Application of Hidden Markov Models to quantify the impact of enrollment patterns on student performance

Shahab Boumi
University of Central Florida
4000 Central Florida Blvd
Orlando, United States
sh.boumi@knights.ucf.edu

Adan Vela
University of Central Florida
4000 Central Florida Blvd
Orlando, United States
adan.vela@ucf.edu

ABSTRACT
Simplified categorizations have often led to college students being labeled as full-time or part-time students. However, at many universities student enrollment patterns can be much more complicated, as it is not uncommon for students to alternate between full-time and part-time enrollment each semester based on finances, scheduling, or family needs. While prior research has established that full-time students maintain better outcomes than their part-time counterparts, little study has examined the impact of mixed enrollment patterns on academic outcomes. In this paper, we apply a Hidden Markov Model to identify students' enrollment strategies according to three different categories: part-time, full-time, and mixed enrollment. According to the enrollment classification we investigate and compare the academic performance outcomes of each group. Analysis of data collected from the University of Central Florida from 2008 to 2017 indicates that mixed enrollment students are closer in performance to full-time students, than part-time students. More importantly, during their part-time semesters, mixed-enrollment students significantly outperform part-time students. Such a finding suggests that increased engagement through the occasional full-time enrollment leads to better overall outcomes.

Keywords
Hidden Markov model, student enrollment mode, academic outcomes

1. INTRODUCTION
In practice, either through choice or necessity [10, 15, 5, 11, 13], students engage in a variety of enrollment patterns over their academic career that includes full-time and part-time enrollment, or halting [13]. Based on a survey conducted at 253 academic institutions, only 18% of students maintain full-time status during all semesters they are enrolled, while 29% of students maintain part-time enrollment over their whole academic career. Meanwhile, the majority of students, 59%, change their enrollment status between part-time and full-time at least once during their studies [1].

To date, part-time enrollment status has been indicated as risk-factor to student success. Feldman [8] shows that on average, at the end of the first academic year, full-time college students have higher retention rates and GPAs when compared to the part-time students. In another study, Pelkey [14] analyzed how race, age, enrollment status, GPA and financial aid can impact a student's persistence. Their analysis indicated that GPA and enrollment status, have the highest impact on persistence at college. Not only is enrollment status a factor but so is course-load, as demonstrated in [6], students with more credits during their first semester are more likely to complete their credits and degrees.

Despite it's perceived importance to student success there is no clear definition of what it means to be a *part-time* student or *full-time* student outside the ephemeral academic label. Given that the majority of students alternate between both enrollment statuses, it appears to be overly simplistic to group students together for analysis based on their enrollment status during a single semester; there is likely value in understanding more complex enrollment dynamics, and it's potential value in understanding student outcome. This assertion is supported by a 2015 nation-wide study indicating that student success can be found through mixed enrollment strategies [16] – the authors report that non-first-time-in-college students that attend college utilizing a combination of part-time and full-time enrollment are less likely to drop out and more likely to complete degrees when compared to full-time students.

In this study, we seek to find a more comprehensive means of identifying and clustering students with regards to their enrollment strategy (e.g. part-time, full-time, etc). Unlike a single-period model in which the students' strategy is equivalent to the observed student status (part-time or full-time), we make use of a multi-period dynamic approach using the Hidden Markov Models. Through application of the model we are able to provide a richer understanding of enrollment strategies, by extending our traditional notions to include not only full-time and part-time enrollment strategies, but also a mixed enrollment strategy. Students who use a mixed enrollment strategy regularly alternate between full-time and part-time status. After categorizing students into three groups of full-time, part-time and mixed enrolment strategy, we examine the student outcomes such as

Shahab Boumi and Adan Vela "Application of Hidden Markov Models to quantify the impact of enrollment patterns on student performance" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 264 - 269

Stu. #	Enroll. Status	Enroll. Strategy
1	F,P,F,F,F,F	F,F,F,F,F,F
2	F,P,F,P,F,P	M,M,M,M,M,M
3	P,F,P,P,F,P,P	P,P,P,P,P,P
4	P,F,F,P,P,P,P	M,M,M,P,P,P,P
Legend	FT=F, PT=P	FES=F, MES=M, PES=P

Table 1: Example enrollment status' over academic career and corresponding enrollment strategies

GPA and graduation rate associated with each strategy.

