Journal of Advanced Academics
jaa
Volume 20 ✣ Number 4 ✣ Summer 2009 ✣ pp. 662–700

Describing Students Involved in the Search Phase of the College Choice Process:

A Cluster Analysis Study

Emily J. Shaw
Jennifer L. Kobrin
Sheryl F. Packman
The College Board

Amy Elizabeth Schmidt
Educational Testing Service

Applying to college is a rigorous process based on a reciprocal relationship between students and institutions, as both actively search to meet their own needs and aspirations for the best education and student body, respectively. The popular media often depict college admission as characterized by two extreme groups of applicants: those who are academically gifted, apply to many competitive institutions, and spend exorbitant amounts of money on college preparation (Rubin, 2008) and those who are underprepared and unaware of the college admission process (Nyhan, 2006). Although these two groups do exist, they are far from the norm of college applicants. They may be better exemplified as at least a few groups of students who can be classified on a variety of characteristics. As such, there arises a need to identify and describe these unique clusters of college applicants in order to guide their academic preparation for college success as well as to rethink col-

Summary

The media often communicates the existence of two distinct types of college applicants: the frenzied, overachieving, anxious student who applies to many institutions and the underprepared, less advantaged student who is not at all familiar with the application process. Although these two groups likely do exist, they are far from the norm of college applicants who are better exemplified as at least a few groups of students who can be classified based on relevant characteristics. We identified five unique clusters of students: Privileged High Achievers/Athletes, Disadvantaged Students, Average Students Needing More Guidance, Mostly Female Academics, and Privileged Low Achievers. These clusters differed from each other based on variables including: academic performance, demographic characteristics, home and school characteristics, participation in school activities, and the number and types of higher education institutions to which they apply. An understanding of these descriptive clusters, comprised of students with similar backgrounds and goals for higher education, is a necessary first step in developing more thoughtful and inclusive enrollment management and college preparation practices.

Shaw, E. J., Kobrin, J. L., Packman, S. F., & Schmidt, A. E. (2009). Describing students involved in the search phase of the college choice process: A cluster analysis study. *Journal of Advanced Academics, 20,* 662–700.

lege recruitment and admission policies in order to maximize the benefits of higher education for both students and institutions.

Admission directors, institutional researchers, enrollment managers, and guidance counselors around the country are examining questions related to the changing nature of college admission in the 21st century. There have been demographic shifts by ethnicity in high school and college attendance (Western Interstate Commission for Higher Education, 2008), technological advances have improved the dissemination of information about colleges and universities (MacAllum, Glover, Queen, & Riggs, 2007), rankings are playing a larger role in driving postsecondary educational goals (Thacker, 2008), debates about merit- versus need-based aid remain heated (Marklein, 2007; Wilkinson, 2005), and the admission profession is discussing the need to redefine its purpose (College Board's Task Force on Admissions in the 21st Century, 2008). In an effort to understand where college admission is headed, it would seem important to reexamine who the college applicants are in the U.S. and how they can best be understood within a framework of important characteristics related to college choice and admission decisions. Ultimately, this can lead to targeted, innovative efforts to assist students in preparing for and successfully completing college.

Related Literature

Both student and institutional characteristics form the basis for decision-making about where students apply to college and how college admission officers recruit students. The major theory of college choice avers that selecting a postsecondary institution is a dynamic process with three distinct phases: predisposition, search, and choice (Hossler & Gallagher, 1987). During the predisposition phase, an individual's aspiration to attend higher education is explored, which is influenced by the individual's gender and ethnicity; socioeconomic status (SES); academic achievement and ability; parental education and expectations; peer support and peer college choice; high school counselor and teacher

support; career plans; involvement in extracurricular activities; location of family residence; and high school quality and curriculum track. Once college aspirations are formulated, the search phase, in which students locate institutions that meet their most important criteria, begins. After applications are sent and decision letters are received, the final phase of choice occurs and the process is completed with enrollment. Research has shown that the characteristics considered during the predisposition stage are present throughout the entire process. Given that these characteristics are the most influential indicators of college aspirations, search, and choice, identifying how these variables influence college application and enrollment becomes a pertinent question for investigation. Prior research in the realm of college choice has primarily focused on the predisposition and choice phases of the process (Gonzalez & DesJardins, 2002; Weiler, 1994). This study will focus on describing those students engaged in the search phase in order to gain increased awareness and comprehension of this understudied phase.

Although students applying to college have changed quite dramatically since the 1980s, theorists and practitioners continue to rely on older models of college choice that are in need of updating to reflect the current student population (Southerland, 2006). Certainly many of the same variables included in these older models provide a foundation to understand the current college choice process (e.g., academic achievement, finances), but researchers are also aware that additional factors are more or less important to different groups of students such as first-generation college students (Cho, Hudley, Lee, Barry, & Kelly, 2008) or nontraditional or adult students (Southerland, 2006). For example, first-generation students, because they tend to receive less assistance from parents and counselors in searching for colleges, often end up choosing to attend local, public 2-year institutions (MacAllum et al., 2007). Another more recent trend likely influencing the college choice process for certain students is the establishment of higher education access programs across the U.S. There has been a movement for campuses to sponsor programs designed to increase college access and enrichment

opportunities for historically underserved, economically or educationally disadvantaged students (Perna, 2002). These programs likely play a strong role as to where the participants choose to attend college (Bergin, Cooks, & Bergin, 2007).

Factors Associated With College Choice and Application Behaviors

Academic Achievement. The most apparent influence on college choice is academic achievement, which has been positively associated with college enrollment and generally guides the college search process (DesJardins, Ahlburg, & McCall, 2006; Manski & Wise, 1983). Specifically, researchers found that the best predictors of whether a student applied to any college or university were high school grade point average (GPA) and SAT scores (Hossler, Braxton, & Coopersmith, 1989; Manski & Wise, 1983). This finding can be explained by the fact that most colleges and universities utilize measures of high school achievement and standardized test scores as a heuristic for selecting students. In addition, many colleges and universities advertise the mean and/or range of these measures for their most recently admitted freshman class in an effort to advise students on the probability of gaining admission based on their academic credentials.

Gender. Gender and ethnicity both appear to have a role, primarily mediated by cultural values and SES, in how students approach the college application process (Hossler & Gallagher, 1987). A recent survey on college enrollment indicated that there are more females in higher education than males (National Center for Education Statistics, 2005). However, there are a greater proportion of females in 2-year institutions compared to females enrolled in 4-year institutions and a greater proportion of males in 4-year institutions in comparison to males enrolled in 2-year institutions (National Center for Education Statistics, 2005).

Race/Ethnicity and Best Language. Similarly, statistics on enrollment by race/ethnicity showed that an overwhelming majority of enrollees in both 2- and 4-year institutions were White (National Center for Education Statistics, 2005). Asian Americans seem to present a unique minority group as their college enrollment rates rival those of White students (Goyette, 1999; Michaelides, 2002). Latino students represent the highest proportion of students not applying to college by the end of 12th grade followed by African American students (Hurtado, Kurotsuchi Inkelas, Briggs, & Rhee, 1997). Asian American students were more likely than any other racial/ethnic subgroup to apply to five or more colleges, which is indicative of some strategic planning in the college choice process (Hurtado et al., 1997).

Related to race/ethnicity, students' best language (English versus another language) plays a role in their college choice process. There has been tremendous growth in the resident population in the U.S. over the past 10 years—particularly in the South, Southwest, and West (College Board, n.d.). Recent immigration has been primarily dominated by individuals from Latin American countries (Western Interstate Commission for Higher Education, 2008). For students for whom English is not their best language, the college search process can be quite complex. These students tend to be limited in the colleges they consider during the college search process, focusing mostly on nearby 2-year institutions (MacAllum et al., 2007). Most high schools and colleges are lacking bilingual materials and recruiters that could help explain and demystify the college application process for these students and families (College Board's Task Force on Admissions in the 21st Century, 2008).

