Visualizing and predicting the path to an undergraduate physics degree at two different institutions

John Stewart ,[1,*] John Hansen ,[1] and Eric Burkholder[2]

[1]*Department of Physics and Astronomy, West Virginia University, Morgantown, West Virginia 26506, USA*
[2]*Department of Physics and Department of Chemical Engineering, Auburn University, Auburn, Alabama 36849, USA*

(Received 26 March 2022; accepted 30 August 2022; published 15 September 2022)

This study examined physics major retention to degree at two institutions with substantially different admissions selectivity. Two modes of leaving the physics major were examined: leaving college and changing to another major while staying in college. The risk of leaving college while still enrolled as a physics major was highest in the spring freshman semester. The changing major risk was substantially different between the two institutions. For the less selective institution, the students changed major at the highest rate in the fall sophomore semester. For the more selective institution, the risk of changing major was high through the first two years of college with highest risk in the fall freshman semester and the fall junior semester. Different features were important in predicting the two modes of leaving; these also differed between institutions. For the less selective institution, math readiness (being academically prepared to enroll in Calculus 1 in the fall freshman semester) was the most predictive feature for leaving the physics major while staying in college; high school GPA was the most important feature for predicting both leaving college and graduating with a physics degree. For the more selective institution, ACT composite scores were the only significant predictor of retention. The role of math readiness was dramatic at the less selective institution with 41% of students not math ready upon enrolling in college as physics majors; 59% of these students failed to enroll in the first required physics class.

DOI: 10.1103/PhysRevPhysEducRes.18.020117

I. INTRODUCTION

Since its inception, physics education research (PER) has investigated issues of critical importance to university physics departments and to the physics community in general. Much of this research has explored issues specific to the teaching and learning of physics [1]. A second more recent strand has explored another central issue, the promotion of diversity, equity, and inclusion in physics programs and physics classes [2,3]. A third issue of central and sometimes existential importance to physics departments is the retention of physics majors to degree. While the American Institute of Physics maintains detailed data on the number of physics graduates [4] as well as junior and senior undergraduate physics enrollment, little is known about how many students enter physics programs and fail to complete the degree. For many programs, because of the relation between the number of physics majors and university economic support for the department, the retention and recruitment of physics majors represents one of the most important departmental priorities. For some programs, because of state laws closing smaller academic units, retention of majors is a matter of survival [5].

A. Research questions

This work explores physics major retention at two institutions with student populations with differing levels of high school academic preparation. This work investigates factors influencing students departing physics programs through two modes: leaving college entirely and changing to a different major while staying in college.

RQ1: At which point in their undergraduate physics career are students most at risk of leaving the physics major? How does this differ by modes of leaving the major? How similar are these risks across different institutions?

RQ2: What precollege academic factors influence a student's risk of leaving the major through each mode? How similar are these factors across different institutions?

This work focuses on precollege academic factors because these factors largely control the student's progression

[*]jcstewart1@mail.wvu.edu

through the first year of college which will be shown to be key to retaining physics majors. These factors determine the first mathematics classes in which a student enrolls which largely sets the progression of future courses the student must take. Precollege factors such as ACT scores also form the primary data available to physics programs to inform the adjustment course structures and the placement of students in those structures to allow more students to succeed.

This work also introduces a number of methods to visualize physics retention which may be useful for physics departments to understand and improve the retention of majors.

B. Results of prior research

While little research into physics major persistence has been performed within PER, substantial research has investigated general college persistence and success as well as persistence in science, technology, engineering, and mathematics (STEM) majors. Within PER, a substantial research strand has investigated factors influencing student success in physics classes, a key component of college retention.

1. Physics retention

Some studies have explored the issue of retention in physics including retention of majors to physics degrees, retention within the introductory sequence, and intention to persist in physics. Aiken *et al.* used a random forest machine learning model to examine the factors most important in predicting whether a student would earn a physics degree [6]. They found that taking Modern Physics and taking an engineering class were the variables most important in the prediction.

Another study examined the retention of students in a physics course sequence, which included other scientists and engineers. Zwolak *et al.* used network analysis to determine students' social and academic integration which was used to predict if students who enrolled in the first course of an introductory physics course sequence would persist to the second course in the sequence [7]. They found that by using a student's centrality measures in the integration network, they could predict a student's persistence in the sequence at a rate of 75%. This is similar to work done by Forsman *et al.* who used complexity science in analyzing social and academic networks of students in physics classes to explain student retention [8].

A largely qualitative study by Stiles-Clark and MacLeod surveyed students after the second course of a two-course calculus-based introductory physics sequence and asked about factors that influenced the decision to continue in the physics program or a different program at the university. They found that the primary reasons for persistence were the students' interest in the subject matter, the quality of their physics instructors, and their perceived career opportunities with a physics degree [9]. The researchers noted the need for physics faculty to engage students in research-based classroom and lecture techniques, as well as the need to combat misconceptions about career opportunities for physics degrees.

2. General college retention

College retention and college persistence are major research strands in general education research. High school academic preparation is an important predictor of college success. Composite SAT scores are highly correlated with GPA in the first year of college [10]. Benchmarks for ACT composite scores have been created indicating the score required for a 50% chance of earning at least a B in introductory college classes [11]. High school GPA is more variable due to the variety of high school curriculum [12] but still a strong predictor of first year GPA [13,14] and overall college GPA [15]. One educational data mining study found that factors associated with the socioeconomic status and first generation status were highly predictive of retention after a student's third year as was a lack of academic preparedness based on ACT and COMPASS scores [16]. The COMPASS tests are administered by ACT Inc.; COMPASS scores are designed to help place students in the appropriate college classes.

The amount of literature available on the subject of college student retention is staggering. A book with a forward by Tinto [17] reviews the history of the field including differing models of student retention, economic considerations of student retention, retention in less traditional colleges such as community colleges and online colleges, as well as suggested actions to improve student retention. Although several models of student retention have been postulated, the most widely applied model was developed by Tinto [18,19]. Tinto proposed that a student's persistence depends on their skills, attributes, intentions, commitment, and interactions with students and faculty within the college. He claimed the most important factor in student retention was the student's experiences in the college, and as a student became more integrated into the academic and social communities at the college the more likely they were to persevere until graduation. Social integration refers to student-to-student interactions and extracurricular activities available at the college. Academic integration is described as the congruence of a student's abilities, skills, and interests with the academic demands of the institution and also interactions between the student and faculty and staff. In 2012, Tinto introduced a framework for institutional actions to improve student retention [20]. His framework focused on improving teaching methods and classroom interventions as this is the primary interaction between students and faculty and thus the primary way they can become integrated into the college's academic community. While improving retention is often an institutional priority, a study by Henderson *et al.* [21] showed that among physics faculty only 48% use methods that have been empirically proven to improve student learning, and only 23% used them at a high level.

3. STEM retention

The demand for employees having at least a bachelor's degree in a STEM discipline continues to grow [22]. In 2012, the President's Council of Advisors on Science and Technology [23] called for improved STEM retention to prevent an estimated one million student shortfall in STEM employment. Despite the critical need, only 40% of STEM students graduate with a STEM degree [23]. In a 2014, the U.S. Department of Education reported wide variation in the attrition rates (defined as leaving the university or the degree) of different STEM disciplines with an average rate of 48%. Attrition was highest for computer and information science majors (59%) and lowest for mathematics majors (38%) [24]. This attrition rate was lower than the attrition rate of students in the humanities or education (56%–62%) and approximately equal to the rate for students in business and social or behavioral science [24].

Many studies have investigated STEM degree retention and methods to improve retention [24–31]. In general, measures of prior high-school preparation (high school GPA and ACT or SAT scores) as well as college performance metrics such as credit completion and college GPA were important factors in predicting student retention. Other factors that have been found to be important include relationships between faculty and students [32,33], the use of learning communities [34], the implementation of a career planning seminar or career planning course [35,36], a scientific thought and methods course [37], and for engineering students their grades in introductory physics courses [38]. A study using self-reported survey data [39] found that an institution's academic environment was important for students deciding to stay in STEM: specifically features such as smaller class sizes, more integration of undergraduate student research, faculty teaching skills, and whether or not students were engaged in active learning strategies were important. A review article by Sithole *et al.* synthesizes many reforms or changes that have been suggested to improve student retention such as improved academic advising, blending courses, peer mentoring, instruction in time management and study habits, and improving high school STEM curriculum and instruction [40].

The current work is a quantitative study that focuses on the role of precollege academic factors on physics retention with the goal of identifying quantitative factors that would allow a physics department to identify incoming students likely to leave the major. As such, this review focuses on quantitative studies of retention. A large and excellent qualitative body of research exists that both examines STEM retention in general [41–43] and the retention of specific demographic groups underrepresented in physics in particular [44–47]. This body of research can provide greater context into the factors that ultimately cause a student to leave the sciences.

4. Educational data mining

With the advent of university learning management systems and increases in computing power, a very substantial branch of education research has attempted to use these large data systems and emerging computer technologies to predict both in-class success and retention to graduation. These techniques are called educational data mining (EDM) or learning analytics. Multiple reviews have summarized the efficacy of the numerous algorithms used by EDM to predict both in-class and overall student performance [48–55]. The application of data mining to the university retention problem began in the early nineties; Nandeshwar *et al.* provides a review of this work [16]. They report that college performance, high school GPA, ACT scores, and some sociofamily factors affect student retention.

These techniques have been used in multiple studies to predict student first year retention and persistence through graduation for engineering students [35,56–61]. Engineering students form the majority of the students in the introductory physics classes taken by physics majors. Machine learning has recently been applied in PER to understand student performance in physics classes [62,63].

5. Physics course success

Many PER studies have examined factors that influence student success in physics courses (generally introductory classes) using metrics such as final exam grade, course grade, and conceptual post-test scores. A certain level of success in physics classes is typically required for persistence in the major. One would also hypothesize that students who are more successful in their introductory physics courses are more likely to persist in the physics major. Much of this research has examined either instructional methods to increase success or conceptual barriers (misconceptions) that prevent success. Meltzer and Thornton provide an extensive review of research into interactive instructional methods and the efficacy of these methods [64]. Research into student misconceptions spans the history of PER [65–69]. In 2014, the National Academy of Sciences published a synthesis of results from many disciplines showing interactive instruction improved conceptual performance as well as course outcomes [70]. A further meta-analysis demonstrated the efficacy of these methods both at the college and precollege level [71].

