The model worked because students, once so disaggregated, appeared to behave in remarkably consistent ways (Zemsky & Oedel, 1983, p. 4).

SAT test-takers were categorized into one of four different *market segments* – local, in-state, regional, and national – based on SAT score-sending behavior. For a given student, each college that receives a score from the student can be defined as "local" (college located in the the same Geomarket as the student), "in-state" (same state but different Geomarket), "regional" (same region but different state), or "national" (different region). In turn, a test-taker is categorized in the "local" market segment if they submit more SAT scores to local institutions than they do to in-state, regional, or national institutions. An "in-state" student submits more SAT scores to in-state institutions than they do to local, regional, or national institutions, etc.

The two primary outputs of the Market Segment Model are the (1) Market Segment Profile and (2) the Institutional Profile. Appendix A describes these outputs in more detail. Both outputs are created separately for each Geomarket. For each market segment (local, in-state, regional, national), the Market Segment Profile shows the number of students, their average SAT score, the percent aspiring to a BA+, percent with family in come greater than $35,000, percent with both parents having a BA, etc.[5] The Institutional Profile shows the number of students who send scores to a particular institution – separately by market segment – and which majors these students are interested in. A university could obtain this for their own institution or for a competitor.[6] Taken together, the Market Segment Profile shows a colleges which Geomarkets possess attractrive student market segments while the Institutional Profile shows the extent to which students in a particular Geomarket are interested in your college or a competitor college. These outputs became the basis for the Enrollment Planning Service (EPS) software.

Correlations. Zemsky & Oedel (1983, Chapter 3), *A Sense of Place: Students, Families, and Communities*, identifies the student characteristics correlated with being in the local, in-state, regional, or national market segment. The analyses identify four variables – educational aspirations, parental education, scholastic aptitude, and family income – that predict score sending behavior, both individually and in combination. Because these four variables "reflect the basic social patterns of the nation, it would have been surprising if these were not the four social variables that best explained the patterns of college choice" (Zemsky & Oedel, 1983, p. 33).

The thesis of Chapter 3 is that student demand for higher education is a function of social origin. For the authors (p. 33) state that "these data allow us to say with considerable confidence that local and in-state students are not likely to come from families in which both parents have received

[5] For example, Appendix Table A1 shows how many students in "CT3 – Fairfield County" are defined as regional or national based on their SAT score-sending behavior.

[6] Appendix Table A3 reproduces a partial, simplified version of Zemsky & Oedel (1983, fig. 2.3), the Institutional Profile of an anonymous college for students from Fairfield County, CT. This table shows that 69 Regional students and 109 National students from Fairfield County sent SAT scores to the college.

college educations" and that "the implication is simply that college-educated parents instill in their children more wide-ranging educational aspirations." Commenting on family income, Zemsky & Oedel (1983, p. 33) write that "we could predict that all local students would come from moderate-income or low-income families and be wrong only 5.5 percent of the time."[^7] Zemsky & Oedel (1983) [p. 42] conclude that "our research has simply demonstrated what everyone has always known: communities with high levels of family income and parental education are also communities in which students have higher than average SATs and more far-reaching aspirations." The authors also find that these four variables predict student score-sending behavior at the Geomarket-level. However, Geomarkets differ in the relative abundance of students with particular socioeconomic characteristics, which has practical implications for recruiting. Therefore, Zemsky & Oedel (1983, p. 44) recommend that colleges target Geomarkets with desirable compositions of socioeconomic characteristics in order to reach students from desired student market segments:

> On occasion, senior spokespersons for the profession worry that students outside the main [Geo]market areas remain forgotten and hence, unchallenged. Inevitably, the increasing competition for students, the expense of travel and mailings, and internal political constraints compel institutions to concentrate their efforts where they will do the most good. The result is a natural reinforcing of the basic socioeconomic patterns that gave shape in the first place to the structure of college choice.

Homophily. *The Company We Keep: Colleges and Their Competition* (Zemsky & Oedel, 1983, Chapter 4) conducts a network analysis to determine which institutions are in competition with one anothyer.[7] Based on these analyses, Zemsky & Oedel (1983, p. 46) state that competition between colleges is characterized by homophily: "we draw a fundamental conclusion about the structure of college choice: collegiate competition occurs principally between like institutions." Subsequent

[7] This is a "two-mode" social network in which students (mode 1) send SAT scores – the network tie – to colleges (mode 2). The authors turn this into a one-mode college network that defines two institutions as being in competition with one another – the network tie – if at least 15% of students who sent SAT scores to one institution also sent scores to the other institution and vice-versa. Next, the authors develop "tinker toy" diagrams that show which institutions are connected to one another. These diagrams are drawn separately for each student segment – local, in-state, regional, and national – and separately for each geomarket, such that the analyses convey which institutions compete with one another for which student segments in each local market. For example, describing the Figure 4.4 "Structure of Fairfield County Regional Market," (Zemsky & Oedel, 1983, p. 54) state that "competitive overlap, moreover, is often confined to institutions belonging to the same [Carnegie] type as well sector. For example, public flagships compete primarily with other public flagships; private standard colleges, with other private standard colleges; Catholic institutions, with other Catholic institutions."

analyses investigate the tuition price and the socioeconomic composition of institutions in competition with one another. Private selective colleges and private flagship universities compete directly for students, charge the highest prices, and enroll students with the highest socioeconomic status. The authors argue that like-colleges compete for like-students as defined by socioeconomic characteristics. Zemsky & Oedel (1983, p. 72) describe observed patterns as a natural process in which a vertical socioeconomic hierarchy of students is matched to a vertical hierarchy of universities:

> Students describe themselves socially simply by telling us the colleges and universities in which they are interested. The layering of collegiate competition is primarily a socioeconomic layering. The hierarchical structure of collegiate competition largely reflects the stratified social and economic dimensions of the communities from which colleges draw their students. Competition among colleges, as admissions officers have told us for so long, is in fact, a matter of keeping company with one's peers.

