Research article

Associations between majors of graduating seniors and average SATs of incoming students within higher education in the U.S.

Laura H. Gunn a,*, Enrique ter Horst c, Talar Markossian d, German Molina e

a Department of Public Health Sciences, 9201 University City Blvd, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
b School of Public Health, Faculty of Medicine, Imperial College London, UK
c Faculty of Business Administration, Universidad de los Andes, Colombia
d Parkinson School of Health Sciences & Public Health, Loyola University Chicago, USA
e Idalion Capital Group, UK

A R T I C L E   I N F O

Keywords:
Education
SAT percentiles of incoming student cohorts
Majors of graduating student cohorts
Beta regression

A B S T R A C T

While Scholastic Assessment Test (SAT) scores are becoming less widespread than in previous years, they continue to be used as an input by many higher education institutions in the United States to select which students to accept among applicants. This paper explores the association between average SAT scores of incoming undergraduate cohorts and major completions of graduating student cohorts. College Scorecard data from 2019 is collected from all U.S. undergraduate degree-granting, higher education institutions reporting average SAT scores of incoming cohorts (n = 1,389). A multivariate beta regression approach, which allows for overdispersion and unit-interval responses, is proposed to explore associations between graduation rates by major (explanatory variables) and SAT percentiles of new student cohorts (response). Forty-nine percent of the variability in average SAT percentiles of incoming cohorts can be explained by the graduation proportions by major within institutions. Results show strong concurrent positive associations between average SAT percentiles of incoming cohorts and proportions of students graduating in: STEM fields; ethnic, cultural, and gender studies; social science; or languages, among others (p < 0.01). A negative association is found between average SAT percentiles of incoming cohorts and graduating cohorts in degrees like security law enforcement or parks & recreation and fitness, as well as some traditional major choices, such as theology and psychology (p < 0.01). Results are consistent by institution size, as well as public versus private, across most clusters. A statistical framework is introduced for analysis of the expected impact on average SAT percentiles of future student cohorts derived from changes in proportions by major of graduating student cohorts. Higher education institutions can benefit from the proposed methodology by adjusting their degree offerings to their target cohorts. While illustrated using SAT scores due to their historical prevalence and availability across institutions, the proposed approach can utilize any alternative quantitative measure of student preferred characteristics.

1. Introduction

Education can be categorized as a service (Ng and Forbes, 2009), with students often seen as customers (Bunce, 2017). Administrators and researchers increasingly embrace the concepts of branding and marketing applied to the delivery of this service (Litten, 1980; Dolinsky, 1994; Ng and Forbes, 2009), and, with all the controversy surrounding the topic, self-assess their product with tools sometimes questionable such as student course evaluations. Such tools ultimately measure customer satisfaction and engagement (and willingness by existing customers to continue demanding the service, in the form of higher retention rates (Mangum, 2017)) more than service quality in the form of learning outcomes (Clayson, 2009), especially as such measures may be invalid and/or biased (Marsh and Roche, 1997; Greenwald, 1997; Hornstein, 2017).

On the supply side, major offerings are sometimes geared to match the workforce demand (Rhoades, 1987; Tribe, 2003), though they can also simply reflect internal institutional politics, available funding, and balance of power among departments (OECD, 2003). However, major offerings remain a tool for institutions to define and differentiate themselves and the characteristics of students they target. With greater nationwide and international opportunities, higher education institutions...
must compete for similar students, since higher education, like most services and industries, is subject to the globalization laws of supply and demand (Spring, 2008; Yavvedkar and Tiwari, 2014). Understanding the nature of student demand is, therefore, a key component for a successful product/service design and a coherent marketing strategy (Kim et al., 2002; Ng and Forbes, 2009; Brown, 2010). A necessary component to understanding the drivers of student demand involves identifying the factors that influence the student’s choice of major, a topic highly researched in the literature (Malgwi et al., 2005; Porter and Umbach, 2006; Keshishian et al., 2010), though the relationship between that choice of major and the choice of higher education institution has not drawn as much attention.

Students’ decisions of which institution to attend are driven by multi-dimensional factors (Bond et al., 2018; Azzzone and Soncin, 2019), and the weight of each dimension is student-specific and clustered by type of institution (Broekemier, 2002; Drewes and Michael, 2006). Universities can define how to market their institution using multiple tools, from pre-enrollment experiences (Secore, 2018) to digital marketing materials (Taylor and Bicak, 2020) or pamphlets (Mentz and Whiteside, 2003) that describe the institutional offerings and competitive advantages.

Higher education institutions may actively seek to enhance specific clusters of students for their future admitted cohorts. Their marketing efforts can become more successful when they align with the demand and characteristics of the student target populations (Molesworth et al., 2010). Administrators may seek, through changes to institutional offerings and services, among others, to enhance gender or racial diversity (Maple and Stage, 1991; Long, 2004); income diversity (Cebula and Lopes, 1982; Dynarski, 2006; Andrews et al., 2016); traditional measures such as ACT (American College Test) and SAT (Scholastic Assessment Test) scores as they relate to perceptions about institutional quality (Grissmer, 2000); or enrollment numbers targeting larger cohorts represented in a different segment of the consumer market (Smeby, 2003).

Market segmentation and target marketing, versus mass marketing, also becomes relevant depending on the aforementioned institutional preferences (Lewison and Hawes, 2007). When market segmentation is preferred, it is more relevant to fully understand not only how to reach the subpopulations of interest, but also what the drivers are of their decisions to attend one institution or another (Basha et al., 2019). This is even more relevant as institutions expand their scope and marketing efforts to historically underrepresented subpopulations (e.g. first generation students, diverse or rural populations, international students, adult learners, etc.), as well as the methods of delivery of educational services (e.g. online learning, hybrid courses, etc.). However, these marketing efforts will be less effective if the factors driving student decisions regarding institutional selections are not fully understood (Hemsley-Brown, 2017). This manuscript focuses on identifying and assessing the relevance of one such factor, namely the institutional major offerings, which have not been widely researched in this context.