Parental Income and Education Level. Parental income or SES seems to be the most cited and influential factor in college application and enrollment (Chapman, 1981; Delaney, 1998; Ganderton & Santos, 1995; Kane & Spizman, 1994; Rivkin, 1995; Somers, Cofer, & VanderPutten, 2002; Wilson-Sadberry, Winfield, & Royster, 1991). Research has shown that students with a higher parental income are more likely to attend a 4-year

and/or a private institution than students with a lower parental income, who are more likely to attend a 2-year, public, and/or in-state institution (Chapman, 1981). A recent study confirmed that lower income students were less likely to apply to more expensive institutions, clearly limiting the opportunities for higher education among this group (Lillis & Tian, 2008). Low-SES students are not only financially limited in the types of schools they can afford to attend (e.g., 2-year vs. 4-year and in-state vs. out-of-state), they are also limited in the number of institutions they can apply to because of the high cost of application fees. In addition, McDonough (1994) wrote that high-SES students with moderate academic ability prepare to get into the "right" college using innovative techniques including hiring independent counselors, receiving assistance with essays, or arranging for educational experiences or trips over summer breaks that might enhance an application.

Research also shows a positive relationship between the parents' education level and the child's education level (Hurtado et al., 1997; Stage & Hossler, 1989), as well as students' educational expectations (Goyette, 1999). First-generation college-bound students, or those students whose parents have no education higher than a high school diploma, tend to be at a distinct disadvantage when it comes to preparing for and applying to college (Pascarella, Pierson, Wolniak, & Terenzini, 2004). They often are lacking in knowledge about the application process, costs, and other information related to attaining a college degree, including necessary high school preparation.

Location of Residence and High School. Location of home residence has also been shown to influence college predisposition and choice. Although research has indicated that students who reside in an urban setting are more likely to enroll in college than students who reside in a rural setting (Dahl, 1982), this difference seems to disappear after controlling for SES. Avery and Hoxby (2004) noted that low-income students, regardless of race or academic achievement, were more likely to respond negatively to a college's distance from home. A report from the

National Postsecondary Education Cooperative (MacAllum et al., 2007) summarized research on the role of geographic location in college choice, stating that proximity to home was of particular importance to African American and Hispanic students who did not want to leave their families and was of lesser importance to White students.

Finally, Litten (1982) noted that high school characteristics, student performance, and high school curriculum influence student aspirations to attend a particular institution. Specifically, social and cultural capital, operationalized by Perna (2000) as high school segregation and high school quality, influence the level of knowledge and available information on applying to college.

Extracurricular Participation. School-based extracurricular activity participation (e.g., school government, newspaper, performing arts, academic clubs) is also associated with 4-year college enrollment (Horn, 1997; Perna, 2000), and higher educational and occupational aspirations exist for students who participate in these activities in comparison to those who do not (Marsh, 1992). Perna (2000) speculated that this association may be due to the increased opportunity for exchange of information related to the college application and enrollment process. Participation in sports was also positively associated with educational attainment after high school (Marsh & Kleitman, 2002).

Given the many influences on college application behaviors, this study aims to understand whether certain behaviors and characteristics of students applying to college can be clustered together to form meaningful groups of students for further exploration.

Method

Participants

The sample for this study was taken from the College Board's 2006 College Bound Seniors database. This database is comprised of the 1,465,744 students who took the SAT or at

least one SAT Subject Test and planned to graduate from high school in 2006.

Materials

Five different databases were merged to conduct this research: the 2006 College Bound Seniors database with SAT Questionnaire data[1], the College Board's Advanced Placement (AP) program participation database, 2005 Quality Education Data (QED) National Education Database[2], and the 2006 Annual Survey of Colleges[3] database. The variables examined in this study can be found in Table 1.

Design and Procedure

Cluster analysis seeks to identify homogeneous subgroups of cases in a population. The method includes a wide variety of procedures that are used to empirically form "clusters" or groups of highly similar entities (Aldenderfer & Blashfield, 1984). This study employed two-step clustering, which is the clustering procedure most suitable for large datasets that include both categorical and continuous variables. In the first step, cases were assigned to "preclusters" to reduce the size of the data matrix. In the second step, the preclusters were clustered using the agglomerative hierarchical clustering algorithm whereby each precluster began as its own cluster, and at successive steps, the preclusters were merged until all preclusters were formed into one total cluster. All variables were standardized to have the same mean and standard deviation to equalize variables measured on different scales. The solutions based on a number of different clusters were examined and the best solution was chosen based on the ease of interpretation.

A number of cluster analyses were performed on the data to determine the optimum number of clusters. Cluster solutions were examined based on different combinations of variables guided by Hossler and Gallagher's (1987) theory of college choice. Many of the variables from the 2006 Annual Survey of Colleges (ASC) had a large number of missing values (the percentage of miss-

Table 1

Variables for Cluster Analysis

Variable Category	Variable Name	Variable Levels/Descriptive Statistics
Academic Achievement	SAT Composite (Critical Reading and Math)	*M* = 1027.03, *SD* = 200.70 *Min* = 400, *Max* = 1600
	HSGPA	*M* = 3.34, *SD* = 0.62 *Min* = 0.00, *Max* = 4.30
	Number of AP exams taken	0 = None 1 = 1 or 2 2 = 3 or more
Demographic Characteristics	Gender and Minority Status	1 = Male Minority 2 = Male Nonminority 3 = Female Nonminority 4 = Female Minority
	English Best Language	0 = English is best language 1 = English is not best language
Parental Income and Education Level	Socioeconomic Status (SES)	1 = Low income, low education[a] 2 = Low income, medium education 3 = Low income, high education 4 = Medium income, low education 5 = Medium income, medium education 6 = Medium income, high education 7 = High income, low education 8 = High income, medium education 9 = High income, high education
	First-Generation Student	0 = No (at least one parent has an Associate's Degree or higher) 1 = Yes (highest level of parental education is less than an Associate's Degree)
Location of Home Residence and School	Percent of HS eligible for free lunch	1 = Low (under 20%) 2 = Moderate (20–30%) 3 = High (31–50%) 4 = Very high (greater than 50%)
	Size of HS	1 = Small (less than 500 students) 2 = Medium (500–999 students) 3 = Large (1,000 or more students)
	Region of country	1 = West 2 = Southwest 3 = South 4 = Midwest 5 = Mid-Atlantic 6 = New England
	Majority of HS is college-bound	0 = No (less than 50% of HS is college-bound) 1 = Yes (more than 50% of HS is college-bound)

Variable Category	Variable Name	Variable Levels/Descriptive Statistics
	Majority of HS is comprised of minority students	0 = Yes (minority students are more than 50% of student body) 1 = No (minority students are less than 50% of student body)
Extracurricular Participation	Number of activities the student participated in for 2 or more years	*M* = 1.66, *SD* = 1.38 *Min* = 0, *Max* = 5
	Number of types of activities the student participated in during HS	*M* = 3.35, *SD* = 2.52 *Min* = 0, *Max* = 10
	Student participated in varsity sports for 2 or more years	0 = No 1 = Yes
Institutions of Interest	Number of institutions to which student sent scores	*M* = 4.12, *SD* = 3.46 *Min* = 0, *Max* = 30
	Primary selectivity level of institutions sent scoresc	1 = Mostly[b] highly selective 2 = Mostly selective publics 3 = Mostly selective privates 4 = Mostly moderately selective publics 5 = Mostly moderately selective privates 6 = Mostly nonselective publics 7 = Mostly nonselective privates 8 = Mostly 2-year institutions
	Variability of selectivity level of institutions sent scores	0 = Selectivity does not vary 1 = Selectivity varies
	Primary type of campus of institutions sent scores	1 = Mostly urban 2 = Mostly suburban 3 = Mostly rural
	Variability of type of campus sent scores	0 = Type of campus does not vary 1 = Type of campus varies
	Primary distance of institutions sent scores from student's home	0= Mostly in-state or bordering home state 1 = Mostly out-of-state
	Variability of distance of institutions sent scores from student's home	0 = Distance does not vary 1 = Distance varies

Note. **[a]Low education is defined by less than a Bachelors degree, medium education is defined by a Bachelors degree, and high education is defined by more than a Bachelors degree. Low income is defined by a family income of less than $35K, a medium income is defined by $35K to $100K, and high income is defined by more than $100K. [b]In this table, "mostly" refers to the mode among institutions. If there was no mode, then the student was coded as missing data in this category. [c]The selectivity of the institutions was determined based on data from the 2004–2005 Annual Survey of Colleges (percentage of applicants admitted, mean SAT and ACT scores). In this table, "varies" is defined by the students sending their scores to at least two different types within the category.**

ing cases ranged from 63% to 96%), thus the cluster solutions including these variables involved only a small percentage of the cases (students) in the 2006 SAT cohort file. The students with valid values on the ASC variables were not representative of the college-bound senior population, so the decision was made to exclude the ASC variables when deriving the clusters and to use these variables to describe the final clusters.

