8-1-2021

# Academic and Demographic Cluster Analysis of Engineering Student Success

Farshid Marbouti
*San Jose State University*, farshid.marbouti@sjsu.edu

Jale Ulas
*San Jose State University*

Ching Ho Wang
*San Jose State University*

# Academic and Demographic Cluster Analysis of Engineering Student Success

Farshid Marbouti, Jale Ulas, and Ching-Ho Wang

***Abstract*—*Contribution:* This article uses student semester grade point average (GPA) as a measure of student success to take into account the temporal effects in student success. The findings highlight the student performance based on their demographic status and use of university resources such as financial aid. College campuses should not only increase current resources but also raise awareness of current resources and make them more accessible (e.g., easier to apply or automatic applications). This is especially important for some demographics such as Hispanic first-generation students.**

***Background:* Higher education institutions are facing retention and graduation problems. One way to improve this is by understanding why students are not academically successful.**

***Research Questions:* In this study, demographic information and past academic records were analyzed to understand patterns of student success.**

***Methodology:* A cluster analysis was conducted to understand groups of students based on academic performance and demographic information. Examples of these factors are enrollment status, financial status, first-generation status, housing status, and transfer status. For the purpose of getting more accurate results, the students were separated into two different groups according to their admission status: 1) freshman and 2) transfer.**

***Findings:* The results indicate Hispanic, first-generation, low-income students are not likely to apply for financial aid although they are eligible. They have lower GPA and take fewer units per semester than other students. This can cause delayed graduation and accumulating more debt.**

***Index Terms*—Clustering methods, data mining, engineering education, machine learning, student success.**

## I. INTRODUCTION

HIGHER education is becoming more significant overtime as more people begin pursuing a bachelor's or even masters' degree in their lifetimes. Moreover, one of the America's goals is to achieve the highest proportion of college graduates in the world [1]. However, institutions are facing the problems with students' academic retention and graduation. Within these challenges, educational data mining has recently become a popular research topic. Institutions are interested in figuring out the potency of data to predict students' academic performance. In order to achieve a better educational performance, universities are interested in graduating students based on a plan. Most students enter college with a four-year graduation plan. However, the plan may be extended due to several reasons, such as a change of major, academic suspension, or poor learning performance [2]. In this situation, students may have less work experience compared to those who will graduate on time, and they may need to pay more tuition in their academic life. Therefore, understanding the student learning performance could help the colleges in assisting their students to academically succeed.

Manuscript received February 6, 2020; revised September 19, 2020; accepted November 1, 2020. Date of publication December 17, 2020; date of current version August 4, 2021. This work was supported in part by the National Science Foundation under Grant IUSE 1827406. *(Corresponding author: Farshid Marbouti.)*

The authors are with the General Engineering Department, San Jose State University, San Jose, CA 95112 USA (e-mail: farshid.marbouti@sjsu.edu; jaleulas@gmail.com; ching-ho.wang@sjsu.edu).

Digital Object Identifier 10.1109/TE.2020.3036824

## II. LITERATURE REVIEW

To perform this study, the authors identified the factors related to the university's engineering context and investigated them in terms of student success. These factors include ethnicity, housing, financial aid, first-generation status, transfer status, and enrollment status (full or part time). In this section, the university's context and the previous research on how these factors relate to student success are reviewed. The elements of student success in the review are identified as grade point average (GPA), retention, and graduation rates.

### A. Ethnicity and Gender

In the college of engineering at San Jose State University (SJSU), 43% students identified themselves as Asian, 21% white, 21% Hispanic, 3% black, 1% pacific islander, 0.2% American Indian, and 3% unknown. The gender breakdown shows that 83% were male, 15% female, and 2% unknown. For the purpose of this study, the students who are identified as Asian, White, and Hispanic are presented since these three ethnicity groups form the majority of the student body in the college of engineering.

In general, research has identified male students to have higher retention rates when compared to their female peers [3]. The probability of persistence is lower for Latino students, and higher for Asian students compared to their White peers [4]. Furthermore, White students have higher retention rates compared to black students [5]. Female Hispanic students are at a greater risk of leaving college compared to Hispanic male or white students [6]. Males and/or nonminority students are more likely to graduate in Science, Technology, Engineering, and Mathematics (STEM)-related degrees compared to females and/or minority students [7].

