Multi-Institution Study of Student Demographics and Stickiness of Computing Majors in the USA

Leila Zahedi, Florida International University

Leila Zahedi is a Ph.D. Candidate in the School of Computing and Information Science (SCIS) at Florida International University. She has two Master's degrees in Information Systems Management and Computer Science from Yazd University and Florida International University. She received her Bachelor's degree in Computer Engineering from the University of Isfahan in 2010. Her research interests include educational data science, machine learning, and deep learning.

Mr. Hossein Ebrahiminejad, Purdue University at West Lafayette (COE)

Hossein Ebrahiminejad is a Ph.D. student in Engineering Education at Purdue University. He completed his M.S. in Biomedical Engineering at New Jersey Institute of Technology (NJIT), and his B.S. in Mechanical Engineering in Iran. His research interests include student pathways, educational policy, and quantitative research methods.

Dr. Monique S Ross, Florida International University

Monique Ross earned a doctoral degree in Engineering Education from Purdue University. She has a Bachelor's degree in Computer Engineering from Elizabethtown College, a Master's degree in Computer Science and Software Engineering from Auburn University, eleven years of experience in industry as a software engineer, and three years as a full-time faculty in the departments of computer science and engineering. Her interests focus on broadening participation in engineering through the exploration of: 1) race, gender, and identity in the engineering workplace; 2) discipline-based education research (with a focus on computer science and computer engineering courses) in order to inform pedagogical practices that garner interest and retain women and minorities in computer-related engineering fields.

Dr. Matthew W. Ohland, Purdue University at West Lafayette (COE)

Matthew W. Ohland is Associate Head and the Dale and Suzi Gallagher of Professor of Engineering Education at Purdue University. He has degrees from Swarthmore College, Rensselaer Polytechnic Institute, and the University of Florida. His research on the longitudinal study of engineering students, team assignment, peer evaluation, and active and collaborative teaching methods has been supported by the National Science Foundation and the Sloan Foundation and his team received for the best paper published in the Journal of Engineering Education in 2008, 2011, and 2019 and from the IEEE Transactions on Education in 2011 and 2015. Dr. Ohland is an ABET Program Evaluator for ASEE. He was the 2002–2006 President of Tau Beta Pi and is a Fellow of the ASEE, IEEE, and AAAS.

Stephanie J. Lunn, Florida International University

Stephanie J. Lunn is a Ph.D. candidate in the School of Computing and Information Sciences at Florida International University (FIU). Her research interests span the fields of computing education, human computer interaction, data science, and machine learning. Previously, Stephanie received her B.S. and M.S. degrees in Neuroscience from the University of Miami, in addition to a B.S. degree in Computer Science from FIU.

Multi-Institution Study of Student Demographics and Stickiness of Computing Majors in the USA

Abstract

Retention and graduation rates in science, technology, engineering, and mathematics (STEM) careers are a worldwide concern because of the shortage of professionals in STEM fields. While there is a high need for computer professionals in industry, enrollment in computing programs has not kept pace with that demand. This is further exacerbated when the data is disaggregated on the basis of race and gender. Exploring patterns regarding race/ethnicity and gender can help education researchers and the computing community to reveal the hidden stories that help them provide guidelines, strategies, and/or mechanisms that lead to enhancing the persistence of underrepresented minority students in these fields. This study was conducted using a subset of a longitudinal database - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) to determine the computing "stickiness" of students in computing fields across multiple U.S. institutions. For the purpose of this study, computing stickiness is defined as the likelihood of graduation (the fraction that "stick" to the program or persist) for students who came into contact with a computing program. Contact is considered any student which declared a major or took a course in a computing discipline, at any time during their studies. Findings confirm variations in disciplinary stickiness by race/ethnicity and gender in computing majors. Results show that not only do White/Asian students dominate the enrollment in these disciplines, but they have the highest stickiness. That means that not only are Black /Latinx students less attracted to these majors but also that when they do explore these majors, they choose not to stay.

Introduction

Retention and graduation rates in science, technology, engineering, and mathematics (STEM) careers are a worldwide concern [1], which has led to a shortage of professionals in STEM fields. Additionally, according to the Bureau of Labor Statistics, computer science (CS) is the only STEM field where there are more jobs available relative to the amount of graduating students. It has also been reported that computing occupations are projected to increase, to nearly half a million new jobs; which is by far more than any other group in STEM (U.S. Bureau of Labor Statistics). More importantly, the disparate representation is of concern because it is socially just that all people have equal access to engineering studies and careers.