The final cluster solution was chosen based on both statistical criteria and ease of interpretation. The auto-clustering procedure in SPSS, using the Schwarz Bayesian Information Criterion (BIC), was used to determine the number of clusters. This procedure determines the number of clusters in which the BIC is small and the change in BIC between adjacent numbers of clusters is small in comparison to that for all other possible numbers of clusters. The auto-clustering procedures chose five clusters for the final solution, with a BIC of 7009466.933 and a BIC change of -185598.544 between a five- and six-cluster solution. A list of the final variables used in this solution can be found in Table 1; however, as mentioned before, none of the variables describing the institutions of interest (ASC variables) were included in the cluster analysis. A separate cluster was created to include outlier cases that did not fit well into any other cluster.

There are several possible methods for validating the results of a cluster analysis. Techniques considered appropriate for the two-step method of clustering include: performing significance tests on variables used to create the clusters, replicating the cluster solution on an independent sample, performing significance tests on variables that are *not* used to create the clusters, and Monte Carlo procedures to generate an artificial data set and compare cluster solutions on the real and artificial data (Aldenderfer & Blashfield, 1984). In this study, where the decision to accept or reject a cluster solution was based primarily on the face validity of the results, two different methods were chosen to validate the cluster solution: performing significance tests on variables used to create the clusters (internal validation) and performing significance tests on variables that are *not* used to create the clusters (external validation). For those variables

included in the cluster analysis, chi-square tests of association were performed for each categorical variable by cluster, and all tests were statistically significant at $p < .01$. A series of one-way analysis of variance (ANOVA) were conducted for each continuous variable included in the cluster analysis and all tests were statistically significant at $p < .01$, which indicated significant differences in the means across the clusters. Post-hoc multiple comparisons using the Bonferroni correction revealed that all pairwise comparisons of clusters were statistically significant for all continuous variables at $p < .01$.

Performing an external validation of a cluster solution is considered to be one of the stronger methods of validation, as Aldenderfer and Blashfield (1984) stated that "the value of a cluster solution that has successfully passed an external validation is much greater than a solution that has not" (p. 66). To perform an external validation, the first two authors arrived at a priori hypotheses regarding expected differences among clusters on the following variables: the number of honors or awards that a student received in high school, whether the student was a first-generation college student, and the student's highest educational degree goal. The results of these validation analyses showed that all *a priori* hypotheses were supported by ANOVA and chi-square results. Specifically, Cluster 1 had significantly more honors received than Cluster 5, $F(5, 728{,}003) = 11211.70$, $p < .001$. Cluster 2 had significantly more first-generation college students than Cluster 4, $F(5, 728{,}003) = 17{,}588.93$, $p < .001$. Cluster 4 had significantly more students desiring the highest degree goal level (doctoral) than Cluster 3, χ^2 (30, $N = 688{,}709$) = 23145.50, $p < .001$ (Cluster 3 standardized residual = -27.4; Cluster 4 standardized residual = 54.1).

Results

Table 2 displays the correlation matrix of the variables included in the cluster analysis. Tables 3, 4, and 5 display the frequency distribution and effect sizes of each categorical variable

for the five clusters; the number and percentage of cases in each cluster is shown in the header rows of these tables. Table 6 displays the means, standard deviations, and effect sizes of the continuous variables by cluster. Variable importance plots, or graphs that show which variables are more or less important in differentiating the five clusters, are provided in the Appendix. For each cluster, there is one plot with categorical variable information and one with continuous variable information. Variables with higher chi-square or *t* values than other variables are more important in differentiating a particular cluster from other clusters.

As a measure of practical significance, effect sizes were computed for all statistical tests to show the strength of the relationship or mean differences between variables and clusters (Trusty, Thompson, & Petrocelli, 2004). Cohen's *d* standardized-difference effect size was calculated for each continuous variable and Cramer's *V* was calculated for categorical variables. Cohen's *d* was calculated by subtracting the total group mean on a variable from the cluster mean and then dividing by the total group (pooled) standard deviation. This value can be negative or positive and a value of zero would indicate no differences between groups. Cohen (1988) provided guidance in the interpretation of effect sizes by characterizing an effect size of 0.2 as small, an effect size of 0.5 as medium, and an effect size of 0.8 as large. In this study, Cohen's *d* ranged from -0.65 to 0.61, with the smallest *d* value being -0.07. Cramer's *V* is an index of the degree of association in a contingency table that is larger than 2 x 2 (Hayes, 1994). It ranges between 0.00 and 1.00, and the higher the value the more strongly related two variables are considered to be. In this study, Cramer's *V* ranged from 0.07 to 0.91. While there is no standard way to interpret Cramer's *V*, values closer to 0.00 indicate a weaker relationship.

Description of the Clusters

Effect Sizes. An examination of the standardized differences among the continuous variables showed some meaningful differences across the clusters. The largest effect sizes were found in

Table 2

Correlation Matrix of Cluster Analysis Variables

*Correlations of Variables**									
		1	2	3	4	5	6	7	8	9
1	AP Tests (number of)									
2	ECA: # Participated in for 2+ Yrs	0.19								
3	ECA: # of Types Participated	0.14	0.81							
4	HSGPA	0.37	0.27	0.20						
5	HS: Percent Eligible for Free Lunch	-0.03	-0.04	-0.03	-0.05					
6	HS: Size	0.08	-0.08	-0.08	-0.05	0.04				
7	Number of Institutions Sent Scores to	0.27	0.16	0.15	0.20	-0.05	0.01			
8	Parental Education	0.21	0.13	0.11	0.17	-0.24	-0.00[+]	0.20		
9	Parental Income	0.10	0.08	0.05	0.11	-0.35	-0.01	0.09	0.38	
10	SAT Critical Reading and Math	0.50	0.24	0.17	0.50	-0.27	-0.01	0.31	0.37	0.31

Note. All correlations are significant at $p < .01$.

[+]Correlations were rounded to the nearest hundredth and this significant correlation is actually larger than 0.00.

Table 3

Frequencies and Effect Sizes of Categorical Student-Level Variables by Cluster

		Cluster 1 Privileged High Achievers/Athletes N = 177,287		Cluster 2 Disadvantaged Students N = 115,784		Cluster 3 Average Students Needing More Guidance N = 162,198		Cluster 4 Mostly Female Academics N = 84,306		Cluster 5 Privileged Low Achievers N = 182,045	
Variable	Cramer's V	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster
Gender x Minority Status	.33***										
Female minority		8,832	5.0	52,788	45.6	17,608	10.9	12,848	15.2	16,521	9.1
Female nonminority		70,405	39.7	15,700	13.6	60,567	37.3	64,491	76.5	67,164	36.9
Male minority		11,666	6.6	32,971	28.5	16,662	10.3	6,967	8.3	12,802	7.0
Male nonminority		86,384	48.7	14,325	12.4	67,361	41.5	0	0.0	85,558	47.0
Best Language	.15***										
English		176,656	99.6	110,456	95.4	160,363	98.9	83,116	98.6	178,926	98.3
Not English		631	0.4	5,328	4.6	1,835	1.1	1,190	1.4	3,119	1.7
Parental Income and Education	.23***										
Low income, low education		4,222	2.4	53,548	46.2	35,940	22.2	7,804	9.3	18,648	10.2
Low income, med education		3,432	1.9	6,611	5.7	6,612	4.1	3,071	3.6	5,467	3.0
Low income, high education		1,702	1.0	3,321	2.9	2,280	1.4	1,620	1.9	2,668	1.5
Med income, low education		34,799	19.6	26,332	22.7	61,686	38.0	16,966	20.1	45,239	24.9
Med income, med education		36,417	20.5	11,607	10.0	26,613	16.4	16,726	19.8	33,599	18.5
Med income, high education		23,782	13.4	6,985	6.0	15,483	9.5	12,497	14.8	19,473	10.7
High income, low education		6,906	3.9	1,715	1.5	6,055	3.7	2,786	3.3	9,634	5.3
High income, med education		25,755	14.5	2,208	1.9	4,601	2.8	8,635	10.2	21,829	12.0

