International Journal of Engineering Education Vol. 35, No. 1(A), pp. 142–155, 2019
Printed in Great Britain

0949-149X/91 $3.00+0.00
© 2019 TEMPUS Publications.

Visualization and Analysis of Student Enrollment Patterns in Foundational Engineering Courses*

DAVID REEPING
Virginia Tech, Department of Engineering Education, 659 McBryde Hall, 225 Stanger St, Blacksburg, VA 24061, USA.
E-mail: dreeping@vt.edu

DAVID B. KNIGHT
Virginia Tech, Department of Engineering Education, 353 Goodwin Hall, 635 Prices Fork Rd, Blacksburg, VA 24061, USA.
E-mail: dbknight@vt.edu

JACOB R. GROHS
Virginia Tech, Department of Engineering Education, 375 Goodwin Hall, 635 Prices Fork Rd, Blacksburg, VA 24061, USA.
E-mail: jrgrohs@vt.edu

SCOTT W. CASE
Virginia Tech, Department of Biomedical Engineering and Mechanics, Collegiate Square Innovation Place (0151), 460 Turner Street, Suite 304, Blacksburg, VA 24061, USA. E-mail: scase@vt.edu

The literature in engineering education and higher education has examined the implications of course-taking patterns on student development and success. However, little work has analyzed the trajectories of students who need to retake courses in the curriculum, especially those deemed to be fundamental to a student's program of study, or sequences of courses. Sequence analysis in R was used to leverage historical transcript data from institutional research at a large, public, Land-Grant university to visualize student trajectories within individual courses—with attention to those who re-enrolled in courses—and the pathways students took through a sequence of courses. This investigation considered students enrolled in introductory mechanics courses that are foundational for several engineering majors: Statics, Dynamics, and Strength of Materials (also called Mechanics of Deformable Bodies). This paper presents alluvial diagrams of the course-taking sequences and transition matrices for the different possible grades received upon subsequent attempts for the Mechanics core courses to demonstrate how visualizing students' paths through sequences of classes by leveraging institutional data can identify patterns that might warrant programs to reconsider their curricular policies.

Keywords: archival data; course-taking patterns; gatekeeper courses

1. Introduction

Shifts toward a more "data-driven" educational ecosystem can help address some of higher education's most pressing challenges, such as expanding access, reducing costs, and enhancing the quality of educational experiences [1]. With the increase in the quantity of data, reduction of computational costs, and development of new tools and techniques to identify and visualize patterns from such data, institutions can enhance the potential for student success by learning from existing data [2]. For example, determining patterns in existing transcript longitudinal data can help colleges and universities assess curricular policy and system level performance [e.g., 3, 4]. Within this research area, authors have largely focused their attention on developing models based on entering student characteristics to identify predictor variables and to forecast future events [5], but the concept of course-taking patterns as a phenomenon has not received as much attention [6]. Decisions related to course-taking, however, are fundamental to a student's experience in an institution and can have important practical implications on students' retention, graduation, and time to degree, and, consequently, their financial burdens. Given the highly sequential nature of undergraduate engineering curricular plans, a single course has the potential to extend students' time to degree by one or more semesters, translating into many thousands of dollars in tuition, fees, living expenses, and lost salary from subsequent employment.

Building upon previous work studying foundational engineering courses [7, 8], our paper demonstrates how we can visualize and learn from existing institutional course-taking data. We focus our analysis on some of the most troubling "gatekeeper" classes for undergraduate engineers—such courses stall progress in the major, and roadblocks emanate from them broadly because they are often fundamental prerequisites for required courses later in the degree program [9]. Within undergraduate engineering, the introductory sequence of mechanics courses (i.e., Statics, Strength of Materials or Mechanics of Deformable Bodies, and Dynamics)

* Accepted 14 September 2018.

are often characterized by large lecture halls and high rates of near-failing, failing, or withdraws (i.e., dropping a class before a grade remains on a student's permanent record)—often referred to as the "D-W-F rate" [10]. Although these three foundational courses are essential for certain majors, prior work has also shown that many students try to find "workarounds" for the courses by using transfer credit to replace one or more of the courses before attempting them at their primary institution [10]. The result of the substitutions may be an added cost for students but may also increase the likelihood of success in that course if the transfer credit pathway is sufficiently easier than the intended path. Thus, as depicted in the conceptual framework presented in Fig. 1, students can take a variety of paths through this sequence of gatekeepers. The path to an engineering degree from the "starting line" to the "finish line" is conceptualized as a path down a roadway. Fig. 1 is not meant to be comprehensive; rather, it is meant to illustrate that paths to an engineering degree are not straight and free of adversity for everyone. The "detours" are meant to capture unforeseen issues that lead to a student needing to retake a course at the institution or a community college and transferring in credit. Similarly, detours could lead to a student switching majors or leaving the institution altogether, which we frame as a "detour" relative to the initial plan (we do not cast any kind of judgment on those decisions, as a variety of reasons can explain such detours). In our analysis, we aim to visualize the paths using institutional data and identify the different academic performance implications of each path.

We address the following research questions:

RQ1: What are the array of pathways that students commonly take through undergraduate engineering gatekeeper courses, specifically the core Mechanics courses?

RQ2: What is the relationship between the pathways in RQ1 and academic performance in subsequent courses?

By addressing these research questions, our analyses demonstrate how leveraging existing institutional data can yield a more nuanced understanding of the implications of these different pathways through the curriculum as academic advisors, instructors, and students navigate recommendations and decisions related to course-taking.

2. Literature review

Of the existing studies that have focused on course-taking patterns, the most predominant focus has been on the role of course-taking in vertical transfer from a two-year to a four-year institution [11–15]. Similar research focused on pathways to four-year degrees has examined the role of pre-college course-taking patterns [16–19]. The common tools of analysis in both research areas include ordinary least squares (OLS) and hierarchical linear modeling (HLM) [20], which have been used to create predictive models for a variety of outcomes, such as transfer or degree attainment. For example, Riegle-

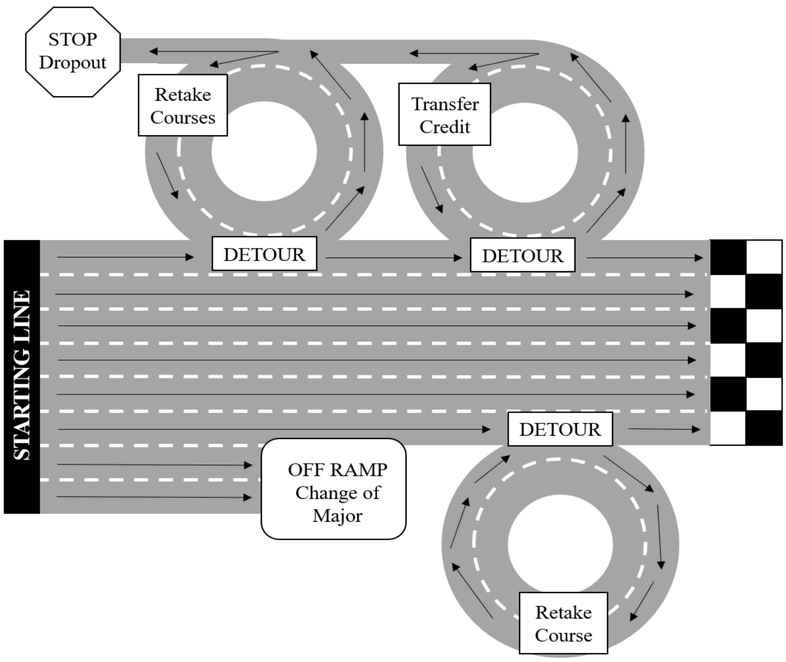

Fig. 1. Conceptual depiction of the multiple paths students may follow through a "gatekeeper" undergraduate Mechanics course.