The discussion of competition by Zemsky & Oedel (1983) exemplifies the concerns about correlation and homophily described by Chun (2021). A correlational analysis of 1980 SAT score-sending patterns finds competition between colleges is defined by socioeconomic homophily. This homophily is presented as a naturally occurring phenomenon. Given these findings, Zemsky & Oedel (1983) recommend that colleges should target Geomarkets that contain a critical mass of students interested in peer-coleges, information that can be discerned from the Institutional Profiles (Appendix Table A3). In itself, Zemsky & Oedel (1983)'s Market Segment Model is a social science depiction of student demand – akin to the status attainment model – that does not consider historic, structural inequalities that cause observed patterns. However, by inscribing the Market Segment Model and Geomarkets into the EPS software, College Board amplified structural inequalities that contributed to homophily observed in 1980 SAT score-sending beahivor. In this way, commodification of the Market Segment Model and Geomarkets "engineers homophily" (Chun, 2021).

Enrollment Planning Service. In 1984, College Board created the Enrollment Planning Service (EPS), based on the Market Segment Model (College Board, 2012; Takamiya, 2005). EPS was an early software-as-service platform that recreated the analyses of Zemsky & Oedel (1983). For each Geomarket, colleges could obtain Market Segment Report for each local market and the Institutional Profile – their own and that of competitors. Based on background conversations with enrollment management professionals, EPS software also provided information about the score-sending behavior of individual high schools within each Geomarket. Therefore, colleges used EPS software to decide which Geomarkets to recruit from and which high schools to visit within targeted Geomarkets. Typical College Board (2005) marketing material describes EPS as,

> The marketing software that pinpoints the schools and Geomarkets where your best prospects are most likely to be found. With the click of a mouse, EPS provides you with comprehensive reports on your markets, your position in those markets, and your competition. Focus your valuable time and resources on the right prospects.

Whereas Zemsky & Oedel (1983) identified 143 Geomarkets covering the New England, Middle States, and South region, EPS created Geomarkets for the remaining U.S. states, with 304 Geomarkets in total. College Board (2023) shows the contemporary Geomarkets. Documentation and promotional material suggests that geomarket borders were chosen based on a combination of formal geographic borders (e.g., counties) as well as proprietary College Board data designed to identify geographic areas with different college-going behaviors.[8] However, Geomarkets for New England, Middle States, and the South are identical to those developed by Zemsky & Oedel (1983).

Berman & Hirschman (2018) argue that quantification has effects to the extent that stakeholders care about the numbers. Market research suggests that EPS was highly salient. Noel-Levitz (1998) reports that in 1995, 37% of 4-year publics and 49% of 4-year privates used EPS, while 41% of 4-year publics and 16% of 4-year privates used ACT's market analysis service product.

EPS software may "discipline" (Espeland & Stephens, 2008) colleges to approach recruiting in a manner consistent with the Market Segment Model. Drawing from promotional literature (e.g., College Board, 2005, 2011a), Takamiya (2005), and background conversations with enrollment managers, the practical purpose of EPS was to inform the "travel schedule" of admissions recruiters. EPS promotional material and user guides encourages users to begin by identifying which Geomarkets they will recruit from.[9] Second, EPS users decide which high schools they will visit within

[8] College Board (n.d.) states that "geomarkets are areas within a state that represent a further segmentation of a population. Students from California don't all share the same college-going behaviors. We have accounted for this variance by segmenting the 50 states into 304 geomarkets to provide further insight into student behaviors within particular areas of individual states."

[9] Software documentation by Oracle (n.d.) states that, "EPS market codes are proprietary market codes owned by the College Board and are used to categorize external organizations and people into geographical areas.... Some

selected Geomarkets. Based on the principle of homophily, the Market Segment Model suggests that selective colleges should target Geomarkets with large numbers of affluent, college educated households, while low-income communities are left to local four-year and community colleges.

EPS may also weaken the authority of local decision-makers. Zemsky & Oedel (1983) sought to develop a concrete, data-driven framework – The Market Segment Model – that replicates the aggregate knowledge of local admissions officers. Once this knowledge was quantified and commodified onto a CD-ROM, the local expertise of admissions officers becomes less valuable. The EPS product increases the ability of a college admissions leader – working with College Board staff or an enrollment management consultant – to plan recruiting efforts centrally. In background conversations, enrollment management professionals told us that EPS software enabled colleges to plan travel without relying on admissions officers having strong local knowledge of their territories. However, EPS software often recommended visiting the same sets of affluent, high-achieving high schools that were receiving visits from other colleges. We were told that savvier, quantitatively-adept admissions offices used EPS to visit schools that EPS recommended ignoring because there would be less competition for the good students who attended these high schools.

Finally, we suspect that Geomarkets affect how admissions offices organize the recruiting territories assigned to each admissions officer. Geomarkets were created to mirror the territories of admissions officers (Zemsky & Oedel, 1983). In turn, as Geomarkets became more salient, admissions offices often structured their territories around Geomarkets. On background, admissions officers and enrollment consultants told us it remains common parlance to hear an admissions officer say something like, "I recruit 'PA 2'," which refers to the "Chester County, PA Geomarket. In states a college recruits heavily from, admissions officers are often assigned specific Geomarkets as their territory. This can be seen in the Chicago-land recruiting territories of University of Chicago [ADD TO APPENDIX], of Lake Forest College, of Saint Mary's College [ADD TO APPENDIX CRYSTAL]

Although we cannot gain access to EPS software, Geomarkets are fundamental to EPS software and (arguably) to the organization of recruiting territories for many colleges. Therefore, we are

admissions offices use EPS market codes to focus their recruiting efforts in geographic areas in which they believe they will be the most successful."

interested in the extent to which this building block is associated with race and class. Research question 1 asks, what is the socioeconomic and racial variation between Geomarkets and how does this variation change over time? We focus on metropolitan areas that are associated with three or more Geomarkets.

We expect substantial socioeconomic inequality between geomarkets in large metropolitan areas. Given the extent of class- and race-based residential segregation in the U.S., it would be surprising to not observe such inequality. Moreover, Zemsky & Oedel (1983) viewed demand for higher education as a function of class and developed Geomarkets with an eye towards identifying geographic areas that differed from one another in terms of class composition.