### B. Housing

The city of San Jose has one of the highest costs of living in the country [8] and anecdotally, due to the high rents, it is known that many of the students are likely to live with their parents, especially if they are local. Other students may also struggle with homelessness. Fortunately, there is evidence of increased student-centered food distribution on campus, new Student Homelessness Association clubs, and providing 24 h of library access to accommodate homeless students. Current university policy mandates freshmen to live on campus during their first year if they live over 30 miles away.

Students' living arrangements have influence on retention, academic performance, and graduation rates. A study involving the housing arrangement of 103 science major students showed that retention rates increased for students who lived in close proximity with other science majors [9]. Living on campus rather than off campus increases the probability of persistence to degree completion [4], [10]. Off campus residency also negatively influences black students' academic performance and participation in advisory and tutorial programs [11]. Results of an interview of 34 first-year students revealed that students who do not live on campus found it more difficult to make friends or have a social life at the university [12]. Compatible friendships help improve retention because it creates an emotional support system.

### C. First-Generation Status

About 25% of the students in the college of engineering have parents who did not receive a college degree. As a result, these first-generation students are likely to have a lower engagement when compared to non-first-generation students. Studies showed that first-generation students have lower interactions with faculty members, contributions in class discussions, and questions asked during class [13]. While some studies show different 4th-year retention rates [14] and persistence to degree along with family income [4], first to second retention rates do not differ for first-generation and non-first-generation students [14]. A one-credit course for freshman students' parents at Wichita State University showed positive results on students' retention [15]. The goal of this course was to enable parents to give more useful advice and support to their college freshman children. Retention rates would significantly increase if students have parents who have successfully completed this course. Parents reported that having insights about the university environment and courses helped them to give advice to the students [15].

### D. Financial Status and Aid

The city of San Jose has some of the highest levels of income inequality in the country; therefore, students vary significantly in terms of financial status. One factor of poverty amongst the students is the level of food insecurity. A recent study found that 43% of students have faced food insecurity, and the students who faced food insecurity were over four times as likely to receive lower grades and drop courses. In response, there has been an increase of food banks aimed at providing food to students on campus.

Family income has been shown to influence graduation rates. Students from higher income families are more likely to graduate when compared to students from lower income families [5]. For female students, the mother's education level and family income are positively related to retention [16]. Also, while parental education is positively related to student success, divorced or separated parents are negatively related [17].

As a measure to determine economic status, need-based financial aid through the Pell grant was explored, which is a U.S. government grant based on student financial need. Financial aid has a significant impact on retention rates. Students who receive gifts, loans, or work-study income have higher retention rates; the higher the amount of the aid, the more likely the student stays in college [7], [14]. Furthermore, a grant or gift type of aid creates a positive impact on retention rates [5]. For these reasons, the student's socioeconomic status is correlated to retention [4].

## III. Research Purpose and Question

This study aims to find patterns of student success by analyzing eight years of student demographic and academic data to enable higher education institutions to improve student success. To be able to take into account temporal changes for each student, the semester GPA was used as a measure of success in the semester. This study aims to answer the following research question.

1) What are clusters and characteristics of student success groups based on their demographic and academic information?
2) What are the practical implications of these clusters?

## IV. Methods

### A. Data

The dataset used in this study consisted of 12 053 students' demographic information and academic performance. These students were enrolled in the College of Engineering from Spring 2009 to Fall 2016. Most of the students in the collected data were male—which contained 85% of the participants—and the remaining 15% were female. Age of the students ranged from 15 to 67 but the majority age was 20. The university has ethnic diversity, including 43% Asians, 21% Hispanics, and 21% Whites. A relatively high percent of students, 37%, were transfer students.

Demographic and academic information were both included in the analysis. The demographic information included gender, age, ethnicity, first-generation, financial aid eligibility, housing, and Pell eligibility. For the academic information, the data involved department, high school GPA, scholastic assessment test (SAT) or American college testing (ACT) scores, and semester GPA. However, the different methods of admission (i.e., freshman or transfer) could affect student success. Therefore, the data were separated into two different groups by students' method of entry into the university (Table I).