Crumb [21] used HLM to predict the highest math course students would have taken by the end of high school based on students' demographic variables. The higher level predictive models are useful for examining the effects of different treatments and determining the impact of predictors.

However, using the aggregate predictive methods of regression can minimize the nuance of the individual units of the analysis. Course-taking sequences are often simplified into bins that capture the outcome of the trajectory rather than the process students go through the curriculum. For example, Kelly [22] outlines the course-taking bins by placing trajectories into categories such as "Greater than Algebra II or Geometry" or "Algebra II or Geometry or not both" to investigate racial differences in mathematics course-taking. More specific examples in engineering education are provided by Tyson [23], who used multinomial logistic regression to assess the effects of achievement in high school and grades in physics and calculus on subsequent outcomes, as well as Lord et al. [24] who studied the relationship between grades in science, technology, engineering, and math courses and subsequent outcomes. Both studies aggregated course-taking trajectories and used them as a variable but removed an underlying structure in the data—its sequential nature. More recent endeavors to capture the "structure" of the trajectories have used clustering techniques to elicit patterns across general categories of classes [25]. Compared to the aggregate predictive methods, the exploration of trajectories beyond a variable in regression is less explored.

Although some progress has been made toward a more thorough treatment of the unit of analysis, few papers have attempted to make sense of more specific course-taking patterns using techniques like sequence analysis—especially in the discipline of engineering education. Mills [26] conducted a sequence analysis to examine course-taking patterns in mathematics concerning Hispanic students at public two-year colleges to infer transfer and degree attainment, but other literature using such an approach is scarce. Within the engineering context, Almatrafi et al. [6] identified course-taking trajectories of engineering students in Data Analytics who were high-achieving, which determined the "structure" of the patterns and is in the spirit of sequence analysis. In addition, Hachey, Wladis, and Conway [27] examined student success in online courses based on their prior experiences, but "retaking" a course was not understood in the sense of re-enrolling in the same course, as is typical for the mechanics sequence.

Heileman, Slim, Hickman, & Abdallah [28] described a curriculum analytics tool that explored the complexity of different curricular arrangements, which can quantify the extent to which curricular sequencing, difficulty, and connectedness between courses relates to students' time to degree. The tool determined "delay factors" of different courses, thus empirically identifying a course's function as a gatekeeper. Finally, Heileman, Babbitt, Abdallah, and Dougher [29] also examined the flow of students through their entire university curriculum using Sankey (or Alluvial) diagrams. The existing work on flow diagrams and sequence analysis inspired our analysis concerning pathways through the gatekeeper engineering courses.

3. Data and methods

In this paper, we focus on a similar sequencing phenomenon but seek to offer a different point of view by exploring the specific patterns of course taking—or the processes students go through as they navigate gatekeeper courses—using ideas from sequence analysis and data visualization. We focus on variations tied to specific gatekeeper courses as experienced by students, whereas Heileman et al. [28, 29] focused on the curricular sequences laid out by programs.

3.1 Data description

This study draws upon a longitudinal transcript data set of students who attended a large, public, Land-Grant university in the United States. The population included students who had taken Statics, Mechanics of Deformable Bodies (Strength of Materials), or Dynamics between 2009 and 2016 for all possible course terms: fall, spring, both summer sessions, and winter. The data encompassed a total of 11008, 7170, and 7811 observations of students who enrolled in Statics, Dynamics, and Mechanics of Deformable Bodies, respectively. As a subset of records maintained by the university's registrar, the data set is compiled from institutional research transcript information. Each observation is an instance of a course that is either one of the three Mechanics courses or a prerequisite of the cluster (e.g., Calculus). An observation is composed of several pieces of administrative information to designate the courses: course term, course subject, course number, and course reference number. The remaining variables directly attached to the student include a student identifier, credit type of the course (taken at the institution versus transferred in), major, overall grade-point average (GPA), institution admit type (i.e., transfer, first-time in college), and first and last term of attendance. One advantageous feature of the data set is that the subset of mechanics course attempts can be considered institution population-level data since all students who took the mechanics courses are represented.

3.2 Data preparation

The study had two phases: (1) analysis of course retaking patterns, and (2) analysis of course-taking with two different classes. The data cleaning procedure was fundamentally the same for both efforts, albeit with slightly different binning procedures. Using the R programming environment, we restructured the data to prepare for curricular sequence analyses. As a first step, all grades were binned into a stock A-F (T/W) grading scale, disregarding plus and minus grades to simplify the structure of the sequences. A "T" grade refers to course credit transferred from a different institution that is not factored into the student's overall GPA. The "W" grade reflects a withdrawal without grade penalty from the course. (Students may select this option before final exams.) For the second phase of the analysis, grades were binned into "Pass" and "Fail" categories so that {A,B,C} denotes {Pass} and {D,F} denotes {Fail}. Withdrawal (W) and transfer (T) grades were not re-binned. The data were originally in a panel format, where each observation was an instance of a student enrolling in a course. Observations for each student were consolidated into a single observation in a "longitudinal" format where the columns indexed the course term, and the observation was a student's series of attempts in the course (Fig. 2). This format allowed for a visual examination of when students were enrolling and re-enrolling in the courses. The "TraMineR" package enabled the functions to run the analyses. The sequence objects, a formal data type in the package used to run the other functions, were then constructed using the "seqdef" function from the "TraMineR" package.

The sequences in the second phase were connected to the degree the student received. If the student received an engineering degree based on the degree code, the observation was designated "Grad w/ ENG." If the student graduated, but not with a degree in engineering, the observation was marked with "Grad w/o ENG." Finally, if the student had no degree recorded, the observation was designated as "No degree." Six-year graduation rates were utilized; therefore, students who enrolled within six years of the data pull were excluded from the analysis. To filter out false positives for "No degree," students who likely would not have graduated by the time of data collection were removed from consideration by excluding students enrolled six years before the date of the data pull. Six years was chosen as a conservative estimate of the engineering time to degree. The filtering lowered the number of students enrolled to a total of 2064, 1124, and 927 observations in Statics, Dynamics, and Mechanics of Deformable Bodies, respectively.