Table 3, continued

		Cluster 1 Privileged High Achievers/Athletes N = 177,287		Cluster 2 Disadvantaged Students N = 115,784		Cluster 3 Average Students Needing More Guidance N = 162,198		Cluster 4 Mostly Female Academics N = 84,306		Cluster 5 Privileged Low Achievers N = 182,045	
Variable	Cramer's V	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster
High income, high education		40,272	22.7	3,457	3.0	2,928	1.8	14,201	16.8	25,488	14.0
First-Generation College	.33***										
No		144,077	81.2	44,383	38.3	80,373	49.6	63,036	74.8	125,289	68.8
Yes		33,280	18.8	71,401	61.7	81,825	50.4	21,270	25.2	56,756	31.2
AP Exams Taken	.40***										
None		93,938	53.0	70,281	60.7	123,712	76.3	164	0.2	160,678	88.3
1 or 2		41,369	23.3	25,965	22.4	18,921	11.7	46,167	54.8	21,367	11.7
3 or more		41,980	23.7	19,538	16.9	19,565	12.1	37,975	45.0	0	0.0
Degree Goal	.08***										
Certificate		939	0.5	1,309	1.2	1,805	1.2	314	0.4	1,790	1.1
Associate		843	0.5	1,772	1.6	2,898	1.9	202	0.3	2,324	1.4
Bachelor's		45,176	25.8	26,816	24.5	47,281	30.8	14,537	18.5	54,254	32.6
Master's		58,638	33.5	36,067	33.0	44,214	28.8	26,086	33.2	214,484	31.1
Doctoral		36,978	21.2	27,702	25.3	26,565	17.3	22,933	29.2	23,835	14.3
Other		384	0.2	704	0.6	753	5.6	208	0.3	688	0.4
Undecided		31,857	18.2	15,033	13.7	30,097	19.6	14,320	18.2	35,705	21.4
Particip. in Varsity Sports (2+ yrs)	.78***										
No		5,194	2.9	88,816	76.7	107,314	66.2	84,306	100.0	182,045	100.0
Yes		172,093	97.1	26,968	23.3	54,884	33.8	0	0.0	0	0.0

***$p < .001$.

Table 4

Frequencies and Effect Sizes of Categorical High School-Level Variables by Cluster

		Cluster 1 Privileged High Achievers/Athletes N = 177,287		Cluster 2 Disadvantaged Students N = 115,784		Cluster 3 Average Students Needing More Guidance N = 162,198		Cluster 4 Mostly Female Academics N = 84,306		Cluster 5 Privileged Low Achievers N = 182,045	
Variable	Cramer's V	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster
Location of High School	.12***										
Mid-Atlantic		52,854	29.8	22,486	19.4	42,532	26.2	18,714	22.2	57,767	31.7
Midwest		24,280	13.7	2,822	2.4	16,760	10.3	9,689	11.5	20,046	11.0
New England		21,523	12.1	3,398	2.9	9,347	5.8	5,455	6.5	22,817	12.5
South		32,509	18.3	31,234	27.0	41,830	25.8	20,880	24.8	32,776	18.0
Southwest		11,995	6.8	22,926	19.8	15,519	9.6	8,706	10.3	14,833	8.1
West		34,126	19.2	32,918	28.4	36,210	22.3	20,862	24.7	33,806	18.6
School size	.13***										
Small (< 500 students)		18,298	10.3	4,441	3.8	15,773	9.7	4,762	5.6	12,937	7.1
Medium (500–999 students)		45,568	25.7	10,239	8.8	38,418	23.7	13,942	16.5	35,069	19.3
Large (> 999 students)		113,421	64.0	101,104	87.3	108,007	66.6	65,602	77.8	134,039	73.6
Students eligible for free lunch	.51***										
Low (< 20%)		165,854	93.6	7,346	6.3	61,844	38.1	71,029	84.3	178,987	98.3
Moderate (20–30%)		9,552	5.4	15,813	13.7	56,827	35.0	8,912	10.6	2,687	1.5
High (31–50%)		1,841	1.0	45,701	39.5	40,281	24.8	4,365	5.2	371	0.2
Very High (> 50%)		40	0.0	46,924	40.5	3,246	2.0	0	0.0	0	0.0
Majority is college-bound	.44***										
No (< 50% college-bound)		22,756	12.8	51,765	44.7	86,455	53.3	16,552	19.6	11,254	6.2
Yes (> 50% college-bound)		154,531	87.2	64,019	55.3	75,743	46.7	67,754	80.4	170,791	93.8
Majority is minority students	.91***										
No (< 50% minority students)		175,246	98.8	9,021	7.8	160,148	98.7	81,558	96.7	180,472	99.1
Yes (> 50% minority students)		175,246	98.8	9,021	7.8	160,148	98.7	81,558	96.7	180,472	99.1

***$p < .001$.

Table 5

Frequencies and Effect Sizes of Categorical Institution-Level Variables by Cluster

		Cluster 1 Privileged High Achievers/Athletes N = 177,287		Cluster 2 Disadvantaged Students N = 115,784		Cluster 3 Average Students Needing More Guidance N = 162,198		Cluster 4 Mostly Female Academics N = 84,306		Cluster 5 Privileged Low Achievers N = 182,045	
Variable	Cramer's V	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster
Number of Institutions Sent Scores	.23***										
None		29,203	16.5	21,226	18.3	42,021	25.9	10,560	12.5	48,156	26.5
1–4 schools		61,257	34.6	52,359	45.2	76,780	47.3	32,815	38.9	76,628	42.1
5-10 schools		74,007	41.7	38,734	33.5	36,159	22.3	38,338	45.5	53,190	29.2
11+ schools		12,820	7.2	3,465	3.0	7,238	4.5	2,593	3.1	4,071	2.2
Institution Selectivity (Majority)	.14***										
Highly selective		2,983	21.4	751	8.3	870	5.4	1,523	23.5	1,051	5.8
Selective publics		0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Selective privates		1,913	13.7	637	7.1	1,126	6.9	896	13.8	1,805	9.9
Moderately selective publics		4,723	33.9	4,862	54.0	7,408	45.7	2,273	35.1	8,254	45.2
Moderately selective privates		1,938	13.9	890	9.9	2,118	13.1	953	14.7	2,373	13.0
Nonselective publics		1,144	8.2	656	7.3	2,037	12.6	484	7.5	1,968	10.8
Nonselective privates		158	1.1	66	0.7	231	1.4	50	0.8	262	1.4
Two-year institutions		1,081	7.8	1,145	12.7	2,430	15.0	302	4.7	2,553	14.0
Institution Selectivity Varies?	.07***										
Varies		52,765	46.5	38,946	56.9	42,014	48.7	25,312	45.3	46,928	47.4
Does not vary		60,592	53.5	29,500	43.1	44,234	51.3	30,533	54.7	52,030	52.6

		Cluster 1 Privileged High Achievers/Athletes N = 177,287		Cluster 2 Disadvantaged Students N = 115,784		Cluster 3 Average Students Needing More Guidance N = 162,198		Cluster 4 Mostly Female Academics N = 84,306		Cluster 5 Privileged Low Achievers N = 182,045	
Variable	Cramer's V	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster	n	% Cluster
Type of Campus	.16***										
Mostly rural		2,776	12.8	826	4.6	4,532	18.1	1,033	8.6	3,343	11.3
Mostly suburban		13,021	59.8	7,194	40.0	11,365	45.4	6,282	52.3	16,559	56.0
Mostly urban		5,969	27.4	9,955	55.4	9,146	36.5	4,687	39.1	9,659	32.7
Campus Type Varies?	.08***										
Varies		18,529	16.6	8,442	12.2	11,948	14.1	7,426	13.3	14,320	14.5
Does not vary		93,358	83.4	61,033	87.8	72,516	85.9	48,242	86.7	84,458	85.5
Campus distance from home	.18***										
Mostly in-state/bordering state		108,594	78.2	81,421	90.5	102,148	89.4	55,238	79.8	110,026	86.8
Mostly out-of-state		30,279	21.8	8,550	9.5	12,144	10.6	13,989	20.2	16,774	13.2
Campus Distance Varies?	.16***										
Varies		1,352	1.0	1,396	1.6	1,164	1.0	718	1.0	1,771	1.4
Does not vary		136,662	99.0	87,875	98.4	112,261	99.0	68,000	99.0	123,974	98.6

