Credit-Hour Analysis of Undergraduate Students Using Sequence Data

Tushar Ojha, University of New Mexico

Tushar Ojha is a graduate (PhD) student in the Department of Electrical and Computer Engineering at the University of New Mexico (UNM). His work is focused on researching and developing data driven methods that are tailored to analyzing/predicting outcomes in the higher education space. He works as a Data Scientist for the Institute of Design & Innovation (IDI), UNM.

Don Hush, University of New Mexico

Dr. Hush has worked as a technical staff member at Sandia National Laboratories, a tenure-track professor in the ECE department at the University of New Mexico, a staff scientist at Los Alamos National Laboratories, and is currently a Research Professor in the ECE department at the University of New Mexico. He has a technical background in Machine Learning, Signal Processing, Theoretical Computer Science, Pattern Recognition, and Computer Vision. He is the coauthor of a 2009 text entitled "Digital Signal Analysis with Matlab" and is the author of over 100 peer-reviewed scientific publications.

Credit Hour Analysis of Undergraduate Students using Sequence Data

Abstract

Representing credit accumulation as a function of time (a student's terms), rather than as a single cumulative number, unlocks potential for uncovering patterns in the accrual of credits. In this paper, we present an analysis into the credit hour usage pattern of university students as a function of time. However, not all credits accumulated by the students are usable towards their degree program of study. Even if they are usable, they may not be applicable. In this work we use a custom-built specialized audit tool to (automatically) decompose student credits in the following way: *unusable* credits that do not match any degree requirement, *excess* credits that can be removed without changing the requirement satisfaction, and *applied* credits that contribute to requirement satisfaction without excess. We would like to remark that there is a dearth of studies that seek to explain the credit hour usage pattern of university students. The paucity of studies can be attributed to the use of inflexible and/or opaque commercial degree audit tools at universities, which curtails the possible scope of analytics on degree audit data. The credit decomposition allows us to consider the *usability* and *applicability* of credit hours towards the student's degree program, enabling us to take a step further than most analyses concerning credit hours. At US universities, a large number of degree-seeking undergraduate students graduate with a higher number of credit hours than is required for graduation, thus incurring *extra* credits. Excess and unusable credit categories make up the extra credits. Preliminary analysis reveals that excess credits are the dominant extra credit category. Naturally, analysis of excess credits as a function of time, i.e., the excess credit sequences, as a means to understand the student's credit efficiency serves as the focal point of this paper. The results presented reveal interesting excess credit accumulation patterns that help explain some of the reasons behind excess credits. In particular, excess credit sequences for different student groups were used to investigate the widely held notion of "transfer credit loss" and "program (major) change" as significant contributors to extra credits. The results present a more nuanced view of this notion. Moreover, we propose a novel feature engineering method as a way to study cooperation between a student feature sequence (e.g., financial aid, program change, etc.) and an outcome feature sequence (e.g., excess credits). As a result, each relevant student feature sequence is mapped into a feature value that attempts to capture information that is relevant to the outcome. This enables a data-driven way to analyze the effect of a large number of student features on excess credit accumulation.

1 Introduction

The credit hour was born out of the need to standardize learning for all students, to improve efficiency of institutions, to facilitate cross-institutional transfer, and to keep tabs on curriculum quality [20]. Recently, it has additionally grown into an instrument of administrative budgeting, accountability measures, and outside reporting [22]. A bachelor's degree in the US is formed using a minimum of 120 credit hours, congruent with the requirements of all the regional accrediting agencies. However, a large number of degree-seeking undergraduate students graduate with a higher number of credit hours than is required for graduation. According to a study done by Complete College America(CCA) [4] on data provided by 33 states, students accumulated 16.5 credits (a semester of coursework) in *excess* of the 120-credit-hour requirement. Credit hours are "accumulated" by attempting courses in subject areas specified by the degree program requirements, whereas credit hours are "earned" by successfully completing these courses and thus satisfying the requirements. One way to assess the *progress* of the student would be to compare the student's total accumulated credits to the credits required by the student's specific degree program at graduation. Such credits are generally referred to as *excess* credits in the higher-education literature [13].

This paper seeks to explain the credit efficiency of university students. Not all credits accumulated by students are *usable* towards their degree program of study - a fact that often goes unaccounted for based on the excess credits definitions mentioned above. Even if they are usable, they may not be *applicable*. In this work, we use a custom-built specialized audit tool to decompose student credits in the following way: *unusable* credits that do not match any degree requirement, *excess* credits that can be removed without changing the requirement satisfaction, and *applied* credits that contribute to requirement satisfaction without excess [10]. From this point forward, the term *excess* refers to this updated definition that accounts for the *usability* of credits towards the degree program requirements [18]. We will use the term extra credits to refer to the combination of unusable and excess (Ex) credits. Unusable-credits can be further broken down into *unusable-earned (UE)* and *unusable-unearned (UU)* credit categories. The former includes unusable courses that were successfully completed by meeting the minimum passing requirements of the course, whereas the latter includes courses that were attempted but no credits were earned (e.g., failed courses, remedial courses, etc.). Preliminary analysis using actual student data from a large public (Research 1[2]) university showed that 98.5% of the undergraduate students graduate with non-zero extra credits, and half of them do so with at least 30 credits - equivalent to one additional year of full-time enrollment.