To avoid courses taken concurrently and simplify the visualizations and analysis, we formed sequences of Statics to Dynamics and Statics to Mechanics of Deformable Bodies using the TraMineR package in R [30–31]. The tools provided by the TraMineR package concern the analysis of sequences and allow for plotting sequences in a variety of formats, examining their longitudinal characteristics, and measuring the similarity between sequences [31]. However, we used the "alluvial" package to create cleaner diagrams after data exploration with TraMineR. The Alluvial diagrams were created by forming sequences using the "seqdef" function on the longitudinal observations like in the first phase, then using the "seqtab" function to obtain frequency counts of popular sequences. Patterns with more than 10 occurrences were considered since frequencies fell exponentially thereafter. The values were entered in a separate spreadsheet, loaded into RStudio, then taken as the argument of the "alluvial" function. By abstracting the course-taking trajectories as categorical sequences, the TraMineR and alluvial packages can be leveraged to make sense of students' curricular decisions at the macro-level for each course.

3.3 Analysis

To examine course-retaking patterns, we used TraMineR's seqtrate function to compute the probabilities associated with transitioning from one state to the next. Each element in the sequence is an attempt at a specific course, and the states are the grades received in the course. The output of the seqtrate function is an n-by-n *transition rate matrix*, where n is the number of states, and each element of the matrix is the probability of moving from state i to state j. More precisely, the probability statement

ID	Term	Grade
1	Fall	D
1	Summer	B

→ Transformation →

ID	Fall	Spring	Summer
1	D		B

Fig. 2. Transformation of observations from panel format to a single longitudinal format.

tells us the following: given the sequences of grades without gaps from the dataset, what is the probability of a student receiving a grade of X given that he or she received a grade of Y on the previous attempt from any point in the sequence [31].

To visualize sequences of courses, we examined patterns from Statics to Deforms and Statics to Dynamics. Although the courses technically follow a chain of three courses in some sense, the sequences were treated as separate chains to avoid difficulties of analyzing sequences with two courses potentially taken in the same semester. The "alluvial" package was used to construct an alluvial diagram of the pathways for the two scenarios. Alluvial diagrams are used to analyze structural changes in a system over time [32]. Alluvial plots are composed of nodes (levels of categorical variables) and flows, which are streams of colored area connecting the nodes together. The nodes are arranged in columns, called steps, such that each column is a categorical variable, and the nodes within it are the associated levels of the variable. Larger, thicker flows correspond to a large proportion of observations following that pathway.

In the case of our study, the alluvial diagrams were used to visualize the pathways of students through the two-course sequences of interest. Each step was a different course, and the nodes were the potential outcomes (e.g., Statics—F, failed Statics). Some pathways only have one or two courses, whereas other pathways have three courses, a characteristic which presents a challenge when constructing the alluvial diagram. To address the challenge of the unequal pathway lengths, NA nodes were added so that each pathway had three elements in the sequence.

3.4 Limitations

The existing data set inherits some limitations. Additional demographic data would have been helpful to include in the analysis beyond student admit status, but the information was withheld because of privacy concerns. The type of transfer credit is also unspecified, so we cannot determine if transfer credit was received at a community college or another four-year institution. Developing an understanding of the effects of specific types of transfer credit would require more information about the other institutions and the students themselves, particularly if the credit comes from a community college. The type of transfer could suggest students have different prior experiences, which may affect their overall success in traversing the core mechanics sequence and beyond. For instance, students transferring from a rural community college may have experienced the same course in title as a student transferring from a large, public four-year institution, but the educational experiences likely differed greatly.

Additionally, our analyses do not explore variation in pathways by students' intended major. The mechanics sequence may be more relevant for certain engineering majors than others, and such variation could explain students' decision-making when determining the appropriate curricular path. For example, students majoring in mechanical engineering may have been compelled to take a more rigorous Statics sequence at the four-year institution, whereas students majoring in industrial and systems engineering may have been more likely to seek workaround options since the content is less relevant for that major.

4. Results

The results are separated into three parts. First, cross-tabulations of student performance are provided by admit-type to contextualize the two prevailing types of students in the dataset. The division of the cross-tabulations was motivated by our observations of student behavior. A surprisingly large pathway taken was one in which transfer credit used on the first attempt of the course by students who were concurrently enrolled at the institution from which the data set originates. The most common sequence in retaking a course was withdrawing from the course, then transferring credit for each of the three courses; in the case of Mechanics of Deformable Bodies, the sequence occurred at least twice as frequently as the other sequences. In other words, students left the course before the end of the semester and opted to take the course at another institution for credit. Considering the high frequency at which students transferred both before attempting the course at their primary institution and after withdrawing, we opt to show the cross-tabulations of student performance separated by admit status (i.e., first-time in college versus transfer student from another postsecondary institution). Second, course-taking patterns are presented to show how students repeat the foundational mechanics courses. Third, we connect the courses in pairs to explore what pathways are common for students between the two courses concerning degree attainment.

4.1 Cross-tabulations by admit-type

The first set of results provides a portrait of the multiple pathways with which students entered the institution before taking the Statics sequence. The cross-tabulations are updated calculations from [8] since more academic terms were added to the data set since the previous publication. Table 1 provides a cross-tabulation of students' performance on the

Table 1. Cross-tabulation of First Attempt in Statics Grades between Admit Status

	Course Grades							
Admit Status	**A**	**B**	**C**	**D**	**F**	**T**	**W**	**Total**
First-Time In College	1202	1614	1819	739	468	806	1368	8016
Transfer	54	67	80	41	38	681	73	1034
Total	1256	1681	1899	780	506	1487	1441	9050

Table 2. Cross-tabulation of First Attempt in Deforms Grades between Admit Status

	Course Grades							
Admit Status	**A**	**B**	**C**	**D**	**F**	**T**	**W**	**Total**
First-Time In College	927	1552	1406	534	309	666	458	5852
Transfer	65	100	115	56	30	440	57	863
Total	992	1652	1521	590	339	1106	515	6715

first attempt of the course sequence based on their admit status. Students who performed well enough in their first attempt of the course by earning a C could move forward in the sequence without needing to consider trying again for a reasonably higher grade.

Of the 8016 first-time in college-admit students, 10% chose to take Statics elsewhere and transfer the credit they earned into their primary institution (denoted by the "T" category). Perhaps even more shocking is the number of students who withdrew from the course. Seventeen percent of first-time in college students selected the "W" option, meaning that they withdrew from the class after the sixth week of the semester but before the final exam, regardless of performance in the course. Only 7% of transfer students withdrew, but most of the observations for Statics with transfer students are within the "T" category (66%)—meaning most transfer students bring in credit for Statics from another institution. Students most frequently earned a C if they took the course at the institution across both admit types on their first attempt.

It was found that 19% of all students retook the course at some point in time by restructuring the full data set and calculating the number of times students repeated Statics. Therefore, nearly one-fifth of all students had to take a course detour for Statics. Understanding re-taking patterns became the next step in our analysis considering the large portion of students retaking the course. A crosstabulation of the earned grades served as the first step of analysis to gain a broad picture of the dataset.

In comparison to Statics, we can make similar observations for two of the next courses in the Mechanics core, Mechanics of Deformable Bodies ("Deforms") (Table 2) and Dynamics (Table 3). Again, a sizable portion of first-time in college students, 11%, transfer in credit for Deforms on their first attempt. Like Statics, about 51% of transfer students bring credit for Deforms with them upon matriculating. In the case of Deforms, 8% of first-time in college students withdrew from the course, and unlike Statics, the most common grade for students was a B.