To better understand the immediate effects of factors in student performance and to consider the temporal effects, the semester GPA was used in this study. Each student data demonstrated the performance within multiple semesters. The datapoints included 32 958 freshman student semester data and 18 330 transfer student semester data, which had 51 288 student semester data in total.

TABLE I
DATA FOR FRESHMAN AND TRANSFER STUDENTS

| **Freshman** | **Transfer** |
|---|---|
| Ethnicity | Ethnicity |
| Gender | Gender |
| First-Generation | First-Generation |
| Pell Eligibility | Pell Eligibility |
| Housing | Housing |
| Department | Department |
| Applied for Financial Aid | Applied for Financial Aid |
| Disbursed Amount | Disbursed Amount |
| Age | Age |
| Units taken | Units taken |
| Semester GPA | Semester GPA |
| High School GPA | Transferred GPA |
| SAT/ACT Score | |

### B. Cluster Analysis

Bisecting $k$-means clustering [18] was used in this study, which is a variated clustering algorithm of the $k$-means algorithm. It combines the idea of $k$-means clustering and hierarchical clustering, which can improve the performance penalty problem caused by the random initialization of centroids in $k$-means clustering. The most different part in bisecting $k$-means is that it takes one cluster with the worst data similarity and separates it into two new different clusters in each iteration. In this situation, sum of squared errors (SSE) is a good validation method to find the data similarity in each cluster. The experiments in [18] show that the bisecting $k$-means algorithm has a better clustering performance than the $k$-means algorithm. The reason is that bisecting $k$-means focuses more on the global minima than the local minima in each cluster.

One important question in clustering is how to choose the appropriate number of clusters while running the algorithm. For the purpose of finding the best number of clusters, the results with different numbers of clusters (from 2 to 10) were compared. As the basic concept of clustering, the goal is trying to make the total within-cluster sum of squares (WSSs) as minimum as possible. For both freshman and transfer student clusters, the total WSS decreases when the number of clusters increases. However, it was difficult to say when to stop adding another cluster. Therefore, to determine the ideal number in clusters, the Elbow method [19] was used. The Elbow method looks at the total WSS and stops when it does not have a great improvement upon adding a new cluster. The slope between each iteration with a different number of clusters was calculated (Table II). The location of the knee, where the slope becomes smoother in the curve, is 5 clusters in the freshman group and 6 clusters in the transfer group (highlighted in Table II). Therefore, five clusters and six clusters were used in this project as the appropriate number of clusters for the freshman group and the transfer group, respectively.

### C. Statistical Tests

It was important to find out whether the clusters were different than each other after clustering. The analysis of variance (ANOVA) [20], an omnibus test, tested the data as a whole and indicated that the mean value of each cluster was different somewhere in the model. The ANOVA test of the five clusters in the freshman group and six clusters in the transfer group results were significant ($p < 0.001$). In this situation, the ANOVA test showed that there was an overall effect of significant differences in the clustering results. However, the test did not point out which two clusters are significantly different. In order to find out where the difference was, the *post-hoc* test was used.

TABLE II
SLOPE WSS (HIGHLIGHTED CELLS SHOW CHANGE IN THE SLOPE)

| **Number of Clusters** | **Freshman** | **Transfer** |
|---|---|---|
| 2 - 3 Clusters | 0.00892 | 0.01137 |
| 3 - 4 Clusters | 0.00648 | 0.00636 |
| 4 - 5 Clusters | 0.00475 | 0.00407 |
| 5 - 6 Clusters | **0.00149** | 0.00372 |
| 6 - 7 Clusters | 0.00380 | **0.00246** |
| 7 - 8 Clusters | 0.00170 | 0.00246 |
| 8 - 9 Clusters | 0.00203 | 0.00181 |
| 9 - 10 Clusters | 0.00188 | 0.00165 |

*Post-hoc* testing, comparing all groups against each other in the clustering results, would help verify which two clusters were significantly different. Bonferroni correction [21] was implemented to help analyze the clustering results. Bonferroni is used to prevent data from incorrectly appearing to be statistically significant. Bonferroni correction would do multiple $t$-tests with a new alpha value, 0.05 divided by the number of clusters in the group. The two clusters were considered different if the $p$-value is smaller than the new alpha value.