Credits are amassed over time, and thus representing credit accumulation as a function of time (a student's terms), rather than as a single cumulative number, opens up ways of identifying trends in the accumulation of credits. By doing so, essentially, we include the temporal properties of credit accumulation in our credit efficiency analysis. Most studies on credit hour accumulation at colleges or universities focus on *extra* credits [6, 12, 8, 9, 24, 23] accumulated by graduated students and the factors that influence that. Our early findings revealed that excess credits are the dominant type of extra credit. Thus, we focus our attention on excess credits in this paper. We analyze the extra credit sequences across all students to reveal interesting insights about points in the student journey that contribute most to excess credits, or in other words, credit inefficiency. The most common reasons for extra credit accumulation in the higher-education literature include: transfer credit loss, program (major) change, financial incentives (e.g., extra credits taken to maintain financial aid), repeated courses, remedial courses, filler courses, skill enhancement, etc. We will assess the influence of these factors by comparing their timing with the timing of excess credit accumulation. To this end we propose a new co-occurrence measure that captures the cooperation between a student feature sequence (e.g., financial aid, program change, etc.) and an outcome feature sequence (e.g., excess credits).

The discussion so far in this section clearly alludes to the primary objective of this study: analyze credit efficiency of undergraduate students through the lens of extra credits, and in particular, excess credit sequence. Following are the research questions that this paper addresses as a part of the above objective: (a) is the sequence data helpful in uncovering patterns of excess credit accumulation, (b) is the sequence data helpful in explaining factors affecting extra credit accumulation, and (c) is it possible to generate a feature that captures the cooperation between student feature sequence and excess sequence in such a way that it is helpful in explaining the outcome, i.e., excess. The main contributions of this work can be succinctly given as:

- A non-traditional approach to analyze the credit efficiency of undergraduate students by treating credit and student data as a term-by-term sequence – in particular the term-by-term excess credit sequence data. This approach, combined with the fact that the analysis takes into account the *usability* of the credits towards the student's degree program requirements, gives interesting insights into the timing of excess credit generation by the student.
- A novel way to map student feature sequences into a feature value that preserves information relevant to the outcome. This method combines data science concepts and first principles knowledge, and in this case helps analyze factors controlling excess credits using sequence data.

This work is limited to graduated students at present, and the term-by-term excess credits are calculated with respect to the final degree program of the student. The remainder of this paper is organized as follows: Section 2 establishes the relevance of the research question and provides a brief summary of background work done in this area. Section 3 gives an overview of the data framework and related tools used to carry out this analysis and also provides a high-level discussion of the methods and definitions used in this analysis. Section 4 presents the various analyses conducted with actual student data, results obtained, and discussion of results. Section 5 presents the definition and application of the proposed co-occurrence feature. Finally, Section 6 presents concluding remarks and future work.

2 Relevance and background

As mentioned in Section 1, a majority of the research on credit hour accumulation focuses on extra credits and the factors influencing them. While it is not certain whether all extra credits are unnecessary or can be avoided, they may lead to extended time to complete a degree, higher educational expenses for both students and taxpayers, postponed entrance into the job market, and reduced graduation rates, among other issues. As a result, these issues are widely regarded as troublesome and necessitate the attention of policymakers at both the institutional and government levels. Adelman reported an increasing time-to-degree trend (4.34, 4.45, and 4.56 years for 1972, 1982, and 1992 cohorts, respectively) that coincided with increased credit accumulation (130.1, 134.3, and 138.4 completed credit hours for 1972, 1982 and 1992 cohorts, respectively) [1]. Cullinane found that four-, five-, and six-year graduates attempted 124, 142, and 146 credit hours, respectively, highlighting the marked difference in credit hour accumulation between on-time (four-year) and delayed (five- and six-year) graduates [6]. The additional time and credit hours typically result in a higher cost of education, except in instances where extra credits are taken due to financial incentives. Conversely, there is also a cost to the public, involving both financial and labor resources. The cost to Americans if every bachelor's student took only 3 additional credits would be $1.5 billion a year [5]. The gravity of the issue to administrators is demonstrated by the *Excess Credit Hour* policies many states use to discourage extra credit accumulation [14]. Methods like implementing a tuition surcharge for surpassing a specified limit of extra credits [13], or offering tuition rebate incentives for adhering to a designated extra credit boundary [12], appear to be prevalent deterrents.