The lack of diversity in the tech industry is a widely remarked phenomenon. On one hand, there is a lack of diversity in industry, and the majority of tech roles are filled by White and Asian workers [2]. In addition to racial/ethnic disparities, CS has one of the most considerable gender disparities of all STEM fields, with significantly less women, both in terms of enrollment and in terms of in the profession [3]. Although industry needs are high in computing fields, so are students' dropout rates [4]. Women drop out of science early in their careers, even if they do well in their science and math courses, in which many factors are involved [5]. The minoritization of female students and some ethnicities/races, especially Black and Hispanics, in computing fields is an important topic that has garnered attention within universities and programs (Digest of Education Statistics).

This shortage of computing professionals and the disparities between groups has made education researchers more reflective about strategies to attract and retain more students in computing fields, so as to keep pace with industry demands [6]. The persistence of students who have a contact in a given program is a promising place to consider, because it not only includes the students who matriculated in a computing discipline, but it also includes transferred students in addition to the ones who ever showed interest in that program during their studies.

Since the number of job openings in computing fields is higher than the graduation rates, and the number of underrepresented graduates (especially women) are on the decline in computing majors, a better understanding of the patterns in the existing population is required. These patterns regarding race/ethnicity and gender can help education researchers and the computing community reveal the hidden stories that enable them to provide revised guidelines, strategies, or mechanisms that will enhance the persistence of minoritized students.

In education context, different methods such as machine learning [7–9] ,stochastic process [10, 11] and matrix factorization based models [12, 13] are used to evaluate students' academic performance. Studies have shown that factors such as student behavior is an indicator of performance [14]. Student performance can be defined as a variety of approaches to measure the persistence and success of students, including graduation rates, completion in a timely manner, and/or academic performance. However, these metrics may overestimate or underestimate the persistence of a subpopulation such as transferred students or part-time students, respectively; Therefore, these populations are usually not considered in many studies, which at some point leads to a lack of understanding on the educational pathways [15].

In an academic degree program, "stickiness" measures the tendency for a program to retain students in the program until they graduate [15]. In this paper, we calculate the computing stickiness to measure the stickiness of the students in any computing program, namely computer engineering (CE), CS, information science (IS), as well as computer and information sciences and support services (CISSS). Stickiness is primarily the likelihood of graduation for students that ever showed interest in/had contact with the computing majors. Contact refers to anytime that a student may have ever declared a computing major or took a course in a computing discipline during their studies. Whereas performance could be defined based on other factors such as final grades [16], written assignments [17], the number of publications, and so on and so forth. Graduation rate, on the other hand, is the percentage of first-year students who graduate within 150% of the published time for their program. Different factors may affect the likelihood of graduation of students. Some of these factors are individual and some are environmental.

In this study, we investigated the following research questions: 1)Are there any gender differences in regards to the stickiness of students in computing programs and other engineering majors?; 2)What are the differences in computing stickiness when considering race/ethnicity and gender?

and 3)Are there differences in stickiness when comparing different computing majors?

Theoretical Framework

Alexander Astin's (1991) Input-Environmental-Output (IEO) theory is a common theoretical framework that is frequently used by different researchers in order to better understand students' retention. To influence the outcome (in this case stickiness), an influence on the student body (input) and/or environment is required. According to this model, schools (environment) with more motivated students are more likely to stay in the program until they graduate (stick to the program). In this model, Astin (1991) classifies inputs as expectations, aspirations, self-perceptions, demographic characteristics, and educational background of the students. Race, gender, high school GPA, age, and even home rurality can be considered as inputs [7, 18–21]. For the environment however, there are different factors which may impact students' retention, such as school's mission, discipline, affiliation, size, and rank. Also, schools that have supportive plans encourage students' retention and, subsequently their graduation. Therefore, it is important to explore both the input (individual) and the environment (institutional) factors, in order to better understand retention and stickiness of the students to the program. In this paper, we disaggregate the data based on the most common individual and institutional variables to have a better understanding of the "computing stickiness" of the students. Therefore, we considered race/gender and discipline as individual and institutional, respectively.

Methods

Studies of student persistence typically use cross-sectional data focusing on a cohort, which can be challenging to interpret [22]. Moreover, single-institution studies neglect the cultural differences that occur among different institutions. In this paper, we will use a longitudinal database that contains data from institutions across the U.S., to perform a quantitative analysis.