2. PROBLEM STATEMENT

We consider the problem of classifying students according to their enrollment strategy as opposed to their enrollment status during any given semester. For many students the distinction between enrollment strategy and actual enrollment is minor. At the University of Central Florida a students is considered as full time student in a given semester if he or she takes more than 12 credits in that semester. For approximately 35% of the student-body at the University of Central Florida, their enrollment status is consistently full-time throughout their academic career, meaning they employ a strategy of enrolling full-time. In contrast, the case for so-called *part-time* students it is not so clear. In any given semester, about 30% of enrollments are part-time, and yet only 7% of students consistently enroll part-time over their academic career. Enrolling part-time in any given semester is not equivalent to the strategy of consistently enrolling part-time. It follows that just because a student enrolls in a single semester part-time, that does not mean they bear similarity to student's who consistently enroll part-time.

The goal of this paper is to recognize and report the distinction between a student's enrollment strategy and enrollment status, and to find a more meaningful way to classify students over their academic career. More specifically, this paper develops a model that takes as its input a sequence of enrollment statuses and returns a sequence of estimated strategies applied over the same time-frame. In recognition that students apply a greater diversity of strategies then just a full-time enrollment strategy (FES) or part-time enrollment strategy (PES), we introduce the notion of a mixed enrollment strategy (MES). For a mixed enrollment strategy, students alternate between part-time and full-time enrollments. Table 1 provides examples of the enrollment status of four different students over their academic career along with the corresponding enrollment strategies. For example, enrollment strategies for student number 1 through number 3 are FES, MES, and PES respectively.

3. METHODOLOGY

In this study, we generate and apply a Hidden Markov Model (HMM) to identify students' enrollment strategy, and to characterize the impact of enrollment strategy on student outcomes. The use of HMM is not new to educational data mining and modeling. Previously it have been used to investigate students' sequential behaviors, decision-making, and performance [2, 3, 4, 9, 12, 7]. As an example Falakmasir et al. [7], classified students into low-performing and high-performing groups and applied and trained two hidden Markov models for each group separately. For each HMM,

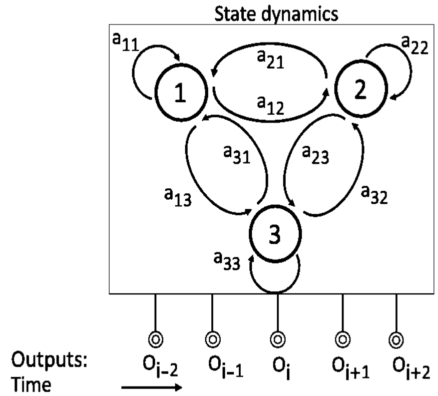

Figure 1: Representation of a simple Hidden Markov Model

they used forward algorithms to compute log-likelihoods for the observation sequences. They continued by applying a linear regression model to explain the difference between the computed log-likelihoods so as to predict post-test scores for the low-performing and high-performing students. Other papers have used HMMs in order to model sequential student behavior. Beal et al [4] colleagues modeled high school students' actions and behaviors using HMMs. By estimating HMM parameters with the Baum-Welch algorithm for each student, the authors clustered the students based on the individual transition matrices to assess differences in behavior and achievement of different clusters.

As depicted in Figure 1, similar to ordinary Markov Models, a HMM represents the dynamics of a system as it moves between operating states or modes (e.g. Modes 1, 2, and 3 in the figure). When operating within a state or mode, the system generates state-related output O_i at each time-step. Unlike Markov Models, in the case of the HMM problem the states are not always directly observable, and as such they can only be estimated by observing a sequence of outputs. For the problem under consideration here, the hidden state corresponds to the enrollment strategy of a student (e.g. full-time enrollment strategy, mixed enrollment strategy, part-time enrollment strategy), and the observations refer to the actualized enrollment in any given semester in the student academic history.

To give a formal definition of Hidden Markov models, we must begin with the following notations: $Q = \{q_1, q_2, \ldots, q_N\}$ represents the set of N possible states in the system; $A = [a_{i,j}] \in \mathcal{R}^{NxN}$ is a transition matrix, where each a_{ij} denotes the probability of transitioning from state i to j at any given time-step; $O = o_1, o_2, \ldots, o_T$ represents a sequence of observations of length T, each drawn from the set of M possible observations $V = \{v_1, v_2, \ldots, v_M\}$; and π represents the distribution of the initial state the system begins in. When a system is operating in a specific state q_i, the output o_t at any given time t is generated according to a unique probability distribution denoted as $B = b_i(o_t)$, the emission probability.