Section 1 mentions the reasons for accrual of extra credits by undergraduate students often cited in higher-education literature as transfer credit loss, program (major) change, financial incentives, etc. One such study that evaluated student academic records reported major change, a high number of transfer credits, and working towards a secondary degree as reasons for extra credit accumulation. Zeidenberg identified several potential causes for *extra* credits at community colleges, including indecision about the chosen program of study paired with inadequate advising, switching majors, problems with course scheduling, and transfer courses [23]. Among these, *course transferability* is arguably the most extensively examined factor that affects *extra* credits [8, 21, 24, 6]. However, many of these studies do not focus on extra credits as the objective, but rather use it as a means to evaluate (important) phenomena such as transfer credit loss, student graduation rate, etc. Kilgore et al. used complex statistical modeling utilizing the commonly studied factors, yet the variables did not significantly contribute to the explanation of extra credits at graduation [13]. Using sequence data to study extra credits, or to study outcomes related to student data in general, is quite elusive in existing research in this area. Furthermore, employing our specialized audit tool [10] enables us to create credit sequences accounting for usability, which we think greatly enhances the ability to investigate factors impacting credit efficiency. Educational researchers in learning and instruction analytics have emphasized the importance of temporality and sequence [16, 11, 19]. Mahzoon et al. provide a framework to explicitly represent temporal relationships in student data, which they claim to have enabled the development of more accurate predictive models for student risk and success [15]. Their data model groups student data items (e.g., GPA, major, etc.) into temporally ordered structures called "nodes" (e.g., terms). Moreover, Mahzoon et al. further state that the sequence representation allows time dependency to be accounted for in analytics. We would like to note here that this representation aligns with our treatment of excess credits and student features as sequences, which shares many of the advantages of such representation discussed in [15]. As described in Section 1, we analyze the excess credit sequences in isolation, and in conjunction with the student features corresponding to factors affecting extra credit accumulation.

3 Methodology

The specialized audit tool solves an optimization problem that matches classes to degree requirements in a way that maximizes the number of applied credits, subject to the constraints that a) each class is matched to at most one requirement and b) no excess credits are attributed to any requirement [10]. As discussed in Section 1, it produces a decomposition of student credits as follows: *applied*, *unusable-earned*, *unusable-unearned* and *excess* (the latter three are constituents of the extra credit category). A concise summary of *surplus* (and *separate*) credits are presented next, followed by a brief overview of the important procedures and definitions related to the analysis carried out in this paper.

3.1 Surplus credits

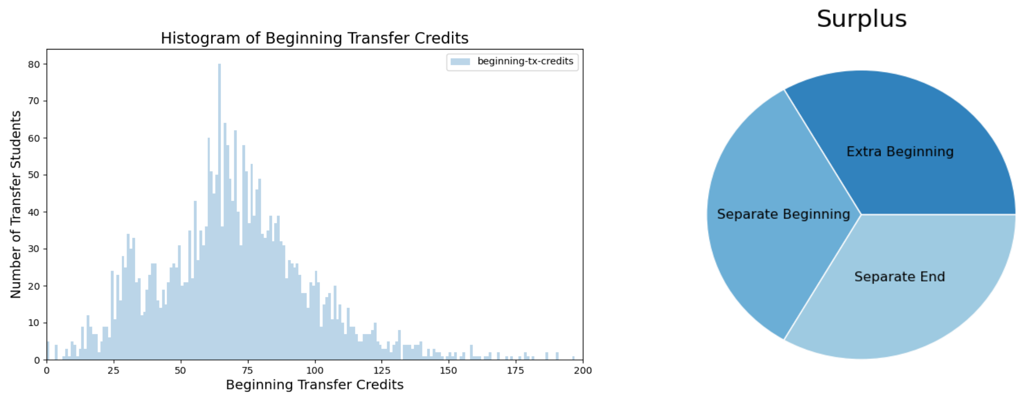


Figure 1: (a) Beginning transfer credits for campus-wide Transfer students (left), and (b) constituents of "surplus" credits (right)[18].

The maximum number of transferable credits at universities can often be calculated by subtracting the credits necessary for the degree program from the credits dedicated to meeting a residency requirement [3], typically resulting in no more than 90 credits at most institutions. Our data shows that about 30% of Transfer students enter with over 90 transfer credits, as seen in the histogram in Fig. 1(a). Irrespective of the credit type and the intended degree program, not all of these transfer credits count towards a Bachelor's degree. This distinction introduces us to the notion of "separate" and "surplus" credits [18]. Credits accumulated for another degree, attempted or completed, that are not plausibly anticipated to be applied to the current degree are defined as "separate" credits. The separate credits may accrue at the beginning or end. Credits from before starting a given undergraduate degree program (e.g., transfer credits exceeding 90) and credits taken towards an advanced degree before officially graduating from an undergraduate degree are examples of beginning- and ending-separate credits, respectively. It is easy to see that separate credits can be a source of bias in extra credit analysis, and therefore should be excluded from the analysis. Transfer students are much more likely than Non-Transfer students to have beginning-separate credits.

Ending-separate credits can be defined as non-applied ending credits potentially applicable to a graduate degree. However, identifying and removing beginning-separate credits can present a challenge. Let us introduce "surplus" credits to help with beginning-separate credit approximation. Surplus credits are those not anticipated to count towards the Bachelor's degree upon the student's admission to the university [18]. They can be determined by running the specialized audit tool [10] on all courses a student has completed before entering the university, with the number of applied credits restricted to to a maximum of

$$max\ \textit{applied}\ credits = \textit{credits required by the degree program} \\ - \textit{credits owed to residency requirement}$$

The audit run produces a set of *applied** and *extra** credits. The *extra** credits are an estimate of the beginning-surplus credits (composed of beginning-separate and beginning-extra credits), and the total surplus can be estimated by summing the beginning-surplus and ending-separate credits. Fig. 1(b) illustrates the constituents of surplus credits.