The metric of stickiness is calculated by the following formula:

$$Stickiness = \frac{\#\, of\ students\ graduating\ from\ discipline}{\#\, of\ students\ who\ had\ contact\ with\ the\ discipline} \tag{1}$$

To further describe this in our context, we are specifically considering computing stickiness, which we define as the likelihood of graduation (the fraction that "stick" to the program or persist) for students who came in contact with a computing program. Thus, we calculate computing stickiness as:

$$Computing\ Stickiness = \frac{\#\, of\ students\ graduating\ from\ computing\ majors}{\#\, of\ students\ who\ had\ contact\ with\ computing\ majors} \tag{2}$$

Additionally, we considered the students who changed their major within any of these disciplines, as sticking to the computing disciplines.

Due to the variability in time for degree completion, we also further reduced the dataset to only include results from respondents enrolled up until 2011, although the data was collected for

MIDFIELD through 2017. The rationale for the exclusion of students who graduated within six years from the end date of the study is that they might have been students who did enroll but were unable to complete their degree at the time the data was collected, which could confound the overall stickiness. The selected six-year time frame is consistent with the definition commonly used by the National Center for Educational Statistics [23], however, in order to include part-time and transfer student duration of graduation is not a matter in this study and all students regardless of their time of graduation are being considered.

In this study, we use a subset of the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) including the computing majors with sufficient participants to do the analysis in a meaningful way (more than 1000 students for each discipline). MIDFIELD includes academic and demographic data from 1988-2017, for more than one and half million undergraduate students among 19 partner institutions across the U.S. After cleaning the data, we only included the students who at some point were enrolled in one of a set of computing disciplines. This subset includes approximately 53,000 students distributed among 14 partner institutions. We disaggregated students' data based on individual (race/ethnicity and gender) and institutional (program) to calculate their stickiness. The initial results of this study are presented in the next section.

Results

As mentioned earlier, stickiness is the number of students who graduated in a program divided by the number of students who ever declared a major in that discipline. The computing stickiness rate of male and female students is shown in Figure 1. and the data markers indicate the stickiness. Overall, students who declared any of the computing majors have an approximate computing stickiness range of 31% to 49% when disaggregated by race and gender. The results also show that male students have higher stickiness than female students for each of the race/ethnicity groups, with the exception of Black male students who have a rather low computing stickiness. This is the case even though earlier studies using MIDFIELD show that female engineering students have a higher stickiness than male students [24].

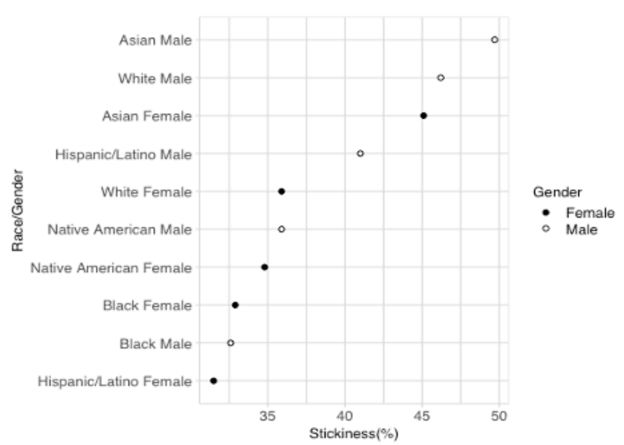

Figure 1: The stickiness of computing major students disaggregated by race and gender

As demonstrated in Figure 1, Asian students and White males tend to have a higher computing stickiness rate in comparison to other race/ethnicity groups. Asian male students have a computing stickiness of 49% and Asian female students have a computing stickiness close to 45%. Moreover, Hispanic females, Native American females, and Black male students have the lowest computing stickiness among all races/ethnicities with a computing stickiness of 31%, 32.5% ,and 33%, respectively. Among all To perform a comparison between different disciplines, we also disaggregated the data based on the discipline and race/gender, and then calculated the computing stickiness within each discipline. Variations of specific stickiness among the four computing disciplines are shown in Figure 2. To add validity to the results, stickiness is not calculated for the groups where the number of enrollments are too small (<15). the Computing stickiness of the students in each discipline is also presented in Table 1.