We also expect substantial racial inequality between geomarkets in large metropolitan areas. Interestingly, the Market Segment Model is explicitly based on socioeconomic stratification, but Zemsky & Oedel (1983) do not mention race once. U.S. cities are characterized be extreme historic and contemporary residential racial segregation (Korver-Glenn, 2022). Structures built upon racialized structures are racialized structures (Norris, 2021). Unless designers intentionally consider racial segregation, selection devices that categorize people based on geographic location are likely to reproduce historical race-based inequality in opportunity (Chun, 2021). Second, geomarket borders may have been drawn along class divides. A strong correlation exists between race and wealth (Kraus et al., 2019). Third, Geomarket borders may have been drawn in a way that follows the countours of racial segregation in residential housing. Examples include: the "South and South Central Los Angeles" geomarket (CA21); the "City of Oakland" geomarket (CA07), which is surrounded by the "Alameda County excluding Oakland" geomarket (CA08); and the "Wayne County Detroit" geomarket (MI01), which is surrounded by the "Detroit's Northern Suburbs" (MI02) and "Ann Arbor" (MI03) geomarkets.

Geomarket Filter in Student Search Service. By 1984, College Board Geomarkets were incorporated as a search filter in the College Board Student Search Service [CITE]. A student list contains the contact information of prospective students who meet the search filter criteria (e.g., test score, GPA) specified by the university. Student lists are the fundamental input for undergraduate recruiting campaigns because purchased names – alongside prospects who reach out on their own – constitute the set of prospects who receive subsequent recruiting interventions (e.g., mail, email) designed to push them toward the application and enrollment stages of the "enrollment funnel." Ruffalo Noel-Levitz (2022b) reports that 86% of public colleges and 87% of private colleges purchase student lists.[10] Historically, dominant list vendors are college board and ACT, which derived student list data from test-takers, but the test-optional movement, advances in technology, and surging private equity investments have contributed to new sources of student list data (Jaquette et al., 2022).

Student lists are positively associated with student outcomes. A College Board research report by Howell et al. (2021) compared SAT test-takers who opted into the College Board Student Search Service – allowing colleges to purchase their contact information – to those who opted out. After controlling for covariates (e.g., SAT score, parental education, school fixed effects), 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 percent change – (41.1–32.8)/32.8 – in the relative probability.[11] Howell et al. (2021) also found that participating in Search was strongly associated with obtaining a BA in four years, particularly for sub-groups that were historically excluded from college.[12]

However, Jaquette & Salazar (2024) argue that student list products exacerbate racial inequality in college access. Jaquette & Salazar (2024) conceptualize student list products as "selection devices" (Hirschman & Bosk, 2020) that enable colleges to select which prospective students they target by incorporating search filters (e.g., high school graduating class, state). Norris (2021) defines "racialized inputs" as ostensibly race-neutral inputs that are systematically correlated with race because marginalized racial/ethnic groups have historically been excluded from the input (Norris,

[10]For public universities that purchased lists, 80% purchase more than 50,000 names annually. Ruffalo Noel-Levitz (2022a) reports that student lists were the top expenditure item in the undergraduate recruiting budget for both private and public institutions in 2022, with the average public institution allocating 15% of its budget to purchasing names.

[11]Participating in Search was associated with a larger change in the relative probability of attending a 4-year college for Black students (24.5%) and Hispanic students (34.4%) than White students (21.6%), and a larger change for students whose parents did not attend college (40.6%) than those whose parents had a bachelor's degree (18.9%). Leveraging a natural experiment in College Board student list purchases, Smith et al. (2022) find that purchasing a prospect profile increases the probability that the student will apply to and enroll at the purchasing college, with larger effects for Black, Hispanic, and lowincome students.

[12]20.6% of students who participated in Search obtained a BA in four years compared to 15.7% of students who opted out, representing a 31.2% (=(20.6-15.7)/15.7) increase in the relative probability of graduation. Furthermore, the relative increase in the probability of obtaining a BA was higher for Black (40.3%), Hispanic (43.3%), and Native American/Alaska Native students (48.3%) than it was for White (27.9%) and Asian (28.2%) students. The relative increase was also higher for students whose parents did not attend college (50.0%) than it was for students whose parents had a BA (25.3%).

2021). Jaquette & Salazar (2024) argue that several frequently utilized student list filters (zip code, AP test score, SAT score) meet the criteria of racialized inputs. Using a national sample of high school students and using data from actual student lists purchased by public universities, Jaquette & Salazar (2024) show that racialized search filters have a strong negative relationship with the selection of Black and Latinx prospects.

Because of the extent of residential segregation in the U.S., geographic borders are a commonly studied racialized input in scholarship about algorithmic bias (Benjamin, 2019; Harcourt, 2007; O'Neil, 2016). Student list products offer many geographic filters. Some geographic filters are based on known geographic borders (e.g., zip code, county, CBSA, state). College Board also created geographic borders that subsequently became filters in the *Student Search Service*, for example the Geodemographic Segment filters.[13] This manuscript focuses on Geomarkets. Drawing from Norris (2021), we argue that geomarkets satisfy the two criteria of racialized inputs. First, they are ostensibly race-neutral inputs in that neither Zemsky & Oedel (1983), nor subsequent promotional material (e.g., College Board, 2011a), mention race. Second, Geomarkets may be systematically correlated with race. For example, Geomarket "CA07" is "City of Oakland" and "CA08" is "Alameda Country, excluding Oakland." A university might purchase prospect contact information by filtering on CA08 in combination with additional filters, such as year of high school graduation and SAT test score. This article extends Jaquette & Salazar (2024) by examining how the racial composition of actual student list purchases changes if universities filter on particular Geomarkets.

Zemsky & Oedel (1983, pp. 42–44) discuss how the Market Segment Model and Geomarkets should be utilized in the context of purchasing student lists. The authors ask,

> Imagine for the moment that you recruit for a college that draws most of its students from regional or national segments. Where would you concentrate your energies? Ideally, you would seek communities [Geomarkets] with a high proportion of students already predisposed toward institutions such as your own. The Market Segment Model would provide this information through segment percentages for the community in question.

[13] For example, the geodemographic segment filters utilize cluster analysis to allocate each census tract and each high school into different categories based past college enrollment and other factors (College Board, 2011b).

> Further classification of students by social attributes allows you to identify a group for mailings or recruiting…If you were to recruit in Boston, only about two out of every ten students with fewer than two attributes would likely listen, while slightly less than half of the students with two or more attributes would be receptive…. Your efforts would surely be better directed toward …Manchester, Hartford, and Fairfield County …Indeed, in Fairfield County alone you could reach more than 40 percent of your "primary target" population – that is, students with a greater than 75-percent probability of concentrating their college choices among institutions like your own.