Demand for educational services is clustered and based on limited, uneven, and imperfect information (Skinner, 2019; Holland, 2020), especially for first generation students from families with no direct experience in higher education (Simoes and Soares, 2010; Walsh et al., 2015; Hemsley-Brown, 2017). Institutional image and reputation, sometimes driven by the perceived quality of specific groups of majors within the institution, and prospective student perceptions and assessments will be two of the factors defining future demand for admission (Landrum et al., 1998; Molesworth et al., 2010; Walsh et al., 2015). Peer effects on decision-making have also been found prevalent in student environments (Sacerdote, 2001; Zimmerman, 2003), and specifically among women (Raabe et al., 2019). Therefore, modifying major offerings can have unexplored multiplier effects on major selections made by both current students and future cohorts. In this paper, effects are explored only on the latter.

Major offerings and availability of desired programs have been found to be a factor in the decision-making process of consumers of higher education (Maringe, 2006; Broekemier, 2002). SAT scores have been shown to relate to college major preferences among intellectually-gifted students (Achter et al., 1999), with further clustering by gender and type of major also explored, as described in Davison et al. (2014).

However, a direct link between major offerings and students’ selections of institution, accounting for SAT scores, has yet to be established in the literature. The primary question in this manuscript is whether there is a direct association between the SAT scores of higher education consumers and the relative major offerings within the corresponding institutions selected by those students (higher education suppliers).

A direct way to observe the relative relevance of majors within each higher education institution is through measurement of the graduation rates by major completed within the institution. This manuscript explores the relationship between major selections made by outgoing student cohorts (graduating undergraduate classes) and the SAT profile of incoming student cohorts. The decisions by academically gifted students regarding majors of choice have been explored at a micro level (Achter et al., 1999), and once the choice of institution has already been made (Malgwi et al., 2005). However, the literature lacks a more macro approach that explores whether that relationship also occurs at the institution level, and prior to the selection of university, upon accounting for all majors offered within each institution. If an association between SAT scores and choices of major exists, maybe students with differing SAT scores also choose institutions on the basis of the majors offered. The choice of major upon entering the institution could be seen as a byproduct of an earlier decision to apply to institutions with a focus on specific majors.

The underlying hypothesis explored in this paper is that new consumers of higher education (in the broader sense, all those influencing the decision-making/application process, including students, families, peers, and mentors) will assess whether the institution will fulfill their requirements and demands, also using information available about decisions made by previous customers (i.e. assessing the nature of the available supply). One such piece of information is the university’s offerings and their relative importance, as described in marketing tools or student visitations to the institution.

Using data from a large number of U.S. higher education institutions reporting both average SAT scores of incoming student cohorts and graduation rates by major of graduating student cohorts, this paper proposes a multivariate beta regression model approach (Ferrari and Cribari-Neto, 2004) to explore potential associations between the observable and measurable decisions undertaken by those two cohorts of graduating and prospective students. This approach allows for the extraction of nonlinear associations across institutions between the major choices of graduating cohorts and the average SAT percentile of the incoming cohorts.

If the aforementioned associations exist, institutions can gear the characteristics of the service supplied, represented by graduating cohorts and affected by institutional preferences and support (i.e. which majors are offered or most institutionally-supported), to attract more desired clusters of demand, represented by entering cohorts of students (Azzzone and Soncin, 2019). A statistical framework is proposed for assessment of the expected changes in average SAT scores of future cohorts as a function of potential institution-led changes (e.g. new major offerings, promotions of certain majors, etc.), that affect major decisions by graduating student cohorts. This could allow institutions to target segments of the incoming student population, as defined in Lewison and Hawes (2007), which may be more aligned with institutional strategic visions and preferences.

This paper is structured as follows: A description of the data and modeling approach is provided in Section 2. Results are described in Section 3. Section 4 concludes with a discussion of the approach and its implications, while addressing strengths and limitations, as well as areas of future research.
2. Methods

2.1. Data

The most recently available Scorecard data, released on December 12, 2019, was collected from the U.S. Department of Education for all U.S. institutions of higher education granting undergraduate degrees. The original dataset includes 7,112 U.S. higher education institutions. This population was restricted to institutions reporting average SATs for their incoming student cohorts, as well as proportions of students graduating by major. Majors were then categorized and grouped according to the 2019 version of the Scorecard dictionary (U.S. Department of Education, 2019), which includes 38 different groups. This reduced the population of interest to n = 1,389 reporting institutions.

Reported average SAT scores were mapped into SAT user percentiles through the CollegeBoard mapping table (CollegeBoard, 2019). SAT percentiles represent the response variable.

In order to reduce potential bias, marginally offered majors were removed. Thus, majors offered in fewer than 5% of the institutions were omitted from the study. While this choice of threshold was arbitrary from a quantitative perspective, the categories excluded were unlikely to be widely chosen by the higher education community as choices to incorporate in their curricula. As an example, among the set of major categories excluded, the one with the highest level of penetration among higher education institutions was personal culinary, with only 3.1% of institutions offering it. The implementation of this threshold reduced the covariate set to k = 31 major categorizations with the most widespread acceptance within the curricula of higher education institutions. Throughout the paper, each of the 31 categorized disciplines are referred to as majors, though they should be understood as groups of majors, since each category contains heterogeneous majors grouped by similarity of discipline. For example, the social science major category can include majors as diverse as economics and anthropology. Proportions of graduating student cohorts within each major for each institution represent the explanatory variables.

Factors such as retention rates, including those who drop out or transfer, can affect results. However, we do not focus on such students, who can be seen as higher education consumers who decide to return the product/service before completion of use. This manuscript focuses on information available regarding sufficiently happy consumers, defined for the purpose of this study as those who chose to complete use of the service provided (graduating students).

2.2. Statistical model

A beta multivariate regression model (Ferrari and Cribari-Neto, 2004) was used to explore the relationship between the doubly-bounded response (percentile of the average SAT score for the incoming cohorts for each institution) and the covariates of interest (percentage of graduates per major for each institution). This model was implemented in R statistical software using the betareg routine within the homonymous package. A comprehensive primer on the betareg package can be found in Cribari-Neto and Zeileis (2010).