Dynamics is exceptional relative to Statics and Deforms because it has the largest percentage of students who transferred in credit on their first attempt at 19%. In fact, the T grade was the most common outcome, ahead of the letter grade C. Transfer students brought in credit for Dynamics 66% of the time. About 11% of first-time in college students withdrew from the course on their first attempt.

4.2 Individual courses: re-taking patterns

As noted, given the high proportions of students who retook the three Mechanics classes, our next step was sequencing the multiple attempts at the same class followed by connecting courses together and examining progression through them. First, we describe re-taking patterns using transition

Table 3. Crosstabulation of First Attempt Dynamics Grades between Admit Status

	Course Grades							
Admit Status	**A**	**B**	**C**	**D**	**F**	**T**	**W**	**Total**
First-Time In College	655	1137	1172	532	283	1059	597	5435
Transfer	26	60	63	27	22	510	69	777
Total	681	1197	1235	559	305	1569	666	6212

Table 4. Transition Matrix for Statics for number of attempts greater than two

Grade	n	[-> A]	[-> B]	[-> C]	[-> D]	[-> F]	[-> T]	[-> W]
[A ->]	0	0	0	0	0	0	0	0
[B ->]	0	0	0	0	0	0	0	0
[C ->]	6	0	0.17	0.17	0.33	0	0.17	0.17
[D ->]	247	0.16	0.30	0.35	0.09	0.05	0	0.04
[F ->]	532	0.04	0.13	0.23	0.14	0.20	0.20	0.06
[T ->]	26	0.15	0.15	0.12	0.08	0.15	0	0.35
[W ->]	1253	0.07	0.16	0.21	0.14	0.12	0.26	0.04

matrices, which were calculated for Statics, Deforms, and Dynamics using the TraMineR package. Table 4 displays the transition matrix for Statics in its entirety. The leftmost column shows the previous grade, and the top row shows the next grade. The n's are row-wise, meaning there are n instances of the previous grade as an antecedent to the next grade. For example, there were 247 instances of D as a previous grade.

An entry in the matrix shows the probability of receiving the next grade given the previous grade was earned. Therefore, the elements of the matrices are conditional probabilities of a sequence transitioning from one state to the next at *any* position in the sequence. In other words, the values represent the probability of receiving a grade of X after receiving a grade of Y, regardless of when the first grade, X, might occur. The matrix can be quickly read to obtain the probability of simple scenarios occurring. For example, consider a student who withdraws from Statics. In the context of the data set, there was a 7% chance of that student receiving an A on their next attempt (i.e., the 0.07 value in the table). Note that probabilities may not add to one row-wise because of rounding for ease of interpretation.

The first salient feature of the transition matrix is the first two rows of all zero probabilities, which should intuitively make sense. A student who receives an A or a B will almost certainly not enroll in the same course again. Next, the C row contains students who passed but were likely not satisfied with their grade. Unfortunately, empirical results indicate that the students who retook the course after earning a C were more likely to perform worse than their previous attempt. (The probability of the sum of withdrawing or earning less than a C was larger than the probability of improving.) The row-wise n is quite small though (only 6), so the finding is more of a curiosity to use as an example in interpreting the matrix. Students who earned a D and then retook the course fared better but were most likely to receive only one letter grade better (35% chance). Students who failed and then retook the course were almost as likely to fail again (20%) as they were to receive a C (23%) Twenty-six students who transferred in credit for Statics re-attempted the course, perhaps attempting to increase their GPA because transfer credit does not factor into cumulative GPA. Students retaking the course after transferring were most likely to withdraw (35%).

The most likely next step for students who withdrew from Statics was found to be transferring in credit from another institution (26% of the time). Students who received an F (failing grade) in the first instance of the course followed by transferring in credit 20% of the time. Such findings raise questions about equity across students of different socioeconomic statuses because students would have to pay additional tuition to take a course at a different institution. Our findings pose the question of whether transferring in credit from a different institution was an accessible option for all students.

Next, Table 5 displays the transition rate matrix for Mechanics of Deformable Bodies. As in Table 4, the sparse first three rows of A-C are unsurprising. For students who received a D and retook Mechanics of Deformable Bodies, the probabilities of receiving the next grade are distributed similarly compared to Statics—a one letter grade improvement is still most likely (40%). Students who decided to retake the course after failing were nearly equally likely to receive another F as they were to receive a C

Table 5. Transition Matrix for Mechanics of Deformable Bodies for number of attempts greater than two

Grade	n	[-> A]	[-> B]	[-> C]	[-> D]	[-> F]	[-> T]	[-> W]
[A ->]	0	0	0	0	0	0	0	0
[B ->]	0	0	0	0	0	0	0	0
[C ->]	2	0.50	0	0.50	0	0	0	0
[D ->]	87	0.09	0.22	0.40	0.14	0.09	0	0.06
[F ->]	270	0.01	0.07	0.21	0.21	0.21	0.23	0.05
[T ->]	7	0	0	0	0.14	0.43	0	0.43
[W ->]	561	0.07	0.10	0.18	0.10	0.06	0.46	0.02

Table 6. Transition Matrix for Dynamics for number of attempts greater than two

Grade	n	[-> A]	[-> B]	[-> C]	[-> D]	[-> F]	[-> T]	[-> W]
[A ->]	0	0	0	0	0	0	0	0
[B ->]	0	0	0	0	0	0	0	0
[C ->]	4	0.75	0	0	0	0.25	0	0
[D ->]	243	0.12	0.33	0.30	0.12	0.08	0	0.03
[F ->]	340	0.05	0.19	0.23	0.16	0.19	0.16	0.02
[T ->]	14	0.14	0.29	0.14	0.07	0.14	0	0.21
[W ->]	523	0.10	0.18	0.21	0.10	0.10	0.31	0.02

(21% and 21%, respectively) but were most likely to transfer in credit by a slight margin (23%). Mechanics of Deformable Bodies exhibited less frequent transitions from W to F (6%) relative to Statics (12%).

Finally, Table 6 displays the transition rate matrix for Dynamics. Once again, the first three rows are sparse since students likely will not retake a course in which they received a passing grade. Unlike Statics and Mechanics of Deformable Bodies, students who retook Dynamics after earning a D were most likely to improve by two letter grades (33%). In the case of Dynamics, students who withdrew from the course were found to transfer in credit for their next attempt 31% of the time. Those who failed did not transfer in credit as frequently (16%) compared to Statics and Mechanics of Deformable Bodies; instead, the students were more likely to try again and fail or receive a C (19% and 23%, respectively).

4.3 Connecting courses: course-taking patterns

Although the transition matrices inform specific course-*retaking* behavior, we can also investigate students' curricular progression regarding course sequences. To extend our discussion beyond viewing courses from an individual perspective, we connected the courses and visualize course-taking behavior *across* two courses logically sequenced after one another. Because of the diversity of requirements in the majors at the institution, some students may only have Statics as a required class, whereas other students will need all three courses. In the case the student did not take a course in the second or third columns in the alluvial diagram, the path travels through an NA box, or boxes. The NA box is simply a placeholder to preserve the structure of the diagram. For example, a student traveling through Statics - P -> NA -> NA -> Grad w/ ENG means the student passed Statics, did not take Dynamics or Deforms, then graduated with an engineering degree. The NA box does not imply a course other than Dynamics or Deforms was taken. For simplicity, we split the sequences into Statics -> Deforms and Statics -> Dynamics since the courses could potentially be taken concurrently.