In both mailings and off-campus recruiting visits, Zemsky & Oedel (1983) recommend focusing on Geomarkets with large populations of students who are deciding between "institutions like your own" (p. 44). Selective colleges primarily draw from students in the "regional" and "national" segments. Zemsky & Oedel (1983) suggest that the student list purchases of selective colleges should focus on Geomarkets with large numbers of affluent, highly educated households. By contrast, the student list purchases of local state colleges and community colleges, which rely on enrollment from students in the "local" and "in-state," should focus on nearby Geomarkets with large numbers of low-SES students. Our analyses attempt to recreate the strategy recommended by Zemsky & Oedel (1983) by showing who is included/excluded when student list purchases filter on particular Geomarkets. Considering the correlation between race and wealth in segregated America, we suggest that a class-based list-buying strategy will increase racial stratification in college access. Furthermore, we anticipate that the students of color from affluent Geomarkets will tend to have college-educated parents and be enrolled in affluent, predominantly white schools.

4 Data and Methods

4.1 Data and Variables

Research question 1, data. RQ1 asks, what is the socioeconomic and racial variation between Geomarkets in metropolitan areas? And, how does this variation change over time? To answer this question, we use census tract-level data from the US Census Bureau. Census tract is the smallest geographic area for which measures of both race/ethnicity and socioeconomic characteristics are available. Utilizing a small geographic area is important for implementing a spatial merge between Census data and Geomarket shape files, described below.

We utilize data from the 1980 Decennial Census, specifically variables from Summary Tape File 1 (STF1), which contains data from the about race and hispanic origin from the "short form" questions answered by all households and data from Summary Tape File 3 (STF3), which contains data about socioeconomic characteristics – including income, education, and poverty – derived from the "long form" questionnaire completed by a sample of households. The 1980 Decennial Census data collection period closely matches the data collection period of SAT score-sending behavior that Zemsky & Oedel (1983) utilized to create Geomarkets and the Market Segment Model. In order to examine how the characteristics of Geomarkets changed over time, we utilize tract-level data from the 2000 Decennial Census, similarly using data from Summary File 1A (SF1A) to obtain measures of race and Hispanic origin and using data from Summary File 3A (SF3A) to obtain measures of income, education, and poverty.[14] Beginning in 2010, the Decennial Census no longer collected a "long form" questionnaire. These questions were replaced by the American Community Survey (ACS). We also utilized data from the 2020 5-year ACS, which includes data collected from 2016-2020.[15] This data collection period mirrors the period for our primary data collection about student list purchases used to analyze RQ2, as described below.

We assigned each census tract to College Board Geomarkets by implementing a partial spatial merge. Spatial manipulations utilized functions from the *sf* R package. Data to create the shapefiles for the College Board Geomarkets were obtained from a 2012 R-bloggers post.[16] Essentially, the data file had one observation per zip code and a column that assigned each zip code to one Geomarket. Using these zip code-level data, we utilized the `aggregate()` function to create shapefile for each Geomarket.

Next, we utilized the `st_intersection()` function to assign each census tract to a Geomarket, based on the intersection of their associated shapefiles. Census tract shapefiles that intersected with multiple Geomarkets were broken into smaller shapefiles wholly contained within a single

[14] Data for the 1980 and 2000 Decennial Census were retrieved from the IPUMS National Historical Geographic Information System, created by Manson et al. (2024).

[15] Data for the 2020 5-year ACS were retreived from the tidycensus R package, created by Walker & Herman (2024)

[16] The author of the post reported that he found the zip codes associated with Geomarkets from a Google search.

Geomarket. The variables in these shapefiles were then replaced with the original variable value multiplied by the proportion of land area of the census tract that was contained within the Geomarket. For a particular census tract, imagine that 60% of its land area was contained in Geomarket A, 40% of its land area was contained in Geomarket B, and the census tract reports 1,000 people who identify as Hispanic non-white. The partial spatial merge splits this census tract into two observations. The observation assigned to Geomarket A reports $.60 * 1000 = 600$ Hispanic non-white people and the observation assigned to Geomarket B reports $.40 * 1000 = 400$ Hispanic non-white people.

Research question 1, variables. RQ1 asks, what is the socioeconomic and racial variation between Geomarkets in metropolitan areas? And, how does this variation change over time? We created measures of race and ethnicity, specifically the number and proportion of people (all ages) in the following categories: White, non-Hispanic; Black, non-Hispanic; Asian, non-Hispanic; Pacific Islander, non-Hispanic; American Indian and Alaskan Native, non-Hispanic; and two+ races, non-Hispanic. Measures of XXX were not available for 1980 Decennial Census. We also created socioeconomic measures that approxmiate the predictors of market segment identified by Zemsky & Oedel (1983). These variables are median household income, mean household income, percent of households below the poverty line, and people age 25+ with a BA.[17] All measures were created at the census tract-level and the Geomarket-level by aggregating tract-level data to Geomarkets.

Research question 2, data. Geomarket is a search filter available in the College Board Student Search Service product. RQ2 asks, how does the racial and socioeconomic composition of included versus included prospects vary when student list purchases filter on particular geomarkets? We answer this question using actual student lists purchased by public universities.

In February 2020, we began issuing public records requests to public universities about student lists purchased from 2016 through 2020. We narrowed the scope of our request to student lists purchased from College Board, ACT, and the National Research Center for College and University Admissions (NRCCUA), the three largest student list vendors at the time. For each student list purchased from 2016 through 2020, we requested two related pieces of data: (1) the order summary,

[17]STATE THAT MEAN HOUSEHOLD INCOME CREATED BY DIVIDING AGGREGATE INCOME BY NUMBER OF PEOPLE. NEED FIX THIS MEASURE THO.

which specifies search criteria for the student list purchase; and (2) the de-identified prospect-level list produced from the search criteria.

The data collection sample for this project was all public universities in CA, IL, MN, and TX. Utilizing public records requests to obtain public records is a painstaking process. Initially, the majority of universities did not respond to our request or denied our request. Subsequently, we obtained pro bono representation from four law firms, which substantially increased the success of data collection. We collected data from Arizona State University and Northern Arizona University, because we were able to obtain pro-bono legal representation for these two universities. However, we were unable to obtain legal representation for TX.