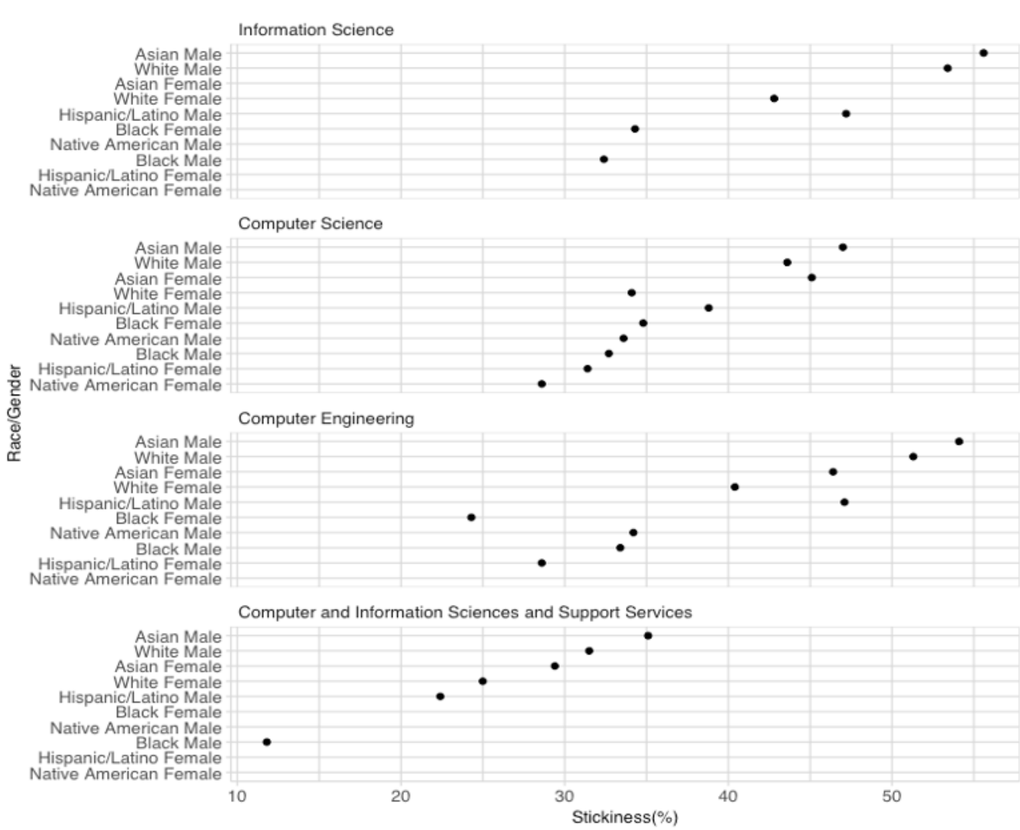

Figure 2: The stickiness of students disaggregated by discipline. No stickiness is computed for the programs with low number of enrollments (<15) of a specific race/gender.

As it is shown in the Figure 2, in all four disciplines, Asians (both genders) and White males stick more to their discipline in comparison to other races (stickiness is shown as a percentage). Also, in all majors, female students, have lower computing stickiness in comparison to male students (Figure 1), however, in general, Asian female students stick to their programs the most, followed by Asian and White male students (Figure 1). Among male students, Blacks have the lowest computing stickiness when compared to their peers. With that being said, all 4 disciplines almost follow the same pattern, however, students majoring in information science and computer engineering demonstrate more stickiness to their programs in comparison to CS and computer and information sciences disciplines. Computing stickiness of students in each discipline is presented in Table 1 (N is the number of enrollments in each program). Our results also delineate

Discipline	Stickiness (%)	N
Computer and Information Sciences and Support Services	28	1028
Computer Engineering	49	14399
Computer Science	41	33178
Information Science	48	1652

Table 1: Computing stickiness (%) of four computing majors. N is the number of students who ever declared a major in that discipline.

that unlike what is observed in other STEM majors, where female students are usually more likely to stick to their programs [15], the stickiness of female students in computing majors is less than that of their male counterparts. Results are shown in Figure 3.

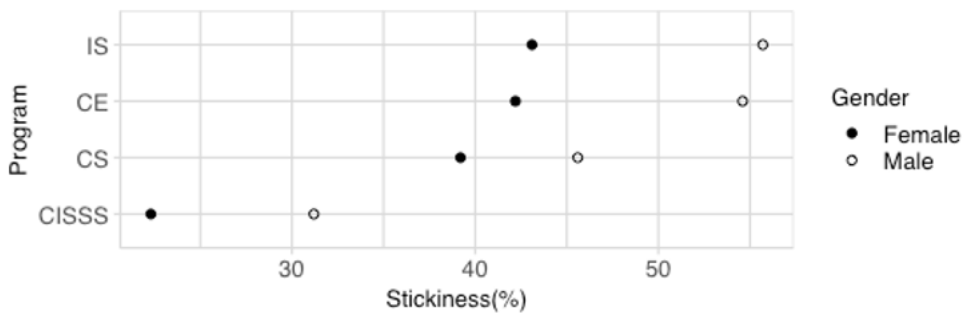

Figure 3: The stickiness of four computing disciplines disaggregated by gender.