Figure 3 displays the alluvial diagram for course sequences involving Statics and Mechanics of Deformable Bodies that occur at least ten times. Flows through those courses are then connected to a graduation outcome. For instance, most students pass Statics on the first attempt (i.e., the largest box on the left side of the figure), then about half of those students pass Deforms on the first try and graduate with an engineering degree. Another set of flows indicate a subset of the students withdraw or fail Deforms and splinter off into other paths, including transferring in credit. Some majors do not require Deforms, as can be observed by students who passed Statics and then flowed through the NA box ultimately to graduate with an engineering degree. We note the numerous paths after passing Statics that still end with an engineering degree. The streams flowing into the engineering degree box include students who have failed each class, withdrawn from each class, and transferred credits from other institutions for each class. Such a result could be shown to students who do not perform well in their first instance taking Statics—it could help them recognize that they still could earn an engineering degree, albeit a much less common path as depicted by the diagram.

The diagram also helps visualize some of the different implications of the decisions to withdraw from or transfer in course credit. About half of the students who transferred in credit for Statics ultimately graduated without an engineering degree. That same pattern did not hold for the students who decided subsequently to transfer in credit for Deforms on either their first or second attempt, as those students tended to graduate with an engineering degree. About two-thirds of students who decided to withdraw from Statics on their first attempt ultimately did not earn an engineering degree. Most students who withdrew from Deforms, however, continued and passed Deforms during their re-take attempt. As noted, Statics and Deforms may have different levels of importance for different majors in engineering, but considering course sequences visually in this manner helps illuminate how the withdraw or transfer curricular policies ultimately may play out differently based on the course under investigation.

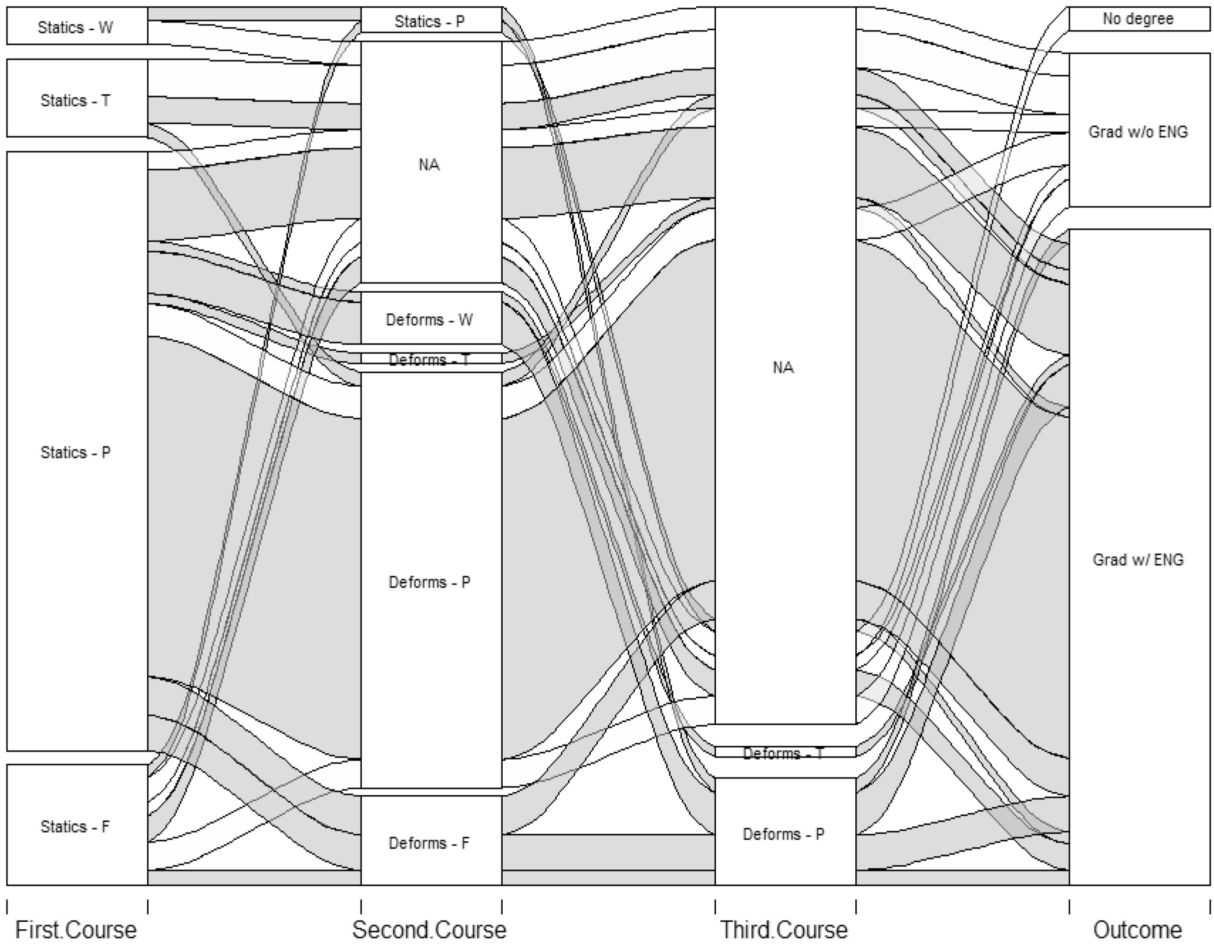

Fig. 3. Alluvial diagram of Statics to Deforms sequences occurring > 10 times.

Figure 4 displays the alluvial diagram for course sequences involving Statics and Dynamics for sequences that occur at least ten times. As before, the flows through the two courses feed into graduation outcomes. About half of students who pass Statics on the first attempt pass Dynamics on the first try and graduate with an engineering degree. Like Deforms, not all majors require Dynamics, which is represented by the path of students who passed Statics on the first attempt, flow into the NA box, then graduate with an engineering degree.

We can still see the pattern where about half of the students transferring in credit for Statics graduated without an engineering degree, so observations concerning Statics from Fig. 3 are still visible in Fig. 4. Like Deforms, students who also transferred in credit for Deforms on any subsequent attempt were able to graduate with an engineering degree, contrasting the previous pattern with only Statics. Unlike Deforms, all flows through Dynamics lead to the student earning an engineering degree—even failures. In other words, any student who enrolled in Dynamics earned an engineering degree for the frequent sequences—i.e., 10 or more occurrences.

5. Discussion

Using the longitudinal plots in Figs. 3 and 4, patterns of student course-taking patterns could be gleaned from the flows. The span of the possible paths students took was a testament to the diversity of student experiences, even in such a specific context as taking a sequence of two courses. The transition matrices proved to be a valuable tool for analysis that provided another layer of detail at the single course level, as it allowed the entire collection of sequences to be described as a single stochastic matrix capturing the various ways students can navigate the courses. In fact, the matrices are probabilistic expressions of revealed preferences [33], which show the points at which students are willing to invest more time in receiving a better grade in a course. While the underlying rationale explaining "why" students behave is obscured by the purely quantitative interpretation, valuable insights into how students choose to retake their courses can be gleaned from examining how students respond to each state.