Note. **The standardized-difference effect size Cohen's d was calculated for each variable by subtracting the total group mean on that variable from the cluster mean and then dividing by the total group standard deviation.**

***$p < .001$.

Table 6

Means, Standard Deviations, and Effect Sizes of Continuous Variables by Clusters

	Cluster 1 Privileged High Achievers/ Athletes $N = 177,287$			Cluster 2 Disadvantaged Students $N = 115,784$			Cluster 3 Average Students Needing More Guidance $N = 162,198$			Cluster 4 Mostly Female Academics $N = 84,306$			Cluster 5 Privileged Low Achievers $N = 182,045$		
	M	*SD*	*d*	*M*	*SD*	*d*	*M*	*SD*	*d*	*M*	*SD*	*d*	*M*	*SD*	*d*
HSGPA	3.47	0.58	0.21	3.18	0.66	-0.25	3.29	0.63	-0.08	3.69	0.46	0.56	3.17	0.61	-0.27
SAT (Critical Reading and Math)	1100.59	184.02	0.37	896.82	188.95	-0.65	983.52	189.27	-0.22	1149.89	164.12	0.61	1009.96	178.59	-0.09
Number of AP exams taken	1.45	2.15	0.18	1.11	1.91	0.00	0.72	1.67	-0.20	2.86	2.06	0.92	0.16	0.48	-0.49
Number of extracurricular categories participated in for 2 or more years	2.13	1.26	0.34	1.42	1.35	-0.18	1.56	1.35	-0.07	2.02	1.47	0.26	1.27	1.31	-0.28
Number of extracurricular categories participated in during HS	4.12	2.21	0.3	3.08	2.5	-0.11	3.16	2.47	-0.08	3.87	2.72	0.21	2.67	2.5	-0.27

Note. **The standardized-difference effect size Cohen's *d* was calculated for each variable by subtracting the total group mean on that variable from the cluster mean and then dividing by the total group standard deviation.**

SAT scores, with Cluster 2 showing significantly lower performance than the other clusters (d = -0.65) and Cluster 4 showing significantly higher performance (d = 0.61). An examination of effect sizes for the categorical variables across clusters using Cramer's V showed that there were a number of major differences among these groups. The largest effect sizes were found in the following variables: Majority of Students in High School is Racial/Ethnic Minority (V = 0.91, p < .001), Participated in Varsity Sports for 2+ Years (V = 0.78, p < .001), Percent of Students in High School Eligible for Free Lunch (V = 0.51, p < .001), Majority of Students in High School is Considered College Bound (V = 0.44, p < .001), Number AP Exams Taken (V = 0.40, p < .001), First-Generation College Student (V = 0.33, p < .001), and Gender by Minority Status (V = 0.33, p < .001).

Cluster Names. All clusters were given descriptive labels based on the results of the analyses. Cluster 1 was labeled Privileged High Achievers/Athletes. Nearly half of the students in this cluster (49%) were male nonminority students who attended affluent high schools with other mostly nonminority students. Many of the students in this cluster resided in the Midwest, Mid-Atlantic, or New England. Students in this cluster sent their SAT scores to the highest number of institutions compared to the other clusters. This cluster also had the largest percentage of students participating in varsity sports for 2 or more years (97%), as well as the smallest percentage of first-generation college students (19%).

Cluster 2 was labeled Disadvantaged Students. Nearly three quarters of the students in this cluster were minority students (29% male minority and 46% female minority) who attended mostly large high schools that had a large proportion of minority students and students eligible for free lunch. This cluster also had the highest percentage of students reporting that their best language was something other than English, as compared to the other clusters (5%). These students tended to come from families with low income and low parental education, and many resided in the West, Southwest, and South. Fifty-five percent of the students in this cluster attended high schools where the majority

of students were college-bound. The students in this cluster had one of the lowest mean high school GPAs (M = 3.18, SD = .66, d = -0.25) and the lowest mean combined SAT score (M = 897, SD = 189, d = -0.65). This cluster also had the highest percentage of students (83%) among the five clusters who indicated a plan to earn a bachelor's, master's, or doctoral degree as well as the highest percentage of students considered to be first-generation college students (62%).

Cluster 3 was labeled Average Students Needing More Guidance. This cluster was comprised of mostly nonminority students (42% male and 37% female) who largely attended high schools with similar racial/ethnic compositions in the West, South, or Mid-Atlantic. Many of the students in this cluster (53%) attended high schools where the majority was *not* college-bound. Half of the students in Cluster 3 are considered first-generation college students. Cluster 3 also had the second highest percentage of students reporting their parents' income and education level to both be low (22%). The average high school GPA for this cluster (M = 3.29) was the median among the five clusters, and the average combined SAT critical reading and math score (M = 984) was only higher than that of Cluster 2. The effect size of the SAT for Cluster 3 was -0.22 compared to the total group including all clusters. Cluster 3 was also the median among the five clusters in terms of the average number of extracurricular activity types for 2 or more years as well as the average number of extracurricular activity categories overall during high school.

Cluster 4 was labeled Mostly Female Academics. Nearly 77% of the students in this cluster were female, nonminority students. The students in this cluster also largely attended high schools located in the West, South, or Mid-Atlantic with mostly nonminority students. The students in this cluster tended to take the most AP exams and did not participate in varsity sports. These students had the highest mean high school GPA among the clusters, with the smallest standard deviation (M = 3.69, SD = 0.46, d = 0.56), as well as the highest average combined SAT critical reading and math scores with the smallest variability (M

= 1150, *SD* = 164, *d* = 0.61). Most of these students (75%) were not first-generation college students.

Cluster 5 was labeled Privileged Low Achievers. This cluster was comprised of mostly nonminority students (47% male and 37% female) who attended wealthy high schools with mostly nonminority students. This cluster had the largest percentage of cases across clusters attending schools where the majority of students were college-bound (94%). Yet, this cluster also had the highest percentage of students who did not take any AP exams (88%) and the highest percentage that did not send their SAT scores to any colleges (27%). In addition, the students in this cluster had the lowest mean high school GPA among the clusters (*M* = 3.17, *SD* = .61, *d* = -0.27), and participated in the least number of extracurricular activities for 2 or more years (*d* = -0.28), as well as the least number of extracurricular activities overall during high school (*d* = -0.27). This group had the smallest percentage among the five clusters (76%) indicating their aspirations to earn a bachelor's, master's, or doctoral degree. Approximately one third of Cluster 5 students are considered first-generation college students.

SAT Score-Sending Behaviors. Cluster 1 had the most variability in the type of campus to which students sent their SAT scores. Most of the students in all five clusters sent their SAT scores to only one type of campus setting (rural, urban, or suburban). However, approximately 17% of the students in Cluster 1 sent their scores to a variety of campus settings compared to 12–14.5% of students in the other clusters. Clusters 1 and 4 showed the largest percentage of students sending their SAT scores to institutions out of their home state (22% and 20% respectively), compared to 9.5–13% of students in the other clusters. Cluster 2 had the largest percentage of students (91%) sending their SAT scores to institutions mostly in state or in a state bordering their state.

Cluster 2 showed the most individual variability among the clusters in the types of institutions students chose with regard to control (public or private) and selectivity. Approximately 57% of the students in Cluster 2 sent their scores to a variety

of types of institutions compared to 45–49% of students in the other clusters. Students in Cluster 2 also tended to send their scores to more moderately selective public institutions. Cluster 3 had the largest percentage among the clusters of students who sent scores to nonselective public institutions or 2-year colleges (29%). Cluster 4 students were the most likely to send their scores to similar types of institutions in terms of control and selectivity, and many of these students sent their scores to highly selective or moderately selective public institutions.

There was virtually no variation between the clusters in whether or not the institutions to which students sent scores varied in distance from the students' homes. With regard to campus environment, the students in Cluster 2 tended to prefer mostly urban settings (55%), while the majority of students in Clusters 1, 4, and 5 sent their scores to mostly suburban campuses. Cluster 3 showed the most variation in the type of campus environment and had the largest percentage of students among the clusters (18%) who sent their scores to mostly rural colleges.