Even with firm representation, data collection remained difficult. Some universities provided records that were not usable for quantitative analyses (e.g., summary statistics across multiple orders; or data did not contain important fields). Some universities did not provide records based on legitimate grounds (e.g., data not in university possession; not required to create records that do not currently exist). We learned that many universities outsourced student list purchases to a third-party consulting firm. Unfortunately, we were rarely able to obtain usable data from these universities. A small number of universities denied requests based on potentially questionable legal rationale, but we lacked the resources to litigate.

This article analyzes student lists purchased from College Board. We exclude student list purchases by MN public universities from this report because Minnesota is predominantly an "ACT state." 14 public universities provided PDF order summaries containing the search criteria for 830 student list purchases. We utilized the XXXX Python/R package to convert these order summaries into tabular data. About 16% of student list purchases utilized the Geomarket filter.

We received both (a) the order summary data and (b) the de-identified prospect-level student list data for 414 orders associated with 2,549,085 prospects. We draw from these data to answer RQ2. Although these lists were purchased by individual universities, the set of prospects included in each list is a function of the search criteria specified for that student list purchase.[18]. Therefore, we utilize these data not to analyze the behavior of individual universities, but to identify which

[18]ADD CAVEAT THAT PURCHASES EXCLUDE PROSPECT PROFILES THAT WERE PREVIOUSLY PURCHASED

prospective students are included when a particular set of search criteria are selected.

In particular, we answer RQ2 by analyzing particular student list purchases that did *not* utilize the Geomarket search filter, but did purchase prospects from all Geomarkets in a particular metropolitan area. For each prospect we know their home zip code and, therefore, which Geomarket they belong to. We answer RQ2 by simulating which prospective students would have been included or excluded from the student list purchase had the purchase filtered on partiulcar Geomarkets.

Research question 2, variables. The variables of interest for RQ2 the characteristics of prospective students who would have been included/excluded by our simulations and the characteristics of their schools and neighborhoods. Prospect characteristics are derived from the pretest questionnaire administered to College Board test-takers. The de-identified student lists we received contain a standard subset of variables (e.g., HS code, HS GPA range, intended major). We create detailed measures of race and ethnicity. Information about student socioeconomic status is limited. However, we create a measure of whether the prospect is first-generation college student. We can also measure the socioeconomic characteristics of the students' home zip code and characteristics of their high school [EXPAND OR DELETE THIS SENTENCE DEPENDING ON WHETHER WE USE THESE DATA].

4.2 Methods

Research design. We utilize a multiple, quantitative case study design in which metropolitan areas are cases. Following XXX [CITE], we focus on a small number [HOW MANY] of metropolitan areas in order to provide sufficient analytic depth while also situating each case within the context of historic segregation and contemporary gentrification. Online Appendix XXX presents results for a larger number of metropolitan areas.

Choice of metropolitan areas is informed by several factors. Based on our read of Zemsky & Oedel (1983), we are interested in "major metropolitan area[s] ...composed of several markets, usually corresponding to the inner city, a first ring of suburbs, and an outer ring of suburbs" (p. 11-12). Building on our conceptual framework, we are interested in racial and socioeconomic inequality between geomarkets within a metropolitan area (RQ1) and how this inequality contributes to inequality in which prospective students are targeted by student list purchases (RQ2). [DISCUS-

SION OF WHICH CASES TO SELECT. KARINA]. Choice of metropolitan areas is informed by data availability. Although Census data are available for all metropolitan areas (RQ1) we do not have good candidate student list purchases for all metropolitan areas (RQ2). [SAY SOMETHING ABOUT WHAT IS A GOOD CANDIDATE STUDENT LIST PURCHASE – PURCHASE WITH NOT TOO MANY CRITERIA THAT INCLUDE ALL GEOMARKETS IN A METRO].

Analyses. Analyses are simultaneously descriptive and spatial. We answer RQ1 by producing Geomarket-level tables and graphs that show how the Geomarkets in a selected metropolitan area vary on racial and socioeconomic characteristics and how they vary over time. For example, in the San Francisco Bay Area, how does the City of Oakland Geomarket (CA07) differ from the Alameda County Exluding Oakland Geomarket (CA08). We also produce interactive maps at the census tract-level to show more granular variation within and between Geomarkets.

We answer RQ2 by analyzing student list purchases that encompassed all Geomarkets in a selected metropolitan areas. Descriptive tables and graphs describe the racial and socioeconomic characteristics of prospects that would have been included/excluded had the purchase filtered on particular Geomarkets. Interactive maps provide a more granual visualization of where prospects lived and went to high school. For example, for purchased profiles who lived in the Chester County Geomarket (PA02) [OR PHILADELPHIA PA05] and identified as Black, where in the county did these students live and which high schools did they attend?

NEXT STEPS: CREATE TABLE FOR PHILLY STUDENT LIST PURCHASES THAT FOCUSES ON FIRST GEN STATUS.

Our conceptual framework suggests that racial and socioeconomic inequality is Geomarket regions is associated with the number of geomarkets in the metropolitan area. and is associated with the patterns of racial segregation in the metropolitan area. comment more on this.

Below, we use simulations and actual student list purchases that filtered on geomarkets. In purchases that filtered on test-score and/or GPA thresholds, we can simluate who would be included and excluded had certain Geomarkets were been selected. In purchases that filter on geomarkets, we can get an initial sense of who is included in geomarkets targeted by regional state colleges versus research universities.

5 Discussion

In their analysis of quantifying school quality in England, McArthur & Reeves (2022, p. 517) observe that "one problem with school league tables …is that the measures of school quality often merely reflect the social origins of those who attend a particular school." Similarly, considering prior research showing that SAT scores are substantially a function of social origin (Sewell & Shah, 1967), the Market Segment Model argues that student demand for higher education is mostly a function of social origin.

Institutional theory defines the organizational field as "those organizations that, in the aggregate, constitute a recognized area of institutional life: key suppliers, resource and product consumers, regulatory agencies, and other organizations that produce similar services and products" (DiMaggio & Powell, 1983, p. 143).