Recent studies have examined how general high school preparation metrics (ACT and SAT scores) and prior preparation in physics measured by conceptual pretest scores affect course outcome measures including final exam grades, overall course grades, and conceptual post-test scores [72–74]. These studies show that both general high school preparation and specific preparation in physics are important in predicting student outcomes; they also show that different factors are of varying importance for different demographic groups. Studies have also investigated the details of high school physics preparation as well as noncognitive variables such as parental support as predictors of success in college physics classes [75].

Success in calculus-based introductory physics courses is also key for engineering and other science majors, who generally make up the majority of the students in an introductory physics class. A recent study by Wingate *et al.* [38] found that success in introductory physics courses was predictive of success in later engineering courses and persistence to an engineering degree. Most students who received a high grade in the introductory physics sequence continued to achieve high grades through the rest of the engineering coursework, while those who received a lower grade continued to struggle through their remaining classes.

The current work examines the retention of physics majors with a variety of methods that have yet to be applied to the problem of retention in the PER literature. This work also investigates two different institutions with student populations with differing levels of high school preparation allowing the determination of which results are institutionally independent. Many of the factors identified as important to retention in general are found to be important for physics majors; however, the analysis will also allow us to quantify when students are most at risk and to provide a general overview of retention at both institutions.

II. METHODS

A. Sample

This study investigates retention using samples drawn from two institutions. These are called Institution 1 and Institution 2 in this work.

Institution 1: Institution 1 is a large land-grant research university in the eastern United States with total undergraduate enrollment in Fall 2020 of 20 500 students. The overall demographics of the undergraduate population were 82% White, 4% Black or African American, 4% Hispanic or Latino, 4% nonresident alien, 4% two or more races, with other groups 2% or less. The ACT composite scores range was 21 to 27 for the 25th percentile to the 75th percentile of students scores [76]. This range of ACT composite scores represents a range of ACT percentile scores of 21 (59%) to 27 (85%). Thirty-one percent of undergraduate students were eligible to receive Pell grants. Pell grants are only given to students of lower socioeconomic status (SES) and are a common measure of the fraction of low SES students.

The dataset included all students who elected a physics major at any point in their undergraduate career from the Spring 2001 semester to the Fall 2019 semester. The university undergraduate population grew during this time from 16 000 in 2001. The university because more diverse over the time period; White students formed 90% of the undergraduate population in 2001. The ACT score range increased slightly over this period. The details of the filtering of this raw dataset to the analysis dataset are given in Sec. III A to show some of the complexities of working with institutional data.

Institution 1 presents both introductory physics and mathematics classes using a variety of active learning techniques. The institution has a strong commitment to instruction and supports many professional teaching professors both in mathematics and physics. These teaching professionals prioritize student success through active learning in their classrooms. Mathematics classes are presented as a combination of large lecture and discussion sections. Physics classes are presented in the large lecture format with a required co-requisite laboratory session. All freshman are required to take a freshman seminar course in their discipline.

Institution 2: Institution 2 is a large land-grant research university in the southern United States with total undergraduate enrollment in Fall 2020 of 24 500 students. The overall demographics of the undergraduate population were 81% White, 5% Black or African American, 4% Hispanic or Latino, 6% nonresident alien, 3% two or more races, with other groups 2% or less. The ACT composite scores range was 25 to 31 for the 25th percentile to the 75th percentile of students scores [76]. This range of ACT composite scores represents a range of ACT percentile scores of 25 (78%) to 31 (95%). Seventeen percent of undergraduate students were eligible to receive Pell grants.

The dataset included all undergraduate students who at some point declared a physics major at the university between the fall 2011 semester and the fall 2021 semester. The university grew somewhat in this time from 20 500 undergraduates in 2011 and became somewhat more diverse; the undergraduate population was 85% White in 2011. The ACT score range was consistent over this period.

Institution 2 presents all introductory mathematics and physics courses using primarily lecture-based pedagogy. All courses have multiple instructors teaching different sections of each course, but the style of teaching is up to the individual instructor. Approximately 20% of the physics faculty teach in specialized active learning classrooms recently built by the university. Mathematics classes are taught in small sections of roughly 30 students and meet three times per week for lecture and once per week for discussion. Physics courses are presented primarily in a large lecture format with a required co-requisite laboratory section. The institution recently implemented a required first-year seminar course, but none of the students in the data presented here will have taken the course.

This work discusses four classes commonly taken by physics majors. Calculus 1 is the single semester calculus course introducing integration and differentiation. Physics 1 is the introductory calculus-based mechanics class taken by physical scientists and engineers. Physics 2 is the introductory calculus-based electricity and magnetism course. Modern Physics is taken primarily by physics majors and covers multiple topics including relativity, quantum mechanics, and statistical mechanics.

B. Variables

This work uses a set of variables drawn from institutional records. This study used high school GPA (HSGPA), ACT or SAT mathematics percentile score (ACTM, Institution 1), ACT or SAT verbal percentile scores (ACTV Institution 1), ACT or SAT composite score (ACTC, Institution 2) and a dichotomous variable MathReady. MathReady was one if the student enrolled in Calculus 1 or a more advanced mathematics class his or her first semester of college, zero otherwise. For most students studied, taking Calculus 1 the first semester of college was required by their four-year degree plans.

C. Statistical and graphical methods

This work utilizes a number of graphical representations of retention and statistical methods to characterize retention. Each will be introduced as it is used. All analyses were performed with the R software system [77].

Sankey plots: Sankey plots give an overall visual picture of retention in physics drawing retention patterns as flows through a series of semesters. The Sankey plots were drawn with the "ggalluvial" package [78] in "R."

Survival analysis: Survival analysis is used to calculate a student's risk of leaving the physics major each semester.

Logistic regression: Logistic regression is used to predict the probability of a number of outcomes including graduation, one-year persistence, and persistence from Calculus 1 to Modern Physics.

III. RESULTS

A. Descriptive analysis

This section presents basic descriptive statistics for the various datasets used in the study. To study retention, one must restrict the temporal range of the data to allow time for persistence or graduation. Different time windows were applied for different outcomes (i.e., graduation or first-year retention) generating datasets with different overall averages. Further, not all variables were available for all students; restricting to complete records may change the overall average of some variables. The general descriptive statistics for Institution 1 and 2 are shown in Table I.

One goal of this work is to inform readers interested in replicating this work about some of the complexities they may encounter in working with institutional data. The Institution 1 dataset studied included all students who elected a physics major at any time during their undergraduate career from the spring 2001 semester to the fall 2019 semester and course taking data for the same time period, a total of $N = 659$ students. For students early in the dataset, additional course records were obtained to ensure a complete academic record was available for all students. Of these, 30 students elected the physics major prior to attending the university but were never enrolled as physics majors for a semester in which they took classes; 23 students never took a class in a semester where they were enrolled as a physics major. These students were removed leaving 606 students. An additional 20 students elected a physics major only after completing a degree in another discipline and did not complete the physics major. These students were also removed leaving 586 students. Descriptive statistics for this set of students are included in the complete dataset section of Table I (dataset 1.1).

At Institution 1, students were admitted to the university under 11 different admission codes (admit codes). The largest group was first-time freshman (FTF), 356 students, followed by students readmitted to the university, 76 students, and transfer students, 70 students. Students with admit codes suggesting they might have academic trajectories distinct from other students were removed to form the admit code dataset in Table I. Students without an admit code ($N = 7$) were removed as well as visiting students ($N = 5$), transfer students ($N = 70$), nondegree students ($N = 13$), and second degree students ($N = 18$). This resulted in a dataset with 463 records (dataset 1.4, Table I). Transfer students would be a fascinating cohort to study, but there were not enough of them in the dataset for statistical analysis.

High school academic control variables, HSGPA, ACTM, and ACTV, were not available for all students. Descriptive statistics for students for which these variables were available are shown in the HS rows of Table I. To investigate graduation or persistence to either sophomore year (1-year persistence), junior year (2-year persistence), or Modern Physics (3-year persistence), the latest records must be removed so all students have the same time to either graduate or persist; the data must be windowed. On sequence students should take Modern Physics in the spring sophomore semester; however, Modern is only offered once per year, and therefore off sequence students must often wait until their junior year to take the class. Removing these records changed the overall statistics of the sample little, as shown in Table I. A six-year window was used to investigate graduation. With this window applied, the percentage of students graduating with a physics degree (Grad Phys %), graduating with a degree in another discipline (Grad Other %), and not graduating (Not Grad %) was calculated. Each of these outcomes is approximately equally likely in both the complete dataset and the

TABLE I. Descriptive statistics applying a variety of filters for Institutions 1 and 2. Filters are abbreviated: HS (high school) for students with HSGPA and ACT or SAT scores, P1 (Physics first) for students whose first declared major is physics, FTF (first-time freshman) students admitted as first-time freshmen, Fall First, students whose first semester was the fall semester. Different windows were also applied to investigate persistence and graduation. Grad (graduation) removes the last six years of records, 1Year (one year) removes the last year of records, 2Year (two year) the last two years, and 3Year (three year) the last three years. Columns are abbreviated ACTC% (ACT composite %), ACTM% (ACT or SAT mathematics %), ACTV% (ACT or SAT verbal %), HSGPA (high school GPA), CGPA (college GPA), Grad Phys % (percentage of student graduating with a physics degree), Grad Other % (percentage of student graduating with a degree other than physics), Not Grad % (percentage of students who do not graduate with any degree), Surv Soph % (percentage of students enrolled as physics majors in their sophomore year), and Surv Junior % (percentage of students enrolled as physics majors in their junior year). Note, Grad Phys %, Grad Other %, and Not Grad % should add to one; for rows in which they do not it, is a result of the cumulative rounding of the numbers.