3.2 Analysis overview

We used the procedure described in Section 3.1 (with the maximum applied credits set to 90) to identify and extract surplus credits from our dataset. We then ran the specialized audit tool against the *filtered class list* (original student class list with all surplus classes removed) to create a new set of unusable-earned, unusable-unearned, and excess credits. This removes the bias separate credits cause. To put it in perspective, once surplus was removed, the non-zero median credits that the Transfer students began with fell from 9 to 3 (and the fraction of students contributing to the non-zero median reduced to 29% from 72%). An initial investigation of the unusable-unearned category found that courses taken more than once contributed meaningfully to the accumulation of unearned credits. Most of these repeated courses confer credit for only one completion (yet many were successfully completed multiple times). These credits are excluded from the analysis and examined independently [18]. The credit and student feature sequences were then formed with one sequence entry for each *adjusted* term of the student (the words term and semester are used interchangeably in this paper). The terms were *adjusted* to account for the fact that *Summer* is a short semester at the university under study. Compared to regular Fall and Spring semesters, Summer has considerably fewer course offerings and course enrollments. It follows that considering Summer as an independent term may lead to imprecise comparison between terms. Consequently, Summer terms for a student were either merged into Fall or Spring, or left alone, based on the following criteria (in the listed order):

- Summer was combined with Spring if there was an immediately preceding Spring semester available for the student,
- Summer was combined with Fall if there was no immediately preceding Spring semester but an immediately subsequent Fall semester available for the student,
- Summer was left independent if none of the previous conditions were met.

We should mention here that there are other ways to resolve this issue without having to combine Summer with regular terms. For instance, normalizing the credits for each term by a suitable denominator (e.g., earned credits in the case of excess sequence) is another potential solution. Following are some definitions and procedures relevant to the analyses carried out in this paper:

- *Excess Credits*: Usable credits that do not contribute to the completion of the degree program because their contribution to individual requirements exceeds the number of credits required to satisfy these requirements [18].

- *Program Change*: All program (major) changes, except the one from a pre-major to its corresponding major under the same degree (e.g., BA, BS), constitute a program (major) change [18]. Thus, transitions such as pre-major to a different major, pre-major to another pre-major, major to another major, major to a different pre-major, pre-major to a corresponding major but involving a change of degree (e.g., the pre-major was BA and the major is BS), etc., result in a program change.
- *Non-zero statistics (NZStats)*: Aggregate credit hour values for any given cohort of students are represented using the non-zero median simultaneously with the percentage of non-zero credits in this analysis. We refer to this non-zero statistics pairing as NZStats in this paper. NZStats stems from the wider use of non-zero order statistics in our credit hour analysis studies, which we found to be a more representative measure of our credit hour data given the high occurrence of both zeros and large numbers in the various credit hour categories.

4 Experimentation

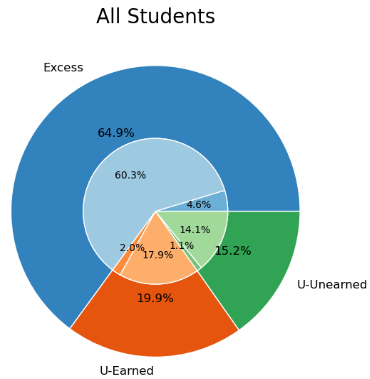
This section presents results and discussion based on data analysis carried out using sequence data to study the credit accumulation patterns of undergraduate students, and using the excess credit sequences and the co-occurrence derived variable proposed in Section 5.1 to understand the effects of some of the common factors influencing extra credit accumulation. We begin our analysis with an investigation into the composition of the *extra* credit category, followed by term-by-term analysis of the excess credit sequence. In particular, excess credit sequences for different student groups were used to investigate the widely held notion of "transfer credit loss" and "program (major) change" as significant contributors to extra credits. Thereafter, we evaluate financial incentives (with the help of various financial aid variables), which are among the prevalent notions of extra credit accumulation, using the derived co-occurrence variables.

4.1 Dataset

We experimented on real transcript data of Bachelor's-degree graduated students from a large public (R1) university. The campus-wide dataset includes 11038 students that graduated between Fall 2015 and Summer 2022. The campus-wide (all) student cohort included students from 114 distinct degree programs. The number of campus-wide *Transfer* students (students that transferred into the university) was 3015. Some noteworthy raw statistics for the undergraduate student group are: 9661 graduated with greater than 0 extra credits, 8841 graduated with greater than 0 excess (Ex) credits, 4100 graduated with greater than 0 unusable-earned (UE) credits, 3389 graduated with greater than 0 unusable-unearned (UU) credits, and 7234 had at least one program (major) change before graduating.

4.2 Composition of extra credits

It can be observed from Fig. 2 that excess constitutes almost two thirds of the extra credit category in terms of sum total credits. The dominance of excess can be further ascertained from the NZStats in Fig. 2. 80% of the students graduate with non-zero excess credits - more than twice the fraction of students graduating with non-zero unusable-earned and unusable-unearned categories, at 37% and 31%, respectively. Half of the students with excess credits accumulate at



	Ex-Tot	Ex-Tx	Ex-NonTx	UE-Tot	UE-Tx	UE-NonTx	UU-Tot	UU-Tx	UU-NonTx
NZStats	10–0.8	3–0.12	9–0.79	4–0.37	3–0.06	4–0.36	6–0.31	4–0.03	6–0.3

Figure 2: A break down of extra credits into its constituent categories for All students. The outer ring in the chart represents the *total* credits, whereas the inner pie breaks down the *total* credits into *transfer* and *non-transfer* credits. The data table displays NZStats with each cell formatted as "$x - y$", where y is the fraction of students with non-zero credit value and x is the median of these non-zero credit values.