Discussion

College applicants enter the college admission process with distinct characteristics and college aspirations. As such, examining the different combinations of these variables to cluster similar students may be a useful way to understand the population of college-bound students and serve their unique needs. The cluster analysis in this study demonstrated that different patterns of characteristics and behaviors were associated with different clusters of students. Five unique clusters of students were identified: Privileged High Achievers/Athletes, Disadvantaged Students, Average Students Needing More Guidance, Mostly Female Academics, and Privileged Low Achievers. Given that the media has focused its attention on the very high achieving students that apply to 20 or more schools with stellar test scores and the students who are ill-prepared to take on academic challenges of college (e.g., Pappano, 2007), one would expect

to find two clusters of college applicants—the Privileged High Achievers/Athletes (Cluster 1) and the Disadvantaged Students (Cluster 2). However, this study found five. In particular, there is little mention in the media or elsewhere regarding students in the Average Students Needing More Guidance (Cluster 3) or the Privileged Low Achievers (Cluster 5) clusters.

The Privileged High Achievers/Athletes cluster (Cluster 1) was comprised of many nonminority males attending affluent high schools. The students in this cluster are similar to the "all-around" students we hear so much about. They have many resources at their fingertips to navigate the college search and choice process. Given that this cluster has the lowest percentage of first-generation college-bound students, it is likely that they are able to rely on their families for information about applying to and attending college.

The Disadvantaged Students cluster (Cluster 2) was comprised of students who appeared to be lacking in financial, social, and educational resources and were lagging in academic qualifications as they embarked on the college search and choice process. Not surprisingly, this cluster had the largest percentage of first-generation college-bound students as well as those reporting their best language to be other than English.

The Average Students Needing More Guidance cluster (Cluster 3) was comprised of students that would seem to benefit from the knowledge of guidance counselors with regard to navigating the college search, choice, and financial planning process. Given that this cluster had the median high school GPA, it is surprising that, on average, the slight majority of the students in the high schools they attended were not college bound. However, the large percentage of first-generation college-bound students in this cluster may explain this finding. Also, the students in this cluster were involved in their schools via extracurricular activity participation, which signified an interest in contributing to their high school and interacting with the community—and is known to be related to a host of beneficial academic and nonacademic outcomes during and after high school (Marsh & Kleitman, 2002).

The Mostly Female Academics cluster (Cluster 4) resembled the Privileged High Achievers/Athletes (Cluster 1) except that the students in the Mostly Female Academics cluster did not participate in sports and had a smaller percentage of students reporting high parental income. This cluster was also the most likely to show little variation in the types of schools to which they sent their scores (with regard to control and selectivity) and many of them sent scores to highly selective or moderately selective public institutions.

The Privileged Low Achievers cluster (Cluster 5) was the largest cluster among the five clusters. This was somewhat surprising, as it seems these were students who had an excellent foundation for success in high school, yet they fell short academically and with regard to aspirations for higher education.

We seek to cluster or classify students into like groups in order to better understand their differences and similarities. Although there were similarities between clusters, such as the Privileged High Achievers/Athletes (Cluster 1) and Privileged Low Achievers (Cluster 5) with their similar proportions of minority students and students in higher SES categories, the clusters clearly differed on the basis of academic performance. Thus, merely classifying students by gender and minority status or SES would have lumped students into groups with discrepant achievement patterns and motivations—clearly important parts of the college search and choice process. This exemplifies the utility in exploring these data with cluster analysis techniques.

Implications and Directions for Future Research

It is likely that these clusters have implications for admission directors, institutional researchers, enrollment managers, and school counselors, as well as the popular media, in developing and evaluating interventions and outreach efforts that aid students in the college preparation, search, and choice processes. First, it would be useful to test hypotheses related to tailoring

such efforts for the different subgroups. The Average Students Needing More Guidance (Cluster 3) may have been more successful in college with greater access to educational resources, and in particular, college preparatory classes and college application, search, and choice direction from parents and educators. This cluster had the second highest percentage of first-generation college-bound students (50%), so it is likely that many of the parents of the students in this cluster are not fully aware of the application and enrollment process and, thus, cannot provide the necessary support to their children. The educational system may be failing these students, and that fact is going undetected. Middle schools guidance counselors can play a helpful role with this group by having students, as well as their parents, think about the college process early. Curricula, often available in English and Spanish, exist to help students, along with their parents, prepare for college starting in the seventh grade through counselor-led lessons in study skills, coursework planning, the college application process, and financial aid.

Although we may have expected to find the Disadvantaged Students cluster (Cluster 2) in the data based on national demographic information (e.g., MacAllum et al., 2007), as well as the popular media (e.g., Schworm, 2008), this does not mean that this group of students and their unique challenges should be overlooked or ignored. New and innovative efforts must be directed at this continuously underserved group. This cluster deserves special attention and deeper exploration in the realm of educational interventions that can aid in the academic preparation of these students and start them on an early and smooth path toward higher education. One example of a promising innovative effort aimed at these students is the Posse Program (Glater, 2006), where small, diverse groups of inner-city students are selected and trained to attend a selective college/university as a group to provide support to each other and are graduating at a rate of 90%.

The Privileged Low Achievers (Cluster 5) seem to be the students who are equipped with the resources to succeed; however, they do not take advantage of these resources for various

reasons. Investigations into their academic motivation or expectations may be worthy of further study. It is possible that these students may benefit from taking a year between high school and college to participate in a community service program, for example, in order to grow and mature and determine what they might want to focus on academically.

Future research should attempt to replicate the analyses in this study with different student cohorts to determine the stability of the findings. It would also be useful to arrive at hypotheses regarding cluster-tailored outreach strategies related to the college application and choice process. A follow-up qualitative investigation, potentially with focus groups of students based on cluster membership, may be worthwhile to address the specific needs and issues that these groups of students are confronting. For example, the Mostly Female Academics (Cluster 4) may benefit from learning techniques to handle the academic pressures they face throughout the college application process, while the Average Students in Need of Guidance (Cluster 3) may benefit from greater direct outreach from colleges and universities—whether this entails meetings with admission counselors or more personal invitations to open houses.

Limitations

There are a number of limitations of this study that are related to the nature of the sample. All students in the database examined took the SAT and their institutions of interest were determined from the colleges and universities that received their scores. As such, students who applied to schools that did not require test scores or accepted unofficial score reports would not be captured in our database. Additionally, sending scores to a particular institution does not guarantee that an application will also be sent to that school. Thus, the database used in this study did not include information on where students ultimately applied or enrolled, but rather all of the institutions they considered seriously enough to send their scores.

Conclusion

Those involved in enrollment management are well aware of the changing nature of college admissions and the students who will be "knocking on the door" (Western Interstate Commission for Higher Education, 2008). Colleges and universities cannot continue to search and recruit the same types of students they have been recruiting in the past. Population and demographic shifts in the United States present new landscapes and challenges for those in higher education. This study provides a unique glimpse into who is currently applying to college with data that are not often jointly considered, including students' demographic, geographic, high school, aspirational, and target college characteristics. Ongoing examinations are necessary as the importance of understanding applicants from all angles cannot be underestimated.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Thousand Oaks, CA: Sage.

Avery, C., & Hoxby, C. M. (2004). Do and should financial aid packages affect students' college choices? In C. M. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 239–302). Chicago, IL: University of Chicago Press.

Bergin, D. A., Cooks, H. C., & Bergin, C. C. (2007). Effects of a college access program for youth underrepresented in higher education: A randomized experiment. *Research in Higher Education, 48,* 727–750.

Chapman, D. W. (1981). A model of student college choice. *The Journal of Higher Education, 52,* 490–505.

Cho, S.-J., Hudley, C., Lee, S., Barry, L., & Kelly, M. (2008). Roles of gender, race, and SES in the college choice process among first-generation and non first-generation students. *Journal of Diversity in Higher Education, 1,* 95–107.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

College Board. (n.d.). *Percent change in resident population, 1986–2006, by College Board region.* Retrieved from http://professionals.collegeboard.com/data-reports-research/trends/higher-ed-landscape

College Board's Task Force on Admissions in the 21st Century. (2008). *Preserving the dream of America: A message to a community of educational leaders.* New York, NY: The College Board.