No.	Filter	N	Math Ready %	ACTC %	ACTM %	ACTV %	HSGPA	CGPA	Grad Phys %	Grad Other %	Not Grad %	Surv Soph %	Surv Junior %
						Institution 1—complete dataset							
1.1	None	586	63					2.99					
1.2	Grad	411	68					3.01	38	29	32		
1.3	Grad, HS	352	68		80	77	3.58	3.00	37	31	32		
						Institution 1—admit code dataset							
1.4	None	463	63					3.00					
1.5	Grad	314	68					3.00	36	30	35		
1.6	Grad, HS	296	69		80	77	3.59	2.99	36	30	34		
1.7	Grad, P1	198	68		76	74	3.51	2.91	31	31	38		
1.8	Grad, HS, P1	187	69		81	78	3.60	2.90	31	32	37		
1.9	1Year, HS, P1	247	66		79	78	3.63	2.92				64	
1.10	2Year, HS, P1	231	67		79	78	3.62	2.91				64	46
1.11	3Year, P1	227	66		75	74	3.53	2.93				64	46
1.12	Grad, P1, First Fall, FTF	143	68					2.94	34	28	38	64	43
						Institution 2—complete dataset							
2.1	None	269	87					3.16					
2.2	Grad	145	94					3.28	28	35	37		
2.3	Grad, HS	116	94	95			4.05	3.35	28	35	37		
						Institution 2—admit code dataset							
2.4	None	204	84					3.18					
2.5	Grad	81	91					3.29	22	35	43		
2.6	Grad, HS	81	91	95			4.03	3.29	22	35	43		
2.7	1Year, HS, P1	114	82	93			3.98	3.13				61	
2.8	2Year, HS, P1	100	83	93			3.98	3.13				54	34
2.9	Grad, P1, First Fall, FTF	54	91					3.18	22	51	30	63	33

admit code dataset. One-year and two-year persistence was studied by windowing the data to remove the final one year or two years of records (the codes 1Year and 2Year in Table I). For the one-year, two-year, three-year, and graduation window, the fraction of students surviving to sophomore year as physics majors was calculated (Surv. Soph. %). For the two-year, three-year, and graduation window, the fraction of students surviving to junior year as physics majors was calculated (Surv. Junior %).

At Institution 2, the admission classification was simpler; students were admitted to the university under two admissions codes: FTF and transfer. Transfer students ($N = 53$) and nondegree students ($N = 12$) were removed resulting in a dataset with 204 records (dataset 2.4, Table I)

B. Visualizing retention

College retention is intrinsically a time dependent process. One method of visualizing the transitions students make between majors and into college outcomes is a Sankey plot. The Sankey plots using the admit code filtered datasets with a graduation window (Table I, dataset 1.5;

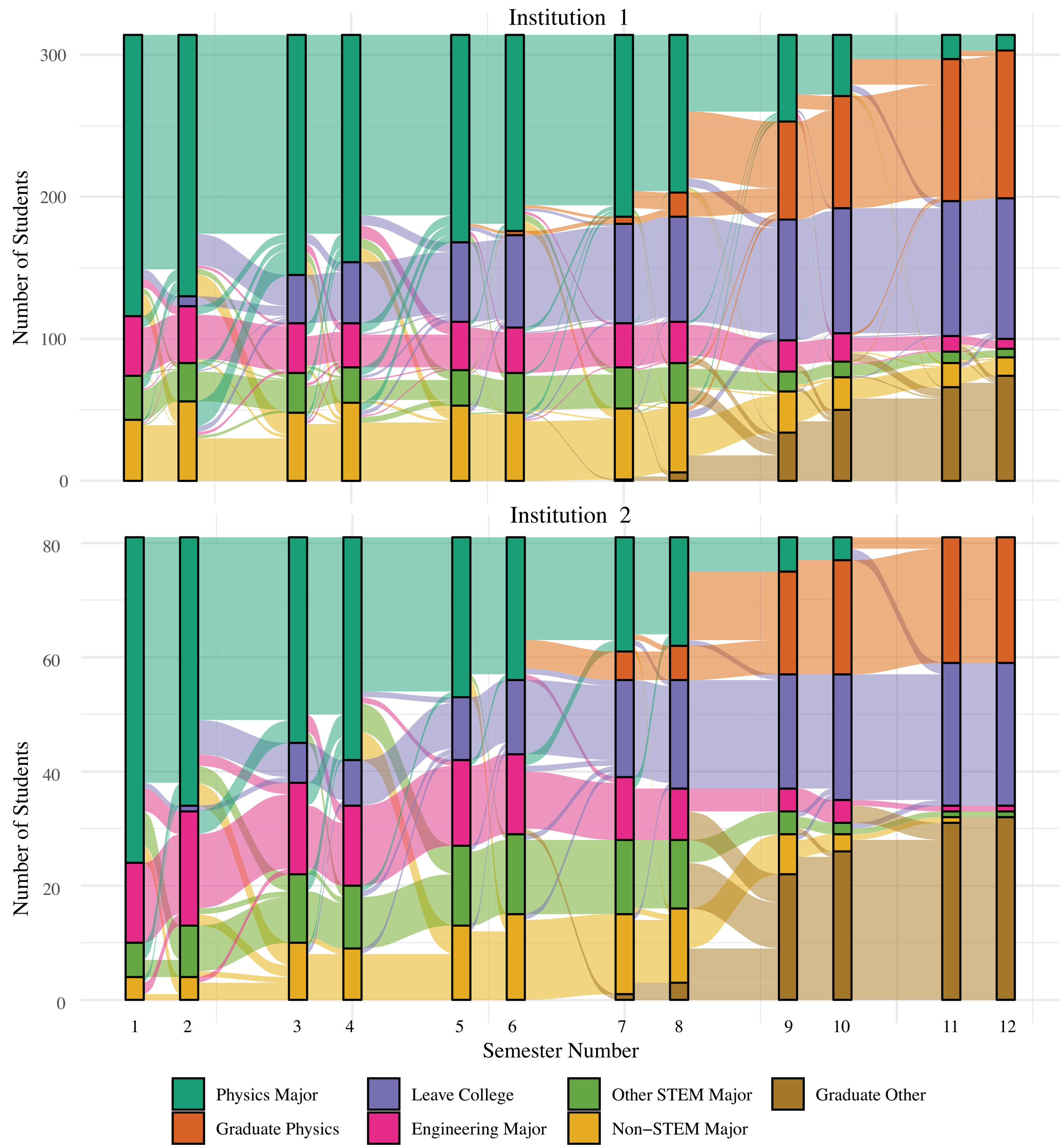

FIG. 1. Sankey plot showing major changing and graduation patterns for students who elect a physics major at any point in their undergraduate career. Each group of two bars represents an academic year; fall semesters are odd numbers, spring semesters even.

Table I, dataset 2.5) are shown in Fig. 1. Students' active majors are classified as physics, engineering, other STEM, and nonSTEM. Students' outcomes are classified as leaving college, graduate physics, and graduate other. The height of the bar in the Sankey plot represents the number of students in each category each semester. Semesters are numbered from 1 (fall freshman) to 12 (spring year 6); summer semesters have been suppressed. Two vertical bars represent an academic year. Curves are drawn showing transitions between semesters; the color of the curve shows the classification in the later semester; the width of the curve represents the number of students making the transition.

The figures for the two institutions are fairly similar. More students who eventually at some point declare a

TABLE II. Institution 1 major election sequences.

Sequence	Grad phys %	Grad other %	Not grad %
Physics	54 (54%)	0 (0%)	46 (46%)
Other—Physics	50 (76%)	1 (2%)	15 (23%)
Other—Physics—Other	0 (0%)	31 (62%)	19 (38%)
Physics—Other	0 (0%)	61 (68%)	29 (32%)
Physics—Other—Physics	8 (100%)	0 (0%)	0 (0%)

physics major, start as physics majors at Institution 2. Both institutions, despite the general difference in academic preparation measured by the ACT score range and SES measured by Pell eligibility of the undergraduate populations, see a similar fraction of students leave college without graduating. A somewhat larger fraction of students graduate with a physics degree at Institution 1 than at Institution 2, while a higher fraction of students at Institution 2 graduate with a major other than physics. These additional graduates come from other STEM disciplines; the fraction of graduates in nonSTEM disciplines is commensurate. The Graduate Other category does not disaggregate STEM and nonSTEM majors; however, examining the enrollment at the end of year four (semester 8) shows the nonSTEM major bars are fairly identical while the other STEM major bar is about twice as tall at Institution 2. Both institutions see substantial outflows from physics majors to leaving college after the second semester. Both institutions see both inflows and outflows to the major which are substantial in the first two years; these seem to continue later at Institution 1.

Tables II and III summarize the patterns observed in Fig. 1. These use the same dataset which was used to construct the Sankey plot. At Institution 1, only 100 of the 314 students are physics majors for their entire undergraduate career; these students graduate with a degree 54% of the time. Unfortunately, 46% of these students do not earn a college degree. This college graduation rate is lower than that of the 90 students who start in physics and leave the major for another degree; these students earn college degrees 68% of the time. A substantial group of students, $N = 66$, begin college in other majors and switch to physics; these students graduate with physics degrees 76% of the time and graduate college 77% of the time.

TABLE III. Institution 2 major election sequences.

Sequence	Grad phys %	Grad other %	Not grad %
Physics	10 (56%)	0 (0%)	8 (44%)
Other—Physics	10 (59%)	0 (0%)	7 (41%)
Other—Physics—Other	0 (0%)	10 (100%)	0 (0%)
Physics—Other	0 (0%)	26 (76%)	8 (24%)
Physics—Other—Physics	2 (100%)	0 (0%)	0 (0%)

One student in the "Other-Physics" pathway earned a degree in another discipline, but not physics. This student was a physics major until the end of their undergraduate career, but applied to graduate with a different major once classes were over.

For Institution 2, the results for students who stay in physics their entire undergraduate career are similar to those of Institution 1. Students who start in other majors and add the physics degree are somewhat less successful at Institution 2. Students leaving the physics major graduate from college at a high rate at both institutions. The small number of students in many of the pathways at Institution 2 means the percentages should be viewed as suggestive only.

C. Survival analysis

The time dependent nature of college retention and retention to major can be thought of the process of surviving to graduation. As such, survival analysis, a statistical analysis method originally developed to model the survival of patients with life threatening diseases, represents a promising method to model the process of successfully graduating with the physics major.