## V. RESULTS AND DISCUSSION

### A. Size of Clusters

In order to have a better understanding of the clusters, the size of each cluster is an important factor to interpret the data distribution. The pie charts in Fig. 1 present the size of each cluster. As described in the methods section, the data used in this study are student data across semesters. The datapoints included 51 288 student semester data in total; 32 958 freshman student semester data and 18 330 transfer student semester data. Each slice has two numbers, the first number is the total number of datapoints (i.e., student semester) in the cluster, and the second number is the percentage in the group. The results were named and sorted by the average of the semester GPA in each cluster. Thus, students in cluster 1 had the lowest average semester GPA and students in cluster 5 (for freshmen students) and cluster 6 (for transfer students) had the highest average semester GPA.

For freshman students, cluster 4 is the largest cluster containing about one third of the student semesters for this group, followed by cluster 3 and cluster 2 which each contains about a quarter of the student semesters. The smallest cluster is cluster 1 which contains only 5% of the student semesters. For transfer students, cluster 2 and cluster 3 were the smallest clusters with 5% and 6% of the student semesters, respectively. The largest cluster was cluster 6 containing about one third of the student semesters in this group. Cluster 1 and cluster 4 consisted of 17% of the student semesters, and cluster 4 contained about one third of them.

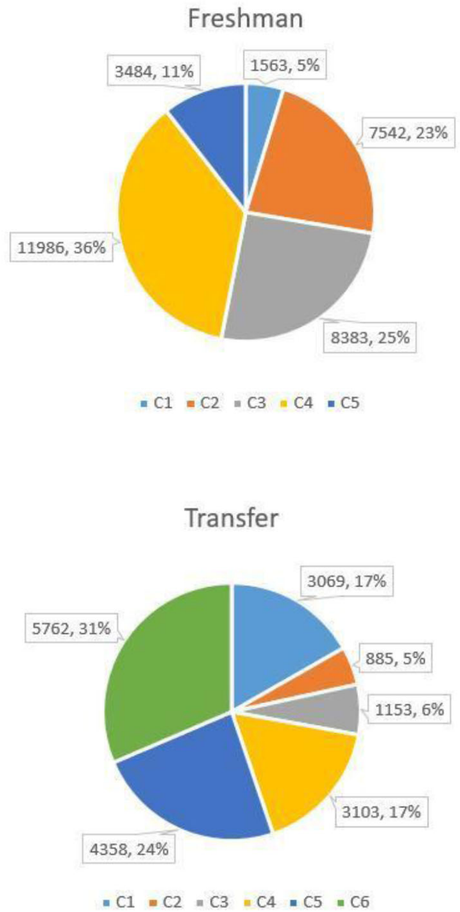


Fig. 1. Size of each cluster for freshman and transfer student groups.For each cluster, the first number is the number of student semester datapoints in each cluster, and the second number is the percentage in that group.

TABLE III
CLUSTERING RESULT IN FRESHMAN GROUP

| **Freshman** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster5 |
|---|---|---|---|---|---|
| % of students | 5% | 23% | 25% | 36% | 11% |
| ETHNICITY | Hispanic | Asian | Asian | Asian | Asian |
| GENDER | Male | Male | Male | Male | Male |
| FIRST GEN. | Yes | No | No | No | No |
| PELL ELIGIBLE | Yes | Yes | No | No | No |
| HOUSING | On Camp | Off Camp | Off Camp | Off Camp | On Camp |
| DEPART. | CompE | CompE | CompE | CompE | CompE |
| APPLY FIN. AID | No | Yes | Yes | No | No |
| AGE | 18.65 | 21.12 | 21.18 | 20.80 | 19.04 |
| UNIT TAKEN | 10.43 | 12.33 | 12.29 | 11.70 | 11.75 |
| DISBURSED AMOUNT | $0 | $13098 | $7404 | $0 | $0 |
| HIGH SCHOOL GPA | 3.39 | 3.34 | 3.37 | 3.32 | 3.38 |
| SAT/ACT SCORE | 20.21 | 20.23 | 22.15 | 22.45 | 23.25 |
| SEMESTER GPA | 2.68 | 2.72 | 2.80 | 2.84 | 2.86 |

### *B. Freshman Students*

Table III shows the clustering results with their optimal number of clusters for the freshman students. As mentioned earlier, the results were named and sorted by the average of the semester GPA in each cluster. Each column in the tables represented the majority identification of students in ethnicity, gender, first-generation status, Pell eligibility, housing, department, application for financial aid, age, units taken, disbursed amount, high school GPA, SAT/ACT score, transfer GPA, and semester GPA.