Discussion and Conclusion

The present study includes the results of a quantitative descriptive analysis to assess similarities and differences among race/ethnicity and gender groups in computing fields. The study was designed to compare the stickiness of these students, by assessing the graduation rates relative to the total number of students who had contact with these disciplines.

Regardless of the varying persistence rates among different computing discipline groups mentioned in this paper, the results presented show that the stickiness of female students in computing majors ranges between 11 and 43 percent. As a comparison, previous research shows that the stickiness of female students in other STEM majors such as industrial engineering, mechanical engineering, chemical engineering, and electrical engineering is in a range of 45 to 55 percent [25]. Therefore, the computing stickiness of female students in computing majors is below the average stickiness of females from other STEM fields.

Furthermore, not only do White/Asian students dominate the enrollment of computing disciplines, but they have the highest computing stickiness. That means that Black/Latinx students are not only less attracted to these majors, but when they do explore these majors, they choose not to stay. This is an indication that future work should revisit research on these minorities to not only seek solutions to overcome the race/ethnicity and gender gaps, but also to investigate solutions to increase the computing stickiness for the groups who are more likely to leave.

We anticipate findings from this ongoing research to be beneficial to the computing and education community, as well as to education researchers. Computing students show different patterns of persistence from engineering students, so it is important to explore the pathways of computing students separately. This research will help these groups to better understand the relative success of computing students, which will be of interest to communities such as Grace Hopper Celebration of Women in Computing (GHC), the TAPIA Conference, the American Society of Engineering Education (ASEE), etc. Moreover, instructors and administrators can better focus on issues that inhibit the success of this subset of students. Also, school deans can benefit from learning about the specific challenges and hopefully target methods that yield increased success – regarding recruiting and graduation [23]. Finally, local and national policymakers can also utilize the findings to improve policies to reduce the minoritization of various groups.

Limitations of the Study

In order to have a better understanding of the MIDFIELD data as a whole, we limited this study to basic demographics (race/ethnicity and gender) to explore the stickiness in computing fields. In the future, we will look at individual and environmental aspects to assess more patterns that would arise from additional information. Also, in this work we looked at computing majors (IS, CS, CE, CISSS) in isolation, and not in relation to any other disciplines. Going forward, it might be interesting to explore other areas, and to compare how other disciplines are doing by demographic as well as other aspects, relative to computing.

Future Directions

This study will provide a basis for more targeted learning engagement strategies to retain and motivate more students in computing fields. The main goals for future directions are to 1) extend the analysis and applying machine learning techniques to find patterns of exclusion and attrition that prevent students from choosing computing fields as their career; 2) evaluate the academic performance of the students who did and did not stick to a computing program to expand the patterns; 3) consider school policies and investigate cultural differences among the institutions available in MIDFIELD database; and last but not least, 4) further explore these results through both quantitative and qualitative studies in order to better understand learning challenges.

References

[1] W.-C. J. Mau, “Characteristics of us students that pursued a stem major and factors that predicted their persistence in degree completion.” *Universal Journal of Educational Research*, vol. 4, no. 6, pp. 1495–1500, 2016.

[2] B. Brinkman and A. Diekman, “Applying the communal goal congruity perspective to enhance diversity and inclusion in undergraduate computing degrees,” in *Proceedings of the 47th ACM technical symposium on computing science education*, 2016, pp. 102–107.

[3] A. Nager and R. D. Atkinson, “The case for improving us computer science education,” *Available at SSRN 3066335*, 2016.

[4] P. Kinnunen, M. Butler, M. Morgan, A. Nylen, A.-K. Peters, J. Sinclair, S. Kalvala, and E. Pesonen, “Understanding initial undergraduate expectations and identity in computing studies,” *European Journal of Engineering Education*, vol. 43, no. 2, pp. 201–218, 2018.

[5] M. Klawe and N. Leveson, “Women in computing: Where are we now?” *Communications of the ACM*, vol. 38, no. 1, pp. 29–35, 1995.