Normally, survival analysis attempts to make predictions about a continuous random variable T which represents the time a state-changing event happens (such as dying or quitting school). The variable has probability density $f(t)$ and cumulative distribution function $F(t) = \int_{-\infty}^{t} f(t)dt = P(T < t)$; $F(t)$ is the probability the event has already happened. The survival function $S(t) = 1 - F(t) = \int_{t}^{\infty} f(t)dt$ is the probability the event happens after t or the probability you have survived to t.

The hazard function $\lambda(t)$ is the probability the event happens in the range $[t, t + \Delta t]$ given the event has not already happened at t, the rate the event is happening at time t is given by

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{1}$$

Survival through college to earn a physics degree is an intrinsically discrete process because information on changing majors and leaving college only exists at the semester level. For the discrete case, Eq. (1) simplifies dramatically. For example, the leaving college hazard in semester j, λ_j^{LC}, is the ratio of the students enrolled in semester j who have left college by semester $j + 1$, $\Delta N_{j,j+1}^{LC}$, to students enrolled in semester j, N_j, given by

$$\lambda_j^{LC} = \frac{\Delta N_{j,j+1}^{LC}}{N_j}. \tag{2}$$

A similar definition can be given for the changing major hazard, λ_j^{CM}. The graduation hazard is the fraction of

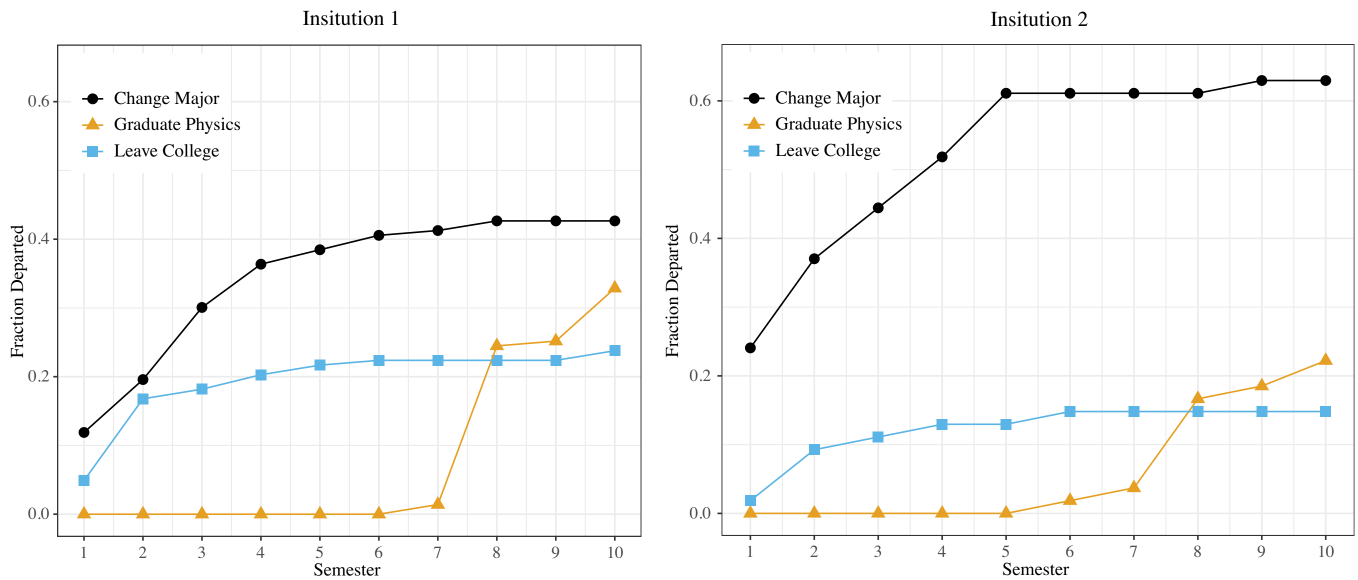

FIG. 2. Fraction departed or graduated for students entering the university declared as physics majors.

students enrolled in semester j who graduate that semester, N_j^G; $\lambda_j^G = N_j^G/N_j$.

For the survival analysis, the data were filtered to a set of maximally homogeneous students after applying a graduation window. The admit code dataset was restricted to include only students who began in the fall semester, who were admitted as first-time freshmen, and who elected physics as their first college major (Institution 1, Table I, dataset 1.12, $N = 143$; Institution 2, Table I, dataset 2.9, $N = 54$). This strong filter was necessary because students who enter in a semester other than the fall have less time to the critical first summer semester. Students who are not initially physics majors may have different course trajectories and require more time to graduate. For this analysis, three modes of leaving the physics major were considered: changing to another major while staying in college (change major), leaving college without earning a degree (leave college), and graduating with a physics degree (graduate physics). The fraction of students in this dataset that leave physics through each mode is shown in Fig. 2. The figure shows that for Institution 1 approximately twice as many students starting with a physics major leave physics by changing to a different major than those who leave physics by leaving college; for Institution 2, 4 times as many students change major as leave college. A higher fraction of physics majors leave college at the less selective Institution 1 than at Institution 2. The fraction of students leaving college is not directly comparable to the Not Grad % in Table I because the plot shows the fraction who leave college while still enrolled as physics majors. Note, these results are somewhat different than those shown in the Sankey plots. These differences are a result of the different datasets used. The students used in the survival analysis are students who have the general academic trajectory (first-time freshmen entering in fall semester) around which the undergraduate physics program was designed and are a particularly interesting subpopulation.

The hazard function for all three modes of leaving physics is shown in Fig. 3. Note, the graduation hazard (rate) is plotted against the right axis. For both Institution 1 and 2, there is a strong peak in the leaving college hazard at Semester 2. This hazard is understandable; students not thriving at college return home after their freshman year and do not return. At Institution 1, there is a peak in the change major hazard at Semester 3, the fall sophomore semester. This likely results from students returning from the summer between freshman and sophomore years and changing their major upon their return. Institution 2 has high changing major hazard throughout the first two years with peaks in Semester 1 and Semester 5. The Semester 1 peak may be the result of students enrolling in physics upon entering the university, discovering what is really involved, and electing a different major. The Semester 5 peak (fall junior year) may result from students who have completed their introductory science core having to make the decision on what major to pursue at this point. The Sankey plots show a strong exchange between physics and other STEM and nonSTEM disciplines at this point. For Institution 1, all semesters plotted in the hazard plot enroll at least 50 students. The smaller Institution 2 dataset would only retain one semester with this criteria; semesters are included that enroll at least 19 students. The students per semester is reported in the caption of Fig. 1. The small enrollment implies the Institution 2 results should be used with caution; however, the high changing major hazard through the fall junior semester is also supported by the linear fraction departed plot for this hazard mode until that time.

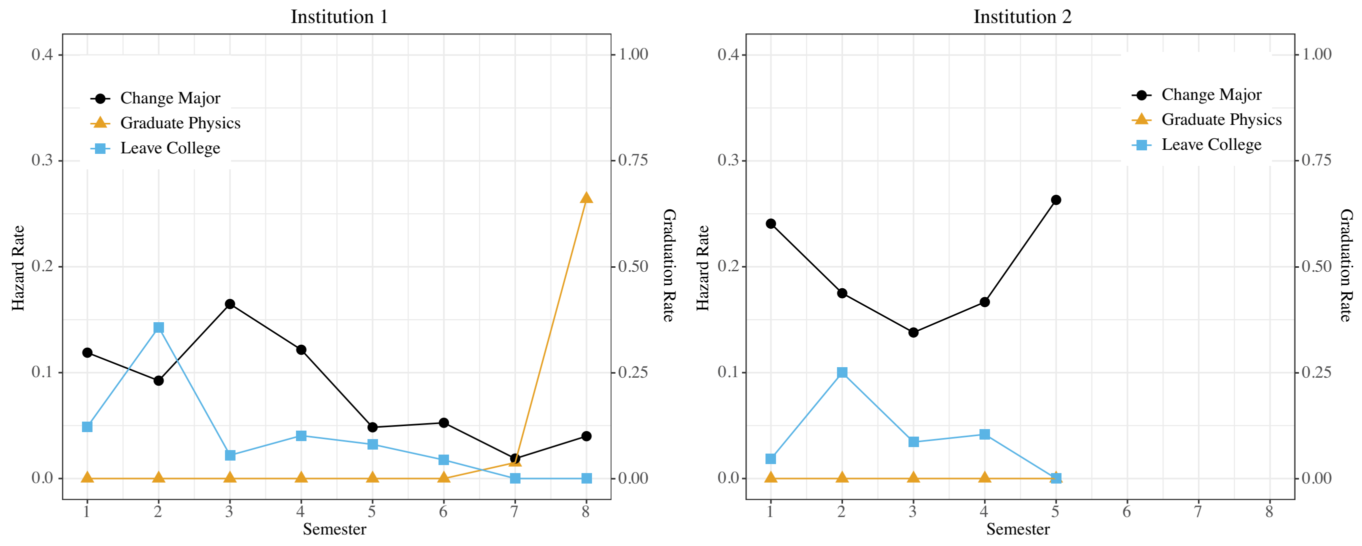

FIG. 3. Hazard functions. The graduation hazard is plotted on a different scale shown by the right vertical axis. For Institution 1, each semester plotted has at least 50 students enrolled as physics majors. Only 5 semesters could be plotted for the smaller Institution 2; these semesters enroll 54, 40, 29, 24, and 19 students, respectively.