6 References

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

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

Berman, E. P., & Hirschman, D. (2018). The sociology of quantification: Where are we now? *Contemporary Sociology-a Journal of Reviews*, *47*(3), 257–266. https://doi.org/10.1177/0094306118767649

Blau, P. M., & Duncan, O. D. (1967). *The american occupational structure* (pp. xvii, 520 p.). New York: Wiley.

Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste* (pp. xiv, 613 p.). Cambridge, Mass.: Harvard University Press.

Bourdieu, P. (1988). *Homo academicus* (pp. xxvi, 344 p.). Cambridge, UK: Polity Press. Retrieved from file://C%3A%2FUsers%2Fozanj%2FDocuments%2Facademic%20papers%2Fphilosophy%20of%20social%20sciences%2Fhomo-academicus-Bourdieu.doc

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

Burrows, R., & Gane, N. (2006). Geodemographics, software and class. *Sociology-the Journal of the British Sociological Association*, *40*(5), 793–812. https://doi.org/10.1177/0038038506067507

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/

Chun, W. H. K. (2021). *Discriminating data: Correlation, neighborhoods, and the new politics of recognition* (pp. xi, 327 pages). Cambridge, Massachusetts: The MIT Press.

Coase, R. (1937). The nature of the firm. *Economica*, *4*, 386–405.

College Board. (2005). Enrollment planning services. College Board. Retrieved from www.collegeboard.com/highered/ra/eps.html

College Board. (2011a). Enrollment planning service. The College Board. Retrieved from https://collegeboardsearch.collegeboard.org/pastudentsrch/support/licensing/college-board-search-services/enrollment-planning-service

College Board. (2011b). *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. (2012). *Welcome to the college board*. College Board. Retrieved from https://secure-media.collegeboard.org/homeOrg/content/pdf/welcome_to_the_college_board_fullversion.pdf

College Board. (2023). *Detailed connections geography maps: Reach more students where they are*. College Board. Retrieved from https://cbsearch.collegeboard.org/media/pdf/connections-geographies.pdf

College Board. (n.d.). *SAT trends dashboard report: Interpretive guide*. College Board. Retrieved from https://satsuite.collegeboard.org/media/pdf/sat-trends-dashboard-interpretive-guide.pdf

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

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160.

Domina, T., Penner, A., & Penner, E. (2017). Categorical inequality: Schools as sorting machines. *Annual Review of Sociology*, *43*, 311–330. https://doi.org/10.1146/annurev-soc-060116-053354

EAB. (n.d.). *EAB: Managing school communication during COVID-19*. Retrieved from https://www.vistaequitypartners.com/spotlight/eab-school-communication-covid19/

EAB. (2019). The 5 key stages of college enrollment—and which metrics to track during each. Retrieved from https://eab.com/insights/daily-briefing/enrollment/the-5-key-stages-of-college-enrollment-and-which-metrics-to-track-during-each/

Espeland, W. N., & Sauder, M. L. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, *113*(1), 1–40. Retrieved from <Go to ISI>://000248485000001

Espeland, W. N., & Sauder, M. L. (2016). *Engines of anxiety: Academic rankings, reputation, and accountability* (pp. xii, 281 pages). New York, New York: Russell Sage Foundation.

Espeland, W. N., & Stephens, M. L. (2008). A sociology of quantification. *Archives Europeennes De Sociologie*, *49*(3), 397–432. Retrieved from <Go to ISI>://WOS:000265147000002

Fishman, S. H. (2020). Educational mobility among the children of asian american immigrants. *American Journal of Sociology*, *126*(2), 260–317. https://doi.org/10.1086/711231

Fourcade, M., & Healy, K. J. (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

Fourcade, M., & Healy, K. J. (2024). *The ordinal society* (pp. pages cm). Cambridge, Massachusetts ; London, England: Harvard University Press.

Harcourt, B. E. (2007). *Against prediction: Profiling, policing, and punishing in an actuarial age*. Chicago: University of Chicago Press.

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

Hirschman, D., Berrey, E., & Rose-Greenland, F. (2016). Dequantifying diversity: Affirmative action and admissions at the university of michigan. *Theory and Society*, *45*(3), 265–301. https://doi.org/10.1007/s11186-016-9270-2

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

Hossler, D., & Bean, J. P. (1990). *The strategic management of college enrollments.* Jossey-Bass.

Howell, J., 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

Huang, T. J. (2023). Translating authentic selves into authentic applications: Private college consulting and selective college admissions. *Sociology of Education.* https://doi.org/10.1177/00380407231202975

Hurwitz, M. (2012). The impact of institutional grant aid on college choice. *Educational Evaluation and Policy Analysis*, *34*(3), 344–363. https://doi.org/10.3102/0162373712448957

Jaquette, O., Han, C., & Castañeda, I. (2024). The private school network: Recruiting visits to private high schools by public and private universities. *Research in Higher Education*, *65*(6), 1269–1315. Journal Article. https://doi.org/10.1007/s11162-024-09783-w

Jaquette, O., & Salazar, K. G. (2024). A sociological analysis of structural racism in ”student list” lead generation products. *Educational Evaluation and Policy Analysis*, *46*(2), 276–308. https://doi.org/10.3102/01623737231210894

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

Karen, D. (2002). Changes in access to higher education in the united states: 1980-1992. *Sociology of Education*, *75*(3), 191–210.

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

Killgore, L. (2009). Merit and competition in selective college admissions. *Review of Higher Education*, *32*(4), 469–488. https://doi.org/10.1353/rhe.0.0083

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

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc Natl Acad Sci U S A*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kraatz, M. S., Ventresca, M. J., & Deng, L. N. (2010). Precarious values and mundane innovations: Enrollment management in american liberal arts colleges. *Academy of Management Journal*, *53*(6), 1521–1545. https://doi.org/10.5465/amj.2010.57319260

Kraus, M. W., Onyeador, I. N., Daumeyer, N. M., Rucker, J. M., & Richeson, J. A. (2019). The misperception of racial economic inequality. *Perspectives on Psychological Science*, *14*(6), 899–921. https://doi.org/10.1177/1745691619863049

Labaree, D. F. (1997). *How to succeed in school without really learning: The credentials race in american education* (pp. x, 323 p.). New Haven, Conn.: Yale University Press.