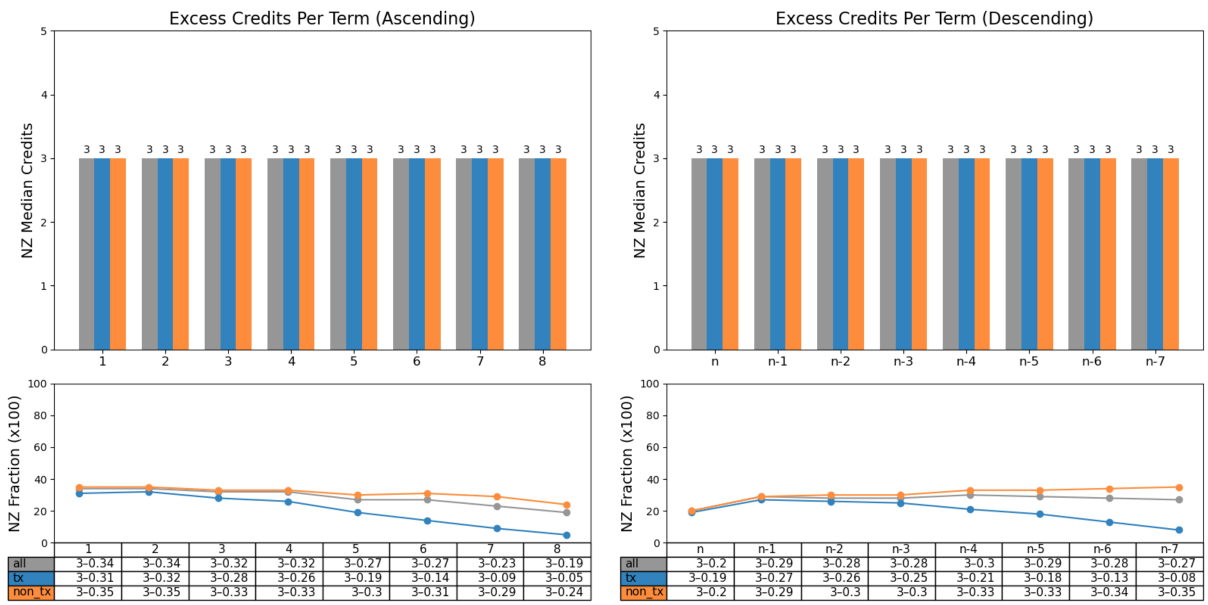
least 10 credits (equivalent to approximately 3 classes). Now that we have established that excess forms the majority of the extra credit category, we will focus our attention on excess credits for the credit efficiency analysis using sequence data presented in the subsequent sections.

4.3 Term-wise credit analysis

This analysis is born of the intuition that distributing a student's *excess* credits over multiple semesters may expose interesting credit accumulation patterns that can help us discover causes of *excess* credits [18]. A secondary aim of this study is to compare the *excess* credit accumulation trends between Transfer students and Non-Transfer students, to explore the idea that transfer credit loss plays a role in the accrual of excess credits. This analysis may also help estimate how many *usable* credits Transfer students enter with. Let us point out here that we are using the *excess* (or *usable*) credits accumulated against the student's degree program of graduation for this analysis. A more direct method (which we plan to implement in the future) would be to perform a credit hour decomposition on a semester-by-semester basis, measured against the program in which students were enrolled in a given semester. Unfortunately, this capability is not yet ready for production use in the *specialized audit tool* at the time of this writing. To conduct this analysis, we can examine the academic terms of a student from either a forward or backward perspective in time:

- *ascending*: starting with the student's entry semester and going forward.
- *descending*: starting with the student's semester of graduation and going back.

This approach situates students in our dataset at a unified initial point in relation to their start or end term, despite their actual starting and concluding semesters varying. Moreover, since most students require more than eight semesters to graduate, tracking them from beginning to end within a manageable number of semesters is impractical. The *ascending* and *descending* method addresses the problem by providing us with insights into students' *excess* credit accumulation pattern from both ends of their undergraduate degree journey. For *ascending*, all the *excess* credits accumulated before the first semester (i.e., E$-$ *excess* credits) are merged with the first semester. For *descending*, if the student's last semester is the one in which the student started, then we merge the *excess* credits accumulated prior to the semester in question into the semester in question.



	1	2	3	4	5	6	7	8
all	3-0.34	3-0.34	3-0.32	3-0.32	3-0.27	3-0.27	3-0.23	3-0.19
tx	3-0.31	3-0.32	3-0.28	3-0.26	3-0.19	3-0.14	3-0.09	3-0.05
non_tx	3-0.35	3-0.35	3-0.33	3-0.33	3-0.3	3-0.31	3-0.29	3-0.24

	n	n-1	n-2	n-3	n-4	n-5	n-6	n-7
all	3-0.2	3-0.29	3-0.28	3-0.28	3-0.3	3-0.29	3-0.28	3-0.27
tx	3-0.19	3-0.27	3-0.26	3-0.25	3-0.21	3-0.18	3-0.13	3-0.08
non_tx	3-0.2	3-0.29	3-0.3	3-0.3	3-0.33	3-0.33	3-0.34	3-0.35

Figure 3: Excess credits accumulated by All, Transfer, and Non-Transfer students per semester (a) going forward (ascending) starting from the first semester (left), and (b) going backward (descending) from the semester of graduation (right). The data table displays NZStats with each cell formatted as "$x - y$", where y is the fraction of students with non-zero credit value and x is the median of these non-zero credit values.