For the freshman group, male students were the majority in all clusters. Thus, gender was not a discriminating factor for freshmen. Cluster 1 was the smallest cluster and was dominantly Hispanic, while other clusters were mostly Asians. Although the Hispanic cluster students were Pell eligible, they did not apply for financial aid. Prior research suggests a variety of reasons for why students may not apply to financial aid. This might be related to lack of knowledge or late notice on their eligibility for financial aid services, difficulties in knowing how to apply them, and complexity of the application process or forms which may cause missing deadlines [22]–[27]. This cluster also had the lowest GPA among all. The average age for this group was 18.65 years, which was also lower than the other groups. Students in this group also took fewer courses/credits compared to other groups, taking about 11 credits per semester on average. This may be due to the fact that students were working to cover school costs. Longerbeam *et al.* [28] found that Latino students are more concerned about financing their education; therefore, they are more likely to work, work longer hours, and even drop out of college due to financial reasons. Thus, it is not surprising that academic progress would be slower for these students than other students.

Students in cluster 1 lived on campus. However, despite what the literature suggests, their performance was not better than other groups of students who lived off campus. This aligns with prior research which indicates that living on or off campus demonstrates a limited effect on student success [29]. Also, as stated earlier, due to the university policy on housing, the freshman students whose families live more than 30 miles away from campus have to live on campus in their first year. This means, the students in cluster 1 were mainly the first-year students who were possibly struggling with the stress of transition from high school to university [30]. Furthermore, Hispanic students may have more stress than other ethnic groups in the university [31]. These two factors might have exacerbated the problem for cluster 1 and caused a lower GPA when compared to other clusters' GPAs.

Similar to cluster 1, cluster 2 students were also Pell eligible, but they applied and received financial aid. They enrolled in more courses compared to the other groups; thus, they were more likely to complete their degree in time. Both cluster 1 and cluster 2 had lower SAT/ACT scores and university GPAs than other clusters. Even though cluster 3 students were not eligible for the Pell grant, they applied and received other

TABLE IV
CLUSTERING RESULT IN TRANSFER GROUP

| **Transfer** | Cluster1 | Cluster2 | Cluster 3 | Cluster4 | Cluster5 | Cluster6 |
|---|---|---|---|---|---|---|
| % of students | 17% | 5% | 6% | 17% | 24% | 31% |
| ETHNICITY | Asian | Asian | Hispanic | Asian | Asian | Asian |
| GENDER | Male | Male | Male | Male | Male | Male |
| FIRST GEN. | Yes | Yes | Yes | No | No | No |
| PELL ELIGIBLE | Yes | No | No | No | Yes | No |
| HOUSING | Off Camp | Off Camp | Off Camp | Off Camp | Off Camp | Off Camp |
| DEPART. | EE | Mech | EE | Mech | Mech | Mech |
| APPLY FIN. AID | Yes | Yes | No | Yes | Yes | No |
| AGE | 26.52 | 27.83 | 27.94 | 26.26 | 26.20 | 25.10 |
| UNIT TAKEN | 11.51 | 11.44 | 9.93 | 11.65 | 11.60 | 11.18 |
| DISBURSED AMOUNT | $9227 | $12646 | $0 | $9839 | $9247 | $0 |
| TRANSFER GPA | 3.04 | 3.00 | 2.97 | 3.01 | 3.05 | 3.03 |
| SEMESTER GPA | 2.81 | 2.83 | 2.84 | 2.90 | 2.90 | 2.92 |