This model was chosen based on the following arguments: (1) It offers a natural and flexible link function between a unit-interval variable (observed SAT percentiles range from 0 to 1) and the variables of interest, as described in Verkuilen and Smithson (2012) within the context of doubly-bounded responses in educational settings; (2) It allows for overdispersion of the observations through an additional parameter for model flexibility, which is shown in the Results section to be statistically significant; (3) It maintains the parametric intuition behind standard generalized linear models to be able to measure the net associations between each of the covariates (after accounting for all other covariates) and the response, and propose institutional changes based on those; (4) It provides a framework for inference both for theoretical new scenarios and for institutional changes; and (5) It is less sensitive to yearly variations or inflation in SAT scores, allowing for inter-year comparisons, as it uses SAT percentiles.

Define $PSAT_i$ as the observed percentile corresponding to the average SAT score for the incoming student cohort for institution $i$, where $i = 1, \ldots, n = 1,389$. Let $P_i$ be an observed k-dimensional vector of covariates corresponding to institution $i$, representing the proportion of students graduating in each of the $k = 31$ majors considered in this study. Following the parametrization in Ferrari and Cribari-Neto (2004),

$$PSAT_i \sim \text{Beta}(\mu_i, \phi);$$

(1)

$$h(\mu_i) = P_i \beta,$$

(2)

where the choice of $h(\cdot) : (0,1) \mapsto \mathbb{R}$ is the standard logistic function. Probit and complementary log-log links were explored with no significant differences found. Under this parametrization, the first two moments (i.e. mean and variance, respectively) are defined as:

$$E[PSAT_i] = \mu_i;$$

$$\text{var}[PSAT_i] = \frac{\mu_i(1 - \mu_i)}{(1 + \phi)},$$

(3)

where $\phi$ is the dispersion parameter. The quantity of interest is the vector $\beta$, which represents the (nonlinear) impact of higher percentages of graduates in each major on the institution's expected SAT percentile for the incoming student cohort. Positive values of $\beta$ for a given discipline imply a positive association between percentages of graduates within that discipline and average SAT percentiles for the institution.

The interpretation of subcomponents of $\beta$ must be such that $P_i$ remains in the k-dimensional closed set $[0,1]^k$, and the sum of $P_i$ remains smaller than or equal to 100% (it can be smaller due to the aforementioned removal of the marginal majors from the analysis).

Residuals of the regression can be interpreted as all other factors not explained by the proportion of graduating students per major within each institution (i.e. the covariates). This includes factors such as the reputation of the educational institution (defined over larger periods of time and with stronger autoregressive characteristics, rather than by the year-to-year operational or policy decisions), location, cost and availability of scholarships, socioeconomic and demographic profiles of the student population, as well as other idiosyncratic factors.

Throughout the paper relationships among key variables are referred to using the conservative term of association, although there is a point for defining it as predictive causality (e.g. Granger causality in the sense that one set of variables is defined earlier in time and, through any existing association, can help predict other set of variables defined later in time. The reverse causal effect is unfeasible.). Graduating cohorts will not be expected to define their major of choice influenced by the SAT characteristics of the incoming cohorts. There is a temporal asynchronicity between the two choices of interest: graduating cohorts define their majors of choice, which are the explanatory variables, often times, if not always, long before incoming student cohort average SATs are known or even before application decisions are made by incoming student cohorts, namely the response. Any association between these variables becomes a tool at the disposal of administrators to gear the nature of the response, that is, the expected SAT scores of incoming cohorts.

2.3. Predictive analysis

Equation (3) allows for building a simple tool for scenario analysis. New values for the explanatory variables, representing future or target proportions of graduating seniors with certain majors, can be used to predict the associated expected changes in average SAT percentiles of incoming students due to the changes in proportions of majors targetted. If the vector $P_i$ represents the initial/current proportions of majors corresponding to graduating seniors for institution $i$, and $P_i + \delta_i$ represents the target/future proportions of graduating seniors' majors for
Table 1

Multivariate beta regression parameter estimates ($\hat{\beta}$), sorted by estimate level (descending), with corresponding standard errors, t-statistics, and p-values. The first column, $n$, represents the number of non-zero observations for each major (i.e., the number of institutions in which the field was offered as a major and at least one student graduating majoring in that field during the 2018-2019 academic year), out of the $n=1,389$ higher education institutions reporting SAT scores. The last two columns highlight, when the variable is significant at the 1% and 5% significance levels, the directionality of the association (positive/negative) of $\hat{\beta}$. R-squared = 0.495, and overdispersion coefficient ($\phi = 9.26$; 95% CI (8.59,9.92); p < 0.0001).

| Major category                      | $n$  | $\hat{\beta}$ | SE  | t    | p    | p < 0.01 | p < 0.05 |
|-------------------------------------|------|---------------|-----|------|------|----------|----------|
| ethnic cultural gender              | 514  | 12.29         | 2.57| 4.14 | 0.00 | POS      | POS      |
| language                            | 975  | 10.48         | 1.76| 5.94 | 0.00 | POS      | POS      |
| mathematics                         | 1128 | 2.74          | 1.22| 3.00 | 0.00 | POS      | POS      |
| physical science                    | 1128 | 2.74          | 1.22| 3.00 | 0.00 | POS      | POS      |
| engineering                         | 1128 | 2.74          | 1.22| 3.00 | 0.00 | POS      | POS      |
| computer                            | 1128 | 2.74          | 1.22| 3.00 | 0.00 | POS      | POS      |
| social science                      | 1068 | 3.18          | 0.38| 6.22 | 0.00 | POS      | POS      |
| English                             | 1068 | 2.09          | 0.97| 2.31 | 0.05 |          |          |
| architecture                        | 1068 | 1.75          | 1.11| 1.68 | 0.11 |          |          |
| resources                           | 1068 | 1.58          | 0.89| 1.80 | 0.08 |          |          |
| legal                               | 1068 | 1.54          | 2.31| 0.93 | 0.35 |          |          |
| philosophy religious                | 774  | 1.16          | 0.38| 3.08 | 0.00 | POS      | POS      |
| communication                       | 993  | 1.08          | 0.43| 2.49 | 0.01 |          |          |
| business marketing                  | 1230 | 0.50          | 0.13| 3.83 | 0.00 | POS      | POS      |
| agriculture                         | 118  | 0.45          | 0.11| 3.66 | 0.00 | POS      | POS      |
| communications technology           | 884  | 0.41          | 0.11| 3.66 | 0.00 | POS      | POS      |
| visual performing                   | 1113 | 0.41          | 0.16| 2.47 | 0.01 |          |          |
| health                              | 804  | 0.20          | 0.09| 2.36 | 0.02 |          |          |
| education                           | 1117 | 0.03          | 0.27| 0.11 | 0.92 |          |          |
| biological                          | 1155 | 0.02          | 0.36| 0.06 | 0.95 |          |          |
| theology religious vocation         | 258  | -0.11         | 0.14| -0.75| 0.45 |          |          |
| transportation                      | 75   | -0.14         | 0.63| -0.22| 0.83 |          |          |
| multidiscipline                     | 811  | -0.73         | 0.32| -2.28| 0.02 | NEG      |          |
| family consumer science             | 314  | -1.01         | 0.75| -1.34| 0.18 |          |          |
| engineering technology              | 287  | -1.07         | 0.69| -1.55| 0.12 |          |          |
| public administration social service| 659  | -1.81         | 0.42| -4.31| 0.00 | NEG      | NEG      |
| parks recreation fitness            | 743  | -2.18         | 0.45| -4.86| 0.00 | NEG      | NEG      |
| psychology                          | 1186 | -2.26         | 0.41| -5.54| 0.00 | NEG      | NEG      |
| history                             | 1050 | -2.41         | 1.69| -1.42| 0.15 |          |          |
| security law enforcement            | 658  | -2.92         | 0.35| -8.39| 0.00 | NEG      | NEG      |