D. Logistic regression

Logistic regression allows the modeling of how factors affect a dichotomous dependent variable. Logistic regression predicts the probability of the high level of the dichotomous variable ($Y = 1$); the variable Y is coded so the low level is zero and the high level is one. The probability that $Y = 1$ is observed for student i is modeled by the probability function $P_i(Y = 1)$. The odds of the $Y = 1$ outcome for student i is then calculated as $\text{odds}_i = P_i(Y = 1)/[1 - P_i(Y = 1)]$, the ratio of probability of $Y = 1$ being observed to the probability of $Y = 0$ being observed. The range of the odds is from 0 to ∞. To project this quantity into an unbounded range, the log-odds is calculated as $\text{log-odds}_i = \ln(\text{odds}_i)$. The log-odds is then predicted with a set of independent variables very much as a continuous dependent variable would be in linear regression (but with differing underlying statistical assumptions). For example, to predict the log-odds using two independent variables X_1 and X_2, an intercept β_0 and two slopes β_1 and β_2 are estimated as follows:

$$\text{log-odds} = \ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2. \quad (3)$$

The intercept predicts the log-odds when X_1 and X_2 are both zero. The slope β_1 is the change in log-odds for a one unit increase is X_1. Log-odds, however, is a fairly difficult quantity to interpret qualitatively. It is much more intuitive to discuss changes in the odds. To calculate the odds, both sides of Eq. (3) are exponentiated, yielding

$$\text{odds} = \frac{P(Y = 1)}{1 - P(Y = 1)} = e^{\beta_0} \cdot e^{\beta_1 X_1} \cdot e^{\beta_2 X_2}. \quad (4)$$

As such, e^{β_0} is the base odds when $X_i = 0$ and e^{β_1} multiplies this base odds when $X_1 = 1$.

Logistic regression was used to explore factors influencing persistence to the sophomore year, the junior year, and to graduation. For this analysis, the admit code dataset was filtered to retain only students electing physics as their first college major for whom high-school-level data were available; the data were then windowed for each outcome variable.

For Institution 1, this produced three datasets shown in Table I: 1-year persistence, dataset 1.9, $N = 247$; 2-year persistence, dataset 1.10, $N = 231$; graduation, dataset 1.8, $N = 187$). Table IV presents the logistic regression results for Institution 1 for a number of outcome variables: leaving college by the sophomore year, leaving college by the junior year, leaving the physics major but staying in college though the sophomore year, leaving the physics major but staying in college through the junior year, and graduating with a physics degree. These models were initially fit using HSGPA, MathReady, ACTM, and ACTV as independent variables. The full regression equation is given by

$$\text{log-odds(Outcome)} = \beta_0 + \beta_1 \times \text{HSGPA} + \beta_2 \times \text{ACTM} + \beta_3 \times \text{ACTV} + \beta_4 \times \text{MathReady}, \quad (5)$$

where β_0 is the intercept, β_i are the slopes, and outcome is one of the following: graduation in physics, leaving college by sophomore year, leaving college by junior year, leaving physics while staying in college by sophomore year, and leaving physics while staying in college by junior year.

For all models, the model using the independent variables was a statistically significant improvement over the

TABLE IV. Institution 1: Logistic regression. All regressions are significant improvements over the null model ($p < 0.001$). β is the normalized regression coefficient, SE is its standard error, z is the z score of the coefficient, p the probability a value larger than z occurred by chance, and e^{β} is the odds ratio.

Variable	β	SE	z	p	e^{β}
Leave college by sophomore year ($N = 247$)					
(Intercept)	−2.12	0.22	−9.62	0.0000	0.12
HSGPA	−0.69	0.19	−3.65	0.0003	0.50
Leave physics stay in college by sophomore year ($N = 247$)					
(Intercept)	−0.39	0.22	−1.73	0.0827	0.68
MathReady	−1.32	0.31	−4.25	0.0000	0.27
Leave college by junior year ($N = 231$)					
(Intercept)	−1.83	0.20	−8.94	0.0000	0.16
HSGPA	−0.72	0.18	−3.93	0.0001	0.49
Leave physics stay in college by junior year ($N = 231$)					
(Intercept)	0.34	0.23	1.47	0.1404	1.41
MathReady	−1.29	0.29	−4.40	0.0000	0.28
Graduate physics ($N = 187$)					
(Intercept)	−1.64	0.40	−4.14	0.0000	0.19
HSGPA	0.91	0.23	4.02	0.0001	2.49
MathReady	0.88	0.45	1.96	0.0504	2.41
Enroll Calculus 1—Pass Physics 2 as major ($N = 132$)					
(Intercept)	0.18	0.20	0.93	0.3527	1.20
HSGPA	1.13	0.24	4.78	0.0000	3.09
Enroll Calculus 1—Pass Modern as major ($N = 132$)					
(Intercept)	−0.67	0.21	−3.14	0.0017	0.51
HSGPA	1.14	0.26	4.37	0.0000	3.14

null model. For logistic regression, the null model is the model including only the intercept term. Once the full model shown in Eq. (5) was fit, it was examined for statistically insignificant independent variables. These variables were removed producing a more parsimonious model. An ANOVA test showed the model removing insignificant independent variables was not significantly less well fitting than the full model in all cases. This model is shown in Table IV. For all models except graduating with a physics degree, only one variable was retained.

For Institution 1, the results for persistence in physics were quite different than the results for persistence in college. Persistence in college while leaving the physics major was most strongly related to math readiness (being able to enroll in Calculus 1 the first semester of college). The base odds of leaving physics while staying in college (the odds, e_0^{β}, of the intercept) was reduced by a factor of 0.27 for the sophomore year and 0.28 for the junior year for math ready students. As such, being math ready decreases the odds of leaving the major by $(1/0.28 - 1) \times 100\% = 260\%$. The relation of math readiness to leaving the physics major but remaining in college is very understandable; nonmath-ready students have to take a sequence of mathematics classes, often a year and a half of mathematics classes, before ever enrolling in their first physics class. They also are very unlikely to complete their degree in four years. These factors make them very hard to retain and add financial pressures to the student to change to a less math intensive major.

The variables important in predicting whether a physics student would leave college by the sophomore or junior year were quite different; HSGPA was the most important variable. While high school classes and curricula are extremely variable, HSGPA provides a measure of how successful a student has been in the high school academic system. This success is an important indicator of whether the student will successfully navigate college. Both MathReady and HSGPA were important in predicting graduation with a physics degree (MathReady was $p = 0.0004$, above the 0.05 significant threshold). A student who graduates with a physics degree must avoid both leaving the major and leaving college, so it is reasonable that both factors are involved. Both factors have similar odds ratios in predicting graduation; math readiness increased the odds of graduating with a physics degree by $(2.41 - 1) \times 100\% = 141\%$ and a 1 standard deviation increase in HSGPA increases the odds by 149%.

The analysis for Institution 2 produced different results. The datasets used are shown in Table I: 1-year persistence, dataset 2.7, $N = 114$; 2-year persistence, dataset 2.8, $N = 100$; graduation, dataset 2.9, $N = 54$. For Institution 2, the variables HSGPA, ACTC, and MathReady were used as independent variables; the full regression equation is given by

$$\log\text{-odds(Outcome)} = \beta_0 + \beta_1 \times \text{HSGPA} + \beta_2 \times \text{ACTC} + \beta_3 \times \text{MathReady}, \quad (6)$$

where β_0 is the intercept, β_i the slopes, and outcome is one of the following: graduation in physics, leaving college by sophomore year, leaving college by junior year, leaving physics while staying in college by sophomore year, and leaving physics while staying in college by junior year. The logistic regression results for Institution 2 are presented in Table V.

The smaller sample size for Institution 2 and the more homogeneous student academic preparation as evidenced by the very high average levels of math readiness, HSGPA, and ACTC shown in Table I reduced statistical power. Many models did not contain a significant regressor. As such, we report the most significant regressor in each model. Because of low statistical power, it may be more appropriate to examine the odds ratios. Only ACTC was a significant regressor in predicting leaving the physics major while staying in college by the sophomore and junior year. Math readiness was the most important regressor in predicting leaving college by the sophomore

TABLE V. Institution 2: Logistic regression. All regressions are significant improvements over the null model ($p < 0.001$). β is the normalized regression coefficient, SE is its standard error, z is the z score of the coefficient, p the probability a value larger than z occurred by chance, and e^{β} is the odds ratio.

Variable	β	SE	z	p	e^{β}
Leave college by sophomore year ($N = 114$)					
(Intercept)	−2.94	0.39	−7.57	0.0000	0.05
MathReady	−1.18	0.67	−1.78	0.0746	0.31
Leave physics stay in college by sophomore year ($N = 114$)					
(Intercept)	−0.47	0.16	−2.89	0.0038	0.63
ACTC	−0.45	0.17	−2.71	0.0067	0.64
Leave college by junior year ($N = 100$)					
(Intercept)	−1.93	0.25	−7.60	0.0000	0.15
HSGPA	−0.44	0.24	−1.86	0.0629	0.64
Leave physics stay in college by junior year ($N = 100$)					
(Intercept)	−0.25	0.17	−1.49	0.1366	0.79
ACTC	−0.51	0.18	−2.89	0.0039	0.60
Graduate physics ($N = 54$)					
(Intercept)	−1.63	0.40	−4.03	0.0000	0.19
ACTC	0.69	0.47	1.47	0.1430	1.99
Enroll Calculus 1—pass Physics 2 as major ($N = 54$)					
(Intercept)	−1.13	0.33	−3.39	0.0007	0.32
ACTC	0.58	0.37	1.55	0.1220	1.79
Enroll Calculus 1—pass Modern as major ($N = 54$)					
(Intercept)	−1.24	0.36	−3.42	0.0006	0.29
ACTC	0.94	0.43	2.21	0.0268	2.56

year; HSGPA predicting leaving college by the junior year. Both were significant at the $p < 0.1$ level. ACTC was the most important variable in predicting physics graduation with an odds ratio commensurate to the odds ratios in the sophomore and junior datasets (the odds ratio for ACTC of not graduating is $1/1.99 = 0.5$).

E. Traversing the course network

As a student persists in college they traverse a network of required courses. For a physics major at Institution 1 or 2, the key sequence of courses early in college is Calculus 1, Physics 1, Physics 2, then Modern Physics. The logistic regression analysis was repeated to explore the factors influencing whether a student who enrolls in Calculus 1 persists to either Physics 2 or Modern Physics.

For Institution 1, HSGPA was the most important predictor of a student who enrolled in Calculus 1 passing either Physics 2 or Modern Physics as a major as shown in Table IV. A 1 standard deviation higher HSGPA increased the odds of staying a physics major through Modern by 200%. For Institution 2, ACTC was the most important predictor of a student enrolling in Calculus 1 completing Modern as a physics major as shown in Table V. For Institution 2, HSGPA was substantially higher than at Institution 1; this may have limited the variation in this variable and its predictive power.