TABLE V
*Post-Hoc* ANALYSIS RESULTS

| **Transfer** | **Freshman P-value** | **Transfer P-value** |
|---|---|---|
| C1 C2 | 0.057 | 0.442 |
| C1 C3 | **0.001** | 0.326 |
| C1 C4 | **0.001** | **0.001** |
| C1 C5 | **0.001** | **0.001** |
| C1 C6 | --- | **0.001** |
| C2 C3 | **0.001** | 0.903 |
| C2 C4 | **0.001** | **0.014** |
| C2 C5 | **0.001** | **0.011** |
| C2 C6 | --- | **0.001** |
| C3 C4 | **0.001** | **0.015** |
| C3 C5 | **0.001** | **0.010** |
| C3 C6 | --- | **0.001** |
| C4 C5 | 0.313 | 0.917 |
| C4 C6 | --- | 0.171 |
| C5 C6 | | 0.167 |

Significant ($p<0.05$) results are shown in bold.

forms of financial aid. This might have helped them to take more courses, similar to cluster 2. Clusters 2 and 3 were similar in many aspects, while the main difference between them was Pell eligibility. Financial status might be one of the main reasons that cluster 2 students performed poorer than cluster 3 students. Clusters 4 and 5 together consisted of almost half of the students. Students in these clusters were not eligible for financial aid. Their university GPA and SAT/ACT scores were highest among all freshman groups.

### C. Transfer Students

Table IV shows the clustering results with their optimal number of clusters for transfer students. The results were named and sorted by the average of the semester GPA in each cluster. Similar to freshman groups, transfer students' gender was not a discriminating factor. All clusters were predominantly male students. One cluster, cluster 3, consisted of mostly Hispanic students while other clusters consisted of mostly Asians. Cluster 3 consisted of Hispanic first-generation students who did not apply for financial aid. Similar to cluster 1 in freshman clusters, this might be related to lack of knowledge or late notice on their eligibility for financial aid services, difficulties in knowing how to apply for them, and complexity of the application process or forms, which may cause missing deadlines [22]–[27]. Furthermore, transfer students in cluster 3 took fewer courses/credits compared to other groups. Students in cluster 3 were older than other transfer students. A combination of age, which is likely correlated with more responsibility, and low-income status were perhaps the main factors for cluster 3 transfer students to perform worse than clusters 4–6.

With the exception of cluster 3, all other clusters were mostly Asians. Students in these clusters also took 11–12 units per semester, which can help them to progress and graduate faster than students in cluster 3. Majority of students in all clusters lived off-campus. Thus, unlike freshmen students, housing status was not a discriminating factor for transfer students.

### D. Cluster Comparison

Table V shows the results of the *post-hoc* analysis in the freshman and transfer groups based on their semester GPA. In the freshman group, according to the *post-hoc* test, all clusters were different, except cluster 1 to cluster 2 and cluster 4 to cluster 5. In the transfer group, clusters 1–3 were different than clusters 4–6. These results separate the freshman students in at least three different performance groups and transfer students into two performance groups. In both cases, Hispanic students who did not apply for financial aid and enrolled in fewer units were in the low performance groups. Most students in these clusters were also first-generation students. This situation points to the lack of cultural capital needed for these students to navigate the university life [32], which means they need more support and mentoring from faculty and staff during this time to be academically successful [33]–[36].

## VI. CONCLUSION

This study utilized a cluster analysis to analyze student success patterns by using their demographic information and academic performance. In order to achieve better results, students were separated into two different groups by their method of admission (i.e., freshman or transfer). Bisecting *k*-means was used to cluster the students into several groups depending on the optimal number of clusters found by the Elbow method. The ANOVA test and *post-hoc* analysis helped to find whether the clusters were significantly different or not. The granularity of our analysis by investigating semester GPA rather than the cumulative GPA helped us understand the influence of temporal factors on student performance on a more detailed level.

The results highlight the importance of not only providing resources on campus to help students, but making these resources accessible to the students especially for the ones who need it the most. For example, although they are eligible, low-income first-generation Hispanic students are not likely to apply for financial aid. For both freshmen and transfer students, Hispanic first-generation students have a lower GPA compared to other students. In addition, these students take fewer units per semester than others, which causes slower academic progress and later graduation. This exacerbates their financial situation by paying more tuition and fees, accumulating more student debt, and delayed post-graduation higher income. Making an effort to reach this portion of students, or making it easier for them to apply for financial aid may help them to be academically more successful and graduate sooner.