that institution, then the expected change in average SAT percentile of incoming cohorts for institution $i$ is

$$E[\Delta P_{SAT}] = h^{-1}(\sum_{j} (P_j + \delta_j)\hat{\beta}_j) - h^{-1}(\sum_{j} P_j\hat{\beta}_j),$$

(4)

where $h^{-1}$ refers to the inverse function and $\delta_j$ is a vector representing the targeted changes (both positive and negative) in proportion of graduating seniors by major for institution $i$. A natural restriction is that all elements of this vector sum to zero ($\delta_i^T = 0$), so that the change reflects a redistribution of proportions among disciplines. The first element of the right-hand-side of Equation (4) represents the expected SAT percentile under the targeted proportions of graduating seniors by major, while the second element represents the expected SAT percentile of incoming students under the current proportions of graduating seniors by major. The difference represents the expected SAT percentile change for incoming students due to the proposed variation in proportions of graduating seniors by major. SAT-maximizing institutions would be interested in exploring feasible values of $\delta_j$ that increase their expected SAT percentile for incoming cohorts.

Equation (4) can serve to explore the potential impact of an array of hypothetical/target changes both at the macro level (expected impact across different institutions) and at the micro level (expected impact for any given institution). Both cases are explored through stylized examples, which cover the expected impact of changes in relative importance of disciplines on the expected SAT scores of incoming cohorts.

3. Results

3.1. Descriptive analysis

Fig. 1 provides an initial exploratory representation of the data. In order to focus on majors that can be seen as most relevant within each institution, and for the purpose of only the heatmap (top panel of Fig. 1), higher education institutions are included where the given major constitutes at least 5% of the graduating class within that institution. Note that for the statistical model, such threshold is not required.

For each major (column), each pixel represents the observed percentage of institutions by SAT decile (with each column adding to 100%, and rows representing the upper bound of each SAT decile). The lighter the color of the pixel, the higher the observed proportion. Columns are sorted by percentage of observations in the top decile, from least (left) to greatest (right). For example, more than 60% of institutions where the graduating cohort in mathematics constitutes at least 5% of the institution’s graduating class have incoming students who, on average, are in the top SAT decile. This number is 66% for ethnic, cultural, and gender studies (ECG). Conversely, none of the institutions where the graduating cohort in communications technology, legal, or transportation constitutes at least 5% of the graduating class have incoming students, on average, in the top SAT decile. Most institutions will be presented in multiple pixels (across different columns of the heatmap), to represent the multidisciplinary nature of their major offerings.

The heatmap shows that when mathematics, ECG, language, or physical science are prevalent majors of graduating seniors within the
institution, average SAT scores of their incoming student cohorts are most likely to be in the top decile compared to other disciplines.

The barplot in the lower panel of Fig. 1 shows the univariate correlations between percentages of the graduating cohort per major and the institution’s average SAT percentiles for the incoming student cohort. Some majors appear to show relatively strong positive associations, while negative associations appear not to be as strong in magnitude.

3.2. Inferential analysis

Table 1 summarizes results from the full model. This table provides the parameter estimates (\(\hat{\beta}\)) associated with each of the disciplines (rows) and sorted descendingly, as well as the associated standard errors, t-statistics, and p-values. The last two columns highlight the significance of each of those parameter estimates, both at the 1% and
5% levels, where 14 and 17 covariates, respectively, were statistically significant among k = 31 total covariates. The first column provides information about the sparsity of the data, indicating the count of non-zero values for each of the disciplines (i.e., the number of institutions in which the field was offered as a major and at least one student graduated majoring in that field during the 2018-2019 academic year). For example, 413 universities, out of the 1,389 within the population, have a degree in ethnic, cultural, or gender studies with at least one graduate within the latest cohort (i.e., the 2018-19 academic year).

Nearly 50% (pseudo R-squared = 0.49) of the variability in the SAT percentile for the incoming cohorts can be explained by the major choices from the graduating senior cohorts (ultimately affected by the major offerings and resource allocations within the institution). This indicates that higher education institutions may have a tool at their disposal through the promotion of majors which are related to a higher likelihood of being the institution of choice by students with higher SAT scores. The dispersion parameter is clearly significant ($\phi = 0.26; p<2e-16$), indicating that there is strong evidence of overdispersion in the data that needs to be captured.

Higher graduation rates in several majors are associated with higher SAT percentiles for incoming cohorts. Larger cohorts of STEM majors (engineering, physical sciences, mathematics, and computing), together with social science, language, and ECG majors, among others, show positive significant associations with percentiles of the average SAT scores of incoming cohorts at the 1% significance level. Conversely, psychology and public administration, among others, show negative significant associations with percentiles of the average SAT scores of incoming cohorts at the 1% significance level. More than half of majors are statistically significant at the 5% level (17 of 31 majors).