The information in the transition matrices and

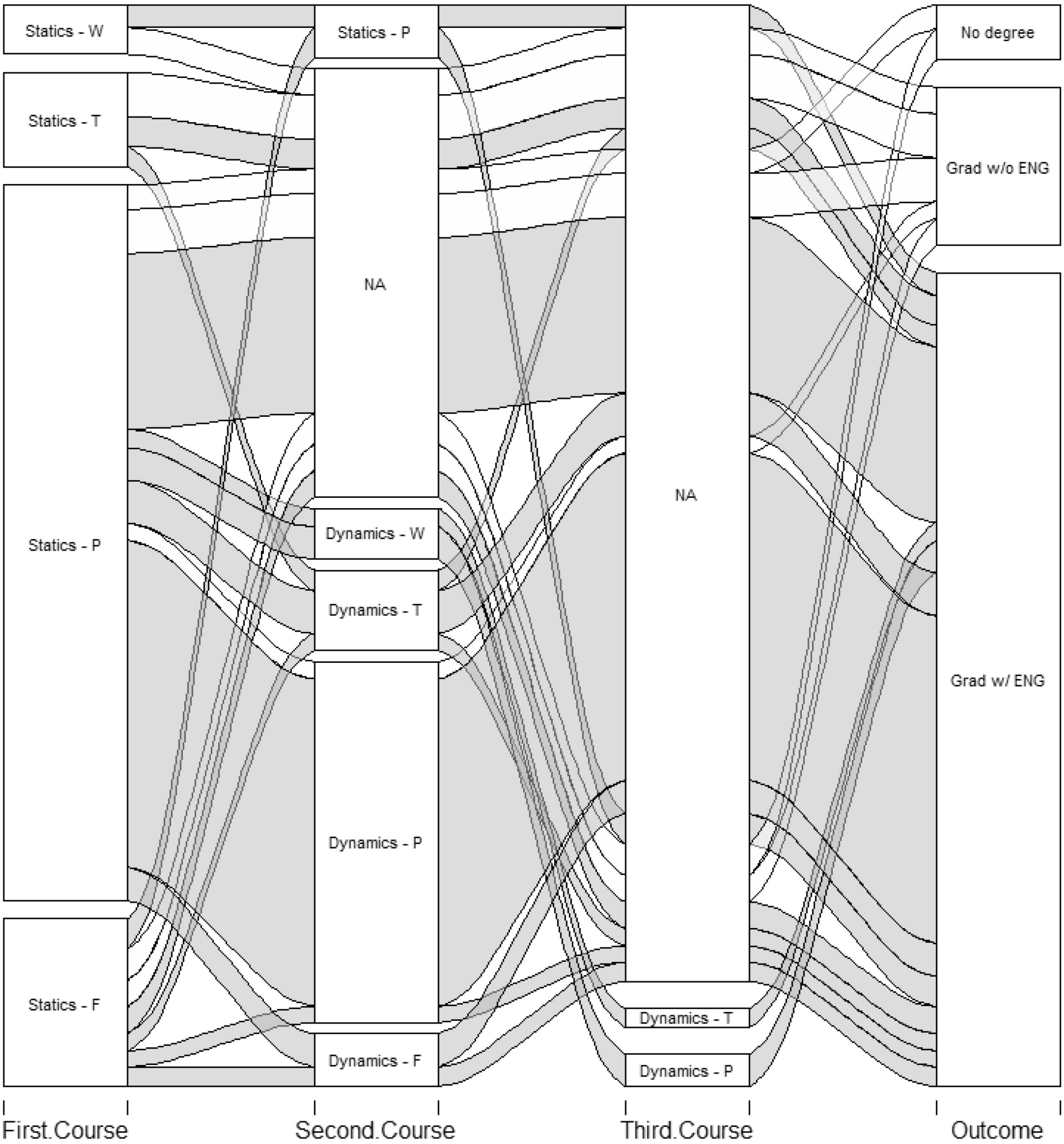

Fig. 4. Alluvial diagram of Statics to Dynamics sequences occurring > 10 times.

alluvial diagrams is useful to academic advisors who are providing recommendations to students who are struggling in any of the three mechanics core courses. Although the aim is not cruelty in the sense of saying "there is no hope of passing the course if you receive an X in class Y," the probabilities can provide a basis for course-taking recommendations and policy [e.g., 14]. Depending on the reader's worldview on the function of universities, limiting the number of times a student can take a course can be a policy that is well-intentioned but can constrain student choice. Certainly two positive outcomes of a limiting policy are to increase classroom efficiency regarding class size and reduce the individual student's investment in higher education [34], which carries other indirect costs when retaking courses. On the other hand, should the student be able to fail a course repeatedly? Should students have complete autonomy with their investments in higher education? By leveraging institutional data to visualize such patterns, curricular policymakers can use existing data and sequence analyses to consider such pressing questions.

Two points of concern emerged from the analyses, as students seemingly found paths to graduation despite poor performance in the foundation courses and used potentially unusual transfer credit paths.

5.1 Flows of failure paving success?

Many students were able to graduate with an engineering degree despite some of the pathways having students who earned exclusively below a C-. This observation led to a search for which disciplines of engineering the students were able to enter with such a performance. Comparing these results with plans of study for each major within the college, the path through which students failed both Statics and Dynamics was associated with instances of students majoring in industrial and systems engineering. Although both courses appear in the required curriculum for the major, prerequisite minimum grades were not enforced for the mechanics courses like they were for math, statistics, and within-major courses. Rather, students must have an in-major and overall GPA of 2.0 or better to remain in good academic standing. It is feasible for a student to receive a D in both courses (which is binned as failing in the diagram) and still be successful, which explains the confusing flows containing only failure but still resulting in graduating with an engineering degree. An implication of this finding could be to re-examine the necessity of requiring certain courses for a degree. If students are still overwhelmingly successful in a major despite such performance in those courses, we might question how relevant the courses are for their success in the major. The visualizations might identify an opportunity for a program to reduce its credit hour requirements or to allocate credits within a degree more effectively.

5.2 Oddities in transfer credit

An odd discovery in the data was the family of transfer transitions across all the courses. One would expect the probability of a student retaking a course after transferring in a course for credit to be zero; yet, the transition rate matrices disagree. A few potential explanations come to mind. First, students may retake the course because of uncertainties related to the transferability of the credit; this would explain the high withdrawal probabilities since students would drop as soon as the credit is approved. A more likely explanation is that the students were attempting to boost their GPAs by enrolling in a familiar course for which they already had credit. The GPA padding strategy may indicate the role that academic eligibility and major change policies have on student behavior, but more work would need to be done to make a causal connection. The high W rates would then be a result of the students realizing they are not performing as well as they had hoped or deciding the course was not worth repeating. The rates could have implications on time-to-degree, as students may have wasted an opportunity to take another course that allowed them to progress toward a degree.