Leeds, D. M., & DesJardins, S. L. (2015). The effect of merit aid on enrollment: A regression discontinuity analysis of iowa's national scholars award. *Research in Higher Education*, *56*(7), 471–495. https://doi.org/10.1007/s11162-014-9359-2

Litten, L. H., Sullivan, D. J., & Brodigan, D. L. (1983). *Applying market research in college admissions* (pp. xxii, 303 p.). New York: College Entrance Examination Board.

Manson, S., Schroeder, J., Van Riper, D., Knowles, K., Kugler, T., Roberts, F., & Ruggles, S. (2024). *IPUMS national historical geographic information system: Version 19.0* (Report). IPUMS. https://doi.org/http://doi.org/10.18128/D050.V19.0

Marcus, J. (2024). Going to a "witch doctor": Colleges' reliance on enrollment management firms grows despite alarm over companies' strategies. In S. J. Burd (Ed.), *Lifting the veil on enrollment management: How a powerful industry is limiting social mobility in american higher education* (pp. 37–53). Cambridge, Massachusetts: Harvard Education Press.

McArthur, D., & Reeves, A. (2022). The unintended consequences of quantifying quality: Does ranking school performance shape the geographical concentration of advantage?¹. *American Journal of Sociology*, *128*(2), 515–551. https://doi.org/10.1086/722470

McKelvey, F. (2022). When the new magic was new: The claritas corporation and the clustering of america. *Ieee Annals of the History of Computing*, *44*(4), 44–56. https://doi.org/10.1109/mahc.2022.3214223

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Mennicken, A., & Espeland, W. N. (2019). What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, *45*, 223–245. https://doi.org/10.1146/annurev-soc-073117-041343

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

Noel-Levitz. (1998). *National enrollment management survey: Findings for fall 1997 four-year institutions*. Noel-Levitz.

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

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (First edition.). New York: Crown.

Oracle. (n.d.). *Loading and assigning EPS market codes*. Oracle. Retrieved from https://docs.oracle.com/cd/E29376_01/hrcs90r5/eng/psbooks/lsad/chapter.htm?File=lsad/htm/lsad40.htm

Poon, M. (2007). Scorecards as devices for consumer credit: The case of Fair, Isaac & Company Incorporated. *Sociological Review*, *55*, 284–306. https://doi.org/10.1111/j.1467-954X.2007.00740.x

Poon, O. Y. A., & Bastedo, M. N. (2022). *Rethinking college admissions: Research-based practice and policy*. Harvard Education Press. Retrieved from https://books.google.com/books?id=IjehEAAAQBAJ

Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Cambridge, MA: Harvard University Press.

Ruffalo Noel Levitz. (2023). *About RNL*. Retrieved from https://www.ruffalonl.com/about-ruffalo-noel-levitz/

Ruffalo Noel-Levitz. (2022a). *2022 cost of recruiting an undergraduate student report*. Ruffalo Noel-Levitz. Retrieved from https://www.ruffalonl.com/thank-you/download-the-2022-cost-of-recruiting-an-undergraduate-student-report/

Ruffalo Noel-Levitz. (2022b). *2022 marketing and recruitment practices for undergraduate students report*. Ruffalo Noel-Levitz. Retrieved from https://www.ruffalonl.com/papers-research-higher-education-fundraising/marketing-and-recruitment-practices-for-undergraduate-students/

Salazar, K. G. (2022). Recruitment redlining by public research universities in the los angeles and dallas metropolitan areas. *The Journal of Higher Education*, *93*, 585–621. 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 by public universities: A first look*. The Institute for College Access and Success. Retrieved from https://ticas.org/wp-content/uploads/2022/09/Geodemographics-of-Student-List-Purchases_A-First-Look.pdf

Sauder, M. L. (2008). Interlopers and field change: The entry of US news into the field of legal education. *Administrative Science Quarterly*, *53*(2), 209–234. Journal Article. Retrieved from <Go to ISI>://000258783000001

Sauder, M. L., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review*, *74*(1), 63–82. Journal Article. Retrieved from <Go to ISI>://000263490200004

Scott, W. R., & Davis, G. F. (2007). The dyadic environment of the organization. In W. R. Scott & G. F. Davis (Eds.), *Organizations and organizing: Rational, natural, and open systems perspectives* (pp. 220–244). Upper Saddle River, New Jersey: Pearson, Prentice Hall.

Sewell, W. H. (1971). Inequality of opportunity for higher education. *American Sociological Review*, *36*(5), 793–793 &. Retrieved from <Go to ISI>://A1971K843800001

Sewell, W. H., & Shah, V. P. (1967). Socioeconomic status, intelligence, and attainment of higher education. *Sociology of Education*, *40*(1), 1–23. Retrieved from <Go to ISI>://A1967ZD96500001

Smith, J., Howell, J., & Hurwitz, M. (2022). The impact of college outreach on high schoolers' college choices: Results from over one thousand natural experiments. *Education Finance and Policy*, *17*(1), 105–128. https://doi.org/10.1162/edfp_a_00334

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

Stevens, M. L., Armstrong, E. A., & Arum, R. (2008). Sieve, incubator, temple, hub: Empirical and theoretical advances in the sociology of higher education. *Annual Review of Sociology*, *34*, 127–151. https://doi.org/10.1146/annurev.soc.34.040507.134737

Takamiya, T. (2005). *Mechanisms for marketing higher education information services: The case of the college board*. University of Pennsylvania.

Taylor, B. J., Rosinger, K., & Ford, K. S. (2024). The shape of the sieve: Which components of the admissions application matter most in particular institutional contexts? *Sociology of Education*. https://doi.org/10.1177/00380407241230007

Walker, K., & Herman, M. (2024). *Tidycensus: Load US census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames*. Retrieved from https://walker-data.com/tidycensus/

Weber, M. (1948). Bureaucracy. In H. H. Gerth & C. W. Mills (Eds.), *From max weber: Essays in sociology* (pp. 196–244). London,: Routledge & K. Paul.

Zemsky, R., & Oedel, P. (1983). *The structure of college choice*. New York: College Entrance Examination Board.