In order to generate a HMM to represent student enrollment strategies, we must learn the optimal model parameters $\lambda = (A, B, \pi)$ that reproduce known observations. The process of learning λ is based on the Baum-Welch algorithm, which is an iterative process that requires calculating the likelihood of any sequence of observations given λ, and decoding relationships between observations and hidden variables. As the model is iteratively updated, the likelihood calculations and the decoding is updated.

4. STUDENT DATA RECORDS

The study presented in this paper makes use of processed undergraduate student records collected from the University of Central Florida, a large public university in the southeast United States, between the years of 2008 to 2017. The total data-set amounts to approximately 170000 records. The data set contains a wide variety of information about students at UCF, Including but not limited to: (1) demographic Information, (2) admission information for students who have been admitted and enrolled, (3) degrees awarded (for bachelor level), (4) courses taken by student at UCF, and (5) family income. Some of the demographic information along with the fraction of students who enroll as full-time and part-time, and admission type (FTIC and transfer) are provided in Tables 2 through 5.

Table 2: Students gender distribution at UCF over 10 years

	Females	Males
Percentage	56%	44%

Table 3: Students ethnicity distribution at UCF over 10 years

	White	Hispanic	African-Am.	Other[1]
Percentage	55%	24%	11%	10%

Table 4: Students admission type distribution at UCF over 10 years

	First-Time-in-College	Transfer
Percentage	41%	59%

The processed student data includes a unique identifier, along with the student's observed academic load for semester they enrolled. Synthetic examples are shown in Table 1. For each student their enrollment sequence is ordered from their first observed enrollment to their last observed enrollment without making note of the semester or year. The data set includes both partial, halted, and graduated enrollment sequences within the indicated 10 years date-range. For the purposes of this study we restrict the problem to enrollment during Fall and Spring semesters, as such information regarding Summer enrollment is excluded when constructing the HMM. It is worth noting that the data-set includes both first-time-in-college students and transfer students.

[1]The other category includes American-Indian, Asian, Native Hawaiian, and Multi-racial ethnicity

Table 5: Enrollment type distribution for different semesters at UCF over 10 years

Semester	Full-time	Part-time
Fall	73%	27%
Spring	71%	29%
Summer	10%	90%

5. APPLYING HMM TO STUDENT DATA

In applying the HMM model to our problem, we begin by identifying the set of hidden states corresponding to three different enrollment strategies: full-time enrollment strategy (FES), part-time enrollment strategy (PES), and mixed enrollment strategy (MES). The probability a student changes his or her enrollment strategy from one semester to the next is represented using a probability transition matrix A. While the probability of observing an enrollment status while using a specific enrollment strategy is given by the emission matrix B. Finally, π is the probability distribution over the students enrolment strategy during their first enrolled semester.

Beginning with an initial guess for A, B, and π, the Baum-Welch algorithm is applied to estimate the true model parameter set (λ). Converging after 20 iterations, the following values for A, B, and π are generated:

$$A = \begin{bmatrix} 0.898 & 0.05 & 0.052 \\ 0.168 & 0.74 & 0.092 \\ 0.007 & 0.12 & 0.873 \end{bmatrix} \tag{1}$$

$$B = \begin{bmatrix} 0.974 & 0.026 \\ 0.611 & 0.389 \\ 0.061 & 0.939 \end{bmatrix} \tag{2}$$

$$\pi = \begin{bmatrix} 0.718 & 0.113 & 0.169 \end{bmatrix} \tag{3}$$

For any two subsequent semesters t and $t + 1$, the rows in the transition matrix A correspond to states FES, MES and PES at semester t, while the columns correspond to states FES, MES and PES at semester $t + 1$. Based on the estimated transition matrix A, most of the students maintain their enrollment strategy with high probabilities from one semester to the next. Reading the diagonal of the matrix, with .898 probability a student employing a FES will continue employing a FES, similarly .74 for PES and .873 for MES. This indicates that most students maintain a static enrollment strategy, even if it is a mixed one.