When the results are clustered by Carnegie classification and grouped by size, consistent patterns exist across multiple majors. The strength of the relationships appears to be aligned with institution size (pseudo R-squared values of 0.49, 0.58, and 0.73 for small, medium, and large institutions, respectively). This indicates that larger institutions appear to be more sensitive, in terms of association with percentiles of the average SAT scores of incoming cohorts, to the impact of major offerings and choices by their graduating cohorts. This can be seen in Fig. 2, which portrays the model fit by observed SAT decile for the overall model and clustered by institution size.

The model fit appears stronger for institutions with the top SAT deciles across all institution sizes. This can be seen in the larger slope for higher deciles in Fig. 2. Each boxplot represents the model fit by plotting the observed SAT decile (x-axis) against the fitted SAT percentile (y-axis). The full model, corresponding to all institution sizes, is shown in the top-left panel. A decomposition by institution size is presented in the remaining panels. Higher-sloping boxplots by decile correspond to better model fits, where a diagonal would represent a perfect fit.

Similarly, when comparing privately-held and publicly-held institutions, the associations remain strong, with pseudo R-squared of 0.58 and 0.40, respectively. The top two panels in Fig. 3 show the aforementioned decile-based representation by institution type, with higher slopes closer to a diagonal again representing better fits. SAT percentiles of entering students at private institutions appear to be more sensitive to major choices by graduating cohorts than those at public institutions. Some students attending public institutions may not have the option financially to attend private institutions which may offer a better fit with respect to major offerings. A higher proportion of incoming students within public higher education institutions would be expected to have financial barriers to freely choose educational institutions, leading to a dilution of these observed associations for lower-income students (and, consequently, for institutions with a higher proportion of such students). Students for whom higher education costs are not a significant
Fig. 3. [Top panel]: Model fit visualization representing observed SAT decile (x-axis) versus fitted SAT percentiles (y-axis) for private institutions (left; beta regression pseudo R-squared = 0.58) and public institutions (right; beta regression pseudo R-squared = 0.40). There were no private institutions reporting average SAT scores in the bottom decile. [Bottom panel]: Density plots of the impact of a 5% redistribution into majors with positive (left) or negative (right) model coefficients in Table 1.

concern will be more likely to place greater weight on other factors, including major offerings across institutions, than students for whom financial costs are among the key relevant factors for electing a higher education institution.

Although this analysis uses all the information available across majors, this approach can also be applied when only subsets of the design matrix are available. For example, if only information about STEM majors is used (combining graduating cohort proportions from all STEM major Scorecard categories), this STEM-grouped explanatory variable can be regressed against the SAT percentiles of entering students. This leads to a highly significant coefficient ($\hat{\beta} = 4.68; p<2e-16$). In this case, even a simpler univariate beta regression (with percentages of combined STEM majors completed by graduating seniors as the explanatory variable) has a reasonable fitting power, with pseudo R-squared = 0.24. Therefore, 24% of the variability in the average SAT percentiles of new student cohorts can be explained by a single variable, namely the proportion of STEM students in their graduating student cohorts.

The R-squared of 0.495 presented in the caption of Table 1 confirms that there is a strong association between disciplines offered within institutions and the average SAT scores of incoming cohorts to those institutions. The demand of higher education appears to be strongly clustered by SAT scores around different disciplines, with offerings of some disciplines serving as magnets for those with higher SAT scores.

3.3. Macro-level inference

Fig. 3 shows the expected SAT percentile increase of incoming students (y-axis) as a function of the starting SAT percentile (x-axis). The bottom panel portrays the distribution of expected changes in SAT percentile of incoming students ($E(\Delta P.S.A.T.)$) across all institutions, for a hypothetical scenario comprising a 5% redistribution in major selection (absolute change in the proportions of graduating seniors across majors upon some intervention). These plots correspond to two stylized examples where the institutions promote majors according to different options.

The bottom left plot in Fig. 3 represents an equally-redistributed 5% increase of graduating seniors amongst the majors with statistically significant and positive estimated coefficients (at the 1% level) in Table 1. This redistributions, however, is at the expense of the remaining disciplines, which share the implied 5% decrease proportionally to the size of their graduating class. This type of change is less traumatic, as it allows an equal proportional burden by major, among both those that see increases and those that see decreases.

Conversely, the bottom right plot in Fig. 3 also represents an equally-redistributed 5% increase of graduating seniors, but this time amongst the majors with statistically significant and negative estimated coefficients (at the 1% level) in Table 1. Again, this is at the expense of the remaining disciplines, which similarly share the implied 5% decrease proportionally to the size of their graduating class.

Since the change in $P.S.A.T$ depends on the target changes in proportions of graduating seniors across majors ($\delta$), as well as the starting level of those graduating major proportions ($P_i$), each institution experiences a different expected impact. Some institutions will benefit (lose) more than others, not only from the direct impact of the positive (negative) increase in graduation rates in the target field, but also from the reduction (increase) of the negative impact from a lower graduation across the remaining fields. The density plots in Fig. 3 show the distribution of the impact on the expected SAT percentile of incoming students for the aforementioned stylized changes $\delta$.

The shape of the density plots aligns with the shape of the mapping between percentiles and SAT scores. Larger moves in SAT scores are needed in the extremes to obtain equivalent moves in SAT percentiles for those with middle-percentile SAT scores.
The biggest impact occurs for institutions whose incoming cohorts occupy the central SAT percentiles (institutions with incoming cohorts with median SAT scores). For these institutions, the redistribution scenarios explored have either a positive impact of around 8 SAT percentile units (redistribution into statistically positively related majors of graduating cohorts) or a negative impact of around 3-4 SAT percentile units (redistribution into statistically negatively related majors of graduating students). In SAT terms, for an institution with a starting average SAT score of 1000 for incoming students (40th SAT percentile), this means that small 5% major redistributions among graduating seniors in the two stylized examples produce expected SAT score ranges for incoming students between 985 and 1045, depending on the majors promoted.

Institutions which are situated more toward the extremes in terms of average SAT percentiles of incoming students will benefit (or lose) less from major redistributions among graduating seniors. The expected impact of changes will be around four times smaller, in percentile unit terms, for those already in the upper and lower SAT deciles for their incoming cohorts.