Dahl, R. W. (1982, May). *College attendance and institutional choice: Results from the Kentucky longitudinal study.* Paper presented at the annual forum of the Association of Institutional Research, Denver, CO.

Delaney, A. M. (1998, May). *Parental income and students' college choice process: Research findings to guide recruitment strategies.* Paper presented at the annual forum of the Association for Institutional Research, Minneapolis, MN.

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2006). An integrated model of application, admission, enrollment, and financial aid. *Journal of Higher Education, 77,* 381–429.

Ganderton, P. T., & Santos, R. (1995). Hispanic college attendance and completion: Evidence from the high school and beyond surveys. *Economics of Education Review, 14,* 35–46.

Glater, J. D. (2006, May 24). In search of standouts who may not stand out enough. *The New York Times,* p. B9.

Gonzalez, J. M. B., & DesJardins, S. L. (2002). Artificial neural networks: A new approach to predicting application behavior. *Research in Higher Education, 43,* 235–258.

Goyette, K. (1999, April). *Application to college: A comparison of Asian American and White high school students*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec.

Hayes, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace.

Horn, L. J. (1997). *Confronting the odds: Students at risk and the pipeline to higher education* (NCES 98-094). Washington, DC: U.S. Department of Education Office of Educational Research and Improvement.

Hossler, D., Braxton, J., & Coopersmith, G. (1989). Understanding student college choice. In F. K. Stage, G. L. Anaya, J. P. Bean, D. Hossler, & G. D. Kuh (Eds.), *College students: The evolving nature of research.* Needham Heights, MA: Ginn Press.

Hossler, D., & Gallagher, K. S. (1987). Studying student choice: A three phase model and the implications for policy makers. *College & University, 2,* 207–221.

Hurtado, S., Kurotsuchi Inkelas, K., Briggs, C., & Rhee, B. S. (1997). Differences in college access and choice among racial/ethnic groups: Identifying continuing barriers. *Research in Higher Education, 38,* 43–75.

Kane, J., & Spizman, L. M. (1994). Race, financial aid and college attendance: Parents and geography matter. *American Journal of Economics & Sociology, 53,* 85–97.

Lillis, M. P., & Tian, R. G. (2008, Summer). The impact of cost on college choice: Beyond the means of the economically disadvantaged. *Journal of College Admission, 200,* 4–14.

Litten, L. H. (1982). Different strokes in the applicant pool: Some refinements in a model of student college choice. *Journal of Higher Education, 53,* 383–402.

MacAllum, K., Glover, D. M., Queen, B., & Riggs, A. (2007). *Deciding on postsecondary education: Final report* (NPEC 2008-850). Washington, DC: National Postsecondary Education Cooperative.

Manski, C. F., & Wise, D. A. (1983). *College choice in America.* Cambridge, MA: Harvard University Press.

Marklein, M. B. (2007, March 15). Colleges taking another look at value of merit-based aid. *USA Today,* p. D8.

Marsh, H. W. (1992). Extracurricular activities: Beneficial extensions of the traditional curriculum or subversion of academic goals? *Journal of Educational Psychology, 84,* 553–562.

Marsh, H. W., & Kleitman, S. (2002). Extracurricular school activities: The good, the bad, and the nonlinear. *Harvard Educational Review, 72,* 464–511.

McDonough, P. M. (1994). Buying and selling higher education: The social construction of the college applicant. *Journal of Higher Education, 65,* 427–446.

Michaelides, M. (2002, April). *Closing the gap in college-going: A study of trends in California public high school graduates' college-going patterns.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

National Center for Education Statistics. (2005). *Enrollment at Title IV institutions, by control and level of institution, student level, attendance status, gender, and race/ethnicity: United States, fall 2004.* Retrieved from http://nces.ed.gov/das/library/tables_listings/show_nedrc.asp?rt=p&tableID=3112

Nyhan, P. (2006, May 22). Dream project helps poor get into college: One student's vision makes a difference. *Seattle Post–Intelligencer*, p. B1.

Pappano, L. (2007, January 7). The incredibles. *The New York Times*, p. A18.

Pascarella, E. T., Pierson, C. T., Wolniak, G. C., & Terenzini, P. T. (2004). First-generation college students: Additional evidence on college experiences and outcomes. *Journal of Higher Education, 75,* 249–284.

Perna, L. W. (2000, April). *Racial/ethnic group differences in the realization of educational plans*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Perna, L. W. (2002). Precollege outreach programs: Characteristics of programs serving historically underrepresented groups of students. *Journal of College Student Development, 43,* 64–83.

Rivkin, S. G. (1995). Black/White differences in schooling and employment. *Journal of Human Resources, 30,* 826–852.

Rubin, B. M. (2008, April 18). From home school to top schools; Chelsea Link feared she might not get into any top schools. She got into seven. Harvard. Yale. Princeton. NU. Stanford. In a tough year, she got into them all. *Chicago Tribune,* p. 1.

Schworm, P. (2008, November 10). Colleges reach out to poorer students: Critics question commitment to diversify. *Boston Globe*, p. B1.

Somers, P., Cofer, J., & VanderPutten, J. (2002). The early bird goes to college: The link between early college aspirations and postsecondary matriculation. *Journal of College Student Development, 43,* 93–107.

Southerland, J. N. (2006, November). *Formulating a new model of college choice and persistence*. Paper presented at the annual meeting of the Association for the Study of Higher Education, Anaheim, CA.

Stage, F. K., & Hossler, D. (1989). Differences in family influences on college attendance plans for male and female ninth graders. *Research in Higher Education, 30,* 301–314.

Thacker, L. (2008). Pulling rank. *New England Journal of Higher Education, 22*(5), 15–16.

Trusty, J., Thompson, B., & Petrocelli, J. V. (2004). Practical guide for reporting effect size in quantitative research in the *Journal of*

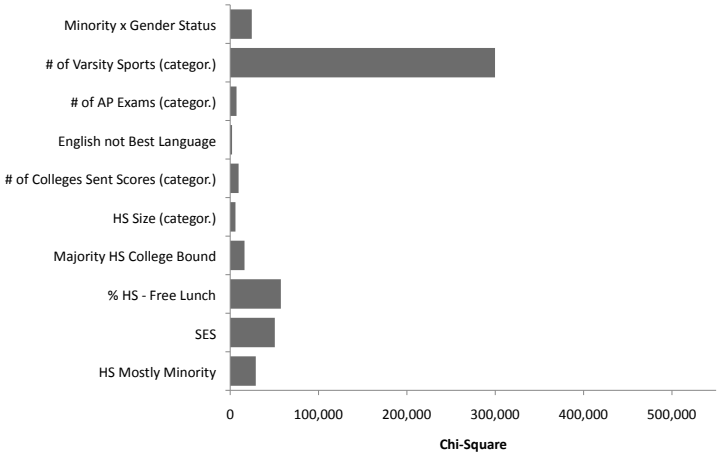
Counseling & Development. Journal of Counseling & Development, 82, 107–110.
Weiler, W. C. (1994). Transition from consideration of a college to the decision to apply. *Research in Higher Education, 35,* 631–646.
Western Interstate Commission for Higher Education. (2008). *Knocking at the college door: Projections of high school graduates by state and race/ethnicity 1992–2022*. Boulder, CO: Western Interstate Commission for Higher Education.
Wilkinson, R. (2005). *Aiding students, buying students: Financial aid in America*. Nashville, TN: Vanderbilt University Press.
Wilson-Sadberry, K. R., Winfield, L. F., & Royster, D. A. (1991). Resilience and persistence of African American males in postsecondary enrollment. *Education and Urban Society, 24,* 87–102.