Examining the progression of students through the network also provides additional insights. This analysis was only performed for Institution 1; the better prepared students at Institution 2 produced a much simpler network, which is not shown to prevent identifying individual student trajectories. Figure 4 shows the progression of students who enter Institution 1 declared as physics majors through Modern Physics and to graduation. For this analysis, a 3-year window was applied to the admit code filtered dataset (Table I, dataset 1.11, $N = 227$). Students first enrolling in Modern Physics or a more advanced physics class were removed (8 students); students who never took a mathematics class were also removed (10 students) leaving 209 students for analysis. The figure uses the abbreviations "< Calc" for students whose first mathematics class is less advanced than Calculus 1, "Calc" for students whose first mathematics class is is Calculus 1, and "> Calc" for students whose first mathematics class is greater than Calculus 1.

The figure starkly shows the importance of math readiness for this population. Of the 209 students, 41% first enroll in a mathematics class less advanced the Calculus 1; 59% of these students leave physics before enrolling in Physics 1. Of the 37% of the students who first enroll in Calculus 1; only 26% of these leave physics before enrolling in Physics 1. Students with AP or transfer credit for Calculus 1 first enroll in a mathematics class more advanced than Calculus 1; only 7% of these students fail to enroll in Physics 1. The advanced math entry students have a persistence advantage over other students through Modern Physics. Once either a nonmath-ready or a Calculus 1 entering student enrolls in Physics 1, they persist to Physics 2 at about equal rates. This indicates that pre-college factors are most important in allowing students to persist to enroll in a physics class; once the student successfully enrolls in physics, pre-college factors become less important. From Physics 2, the nonmath-ready student persists to Modern at a somewhat lower rate than the Calculus 1 entry student. Of the 209 initial physics majors, 19 of the 82 noncalculus-ready students enroll in Modern Physics as a physics major, 23%; 44 of 84 Calculus 1 entry students enroll in Modern Physics, 52%; 28 of the 43 advanced math entry students enroll in Modern Physics, 65%.

For the graduation probabilities after enrolling in Modern Physics in Fig. 4, a 6-year window was applied (Table I, dataset 1.7, $N = 198$). As before, students who first enroll in Modern or a more advanced physics class and students who never enroll in a mathematics class were removed, leaving 181 students. Figure 4 presents the graduation probability of these students once they enroll in Modern Physics. The graduation rates for all math entry

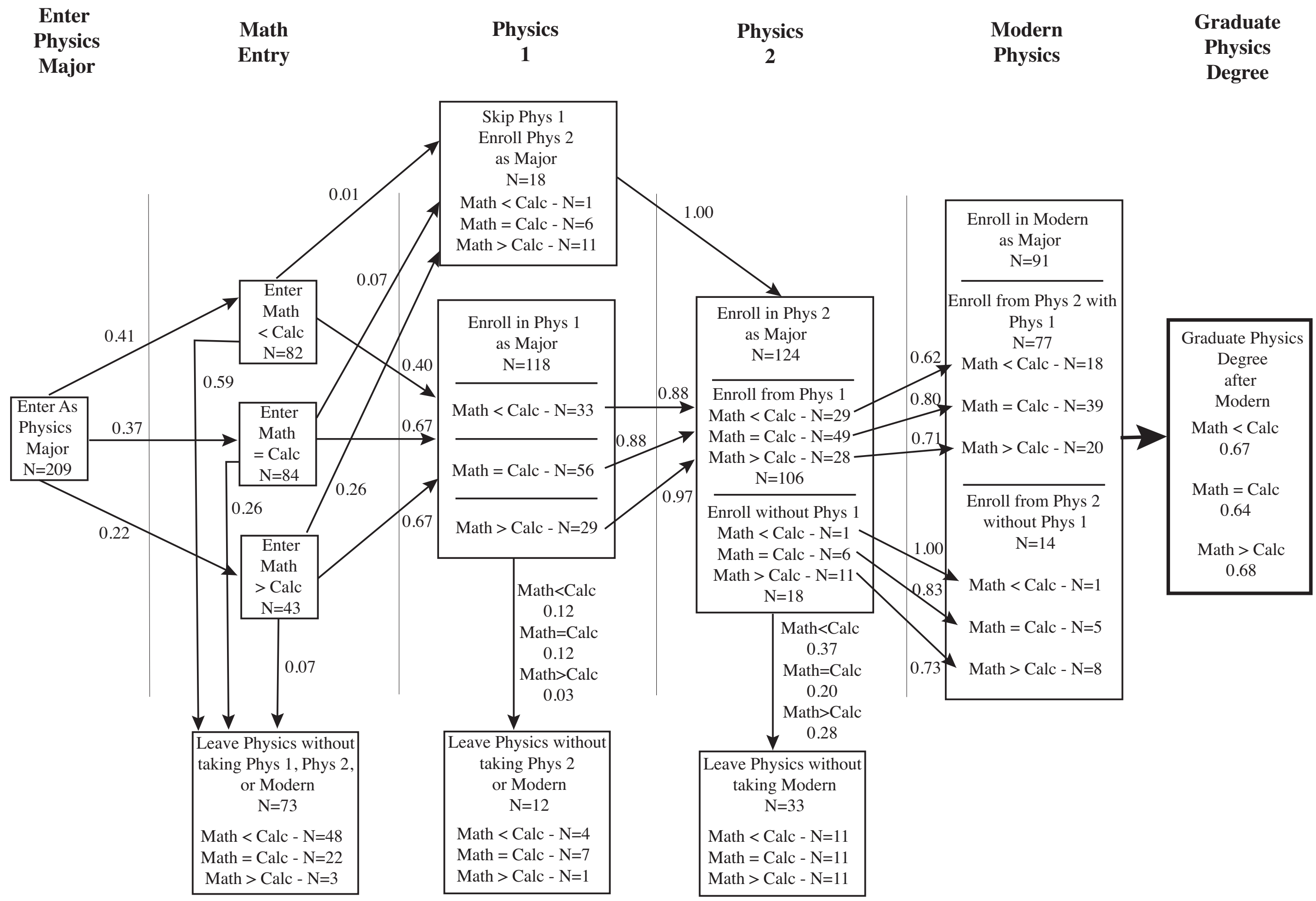

FIG. 4. Traversing the major from entry to Modern Physics for students at Institution 1 who elect a physics major in their first semester. The figure uses the abbreviations < Calc for students whose first mathematics class is less advanced than Calculus 1, Calc for students whose first mathematics class is is Calculus 1, and > Calc for students whose first mathematics class is more advanced than Calculus 1.

points are approximately equal; all students who persist to Modern have an equal chance of graduating with a physics degree.

For the graduation filtered dataset, overall graduation probabilities in physics were calculated for each stage of the progression through the network. Of the 181 students who initially enrolled as physics majors, 31% graduated with a physics degree. Disaggregating by math readiness, of the 65 students not ready to take Calculus 1, 15% graduated; of the 76 students who initially enrolled in Calculus 1, 34% graduated with a physics degree; and of the 40 students who initially enrolled in a mathematics class more advanced than Calculus 1, 53% graduated with a physics degree. Of the 100 students who enrolled in Physics 1 as a physics major, 50% graduated with a physics degree (< Calc 1 42%, Calc 1 46%, > Calc 1 65%). Of the 107 students who enrolled in Physics 2 as a physics major, 53% graduated with a physics degree (< Calc 1 45%, Calc 1 53%, > Calc 1 58%). Of the 79 students who enrolled in Modern Physics as a physics major, 65% graduated with a physics degree (< Calc 1 67%, Calc 1 64%, > Calc 1 68%). As such, the additional advantage confirmed by a more enriched high school STEM experience was important in the early years of college, but ceased to be important once a student progressed to their advanced coursework. We note the 65% graduation rate for students who enroll in Modern Physics is much smaller than the department would like and this will be one target of retention efforts.

IV. DISCUSSION

This study sought to answer two research questions; they will be addressed below. The detailed results were discussed above; the following will synthesize the most important points.

RQ1: At which point in their undergraduate physics career are students most at risk of leaving the physics major? How does this differ by modes of leaving the major? How similar are these risks across different institutions? The risk (hazard) profiles for the two modes of leaving the physics major (leaving college or leaving the major while staying in college) were quite different as shown in Fig. 3. At both institutions, there was a peak in the leaving college hazard in the spring freshman semester as students failed to return to campus for the fall sophomore semester. For both institutions, this hazard decreases dramatically after this point.

The leaving the major while staying in college hazard was quite different for the two institutions. For Institution 1, this hazard peaked in the fall sophomore semester. Students made the major changing decision when they returned to campus for their sophomore year. The hazard declined after this point, but did not reach zero until the fifth year. This hazard was quite different for Institution 2 which saw the highest rate of students changing major in their first semester (fall freshman) and in their fifth semester (fall junior); the hazard of changing major remained substantially higher than at Institution 1 through the first two years of college.

As might be expected, the more academically prepared (higher ACTC, HSGPA, math-readiness) students at Institution 2 left college at a lower overall rate than those at Institution 1 as shown in Fig. 2. Far more students left the physics major by changing to another major at Institution 2 than at Institution 1 as shown in Fig. 2. This may be partially explained by the Sankey plots in Fig. 1. At Institution 2, there are substantially more transitions between different STEM majors and the physics major than at Institution 1; this may be the source of the high levels of leaving the major early in college. At Institution 1, nonSTEM majors approximately equal STEM majors (including engineering) as alternate majors selected by physics students.

RQ2: What pre-college academic factors influence a student's risk of leaving the major through each mode? How similar are these factors across different institutions? The factors influencing different outcomes, one year persistence, two year persistence, and graduation, also differed between the institutions. These factors were explored using logistic regression as shown in Tables IV and V. At Institution 1, leaving the major while staying in college was most strongly related to math-readiness. The odds that a math ready student would leave the physics major for another major were 260% lower than a nonmath-ready student. Not being math ready increases time to degree and delays entry into physics classes, making retention difficult, and other majors with less restrictive mathematics requirements more attractive. Leaving college was more related to general high school preparation and success measured by HSGPA. Each standard deviation increase in HSGPA lowered the odds of leaving college by the junior year by 100%.