While this scenario analysis depicts the expected impact of redistributions of majors across all institutions, offering a glimpse of the distribution of those changes among the study population of institutions, some institutions benefit more than others. This leads to the need for a micro-level framework, to allow administrators of each particular institution to assess the implications for their institutions and assess the best path to achieve their cohort targets.

3.4. Micro-level inference

Equation (4) can also be used to explore the impact on the expected SAT percentiles of entering students, for a given institution, of any set of potential institutional decisions/policies leading to anticipated changes in major selections among future graduating seniors (such as the offering of new majors or relative changes of specific majors). Fig. 4 shows the impacts of some stylized scenarios on the expected SAT percentile changes within a set of three selected institutions, with each line within each plot representing the expected impact of changes along the x-axis for that particular institution.

A series of percent changes in a single major are assumed, at the expense of a proportional reduction in graduation rates in the remaining majors, for a choice of three higher education institutions with different starting SAT scores among their entering student cohort. The four plots in Fig. 4, each corresponding to a different major, portray the impact on expected SAT percentiles of incoming cohorts, for each of the three institutions (corresponding to those with low, medium, and high initial SAT scores) of increases from 0 to 25% in the proportion of graduating students within the majors. The top two plots in Fig. 4 refer to disciplines with positive associations with expected SAT percentiles, while the bottom two plots represent disciplines with negative associations with expected SAT percentiles.

The differing impact, which has a nonlinear nature, depends on the major where the change occurs. For example, the dotted lines represent a higher education institution with average SAT of incoming cohorts close to the median SAT. This institution can see as much as a full decile increase in expected SAT scores from incoming cohorts (approximately 50 SAT points increase for this institution) for a mere 5% increase in graduation rates within the ECG major. However, if this increase were to occur among psychology graduates, the institution should expect a three percentile decrease (approximately 5 SAT points decrease for this institution) for a similar move.

The expected benefits and losses (in terms of expected SAT percentiles of incoming students) from changes in student majors at graduation have diminishing returns (in SAT terms). Larger changes produce
a proportionally smaller impact on expected SAT percentile changes. This occurs both because of the nature of the SAT percentile data (bounded), as well as the nonlinear shape of the relationship.

Finally, higher education institutions with similar average SAT percentiles for incoming cohorts should expect different impacts of equivalent changes in majors (Δ). Those different impacts will depend on the institution’s initial proportions of graduates per major (which are reflected in the design vector \( P_i \)). Therefore, two institutions who receive similar cohorts in terms of their average SAT percentiles do not necessarily benefit equally from similar policies regarding changes in majors selected, since the impact of those depend on the starting proportions of the graduating cohorts. Strategy design should be institution-specific, and a micro-level approach is best to find the optimal target proportion of majors. Each institution will be able to assess not only whether they are above expectations given the relative weights for each discipline, but also have a scenario analysis tool to assess the potential impact of any planned changes. Institutions can also assess expected dynamics of that impact over time, if the scenario comprises changes over multiple periods.

4. Discussion

This manuscript explores concurrent associations between average SAT percentiles for incoming student cohorts and the profile of graduating cohorts by major for U.S. higher education institutions. Analysis results offer a snapshot of where supply and demand meet for each institution (graduating student cohorts) as well as the SAT characteristics of the current demand for each institution (incoming student cohorts).

A multivariate beta regression approach is proposed to account for the large heterogeneity observed among higher education institutions. This approach is intuitive and simple to implement, and the resulting framework allows for scenario analysis of the impact of changes in graduating cohorts on expected SAT scores of incoming cohorts, both at the macro level (across institutions) and the micro level (within institutions).

The results show strong associations (pseudo R-squared of 0.49) between the supply by major (represented by the majors of choice per institution for graduating cohorts) and the SAT profile of the incoming cohorts. Higher proportions of graduating cohorts in some majors, a factor which is, to some extent, under the control of the institutions, are strongly associated with higher average SAT profiles of the incoming cohorts. These associations remain strong among different clusters explored, both by size and ownership.

Some majors with common features show very strong associations, and the methodology allows for such clustering to explore joint associations. For example, nearly a quarter of the variability of the SAT percentiles of incoming cohorts can be explained by the single variable representing the relative size of STEM graduates.

This paper also offers a framework for analysis of theoretical changes in the composition of majors of the graduating cohorts. For example, institutions with median SAT scores for the incoming cohorts can expect increases of up to eight percentile units in incoming cohorts’ SAT scores (from the 50th percentile to the 58th percentile) by targeting a 5% redistribution across majors in their graduating cohorts. However, that same redistribution can induce a drop of four percentile units if geared toward majors that demonstrate negative associations with SAT scores of incoming cohorts.

Study implications include shifting faculty, administrators, and institutional advisory boards to adjust uneven expectations of SAT scores of incoming cohorts and existing major offerings. This adjustment can allow for more realistic expectations of the impact of strategic plans, hiring decisions, institutional collaborations, marketing efforts (within and outside the walls of the educational institution), and resource allocations.

A successful strategy of marketing segmentation and branding by higher education institutions should utilize all the information associated with the decision-making of consumers of higher education. This paper identifies and explores a factor that explains 49% of the variability in SAT scores of incoming student cohorts for the population of SAT reporting institutions in the United States (\( n = 1,389 \)). Unveiling this factor has the potential to provide great explanatory power to future marketing strategies, and serve as a tool for self-definition, branding, differentiation, and customer targeting by higher education institutions. This study shows that incoming students may not be agnostic, directly or indirectly, to the offerings (and relative relevance) of majors by higher education institutions. Therefore, transforming current supply can help institutions gear specific quantitative characteristics of future demand.

Higher education offerings change over time, and this manuscript provides guidance to direct that change objectively if the outcome of interest can be quantitatively measured (e.g. GPA, SAT, or other quantitative measures considered).

4.1. Strengths, limitations, and scope of future research

The proposed approach uses SAT scores since they have been, and continue to be, heavily used for student admissions. Although SAT scores are used for the purpose of demonstrating the formal approach and assessing significant associations, other measures used for student admissions (e.g., average student GPA) could also be used. The proposed methodology can be applied to other current or future widely-acceptable quantitative measures of relevance to higher education institutions. For example, the recently-piloted Environmental Context Dashboard, within the SAT scores, constitutes a novel attempt to quantify qualitative socio-economic factors, providing an additional incoming measure within the scope of the proposed methods.