For emission matrix B, each rows correspond to the probability of full-time and part-time enrollment status in a semester for a given enrollment strategy. Result indicate that students employing FES register full-time with probability 0.974 and as part-time with probability 0.026. While students employing a PES only register full-time with probability 0.061 versus part-time at 0.939. Most interesting are students with MES, as their full-time and part-time enrollment is split between 0.611 and 0.389. The, initial probabilities matrix π indicates that most of the undergraduate students start their first semester with full-time enrollment strategy (with probability 0.718). Moreover, the probability of being at PES

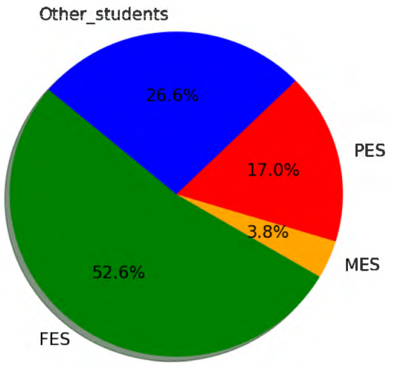

Figure 2: Distribution students' enrollment strategy

and MES at the first semester are 0.169 and 0.113 respectively.

6. ANALYSIS

After estimating model parameters, the next step is to find the strategies (hidden states) for each student in the data set at each semester with the Viterbi algorithm. Based on the estimated hidden states, students are classified into four groups, three of which corresponds to the students who maintain a consistent strategy of FES, PES or MES during their education. The last group corresponds to students who employee a combination of FES, MES and PES over their academic career. Figure 2 shows how students are distributed among these four groups.

Based on Figure 2, most of the students maintain their enrollment strategy during their educational career (Sum of green, red and yellow slices, approx. 73.4%). The most prevalent consistent enrollment strategy is FES, followed by PES and MES groups. For those students that change strategies at some point in the academic career, 73%, change from FES to PES, and 15% move from PES to MES. For virtually all cases of FES to PES, the majority of students adjust their enrollment strategy during their last two semesters. Anecdotally, it appears this shift is due to course scheduling inefficiencies and early entrance into the work place through co-op placements.

Furthermore, Table 6 represents percentage of male and female students for different enrollment strategies. The percentage of female students in FES, MES, and PES groups are 55%, 54%, and 55% respectively, which emphasizes that students enrollment strategy is independent of students gender. Table 7 indicating how students with different ethnicity are distributed among the three enrollment strategy groups. As the table shows, the ratio of students with white and Hispanic ethnicity in FES group are different to MES and PES groups. Hypothesis t-tests are conducted to assess statistical significance of these differences. For FES and PES groups, the p-value is close to 0, implying the difference in the ratios are statistically considerable. However, other complicating factors have not been considered.

Clustering of the students based on enrollment strategy (FES, PES, MES), a number of descriptive statistics are calculated.

Table 6: Female and male ratios for students with different enrollment strategies

Strategy	Female	Male	Number of students
FES	55%	45%	74571
MES	54%	46%	5321
PES	55%	45%	20466

Table 7: Ethnicity ratios for students with different enrollment strategies

Strategy	White	Hispanic	African-Am	Other
FES	56%	22%	12%	10%
MES	50%	27%	13%	10%
PES	50%	27%	13%	10%

They include average cumulative GPA, family income, 6-year graduation rate, and halting. The average GPA for each strategy cluster is shown in Figure 3. Results show that the FES group has the highest average GPA. The lowest GPA corresponds to the PES group, while the MES group's GPA lies in between.

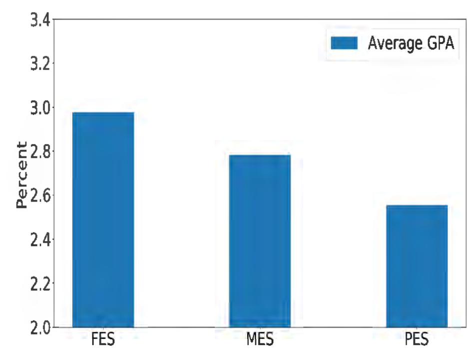

Figure 3: Average GPA for different enrollment strategies

To assess if the average GPA for each group are statistically different from the other groups, statistical hypothesis t-tests are conducted. The result shows that the p-values for all the hypothesis tests are nearly to 0, indicating that the average GPA for each group is statistically different from others. The results are summarized in Table 8.