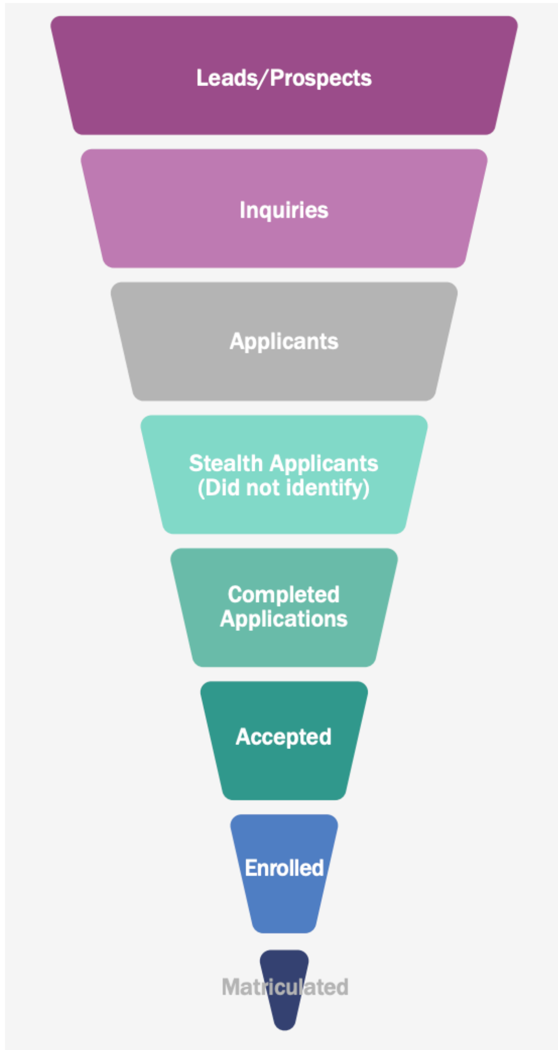

Figure 1: The Enrollment Funnel

A Appendix A

Table A1 reproduces a partial, simplified version of the bottom panel of Zemsky & Oedel (1983, fig. 2.1) which represents the Market Segment Profile for "Connecticut Market 3: Fairfield County." Table A1 has separate columns for each market segment (local, in-state, regional, national) and rows show the number of test-takers and characteristics of test-takers. For example, there were 550 "local" students in Fairfield county and these students submitted SAT scores to 2.8 postsecondary institutions on average. By contrast, there were 1,664 "regional" students who submitted test scores to 4.8 institutions on average. For local students, 11.2% had family income greater than $35,000 and 9% had both parents with a BA. For regional students, 41.9% had family income greater than $35,000 and 34.0% had both parents with a BA. Each Market Segment Profile also present information about the institutions that students from each segment sent scores to.

Table A2 recreates the top panel of Zemsky & Oedel (1983, fig. 2.1) for the Fairfield County, CT local market. For example, of the 550 local students, 277 (50.4%) sent scores to institution #1, which was a private master's granting institution. Of the 1,199 in-state students, 757 (63.1) sent scores to institution #1, a public doctoral granting institution, and 515 (43.0%) sent scores to institution #2, a public master's granting institution.

The *Institutional Profile* describes students who send scores to a particular institution and which majors these students are interested in. For each institution, institutional profiles are created separately for students from a particular local market, for all students in a state, or all students in a region. Table A3 reproduces a partial, simplified version of Zemsky & Oedel (1983, fig. 2.3), the institutional profile of an anonymous institution for students from Fairfield County, CT. Table A3 shows that 58 in-state students submitted SAT scores to the institution. These 58 students represented 4.8% of the total 1,199 in-state students from Fairfield County. 69 regional students sent scores to the institution, representing 4.1% of all 1,664 regional test-takers. Of these 69 regional students, 35 expressed interest in majoring in the liberal arts. These 35 students represent 7.0% of all regional students from Fairfield County who expressed interest in the liberal arts.

Zemsky & Oedel (1983) argue that the Institutional Profile and the Market Segment Profile enable admissions officers to know where to look for students and which institutions are competing for those students (p. 25):

> The Institutional Profile and the Market Segment Profile quantify the admission officers' intuitive grasp of market structure. Structure here carries a dual meaning, connoting both the structure of student choice and the structure of institutional competition...This two-sided interpretation furnishes the essential framework for planning by individual colleges and universities...To draw effectively on its own natural constituency, a college not only must contact the "right" kind of students — that is, students who are predisposed toward that type of institution – but also must persuade them of its special character. This means knowing the competition as well as the clientele.

Table A1: Simplified market segment profile, Connecticut Market 3: Fairfield County

Characteristic	Local	In-state	Regional	National
Total test takers	550.0	1199.0	1664.0	3766.0
Avg SAT (verbal + math)	770.0	850.0	970.0	980.0
Avg # scores sent per test taker	2.8	3.5	4.8	5.3
Percent in top 20% of HS class	27.8	26.1	44.7	45.7
Percent aspiring to more than BA	30.6	41.5	54.5	62.2
Percent family income more than $35,000	11.2	20.6	41.9	43.0
Percent both parents with BA	9.0	16.3	34.0	37.1

Table A2: Top 5 institutions in terms of number of scores sent by segment, Connecticut Market 3: Fairfield County

	Local (N=550)			In-state (N=1,199)			Regional (N=1,664)			National (N=3,766)		
	Num	Pct	Type	Num	Pct	Type	Num	Pct	Type	Num	Pct	Type
1	277	50.4	priv ma	757	63.1	pub doct	610	36.7	pub doct	1226	32.6	pub doc
2	261	47.5	priv ma	515	43.0	pub ma	348	20.9	priv doct	371	9.9	priv doct
3	183	33.3	priv ma	438	36.5	pub ma	272	16.3	priv doct	327	8.7	priv res
4	103	18.7	pub doct	183	15.3	pub ma	248	14.9	pub doct	312	8.3	priv doct
5	100	18.2	pub ma	177	14.8	pub ma	197	11.8	pub doct	308	8.2	priv doct

Table A3: Simplified sample institutional profile for anonymous institution, students from Connecticut Market 3: Fairfield County

	Local	In-state	Regional	National	Total
Total number of scores received	1.0	58.0	69.0	109.0	237.0
Pct of all test-takers in segment	0.2	4.8	4.1	2.9	3.3
lib_arts_num	0.0	25.0	35.0	61.0	121.0
lib_arts_share	0.0	8.4	7.0	5.3	5.9
engineering_num	0.0	2.0	5.0	3.0	10.0
engineering_share	0.0	3.0	5.8	0.8	1.8