5.3 Replicating in other contexts

Those wishing to replicate the methods described in this paper can use existing transcript data if certain data requirements are met. Likely any statistical programming environment is suitable, but this effort used two specific packages in the R environment, "TraMineR" and "alluvial." The student-level dataset should have *at least* the following variables:

An index to sort students—The dataset used in this paper was prepared in consultation with the university's institutional research unit after filing the necessary paperwork to conduct human subject research with the Institution Research Board. Student names were replaced with dummy identification numbers. A similar process likely can be followed at other institutions, but it would be important to consult with the office, department, or individuals maintaining student records as institutions follow different student data policies.

Course name/number—Some variable needs to exist to identify the courses taken by each student, otherwise the sequences will be ambiguous.

Course term—The course term will aid in ensuring the sequences are formed in the correct order and show which courses were taken concurrently. A broader analysis of when students took the courses could be completed as well.

Course grades/marks—The grades are at the center of the analysis, so the grade earned in the course must be in the dataset.

Graduation outcome—If interest lies in graduation outcomes, a graduation indicator is necessary. The variable could be a major or degree the student earned at graduation or simply a yes/no. The dataset for this work specified the major, so categories of "grad w/ ENG" and "grad w/o ENG" could be created.

Time first enrolled—Knowing when a student began at the institution is vital to ensure any assumptions about graduation are valid. The "time first enrolled" variable can be used to omit any students who could not possibly have graduated yet.

Next, cleaning must be done. The dataset provided was in a panel/longitudinal format in the context of this work. Do any necessary cleaning like subsetting observations only containing the courses of interest or removing students who could not have possibly graduated yet with respect to the timing of the data. For example, a student

Table 7. Example Alluvial Data

First Course	Second Course	Third Course	Outcome	Freq
Statics - P	Deforms - P	NA	Grad w/ ENG	337
Statics - P	NA	NA	Grad w/ ENG	70
Statics - P	Deforms - W	Deforms - P	Grad w/ ENG	42

who took Statics last year has a low chance of graduating this year since Statics occurs so early in the curriculum. Therefore, set a time window to consider for the analysis to make sure a "did not graduate" outcome for a given student is a valid assumption. Next, create the graduation variable to connect the course sequences to the outcome. Run any descriptive statistics of interest.

Following these data preparation steps, pick one course of interest and form a subset. Transform the data to make a single observation look like the example in Fig. 2 where the columns are semesters. Then, form the sequences using the "seqdef" function on the transformed dataset by omitting the gaps between enrollments. The observations will then have the form of {C, A} or {D, D, T}, for example. To perform the transition rate analysis, use the "seqtrate" function with the transformed data. The matrices in tables 4 through 6 can then be generated.

To create the alluvial diagrams, pick two or more courses and form a subset of the students progression. Change the grades in the sequence to the course name and a bin for the grades, which are constructed in some meaningful way like "Statics—Pass", "Statics—Fail," and "Statics—Withdraw." Do the same transformation as before, and have the last column be the graduation outcome. Use the "seqdef" function on the data again and omit the gaps between enrollments. Now the observations have the form of {Statics-Pass, Deforms-Pass, Grad w/ ENG}, for example. Use the "seqtab" function to find the most common sequences, then create a spreadsheet and enter them as shown in Table 7. Use "NA" or an equivalent designation if the sequence is shorter than the longest sequence.

Once created, load the csv file back into RStudio and input the data into the alluvial function. The plots shown in this paper will then be generated for your data.

For more information, the user guide and functions of "TraMineR" can be consulted at [30, 31] .

6. Conclusions

This study leveraged institutional data to identify patterns in course re-taking and sequences across courses concerning the Mechanics core courses of Statics, Dynamics, and Mechanics of Deformable bodies as part of a larger effort to study foundational engineering courses. Rather than focusing on outcomes of individual courses, this paper provided examples of data outputs and visualizations that describe how students navigate sequences of courses and multiple attempts of a single course. Few researchers have explored institutional data in this manner, even though determining patterns in existing transcript longitudinal data can help colleges and universities assess curricular policy and system level performance.

This paper presented two strategies for describing course-taking behavior: transition matrices and alluvial diagrams. The transition matrices were calculated for each foundational engineering course and could serve as information that is useful to academic advisors or curricular policy makers within the college to understand the flow of students through bottlenecks in the curriculum. Additionally, visualizing pathways through sequences of courses similarly identified potential areas for further discussion within the college concerning curricular requirements. Leveraging existing data to move toward an understanding of curricular sequences represents a way in which institutions can consider how curricular policies may be adjusted or communicated to students to reduce students' time to degree and overall costs of undergraduate education. This effort is one small example of the insights nonroutine analyses can bring.

Acknowledgements—The authors would like to acknowledge the support provided by the U.S. National Science Foundation (Award # DUE-1712089). Any opinions, findings, conclusions, and/or recommendations are those of the investigators and do not necessarily reflect the views of the National Science Foundation.

References

1. U.S. Department of Education, The President's fiscal year 2015 budget request for education, 2015, http://www.ed.gov/budget15.
2. L. Baer and J. Campbell, From metrics to analytics, reporting to action: Analytics' role in changing the learning environment, In *Game changers: Education and information technologies.* D.G. Oblinger (ed.), 2012, EDUCAUSE.
3. M. W. Ohland, G. Zhang, B. Thorndyke and T. J. Anderson, The creation of the multiple-institution database for investigating engineering longitudinal development (MIDFIELD), 2004, *Proc. Amer. Soc. Eng. Ed.*
4. A. G. Picciano, The Evolution of Big Data and Learning Analytics in American Higher Education, *Journal of Asynchronous Learning Networks*, **16**(3) 2012, pp. 9–20.
5. A. Peña-Ayala, Educational data mining: A survey and a