[6] M. L. Zahedi, “Implications of gamification in learning environments on computer sci-ence students: A comprehensive study,” in *126th Annual Conference and Exposition of American Society for Engineering Education*, 2019.

[7] L. Zahedi, S. J. Lunn, S. Pouyanfar, M. S. Ross, and M. W. Ohland, “Leveraging machine-learning techniques to analyze computing persistence in undergraduate programs,” in *2020 ASEE Virtual Annual Conference Content Access*, no. 10.18260/1-2–34921. Virtual On line: ASEE Conferences, June 2020, https://peer.asee.org/34921.

[8] S. Boumi and A. Vela, “Application of hidden markov models to quantify the impact of enrollment patterns on student performance.” *International Educational Data Mining Society*, 2019.

[9] J. Xu, K. H. Moon, and M. Van Der Schaar, “A machine learning approach for tracking and predicting student performance in degree programs,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 742–753, 2017.

[10] S. Boumi and A. Vela, “Improving graduation rate estimates using regularly updated markov chains,” *arXiv preprint arXiv:2008.13296*, 2020.

[11] N. Thai-Nghe, L. Drumond, T. Horváth, A. Nanopoulos, and L. Schmidt-Thieme, “Matrix and tensor factorization for predicting student performance.” in *CSEDU (1)*, 2011, pp. 69–78.

[12] M. Mirzaei and S. Sahebi, “Modeling students’ behavior using sequential patterns to predict their performance,” in *International Conference on Artificial Intelligence in Education*. Springer, 2019, pp. 350–353.

[13] M. Mirzaei, S. Sahebi, and P. Brusilovsky, “Detecting trait versus performance student behavioral patterns using discriminative non-negative matrix factorization,” in *The Thirty-Third International Flairs Conference*, 2020.

[14] M. Mirzaei, S. Sahebi, and P. Brusilovsky, “Annotated examples and parameterized exercises: Analyzing students’ behavior patterns,” in *International Conference on Artificial Intelligence in Education*. Springer, 2019, pp. 308–319.

[15] M. W. Ohland, M. K. Orr, R. A. Layton, S. M. Lord, and R. A. Long, “Introducing “stickiness” as a versatile metric of engineering persistence,” in *2012 Frontiers in Education Conference Proceedings*. IEEE, 2012, pp. 1–5.

[16] R. McDaniel, R. Lindgren, and J. Friskics, “Using badges for shaping interactions in online learning environments,” in *2012 IEEE international professional communication conference*. IEEE, 2012, pp. 1–4.

[17] M. W. Ohland, C. E. Brawner, M. M. Camacho, R. A. Layton, R. A. Long, S. M. Lord, and M. H. Wasburn, “Race, gender, and measures of success in engineering education,” *Journal of Engineering Education*, vol. 100, no. 2, pp. 225–252, 2011.

[18] H. EbrahimiNejad, H. A. Al Yagoub, G. D. Ricco, M. W. Ohland, and L. Zahedi, “Pathways and outcomes of rural students in engineering,” in *2019 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2019, pp. 1–6.

[19] A. Astin, “What matters in college? four critical years revisited jossey-bass,” *Inc., Publishers: San Francisco*, 1993.

[20] E. T. Pascarella and P. T. Terenzini, *How college affects students: Findings and insights from twenty years of research.* ERIC, 1991.

[21] V. Tinto, *Leaving college: Rethinking the causes and cures of student attrition.* ERIC, 1987.

[22] A. DomíNguez, J. Saenz-De-Navarrete, L. De-Marcos, L. FernáNdez-Sanz, C. PagéS, and J.-J. MartíNez-HerráIz, “Gamifying learning experiences: Practical implications and outcomes,” *Computers & education*, vol. 63, pp. 380–392, 2013.

[23] U. D. of Education, "The integrated postsecondary education data system (ipeds) glossary. graduation rates (gr)," 2017 (accessed June 19, 2019). [Online]. Available: http://nces.ed.gov/ipeds/glossary/index.asp?id=812

[24] S. M. Lord, R. A. Layton, M. W. Ohland, and M. K. Orr, "Student demographics and outcomes in electrical and mechanical engineering," in *2013 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2013, pp. 57–63.

[25] S. M. Lord, R. A. Layton, and M. W. Ohland, "Multi-institution study of student demographics and outcomes in electrical and computer engineering in the usa," *IEEE Transactions on Education*, vol. 58, no. 3, pp. 141–150, 2014.