We begin our discussion with campus-wide trends over the semesters. While considering these results, one should remember that Transfer students are expected to graduate in four semesters and Non-Transfer students are expected to graduate in eight, which can lead to significant fluctuations in trends surrounding their anticipated graduation semester. As demonstrated by Figs. 3(a)and 3(b), students accumulate three credits (one class) worth of *excess* credits per semester (as per the the term-wise non-zero median credits), regardless of whether time is (i.e., *ascending or descending*). However, Fig. 3(a) The data reveals that the proportion of students contributing to the non-zero median is highest (approximately one third of students) in the initial two semesters, subsequently showing a steady decline. This pattern holds true for both Transfer and Non-Transfer students, suggesting that students tend to gather excess credits more frequently at the beginning of their academic programs. This finding is further supported by the analysis of the *descending* data from Fig. 3(b). Here, the share of students adding to *excess* credit accumulation is smallest in the graduation semester (about one fifth, akin to the numbers for the eighth term in the *ascending* analysis), then sharply rises in the second-to-last semester before leveling off and gradually increasing as we move towards the student's earlier terms.

Let us now examine the concept of changing programs as a contributing factor to excess credits. Similar to the case of assessing transfer credit loss, students are divided into two groups based on program change (as outlined in Section 3.2): "Change" students with at least one program change, and "No-Change" students with no program change. Analyzing the trends of *excess* credits over time provides insight into the timing of Change students' accrual of these *excess* credits, and allows for a comparison with students who do not change their program. Like the

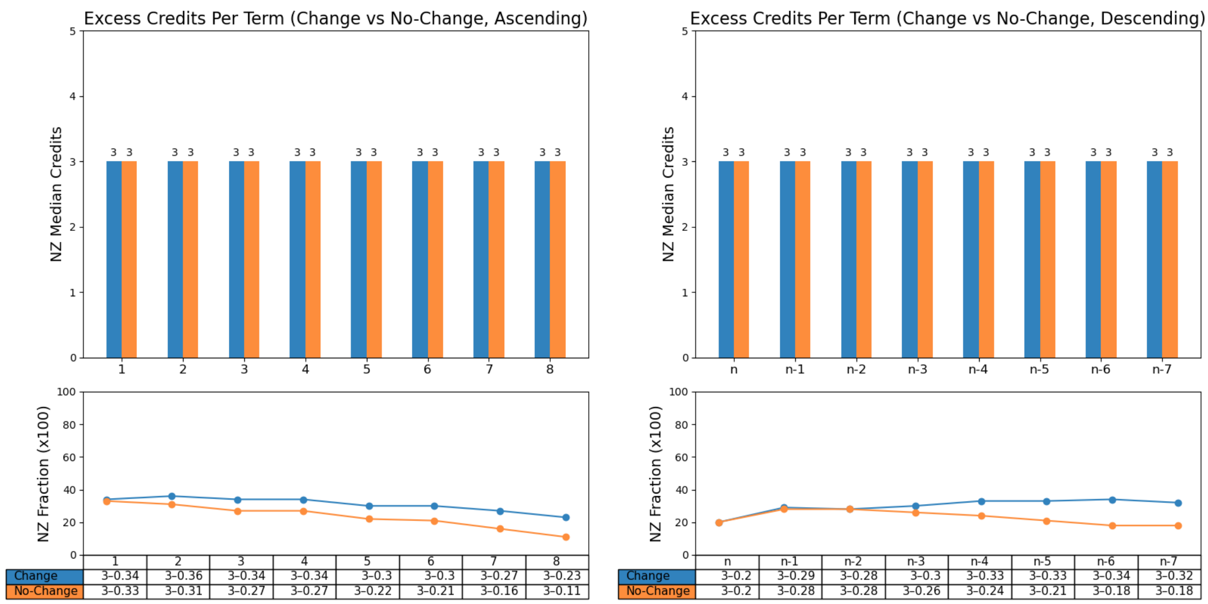


Figure 4: Excess credits accumulated by Change and No-Change students per semester (a) going forward (ascending) starting from the first semester (left), and (b) going backward (descending) from the semester of graduation (right). The data table displays NZStats with each cell formatted as "$x - y$", where y is the fraction of students with non-zero credit value and x is the median of these non-zero credit values.

overarching trends outlined above, Figs. 4(a) and 4(b) demonstrate that both Change and No-Change students accumulate three credits (equivalent to one class) worth of *excess* credits each semester (in terms of the non-zero median), in both *ascending* and *descending* analyses. In line with the overall trends, the *ascending* analysis (Fig.4(a)) reveals that the proportion of students contributing to the non-zero median is highest (around a third of the students) in the initial semester, subsequently beginning to decline. For Change students, this decline does not commence until after the fourth term, while for No-Change students, it begins immediately after the first term. This suggests a tendency for a higher accumulation of *excess* credits during the early terms. The *descending* analysis (Fig. 4(b)) further supports this finding, showing a similar yet slightly more pronounced trend compared to the overall patterns. Additionally, these insights suggest that for Change students, the accumulation of *excess* credits after the final program change is more evenly distributed across several terms.