While cluster analysis of students based on the factors found in the literature is a useful way to understand student success patterns, helping students to be more successful requires understanding the underlying problems. The next step of this research is to interview with students to understand the underlying reasons behind the patterns seen in this article. For example, there may be various reasons for students not applying for financial aid. Some students may not be aware of the opportunity, others may have simply missed the deadline, or have chosen not to apply due to the complexity of the application processes. Deeper dive into the reasons help universities understand and act to help students be more successful.

## REFERENCES


[1] B. Cook and T. Hartle, *First in the World by 2020: What Will it Take. The Presidency* Washington, DC, USA: Amer. Council Educ., 2011, pp. 1–5.

[2] A. Cherif, G. Adams, F. Movahedzadeh, M. Martyn, and J. Dunning, "Why do students fail? Faculty's perspective," in *Proc. Higher Learn. Commission Conf. Qual. Higher Educ.*, Chicago, IL, USA, Apr. 2013.

[3] A. N. Avakian, "Race and sex differences in student retention at an urban university," *Coll. Univ.*, vol. 57, no. 2, pp. 160–165, 1982.

[4] L. Oseguera and B. S. Rhee, "The influence of institutional retention climates on student persistence to degree completion: A multilevel approach," *Res. High. Educ.*, vol. 50, no. 6, pp. 546–569, Sep. 2009.

[5] A. A. Odutola. (1983). *A Longitudinal Study of the Effects of Academic, Demographic, and Financial Aid Factors on Retention for the Freshman Class of 1974 at the Florida State University*. [Online]. Available: http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED230132

[6] B. W. Pidcock, J. L. Fischer, and J. Munsch, "Family, personality, and social risk factors impacting the retention rates of first-year hispanic and Anglo college students," *Adolescence*, vol. 36, no. 144, pp. 803–818, 2001.

[7] D. F. Whalen and M. C. Shelley, II, "Academic success for STEM and non-STEM majors," *J. STEM Educ. Innovat. Res.*, vol. 11, no. 1, pp. 45–60, 2010.

[8] (2020). *Cost of Living in San Jose*. [Online]. Available: https://www.numbeo.com/cost-of-living/in/San-Jose

[9] J. D. Chapple, "Freshmen housing assignments: A road to student retention," *J. Coll. Admissions*, vol. 1, no. 102, pp. 27–28, 1984.

[10] P. A. Murtaugh, L. D. Burns, and J. Schuster, "Predicting the retention of university students," *Res. High. Educ.*, vol. 40, no. 3, pp. 355–371, 1999.

[11] H. F. Giles-Gee, "Increasing the retention of black students: A multimethod approach," *J. Coll. Student Develop.*, vol. 30, no. 3, pp. 196–200, 1989.

[12] P. Wilcox, S. Winn, and M. Fyvie-Gauld, "It was nothing to do with the university, it was just the people': The role of social support in the first-year experience of higher education," *Stud. High. Educ.*, vol. 30, no. 6, pp. 707–722, Dec. 2005.

[13] K. M. Soria and M. J. Stebleton, "First-generation students' academic engagement and retention," *Teach. High. Educ.*, vol. 17, no. 6, pp. 673–685, 2012.

[14] D. Wohlgemuth, D. Whalen, J. Sullivan, C. Nading, M. Shelley, and Y. Wang, "Financial, academic, and environmental influences on the retention and graduation of students," *J. Coll. Stud. Retention Res. Theory Pract.*, vol. 8, no. 4, pp. 457–475, 2007.

[15] W. W. Harmon and J. J. Rhatigan, "Academic course for parents of first-year students impacts favorably on student retention," *J. Freshman Year Exp.*, vol. 2, no. 1, pp. 85–95, 1990.

[16] A. D. Rayle, S. E. R. Kurpius, and P. Arredondo, "Relationship of self-beliefs, social support, and university comfort with the academic success of freshman college women," *J. Coll. Stud. Retention Res. Theory Pract.*, vol. 8, no. 3, pp. 325–343, 2007.