Furthermore, inside each strategy cluster, the average GPA during full-time and part-time semesters are calculated. As indicated in Figure 4, for the FES group the average GPA for full-time and part-time semesters are 3.1 and 2.8. This indicates that students employing a full-time enrollment strategy, tend not to perform as well when registering part-time. The same conclusion is observed for students in the PES group, that is, student utilizing a part-time enrollment strategy perform better when they enroll full-time[2]. Of interest

[2]For both FES and PES, comparison of GPAs between full-time and part-time semesters through difference of means statistical tests rejects the null hypothesis that the means are equal, P=.001<.05

Table 8: Results for the GPA hypothesis tests

Pair of groups	P-value
FES and PES	0
FES and MES	$6.5e^{-84}$
MES and PES	$1.82e^{-86}$

Table 9: Results for the family income hypothesis tests

Pair of groups	P-value
FES and PES	0.7463
FES and MES	0.6518
MES and PES	0.9603

however, is that for students employing a mixed-enrollment strategy hypothesis tests indicate there is no statistical difference in means between the average GPAs of full-time and part-time semesters; in other words, the semester enrollment status for MES students does not significantly impact their GPAs. While the GPA reductions observed for students employing a full-time enrollment strategy and part-time enrollment strategy appear reasonable, the lack of GPA drop for mix enrollment strategy students is somewhat surprising, it suggests potential value in encouraging part-time students to occasionally enroll full-time.

Next, the impact of the family financial status on student enrollment strategy in each group is compared. As shown in Figure 5 the annual family income for students in all three groups of FES, MES, and PES are close to $75000. This implies that at UCF, students enrollment strategy is independent to the family income. Kolmogorov-Smirnov (K-S) test is applied in order to assess if there is statistically significant difference in family income distribution for students with different enrollment strategy. As shown in Table 9, the hypothesis test results p-values greater than 0.05 for all three group pairs of indicating no significant difference in annual family income between students with different enrollment strategies.

The next criteria for comparing student performance between the three different groups is the 6-year graduation rate, summary statistics provided in Figure 6. As the plot shows, PES group has a lower graduation rate (37.3%) when compared to MES and FES groups (with similar graduation rates of approximately 75% adn 74%). While the performance difference between PES and FES follows prior studies, it is interesting to note that employing a mix enrollment strategy does not appear to hinder graduation rates[3].

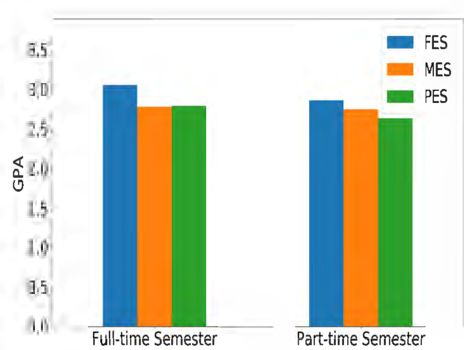

Figure 4: Average GPA for full-time and part-time semesters based on employed strategy

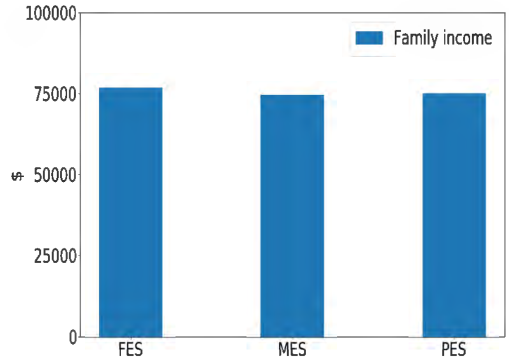

Figure 5: Annually family income for different enrollment strategies

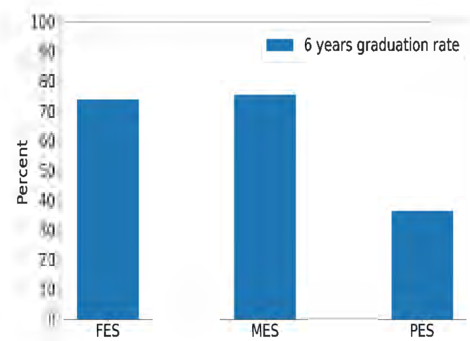

Figure 6: 6 year graduation rate for students utilizing different enrollment strategies