The discussions presented in this section allude to the fact that there is a similar pattern of *excess* credit accumulation followed by both Transfer and Non-Transfer students while at the university under study. The excess sequences further gave an insight into the timing of excess credits - students generally accrue it in the starting few semesters. As a matter of fact, the largest fraction of students with non-zero excess credits is observed in the first semester regardless of Transfer or Non Transfer students, or Change or No-Change students. Furthermore, more credits are amassed by a larger fraction of Change students as compared to No-Change students in the first four semesters.

5 Analysis using co-occurrence variable

5.1 Co-occurrence feature design

We now define a *co-occurrence* (*co-occ*) measure that quantifies the degree to which the timing of events in one sequence are coordinated with the timing of events in a second sequence. In our case the first sequence will be a sequence of student feature values and the second will be a sequence of excess credit values. We are essentially looking for a particular type of correlation between student features and excess credit values. In particular we are seeking a measure that is not dominated by one of the two sequences. For example, consider the co-occurrence measure:

$$co\text{-}occ = number\ of\ sequence\ positions\ where\ both\ sequences\ are\ non\text{-}zero$$

If one sequence contains mostly zeros then this measure will be close to zero regardless of the other sequence values. Furthermore, if one sequence contains mostly non-zeros then this measure will (approximately) count the number of non-zero entries in the other sequence. In both examples this measure is dominated by one of the two sequences.

Keeping this in mind, let us now define the proposed $co\text{-}occ$ feature. Consider two equal length sequences with zero and non-zero entries. Let us define some terms as follows:

$$co\text{-}occ = \frac{n_{11}}{N_1} - \frac{n_{00}}{N_0} \tag{1}$$

$$\begin{aligned}
&where, \\
&\quad n_{00} = number\ of\ co\text{-}occurrences\ of\ zeros \\
&\quad n_{11} = number\ of\ co\text{-}occurrences\ of\ non\text{-}zeros \\
&\quad N_0 = max(n_{0a}, n_{0b}) \\
&\quad N_1 = max(n_{1a}, n_{1b}) \\
&\quad n_{0a} = number\ of\ zeros\ on\ first\ sequence \\
&\quad n_{0b} = number\ of\ zeros\ on\ second\ sequence \\
&\quad n_{1a} = number\ of\ non\text{-}zeros\ on\ first\ sequence \\
&\quad n_{1b} = number\ of\ non\text{-}zeros\ on\ second\ sequence
\end{aligned}$$

The variables n_{00} and n_{11} capture the co-occurrence of $0s$ and $1s$ respectively. The proposed co-occurrence given by Eq. 1 can be interpreted as below:

- zero value when there is no co-occurrence of $0s$ or $1s$,
- positive value when the co-occurrence of $1s$ dominate,
- negative value when the co-occurrence of $0s$ dominate,
- zero value when the co-occurrence of $0s$ and $1s$ is equal.

The proposed $co\text{-}occ$ feature exhibits the following properties:

- quantifies the amount of "cooperation" between the two sequences (or lack thereof),
- independent of the sequence length, and
- is not dominated by either sequence.

5.2 Application of the co-occurrence feature

Moving ahead, we now employ the *co-occ* derived feature to investigate the influence of financial incentives on excess credit accumulation, thereby also demonstrating the efficacy of the derived variable. Recall that the basic idea behind the *co-occ* is quite straightforward: the co-occurrence of certain student behavior with a particular outcome may indicate a possible causal relationship between them. For instance, does receiving gift aid (financial aid that does not need to be repaid [17]) in a given semester coincide with a spike in excess credits? We chose the following financial aid variables:

- Amount of gift aid paid to the student (Gift-aid),
- Amount of self-help aid paid to the student (Self-help),
- Amount of federal work study paid to the student (Fed-WS).

These variables were chosen because together they encompass the three different types (and thus characteristics) of student financial aid defined by the US Federal Student Aid office [7]: Gift aid not requiring repayment (Gift-aid), aid requiring repayment (Self-help), and aid that needs to be earned (Fed-WS; often earned through student jobs). The term-wise sequences for these financial aid variables, and the excess credit variable, were created as described in Section 3.2. These sequences were then used to derive the following three *co-occ* features: c_{GA}, corresponding to gift aid; c_{SH}, corresponding to self help aid; and c_{WS}, corresponding to federal work study. Based on the non-zero median (10 credits) for the excess credit category seen in Fig. 2, the students were divided into low excess (Low-Ex) and high excess (High-Ex) student cohorts, and the corresponding *co-occ* features were derived. Table 1 provides the summary statistics for these *co-occ* features.

The results for all students show that c_{GA} seems to have the most co-occurrences of gift aid with

Table 1: Summary statistics for financial aid-based *co-occ* variables for all (All), low excess (Low-Ex), and high excess (High-Ex) student cohorts.