The regression results for Institution 2 are shown in Table V. The combination of the more restricted range of variation of the continuous variables due to a ceiling effect and the smaller sample size resulted in fewer significant regressors. For this dataset, ACTC significantly predicted leaving physics while staying in college by the junior year, as well as passing Modern Physics as a physics major after enrolling in Calculus 1.

The progression through the major and the role of math readiness at Institution 1 was further explored by examining the progression through the course network in Fig. 4. At this institution, 41% of students enrolled as physics majors their first semester were not ready to enroll in Calculus 1; 59% of these students left physics without ever enrolling in Physics 1. Only 15% of these students graduated with a physics degree. For students whose first mathematics class was Calculus 1, 34% graduated with a physics degree; for students who first enroll in a mathematics class more advanced than Calculus 1 53% graduated with a physics degree. This illustrates the importance of access to advanced high school course offering to success in physics at Institution 1. Some students underrepresented in physics may have limited access to these courses [79]. There were few differences in physics graduation rates for students who remained in the major long enough to enroll in Modern Physics.

V. IMPLICATIONS

The results for the two institutions were similar at some points such as the spring freshman semester being a critical point for retaining students to college; however, the results were quite different in many areas. This suggests the need for each physics department interested in retention to explore their own data to understand the points where there is an opportunity to improve retention. We note that many of the differences between the two institutions studied may have resulted from differences in the selectivity of the admission requirements.

For Institution 1, the analysis suggests three points where retention efforts could be directed. Nonmath-ready students succeed in the major at very low rates and often leave the major before taking Physics 1. Exploring methods to allow these students to begin taking physics while they catch up in mathematics might retain more to the major. This might involve allowing these students to take the algebra-based physics sequence and accepting these for the calculus-based Physics 1 and 2 with successful completion of Modern Physics and Calculus 1. There is a continuous slow attrition of majors after semester 4 (spring sophomore semester) when students are taking their advanced coursework. This attrition is not present in Institution 2. This suggests Institution 1 should examine the features of their advanced undergraduate program that cause students to leave late in the program. Finally, the institution loses majors at the highest rate after the spring freshman semester to the leaving college hazard and after the fall sophomore semester to the changing major hazard (the changing major decision may have been made the semester before). This suggests substantial efforts be focused on retention in the first year of college.

Efforts currently under discussion include a redesigned freshman seminar course focused on retention, a freshman research experience with a cohort building element, and an introductory laboratory section for physics majors taught by faculty.

For Institution 2, physics majors are lost to the program overwhelmingly by changing to another major. The students at Institution 2 are more academically prepared to succeed in college than those on Institution 1 and, therefore, it should be expected that fewer leave college without a degree. Students change to majors other than physics at a high rate through the fall junior semester. After this point, the fraction of students who have departed through this mode becomes approximately constant. This pattern suggests retention efforts should be implemented in both the freshman and sophomore years. Efforts currently under way include a 1-credit seminar course in the first semester to orient students to the physics major and the career paths that are available to physics degree holders. Other strategic planning elements for retention include the development of different concentrations, e.g., Biophysics, Data Science, and Applied Physics, to make the physics degree more flexible and appealing.

VI. LIMITATIONS AND FUTURE WORK

This study was performed at two institutions with relatively small physics undergraduate programs and found strong differences in physics major retention between the programs. This work should be replicated at other programs, both at larger programs and similar programs with different demographic composition, so as to map out the spectrum of physics retention. This work was unable to explore differences in retention of demographic groups underrepresented in physics; these differences should be explored in future studies.

This work chose to focus analysis on pre-college variables available to many physics departments so as to understand the features of incoming students which most predict physics success. Many other variables may also be important to understanding retention including noncognitive variables such as self-efficacy and a sense of belonging. As students progress through college, college-level variables such as course grades or general college GPA may become better predictors of continued success. All these factors should be investigated in future work.

VII. CONCLUSIONS

This work examined the retention of physics majors through multiple points in their undergraduate career at two institutions. Institution 1 had an incoming student population that was less accomplished in high school than Institution 2. The patterns of major changing, risk, and the factors influencing retention were quite different between the two institution. This indicates that physics departments which seek to understand undergraduate retention in more detail should replicate these analyses for their students.

Both institutions experienced a peak in the risk of leaving the physics major by leaving college in the spring freshmen semester. At Institution 1, the changing major risk was highest in the fall sophomore semester. At Institution 2, this risk peaked in both the fall freshman semester and the fall junior semester. At Institution 1, math readiness emerged as the key factor predicting changing to major other than physics while staying in college. Math ready students are prepared to enroll in Calculus 1 or a more advance mathematics class their first semester of college. At Institution 1, 41% of students electing a physics major their first semester were not math ready; only 15% graduated with a physics degree; 37% of incoming physics majors enrolled in Calculus 1 their first semester; 34% graduated with a physics degree. This analysis also suggested advanced high school college preparatory curriculum was important in physics student success; 22% of incoming physics majors had high school credit for Calculus 1 and enrolled in a more advanced class; 53% of these students graduated with a physics major.

Different factors were important in predicting leaving college and graduating. At Institution 1, high school GPA was the most important factor in predicting retention to college and graduation with a physics degree. At Institution 2, math readiness was the most important factor predicting one-year retention to college; high school GPA for two-year retention. At Institution 1, math readiness was the most important factor predicting leaving physics while staying in college; at Institution 2, ACT composite scores were the most important in predicting changing to a major other than physics.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grants No. ECR-1561517, No. HRD-1834569, and No. DUE-1833694.

[1] J. L. Docktor and J. P. Mestre, Synthesis of discipline-based education research in physics, Phys. Rev. Phys. Educ. Res. **10**, 020119 (2014).
[2] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).
[3] I. Rodriguez, E. Brewe, V. Sawtelle, and L. H. Kramer, Impact of equity models and statistical measures on interpretations of educational reform, Phys. Rev. Phys. Educ. Res. **8**, 020103 (2012).
[4] S. Nicholson and P. J. Mulvey, *Roster of physics departments with enrollment and degree data, 2020* (American Institute of Physics, College Park, MD, 2020).
[5] T. Hodapp, The Economics of education: Closing undergraduate Physics Programs, APS news **20**, 8 (2011).
[6] J. M. Aiken, R. Henderson, and M. D. Caballero, Modeling student pathways in a physics bachelor's degree program, Phys. Rev. Phys. Educ. Res. **15**, 010128 (2019).
[7] J. P. Zwolak, R. Dou, E. A. Williams, and E. Brewe, Students' network integration as a predictor of persistence in introductory physics courses, Phys. Rev. Phys. Educ. Res. **13**, 010113 (2017).
[8] J. Forsman, R. Moll, and C. Linder, Extending the theoretical framing for physics education research: An illustrative application of complexity science, Phys. Rev. ST Phys. Educ. Res. **10**, 020122 (2014).
[9] L. Stiles-Clarke and K. MacLeod, Demystifying the scaffolding required for first-year physics student retention: Contextualizing content and nurturing physics identity, Can. J. Phys. **96**, xxix (2018).
[10] P. A. Westrick, J. P. Marini, L. Young, H. Ng, D. Shmueli, and E. J. Shaw, *Validity of the SAT for Predicting First-Year Grades and Retention to the Second Year* (College Board, New York, NY, 2019).
[11] J. Allen, *Updating the ACT College Readiness Benchmarks. ACT Research Report Series 2013 (6).* (ACT, Inc., Iowa City, IA, 2013).
[12] W. J. Camara and G. Echternacht, The SAT I and high school grades: Utility in predicting success in college, The College Board, Research Notes **10**, 1 (2000), https://eric.ed.gov/?id=ED446592.
[13] S. l Geiser and R. Studley, UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California, Educ. Assess. **8**, 1 (2002).
[14] P. A. Westrick, H. Le, S. B. Robbins, J. M. R. Radunzel, and F. L. Schmidt, College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES, Educ. Assess. **20**, 23 (2015).
[15] S. Geiser and M. V. Santelices, *Validity of High-School Grades in Predicting Student Success beyond the Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes. Research & Occasional Paper Series: CSHE. 6.07.* (Center for Studies in Higher Education, University of California-Berkeley, Berkeley, CA, 2007).
[16] A. Nandeshwar, T. Menzies, and A. Nelson, Learning patterns of university student retention, Expert Syst. Appl. **38**, 14984 (2011).
[17] V. Tinto, *College Student Retention: Formula for Student Success* (Greenwood Publishing Group, Santa Barbara, CA, 2005).
[18] V. Tinto, Dropout from higher education: A theoretical synthesis of recent research, Rev. Educ. Res. **45**, 89 (1975).
[19] V. Tinto, *Leaving College: Rethinking the Causes and Cures of Student Attrition* (University of Chicago Press, Chicago, IL, 1993).
[20] V. Tinto, *Completing College: Rethinking Institutional Action* (University of Chicago Press, Chicago, IL, 2012).
[21] C. Henderson, M. Dancy, and M. Niewiadomska-Bugaj, Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process?, Phys. Rev. Phys. Educ. Res. **8**, 020104 (2012).
[22] National Science Board, *Revisiting the STEM Workforce: A Companion to Science and Engineering Indicators 2014* (National Science Foundation: Arlington, VA, 2015).
[23] President's Council of Advisors on Science and Technology, *Report to the President. Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics* (Executive Office of the President, Washington, DC, 2012).
[24] X. Chen, *STEM Attrition: College Students' Paths into and out of STEM Fields. NCES 2014-001* (National Center for Education Statistics, US Department of Education, Washington, DC, 2013).
[25] K. Rask, Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences, Econ. Educ. Rev. **29**, 892 (2010).
[26] E. J. Shaw and S. Barbuti, Patterns of persistence in intended college major with a focus on STEM majors, NACADA J. **30**, 19 (2010).
[27] A. V. Maltese and R. H. Tai, Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students, Sci. Educ. **95**, 877 (2011).
[28] G. Zhang, T. J. Anderson, M. W. Ohland, and B. R. Thorndyke, Identifying factors influencing engineering student graduation: A longitudinal and cross-institutional study, J. Engin. Educ. **93**, 313 (2004).
[29] B. F. French, J. C. Immekus, and W. C. Oakes, An examination of indicators of engineering students' success and persistence, J. Eng. Educ. **94**, 419 (2005).
[30] R. M. Marra, K. A. Rodgers, D. Shen, and B. Bogue, Leaving engineering: A multi-year single institution study, J. Engin. Educ. **101**, 6 (2012).
[31] C. W. Hall, P. J. Kauffmann, K. L. Wuensch, W. E. Swart, K. A. DeUrquidi, O. H. Griffin, and C. S. Duncan, Aptitude and personality traits in retention of engineering students, J. Engin. Educ. **104**, 167 (2015).
[32] B. L. Christe, The importance of faculty-student connections in STEM disciplines, J. STEM Educ. I. R. **14**, 22 (2013), https://www.jstem.org/jstem/index.php/JSTEM/article/view/1797.
[33] Z. S. Wilson, L. Holmes, K. Degravelles, M. R. Sylvain, L. Batiste, M. Johnson, S. Y. McGuire, S. S. Pang, and I. M. Warner, Hierarchical mentoring: A transformative strategy for improving diversity and retention in undergraduate STEM disciplines, J. Sci. Educ. Technol. **21**, 148 (2012).