One key strength of the proposed approach is that it is built on readily-available information. It can serve as a tool for comparison with peer and aspirational institutions. Decision-makers can utilize the model not only to assess the expected SAT scores of incoming cohorts under current major offerings, but also to explore scenario analyses of alternative changes prior to implementing them. This also offers administrators an objective tool to explore and support evidence-based changes, driven by quantitative measures of interest (e.g. SAT scores), that align with institutional visions.

No relationship between SAT scores and any type of universally accepted optimality of entering cohorts is claimed, since optimality is defined by each institution on a broader basis. Every institution will have a differing set of relevant factors defining its optimal student cohort. Additionally, SAT scores may not be an unbiased measure of students’ future academic outcomes. The use of SAT scores in admissions is quite prevalent, as reflected by the large population size of institutions which require them. However, there is mounting evidence of biases in this measure related to race, parental income, and education (Atkinson and Geiser, 2009). Regardless, SAT scores, even if they are less influential, remain relevant and constitute a widely accepted measure of influence in admissions decisions.

There can be confounding and latent factors driving the associations described in this manuscript. However, even if these confounders exist, they do not invalidate the conclusion that relative strength of majors is a differentiating factor between institutions associated with SAT characteristics of incoming student cohorts.

Regarding majors of choice by graduating cohorts, this work describes existing associations between those and the SAT profiles of incoming cohorts unexplored to date in the literature. Relative comparisons of majors are not made in terms of their societal or economic relevance (Chan, 2016), nor whether they are adequate for the students’ preferences, skills, or needs, which would require further qualitative research. The categorization into clusters of majors is defined within the Scorecard provided by the Department of Education. The ideal categorization would have been more granular, to match the granularity of disciplines within institutions. However, this type of study
would require data not readily available. While this paper identifies and describes the associations between major completion of graduating cohorts and institution selections by entering cohorts, the underlying reasoning behind those observable decisions, which is of qualitative nature, remains outside of the scope of this manuscript.

Changes to major offerings are likely to be slow and politically challenging. Changes that can be implemented while keeping departmental structures intact would be more feasible than those that require relinquishing of current/future resources by established departments. Resources are oftentimes distributed taking into account student sizes serviced by department, among other factors. Asking departments hosting specific majors to willingly contribute to a reduction in resources allocated to favor other departments hosting other majors is a complex task, especially if those resources are perceived as a zero-sum game within the institution. Additionally, tenured faculty represent a fixed set of skills that may not be easily transferred from one discipline to another. In practical terms, it may be necessary to create or transfer tenure lines to new (or between existing) disciplines as those tenure lines become open, which is a slower process (e.g. new hires due to growth, retirement of tenured faculty, etc.).

Offering joint degrees can be a practical form of implementation that may be considered a non-zero-sum game while enhancing the visibility and strength of intended disciplines. Other forms of implementation include: (1) marketing specific minors or majors; (2) modifying elective offerings; (3) recommending students to explore specific majors through institutional advising; and (4) modifying the set of majors offered. Creating interdisciplinary programs and hiring interdisciplinary faculty teaching cross-listed, hybrid courses across the more popular and the less popular majors may also be a useful tool to enhance the demand for these latter majors.

Finally, the political nature of implementing these measures requires decision-makers who are willing to endure the challenges of slow, multi-year implementations toward future benefits that may not materialize until their tenure in administrative roles within the institution may already have finalized.

One area of future research is the decomposition and analysis of the sections of SAT scores of entering cohorts. Associations between majors of choice and SAT scores can potentially be further decomposed according to the Evidence-Based Reading and Writing (EBRW) and Math sections within SAT scores. Differences by such cluster may also provide valuable information about differences in entering cohorts based on major availability. Those with higher math SAT scores may lean toward different sets of majors (and institutions) than those with higher EBRW SAT scores.

Another area of future research would be to explore the link between student perceptions and student decisions when it comes to selecting higher education institutions, potentially through more qualitative research. While the associations explored describe the actual decisions of students at the population level, they do not provide a micro-level analysis of the nature of that decision-making process by each individual. When it comes to student decisions, this manuscript offers the what at the macro level. Future research can explore why from a micro level, to further gear marketing efforts by higher education suppliers.

Declarations

Author contribution statement

L.H. Gunn, G. Molina: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. E. ter Horst, T. Markossian: Analyzed and interpreted the data.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Data associated with this study is publicly available from https://collegescorecard.ed.gov/data/

Appendix A

The data used in this study was released by the U.S. Department of Education College Scorecard (https://collegescorecard.ed.gov/data/) on December 12, 2019 and was accessed on January 26, 2020 for analysis.

References

Achter, J.A., Lubinski, D., Benbow, C.P., Eftekhari-Sanjani, H., 1999. Assessing vocational preferences among gifted adolescents adds incremental validity to abilities: a discriminant analysis of educational outcomes over a 10-year interval. J. Educ. Psychol. 91 (4), 777–786.

Andrews, R.J., Imberman, S.A., Lovenheim, M.F., 2016. Recruiting and supporting low-income, high-achieving students at flagship universities. working paper No 22260.

Atkinson, R.C., Geiser, S., 2009. Reflections on a century of college admission tests. Educ. Res. 38 (19), 665–676.

Azzone, G., Soncin, M., 2019. Factors driving university choice: a principal component analysis on Italian institutions. Stud. High. Educ.

Basha, N.K., Sweeney, J.C., Soutar, G.N., 2019. Evaluating students' preferences for university brands through conjoint analysis and market simulation. Int. J. Educ. Manag. 34 (2), 263–278.

Bond, T.N., Bulman, G., Li, X., Smith, J., 2018. Updating human capital decisions: evidence from sat score shocks and college applications. J. Labor Econ. 36 (3), 807–839.

Broekemier, G.M., 2002. A comparison of two-year and four-year adult students: motivations to attend college and the importance of choice criteria. J. Mark. High. Educ. 12 (1).

Brown, R., 2010. Higher Education and the Market. Routledge, New York.

Bunce, L., 2017. The student-as-consumer approach in higher education and its effects on academic performance. Stud. High. Educ. 42 (11), 1958–1978.

Cebula, R.J., Lopes, J., 1982. Determinants of student choice of undergraduate major field. Am. Educ. Res. J. 19 (2), 303–312.