data mining-based analysis of recent works. Expert systems with applications, **41**(4), 2014, pp. 1432–1462.
6. O. Almatrafi, A. Johri, H. Rangwala and J. Lester, Identifying Course Trajectories of High Achieving Engineering Students through Data Analytics, *Paper presented at 2016 ASEE Annual Conference & Exposition*, New Orleans, Louisiana, 2016, doi: 10.18260/p.25519
7. S. Tobias, Stemming the science shortfall at college. *They're not dumb, they're different. Tucson, AZ: Research Corporation*. 1990.
8. J. R. Grohs, D. B. Knight, G. D. Young and M. M. Soledad, Exploring academic performance paths and student learning strategies in a large foundational engineering course, *International Journal of Education in Mathematics, Science and Technology*, **6**(3), 2018, pp. 241–253.
9. J. R. Grohs, G. D. Young, M. M. Soledad and D. B. Knight, Leveraging local data for reflective teaching in large classes, *Innovations in Education and Teaching International*, 2018, pp. 1–11.
10. J. R. Grohs, M. M. Soledad, D. B. Knight and S.W. Case, Understanding the Effects of Transferring In Statics Credit on Performance in Future Mechanics Courses. *Paper presented at 2016 ASEE Annual Conference & Exposition*, New Orleans, Louisiana, 2016, doi: 10.18260/p.27103
11. P. R. Bahr, Deconstructing remediation in community colleges: Exploring associations between course-taking patterns, course outcomes, and attrition from the remedial math and remedial writing sequences, *Research in Higher Education*, **53**(6), 2012, pp. 661–693.
12. N. D. Finkelstein and A. B. Fong, Course-Taking Patterns and Preparation for Postsecondary Education in California's Public University Systems among Minority Youth. Issues & Answers. REL 2008-No. 035, *Regional Educational Laboratory West*, 2008.
13. L. S. Hagedorn, A. Cabrera and G. Prather, The community college transfer calculator: Identifying the course-taking patterns that predict transfer, *Journal of College Student Retention: Research, Theory & Practice*, **12**(1), 2010, pp. 105–130.
14. M. Perry, P. R. Bahr, M. Rosin and K. M. Woodward, Course-Taking Patterns, Policies, and Practices in Developmental Education in the California Community Colleges. A Report to the California Community Colleges Chancellor's Office, *EdSource*, 2010.
15. A. Razfar and J. Simon, Course-Taking Patterns of Latino ESL Students: Mobility and Mainstreaming in Urban Community Colleges in the United States, *TESOL Quarterly*, **45**(4), 2011, pp. 595–627.
16. M. C. Long, D. Conger and P. Iatarola, Effects of high school course-taking on secondary and postsecondary success, *American Educational Research Journal*, **49**(2), 2012, pp. 285–322.
17. K. S. Schiller and C. Muller, Raising the bar and equity? Effects of state high school graduation requirements and accountability policies on students' mathematics course taking, *Educational Evaluation and Policy Analysis*, **25**(3), 2003, pp. 299–318.
18. J. Trusty, Effects of High School Course-Taking and Other Variables on Choice of Science and Mathematics College Majors, *Journal of Counseling & Development*, **80**(4), 2002, pp. 464–474.
19. P. A. White, A. Gamoran, J. Smithson and A. C. Porter, Upgrading the high school math curriculum: Math course-taking patterns in seven high schools in California and New York, *Educational Evaluation and Policy Analysis*, **18**(4), 1996, pp. 285–307.
20. H. Woltman, A. Feldstain, J. C. MacKay and M. Rocchi, An introduction to hierarchical linear modeling, *Tutorials in Quantitative Methods for Psychology*, **8**(1), 2012, pp. 52–69.
21. C. Riegle-Crumb, The path through math: Course sequences and academic performance at the intersection of race-ethnicity and gender, *American Journal of Education*, **113**(1), 2006, pp. 101–122.
22. S. Kelly, The black-white gap in mathematics course taking, *Sociology of Education*, **82**(1), 2009, pp. 47–69.
23. W. Tyson, Modeling engineering degree attainment using high school and college physics and calculus coursetaking and achievement, *Journal of Engineering Education*, **100**(4), 2011, pp. 760–777.
24. S. M. Lord, M. M. Camacho, R. A. Layton, R. A. Long, M.W. Ohland and W. H. Wasburn, Who's persisting in engineering? A comparative analysis of female and male Asian, black, Hispanic, Native American, and white students, *Journal of Women and Minorities in Science and Engineering*, **15**(2), 2009, pp. 167–190.
25. S. B. Sailesh, K. J. Lu and M. Al Aali, Profiling students on their course-taking patterns in higher educational institutions (HEIs), In *International Conference on Information Science (ICIS)*, IEEE, 2016, pp. 160–167.
26. S. R. Mills, *Mathematical Course-Taking Patterns of Hispanic Students at Public Two-Year Colleges and How These Patterns Affect Degree Attainment and Transfer*, Doctoral dissertation, Ohio University, 2016
27. A. C. Hachey, C. W. Wladis and K. M. Conway, Is the Second Time the Charm? Investigating Trends in Online Re-Enrollment, Retention and Success, *Journal of Educators Online*, **9**(1), 2012, n1.
28. G. L. Heileman, M. Hickman, A. Slim and C. T. Abdallah, Characterizing the complexity of curricular patterns in engineering programs, In *Proceedings of the 2017 American Society for Engineering Education (ASEE) Annual Conference, Columbus, Ohio*, 2017.
29. G. L. Heileman, T. Babbitt, C. T. Abdallah and M. J. Dougher, Efficient curricula: The complexity of degree plans and their relation to degree completion, *Higher Learning Commission*, 2014.
30. A. Gabadinho, G. Ritschard, M. Studer and N. S. Müller, Mining sequence data in R with the TraMineR package: A users guide for version 1.8, *Geneva: University of Geneva*, 2018.
31. A. Gabadinho, G. Ritschard, N. S. Müller and M. Studer, Analyzing and visualizing state sequences in R with TraMineR, *Journal of Statistical Software*, **40**(4), 2011, pp. 1–37.
32. M. Rosvall and C. T. Bergstrom, Mapping change in large networks, PloS one, **5**(1), 2010, e8694.
33. P Samuelson, A Note on the Pure Theory of Consumers' Behaviour, *Economica*, **5**(17), 1938, pp. 61–71.
34. M. B. Paulsen, The economics of human capital and investment in higher education, *The finance of higher education: Theory, research, policy, and practice*, 2001, pp. 55–94.

David Reeping is a graduate student pursuing a PhD in Engineering Education at Virginia Tech. He received his B.S. in Engineering Education with a Mathematics minor from Ohio Northern University. He was a Choose Ohio First scholar inducted during the 2012–2013 school year as a promising teacher candidate in STEM. David was the recipient of the Remsburg Creativity Award for 2013 and The DeBow Freed Award for outstanding leadership as an undergraduate student (sophomore) in 2014. He is also a National Science Foundation Graduate Research Fellowship Program awardee. David a member of the mathematics, education, and engineering honor societies: Kappa Mu Epsilon, Kappa Delta Pi, and Tau Beta Pi respectively. He has extensive experience in curriculum development in K-12 and creates material for the Technology Student Association's annual TEAMS competition. David has co-authored two texts related to engineering, Principles of Applied Engineering for Pearson-Prentice Hall and Introductory Engineering Mathematics for Momentum Press. His research interests include: model/method transferability, threshold concepts to inform curriculum development, information asymmetry in higher education processes (e.g., course articulation), and issues in first year engineering

David B. Knight is Associate Professor and Assistant Department Head for Graduate Programs in the Department of Engineering Education at Virginia Tech. He is also Director of International Engagement in Engineering Education and affiliate faculty with the Higher Education Program at Virginia Tech. His research tends to be at the macro-scale, focused on a systems-level perspective of how engineering education can become more effective, efficient, and inclusive, tends to be data-driven by leveraging large institutional data sets, and considers organizational change processes. He has BS, MS, and MUEP degrees from the University of Virginia and a PhD in Higher Education from Pennsylvania State University.

Jacob R. Grohs is an Assistant Professor in Engineering Education at Virginia Tech with Affiliate Faculty status in Biomedical Engineering and Mechanics and the Learning Sciences and Technologies at Virginia Tech. He holds degrees in Engineering Mechanics (BS, MS) and in Educational Psychology (MAEd, PhD).

Scott W. Case is a Professor of Engineering Mechanics at Virginia Tech. He has served as associate department head of Engineering Science and Mechanics and as Interim Associate Dean for Academic Affairs within the College of Engineering.