[34] M. Dagley, M. Georgiopoulos, A. Reece, and C. Young, Increasing retention and graduation rates through a STEM learning community, J. Coll. St. Ret. R. T. P. **18**, 167 (2016).

[35] C. T. Belser, D. J. Prescod, A. P. Daire, M. A. Dagley, and C. Y. Young, Predicting undergraduate student retention in STEM majors based on career development factors, Career Dev. Q. **65**, 88 (2017).

[36] C. T. Belser, M. Shillingford, A. P. Daire, D. J. Prescod, and M. A. Dagley, Factors influencing undergraduate student retention in STEM majors: Career development, math ability, and demographics, Prof. Couns. **8**, 262 (2018).

[37] K. Koenig, M. Schen, M. Edwards, and L. Bao, Addressing STEM retention through a scientific thought and methods course, J. Coll. Sci. Teach. **41**, 41 (2012).

[38] K. A. Wingate, A. A. Ferri, and K. M. Feigh, The impact of the physics, statics, and mechanics sequence on student retention and performance in mechanical engineering, in *Proceedings of the 2018 ASEE Annual Conference and Exposition* (American Society for Engineering Education, Washington, DC, 2018).

[39] Y. J. Xu, Attention to retention: Exploring and addressing the needs of college students in STEM majors, J. Educ. Train. Stud. **4**, 67 (2016), https://eric.ed.gov/?id=EJ1080863.

[40] A. Sithole, E. T. Chiyaka, P. McCarthy, D. M. Mupinga, B. K. Bucklein, and J. Kibirige, Student attraction, persistence and retention in STEM programs: Successes and continuing challenges., High. Educ. Stud. **7**, 46 (2017), https://eric.ed.gov/?id=EJ1126801.

[41] E. Seymour and N. M. Hewitt, *Talking about Leaving: Why Undergraduates Leave the Sciences* (Westview Press, Boulder, CO, 1997), Vol. 34.

[42] S. Tobias, *They're not Dumb, They're Different* (Research Corporation, Tuscon, AZ, 1990).

[43] E. Seymour, A. B. Hunter, H. Thiry, T. J. Weston, R. P. Harper, D. G. Holland, A. K. Koch, and B. M. Drake, *Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education* (Springer International Publishing, Germany, 2019).

[44] B. L. Whitten, S. R. Foster, M. L. Duncombe, P. E. Allen, P. Heron, L. McCullough, K. A. Shaw, B. Taylor, and Heather M. Zorn, What works? increasing the participation of women in undergraduate physics, J. Women Minorities Sci. Engin. **9**, 30 (2003).

[45] K. Rosa and F. M. Mensah, Educational pathways of Black women physicists: Stories of experiencing and overcoming obstacles in life, Phys. Rev. Phys. Educ. Res. **12**, 020113 (2016).

[46] M. Ong, Body projects of young women of color in physics: Intersections of gender, race, and science, Social problems **52**, 593 (2005).

[47] L. T. Ko, R. R. Kachchaf, A. K. Hodari, and M. Ong, Agency of women of color in physics and astronomy: Strategies for persistence and success, J. Women Minor. Sci. Engin. **20**, 171 (2014).

[48] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, Data mining algorithms to classify students, in *Proceeding of the 1st International Conference on Educational Data Mining*, edited by R. S. Joazeiro de Baker, T. Barnes, and J. E. Beck (Montreal, Quebec, Canada, 2008).

[49] A. Peña-Ayala, Educational data mining: A survey and a data mining-based analysis of recent works, Expert Syst. Appl. **41**, 1432 (2014).

[50] A. M. Shahiri, W. Husain, and N. A. Rashid, A review on predicting student's performance using data mining techniques, Procedia Comput. Sci. **72**, 414 (2015).

[51] P. Baepler and C. J. Murdoch, Academic analytics and data mining in higher education, Int. J. Scholarsh. Teach. Learn. **4**, 17 (2010).

[52] R. S. J. D. Baker and K. Yacef, The state of educational data mining in 2009: A review and future visions, J. Educ. Data Mine **1**, 3 (2009).

[53] Z. Papamitsiou and A. A. Economides, Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence., J. Educ. Tech. Soc. **17**, 49 (2014), https://www.jstor.org/stable/jeductechsoci.17.4.49.

[54] A. Dutt, M. A. Ismail, and T. Herawan, A systematic review on educational data mining, IEEE Access **5**, 15991 (2017), https://ieeexplore.ieee.org/abstract/document/7820050.

[55] C. Romero and S. Ventura, Educational data mining: A review of the state of the art, IEEE Trans. Syst. Man Cybern. C **40**, 601 (2010).

[56] R. Alkhasawneh and R. H. Hargraves, Developing a hybrid model to predict student first year retention in STEM disciplines using machine learning techniques, J. STEM Educ. I. R. **15**, 35 (2014), https://www.jstem.org/jstem/index.php/JSTEM/issue/view/125.

[57] N. Misiunas, M. Raspopovic, K. Chandra, and A. Oztekin, Sensitivity of predictors in educational data: A Bayesian network model, in *2015 INFORMS Workshop on Data Mining and Analytics*, CIP (The Institute for Operations Research and the Management Sciences, Catonsville, MD, 2015).

[58] A. McGovern, C. M. Utz, S. E. Walden, and D. A. Trytten, Learning the structure of retention data using Bayesian networks, in *2008 38th Annual Frontiers in Education Conference* (IEEE, Piscataway, NJ, 2008), p. F3D.

[59] A. Sharabiani, F. Karim, An. Sharabiani, M. Atanasov, and H. Darabi, An enhanced Bayesian network model for prediction of students' academic performance in engineering programs, in *Proceedings of the 2014 IEEE Global Engineering Education Conference (EDUCON)* (IEEE, Piscataway, NJ, 2014), p. 832.

[60] C. Lacave, A. I. Molina, and J. A. Cruz-Lemus, Learning analytics to identify dropout factors of computer science studies through bayesian networks, Behav. Inf. Tech. **37**, 993 (2018).

[61] R. Torabi, P. Moradi, and A. R. Khantaimoori, Predict student scores using Bayesian networks, Proc. Soc. Behav. **46**, 4476 (2012).

[62] C. Zabriskie, J. Yang, S. DeVore, and J. Stewart, Using machine learning to predict physics course outcomes, Phys. Rev. Phys. Educ. Res. **15**, 020120 (2019).

[63] J. Yang, S. DeVore, D. Hewagallage, P. Miller, Q. X. Ryan, and J. Stewart, Using machine learning to identify the most at-risk students in physics classes, Phys. Rev. Phys. Educ. Res. **16**, 020130 (2020).

[64] D. E. Meltzer and R. K. Thornton, Resource letter ALIP–1: Active-learning instruction in physics, Am. J. Phys. **80**, 478 (2012).

[65] I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, Am. J. Phys. **53**, 1043 (1985).

[66] I. A. Halloun and D. Hestenes, Common sense concepts about motion, Am. J. Phys. **53**, 1056 (1985).

[67] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, Phys. Teach. **30**, 141 (1992).

[68] D. P. Maloney, T. L. O'Kuma, C. Hieggelke, and A. Van Huevelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, S12 (2001).

[69] J. Wells, R. Henderson, J. Stewart, G. Stewart, J. Yang, and A. Traxler, Exploring the structure of misconceptions in the Force Concept Inventory with modified module analysis, Phys. Rev. Phys. Educ. Res. **15**, 020122 (2019).

[70] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. Pat. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, Proc. Natl. Acad. Sci. U.S.A. **111**, 8410 (2014).

[71] C. M. Schroeder, T. P. Scott, T. Y. Tolson, H. Huang, and Y. H. Lee, A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States, J. Res. Sci. Teach. **44**, 1436 (2007).

[72] S. Salehi, E. Burkholder, G. P. Lepage, S. Pollock, and C. Wieman, Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics, Phys. Rev. Phys. Educ. Res. **15**, 020114 (2019).

[73] R. Henderson, J. Stewart, and A. Traxler, Partitioning the gender gap in physics conceptual inventories: Force concept inventory, force and motion conceptual evaluation, and conceptual survey of electricity and magnetism, Phys. Rev. Phys. Educ. Res. **15**, 010131 (2019).

[74] J. Stewart, G. L. Cochran, R. Henderson, C. Zabriskie, S. DeVore, P. Miller, G. Stewart, and L. Michaluk, Mediational effect of prior preparation on performance differences of students underrepresented in physics, Phys. Rev. Phys. Educ. Res. **17**, 010107 (2021).

[75] Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, Sci. Educ. **91**, 847 (2007).

[76] National Center for Education Statistics, https://nces.ed.gov/collegenavigator.

[77] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2017).

[78] J. C. Brunson and Q. D. Read, Ggalluvial: Alluvial plots in "ggplot2," (2020), R package version 0.12.3.

[79] P. R. Aschbacher, E. Li, and E. J. Roth, Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine, J. Res. Sci. Teach. **47**, 564 (2010).