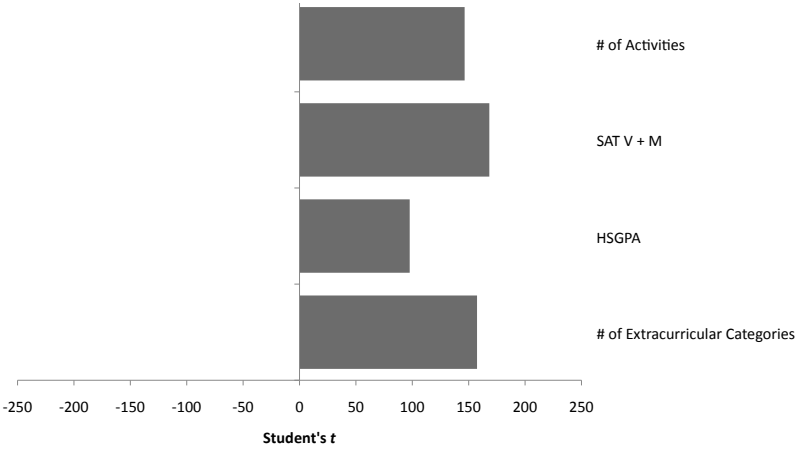
End Notes

1 This database is comprised of the students who participated in the SAT program and reported to graduate from high school in 2006 and includes student responses to the questionnaire on their demographic and academic background, as well as higher education preferences, completed at the time of SAT registration.
2 The QED database includes public and private school data from the Common Core of Data from the National Center for Education Statistics (e.g., percent eligible for free lunch, high school size, etc.).
3 The Annual Survey of Colleges is a yearly College Board survey of colleges, universities, vocational/technical, and graduate schools, the objective of which is to obtain information that is important for potential students. It covers both the information that potential students want to know and the information that the institutions feel that students should know, such as admission requirements, the number of applicants and admitted students, the most popular majors on campus, etc.

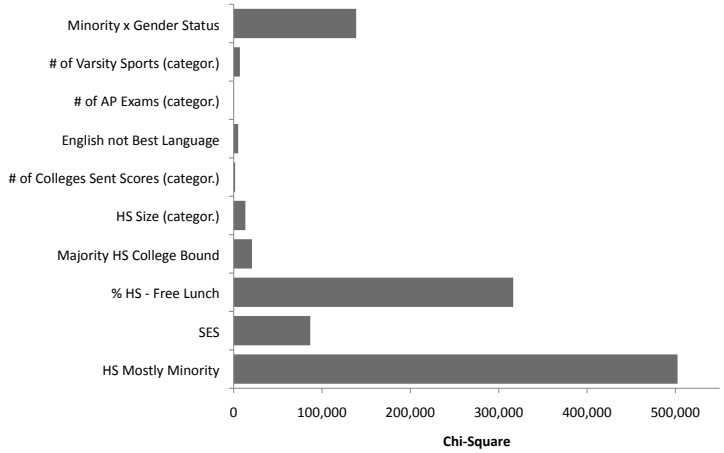
Appendix
Variable Importance Plots

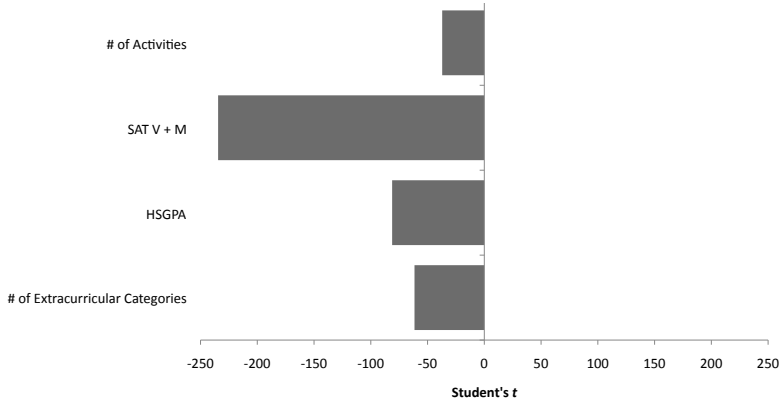
Cluster 1: Privileged High Achievers/Athletes



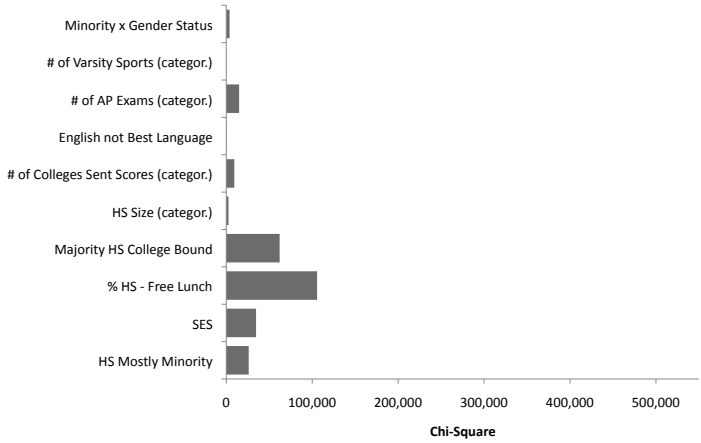


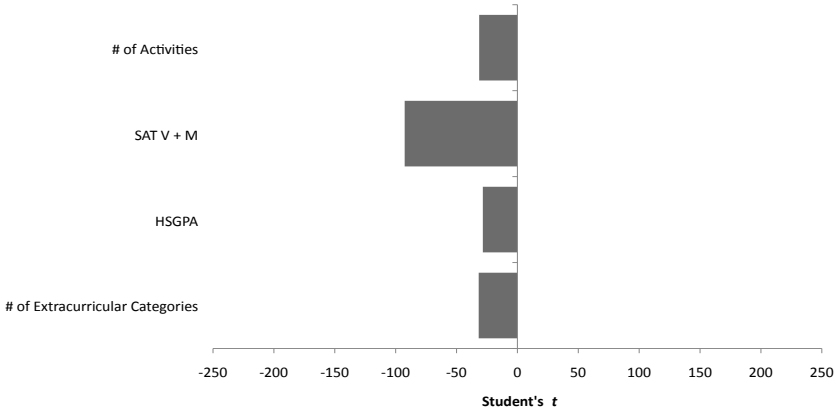
Cluster 2: Disadvantaged Students



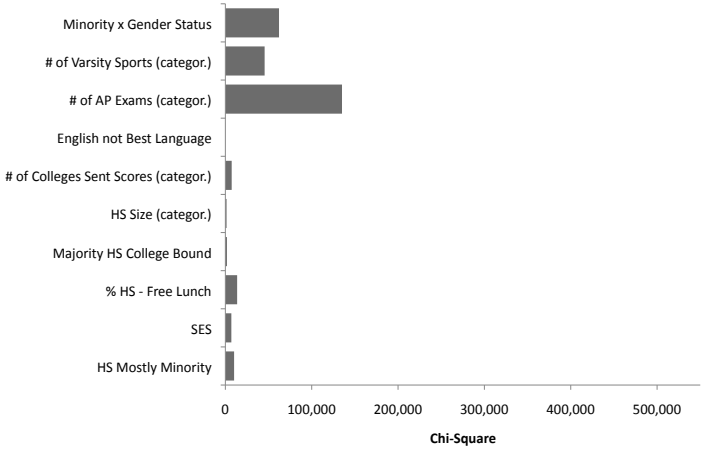


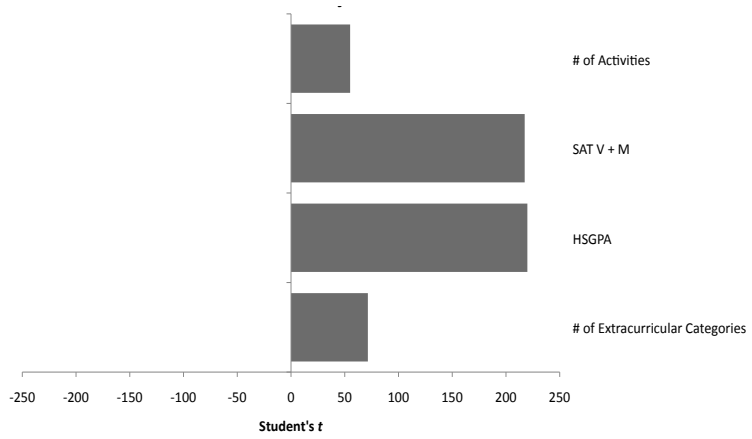
Cluster 3: Average Students Needing More Guidance





Cluster 4: Mostly Female Academics





Cluster 5: Privileged Low Achievers