7. CONCLUSION

The long-term vision of this research is to help identifying strategies that engender student success. Towards that end, this paper examined different enrollment strategies students apply over their academic career. Through application of Hidden Markov Models on a large student data set, we noted three dominant consistent strategies: full-time enrollment strategy, part-time enrollment strategy, and mixed enrollment strategy. The resulting HMM and its application leads to the conclusion that most of the students have a full-time enrollment strategy. When assessing different features of the three different enrollment strategies, we observe that the average GPA for FES students is the highest, followed by MES and PES students. While graduation rates indicate that students employing the PES are more at risk of

[3]Difference of proportions fails to find a difference between FES and MES graduation rates, P=.112>.05

not graduating college. Also, financial analysis shows that there is no statistically significant difference between family income distributions for students with different enrollment strategies.

The major contributions of this research is twofold. Firstly, we provide a powerful tool for identifying students enrollment strategy as FES, PES or MES, based on their historical enrollment status. Secondly, our multi-aspect assessments on each group of students, emphasizes the vulnerability of the PES group, while encouraging university to policymakers identify such students early during their studies and help them shift towards a mixed enrollment strategy by providing them with financial, educational, and social support.

8. REFERENCES

[1] Center for community college student engagement. even one semester: Full-time enrollment and student success. the university of texas at austin, college of education, department of educational administration, program in higher education leadership. 2017.

[2] S. Abdi, H. Khosravi, and S. Sadiq. Predicting student performance: The case of combining knowledge tracing and collaborative filtering.

[3] C. Beal, S. Mitra, and P. Cohen. Modeling learning patterns of students with a tutoring system using hidden markov model, in proceedings of the 13th international conference on artificial intelligence in education (aied), r. luckin et al.(eds), marina del rey, july 2007.

[4] K. E. Boyer, R. Phillips, A. Ingram, E. Y. Ha, M. Wallis, M. Vouk, and J. Lester. Investigating the relationship between dialogue structure and tutoring effectiveness: a hidden markov modeling approach. *International Journal of Artificial Intelligence in Education*, 21(1-2):65–81, 2011.

[5] A. F. Cabrera, K. R. Burkum, S. M. La Nasa, and E. W. Bibo. Pathways to a four-year degree. Technical Report ED 482 160, 2003.

[6] R. Darolia. Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review*, 38:38–50, 2014.

[7] M. H. Falakmasir, J. P. González-Brenes, G. J. Gordon, and K. E. DiCerbo. A data-driven approach for inferring student proficiency from game activity logs. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*, pages 341–349. ACM, 2016.

[8] M. J. Feldman. Factors associated with one-year retention in a community college. *Research in Higher Education*, 34(4):503–512, 1993.

[9] D. Halpern, S. Tubridy, H. Y. Wang, C. Gasser, P. O. Popp, L. Davachi, and T. M. Gureckis. Knowledge tracing using the brain. In *Proceedings of the 11th International Conference on Educational Data Mining, EDM*, 2018.

[10] J. C. Hearn. Attendance at higher-cost colleges: Ascribed, socioeconomic, and academic influences on student enrollment patterns. *Economics of education review*, 7(1):65–76, 1988.

[11] J. C. Hearn. Emerging variations in postsecondary attendance patterns: An investigation of part-time, delayed, and nondegree enrollment. *Research in Higher Education*, 33(6):657–687, 1992.

[12] N. Hoernle, K. Gal, B. Grosz, P. Protopapas, and A. Rubin. Modeling the effects of students' interactions with immersive simulations using markov switching systems. *Proceedings of Educational Data Mining*, 2018.

[13] D. M. O'toole, L. S. Stratton, and J. N. Wetzel. A longitudinal analysis of the frequency of part-time enrollment and the persistence of students who enroll part time. *Research in Higher Education*, 44(5):519–537, 2003.

[14] D. Pelkey. Factors supporting persistence of academically underprepared community college students. 2011.

[15] S. F. Reardon, R. Baker, and D. Klasik. Race, income, and enrollment patterns in highly selective colleges, 1982-2004. Technical report, Center for Education Policy Analysis, Stanford University., 2012.

[16] I. Track. National study of non-first-time students shows full-time enrollment may not be appropriate for all, 2015.