[17] J. L. Zheng, K. P. Saunders, M. C. Shelley, II, and D. F. Whalen, "Predictors of academic success for freshmen residence hall students," *J. Coll. Student Develop.*, vol. 43, no. 2, pp. 267–283, 2002.

[18] M. S. G. Karypis, V. Kumar, and M. Steinbach, "A comparison of document clustering techniques," in *Proc. TextMining Workshop at KDD* May 2000, pp. 1–2.

[19] P. Bholowalia and A. Kumar, "EBK-means: A clustering technique based on elbow method and *k*-means in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 17–24, 2014.

[20] H.-Y. Kim, "Analysis of variance (ANOVA) comparing means of more than two groups," *Restor. Dent. Endodont.*, vol. 39, no. 1, pp. 74–77, 2014.

[21] P. Sedgwick, "Multiple significance tests: The Bonferroni correction," *Brit. Med. J.*, vol. 344, p. e509, Jan. 2012.

[22] E. P. Bettinger, B. T. Long, P. Oreopoulos, and L. Sanbonmatsu, "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment," *Quart. J. Econ.*, vol. 127, no. 3, pp. 1205–1242, 2012.

[23] Zarate, M. Estela, and H. P. Pachon, *Perceptions of College Financial Aid Among California Latino Youth. Policy Brief*, Tomas Rivera Policy Inst., Los Angeles, CA, USA, 2006.

[24] S. Dynarski and M. Wiederspan, *Student Aid Simplification: Looking Back and Looking Ahead*, document w17834, Nat. Bureau Econ. Res., Cambridge, MA, USA, 2012.

[25] J. Scott-Clayton, *Information Constraints and Financial Aid Policy*, document w17811, Nat. Bureau Econ. Res., Cambridge, MA, USA, 2012.

[26] C. K. Carruthers and J. G. Welch, "Not whether, but where? Pell grants and college choices," *J. Public Econ.*, vol. 172, pp. 1–19, Apr. 2019.

[27] S. M. Dynarski and J. E. Scott-Clayton, *The Cost of Complexity in Federal Student Aid: Lessons From Optimal Tax Theory and Behavioral Economics*, document w12227, Nat. Bureau Econ. Res., Cambridge, MA, USA, 2006.

[28] S. D. Longerbeam, W. E. Sedlacek, and H. M. Alatorre, "In their own voices: Latino student retention," *NASPA J.*, vol. 41, no. 3, pp. 538–550, 2004.

[29] F. Marbouti, J. Thompson, and J. Ulas, "Contextual analysis of engineering student academic success at a hispanic serving institution (HSI)," in *Proc. IEEE Front. Educ. Conf. (FIE)*, San Jose, CA, USA, 2018, pp. 1–6.

[30] C. S. Hurst, L. E. Baranik, and F. Daniel, "College student stressors: A review of the qualitative research," *Stress Health*, vol. 29, no. 4, pp. 275–285, 2013.

[31] J. D. Llamas, M. L. M. Consoli, K. Hendricks, and K. Nguyen, "Latino/a freshman struggles: Effects of locus of control and social support on intragroup marginalization and distress," *J. Latina/o Psychol.*, vol. 6, no. 2, pp. 131–148, 2018.

[32] J. P. Berger, "Optimizing capital optimizing capital, social reproduction, and undergraduate persistence," in *Reworking the Student Departure Puzzle*. Nashville, TN, USA: Vanderbilt Univ. Press, 2000, pp. 95–124.

[33] V. Torres, "A mixed method study testing data-model fit of a retention model for Latino/a students at urban universities," *J. Coll. Student Develop.*, vol. 47, no. 3, pp. 299–318, 2006.

[34] V. Tinto, *Leaving College: Rethinking the Causes and Cures of Student Attrition*, S. Ellis Avenue, Ed. Chicago, IL, USA: Univ. Chicago Press, 1987.

[35] A. L. Suarez, "Forward transfer: Strengthening the educational pipeline for Latino community college students," *Community Coll. J. Res. Pract.*, vol. 27, no. 2, pp. 95–117, 2003.

[36] B. D. Cejda and J. H. Rhodes, "Through the pipeline: The role of faculty in promoting associate degree completion among Hispanic students," *Community Coll. J. Res. Pract.*, vol. 28, no. 3, pp. 249–262, 2004.