Chan, R., 2016. Understanding the purpose of higher education: an analysis of the economic and social benefits for completing a college degree. J. Educ. Polic. Plan. Admin. 6, 40.

Clayson, D.E., 2009. Student evaluations of teaching: are they related to what students learn?: a meta-analysis and review of the literature. J. Mark. Educ. 31 (1), 16–30.

Cribari-Neto, F., Zelnits, A., 2010. Beta regression in R. J. Stat. Softw. 34 (2).

Davison, M.L., Jew, G.B., Davenport, E.C., 2014. Patterns of sat scores, choice of stem major, and gender. Meas. Eval. Couns. Dev. 47 (2).

Dolinsky, A., 1994. A consumer complaint framework with resulting strategies: an application to higher education. J. Serv. Mark. 8 (3), 27–39.

Drewes, T., Michael, C., 2006. How do students choose a university?: an analysis of applications to universities in Ontario, Canada. Res. High. Educ. 47 (7).

Dynarski, S.M., 2000. Hope for whom? Financial aid for the middle-class and its impact on college attendance. Natl. Tax J. 53 (3), 629–662.

Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31 (7), 799–815.

Greenwald, A., 1997. Validity concerns and usefulness of student ratings of instruction. Am. Psychol. 52 (11), 1209–1217.

Grissmer, D.W., 2000. The continuing use and misuse of sat scores. Psychol. Public Policy Law 6 (1), 223–232.

Hemley-Brown, J., 2017. Encyclopaedia of International Higher Education Systems and Institutions. Springer, Netherlands.

Holland, M., 2020. Framing the search: how first-generation students evaluate colleges. J. High. Educ. 91 (3), 378–401.

Horstmann, H.A., 2017. Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. Cog. Educ. 4 (1), 1204016.

Keshishian, F., Brocavich, J.M., Boone, R.T., Pal, S., 2010. Motivating factors influencing college students' choice of academic major. Am. J. Pharm. Educ. 74 (3), 46.

Kim, D., Markham, F.S., Cangelosi, J.D., 2002. Why students pursue the business degree: a comparison of business majors across universities. J. Educ. Bus. 78 (1), 28–32.

Landrum, R., Turrisi, R., Harless, C., 1998. University image: the benefits of assessment and modeling. J. Mark. High. Educ. 9 (1), 53–68.

Lewison, D., Hawes, J., 2007. Student target marketing strategies for universities. J. Coll. Adm. 196.

Litten, L.H., 1980. Marketing higher education: benefits and risks for the American academic system. J. High. Educ. 51 (1), 40–59.
Long, M.C., 2004. Race and college admissions: an alternative to affirmative action. Rev. Econ. Stat. 86 (4), 1020–1033.
Malgvi, C.A., Howe, M.A., Burnaby, P.A., 2005. Influences on students’ choice of college major. J. Educ. Bus. 80 (5).
Mangum, E., 2017. Teaching and student success: ACUE makes the link. Change Mag. High. Learn. 49 (5), 17–25.
Maple, S.A., Stage, F.K., 1991. Influences on the choice of math/science major by gender and ethnicity. Am. Educ. Res. J. 28, 37–60.
Maringe, F., 2006. University and course choice: implications for positioning, recruitment and marketing. Int. J. Educ. Manag. 20 (6), 466–479.
Marsh, H., Roche, L., 1997. Making students evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. Am. Psychol. 52 (11), 1187–1197.
Menta, G., Whiteside, R., 2003. Internet college recruiting and marketing: web promotion, techniques and law. J. Coll. Adm. 181, 10–17.
Molesworth, M., Scullion, R., Nixon, E., 2010. The Marketisation of Higher Education and the Student as Consumer. Routledge, New York.
Ng, I.C.L., Forbes, J., 2009. Education as service: the understanding of university experience through the service logic. J. Mark. High. Educ. 19 (1), 38-64.
OECD, 2003. Education Policy Analysis. Organization for Economic Co-operation and Development (OECD), Paris.
Porter, S.R., Umbach, P.D., 2006. College major choice: an analysis of person-environment fit. Res. High. Educ. 47 (4).
Raebe, I., Boda, Z., Stadfeld, C., 2019. The social pipeline: how friend influence and peer exposure widen the stem gender gap. Sociol. Educ.
Rhoades, G., 1987. Higher education in a consumer society. J. High. Educ. 58 (1), 1–24.
Sacerdote, B., 2001. Peer effects with random assignment: results for dartmouth roommates. Q. J. Econ. 116 (2), 681–704.
SAT CollegeBoard, 2019. Sat understanding scores. https://collegereadiness.collegeboard.org/pdf/understanding-sat-scores.pdf. (Accessed 24 January 2020).
Secore, S., 2018. The significance of campus visitations to college choice and strategic enrollment management. Strat. Enroll. Manag. Q.
Simoes, C., Soares, A., 2010. Applying to higher education: information sources and choice factors. Stud. High. Educ. 35, 371–389.
Spring, J., 2008. Research on globalization and education. Rev. Educ. Res. 78 (2), 330–363.
Taylor, Z.W., Bicak, L., 2020. Buying search, buying students: how elite U.S. institutions employ paid search to practice academic capitalism online. J. Mark. High. Educ.
Tribe, K., 2003. Demand for higher education and the supply of graduates. Eur. Educ. Res. J. 2 (3), 463–471.
U.S. Department of Education, 2019. College scorecard data. https://collegescorecard.ed.gov/data/. (Accessed 26 January 2020).
Verkuilen, J., Smithson, M., 2012. Mixed and mixture regression models for continuous bounded responses using the beta distribution. J. Educ. Behav. Stat. 37 (1), 82–113.
Walsh, C., Moorhouse, J., Dunnett, A., Barry, C., 2015. University choice: which attributes matter when you are paying the full price. Int. J. Consum. Stud. 38, 670-681.
Yeravdekar, V.R., Tiwari, G., 2014. Internationalization of higher education and its impact on enhancing corporate competitiveness and comparative skill formation. Proc., Soc. Behav. Sci. 157, 203–209.
Zimmerman, D.J., 2003. Peer effects in academic outcomes: evidence from a natural experiment. Rev. Econ. Stat. 85 (1), 9–23.