		Mean	Std	Median	Q1	Q3	IQ
All	c_{GA}	0.11	0.39	0.12	-0.06	0.38	0.43
	c_{SH}	-0.36	0.46	-0.43	-0.75	0.00	0.75
	c_{WS}	-0.65	0.28	-0.70	-0.88	-0.50	0.38
Low-Ex	c_{GA}	-0.03	0.35	0.00	-0.17	0.20	0.37
	c_{SH}	-0.51	0.43	-0.67	-0.88	-0.15	0.72
	c_{WS}	-0.80	0.20	-0.83	-1.00	-0.71	0.29
High-Ex	c_{GA}	0.31	0.36	0.36	0.10	0.57	0.47
	c_{SH}	-0.16	0.42	-0.25	-0.50	0.13	0.63
	c_{WS}	-0.44	0.23	-0.46	-0.60	-0.30	0.30

excess credits, i.e., high gift aid would suggest high excess. This may be due to the notion that students take additional courses to help meet the requirements for maintaining gift aid, such as GPA or semester credit hours. Also, the low interquartile (IQ) range of c_{GA} suggests consistency in this behavior. Moreover, the median c_{GA} is considerably higher for High-Ex students as compared to Low-Ex students, further reinforcing the above determinations. On the other hand, it can be inferred from the negative median values of c_{SH} and c_{WS} recorded in Table 1 that self-help aid, and especially work-study awards, do not contribute to excess credits. In fact, c_{WS} has the most negative median, as well as a low IQ range, providing the interesting insight that work-study students are more efficient in their credit accumulation pattern (i.e., they cooperate well, with no excess credits). A probable reason may the low number of students with federal work-study awards. Even though C_{SH} has a negative median, it has the highest IQ range (0.75 for all students), which suggests that the co-occurrence of self-help aid with excess credits cannot be assumed to always be low. In other words, self-help aid cooperates neither well nor poorly with excess credits. Again, all these results are strengthened when evaluating the Low-Ex and High-Ex student numbers. Over all, the results from this section reinforced the claim that financial incentives influence excess credits. The derived $co\text{-}oc$ features captured well the amount (or lack) of "cooperation" between financial aid sequences and the excess sequence. It should be noted here that although using the outcome in deriving these features makes them unsuitable for prediction tasks, we are concerned with the explainability of the outcome in terms of student features. The findings in this section also offer promising indications that these derived features may add to the quality of statistical models trying to understand the interaction of different factors affecting excess credits.

6 Conclusion

This paper evaluated the credit hour usage pattern of undergraduate university students using the term-by-term credit sequence data. The focus was on understanding the credit efficiency in terms of the extra credits accumulated by students, a well established problem in higher-education literature due to its undesirable consequences such as delayed time-to-graduation, increased financial burden on the student as well as the public, postponed entrance into the job market, etc. We used revised definition of excess credits that takes into account the *usability* of courses towards the student's degree program requirement. Student transcript-level data was used to produce credit decomposition using a specialized audit tool [10] that (automatically) allocated credit to the applied, excess, and unusable categories. Excess, along with the subcategories of unusable (unusable-earned and unusable-unearned), formed the extra credit category. Given the dominance of excess credits within the extra category (making up almost two thirds of the extra category) apparent from initial analysis, the study was realigned to focus on the accumulation of excess credits, and the common factors perceived to be affecting it, such as transfer credit loss, program (major) change, financial incentives, etc. The term-by-term analysis of excess credit sequence revealed the similar excess credit accumulation pattern of Transfer and Non-Transfer students, and that the students accumulate excess credits earlier in their degree. Contrasting the excess credit sequences of students with no program change (No-Change) to those with at least one program change (Change) revealed that a larger fraction of Change students accumulate excess credits in the first few semesters. A feature derivation method was proposed to map a relevant sequence into a feature value that attempts to capture information that is relevant to the current outcome. This derived feature was termed *co-occurrence* (*co-occ*) and was used to study the impact of various types of financial aid on excess credit accumulation. The results suggested that gift aid was instrumental in generating excess credits.

Limitations and future work: This study is limited to only graduated students at present, and used the excess credits accumulated against the student's degree program of graduation to perform this analysis. We plan to carry out this analysis more directly – via a credit hour decomposition on a semester-by-semester basis measured against the program in which students were enrolled in a given semester. The derived co-occurrence variable was only used to study the influence of financial aid on excess credits. There are opportunities to employ it in statistical modeling aimed at explaining the factors in the accumulation of excess credits, and the relationships among those factors. This is a direction that we are actively pursuing.

References

[1] C. Adelman. Principal indicators of student academic histories in postsecondary education, 1972-2000, January 2004. Retreived from `https://files.eric.ed.gov/fulltext/ED483154.pdf`.

[2] American Council on Education. Carnegie classification of institutions of higher education. Retrieved from `https://carnegieclassifications.acenet.edu/`.

[3] CollegeTransfer.Net. What are degree requirements? Retrieved from `https://www.collegetransfer.net/AskCT/What-are-Degree-Requirements`.

[4] Complete College America. Time Is the Enemy: The Surprising Truth about Why Today's College Students Aren't Graduating... and What Needs to Change, 2011. Retrieved from `https://files.eric.ed.gov/fulltext/ED536827.pdf`.

[5] Complete College America. Four-Year Myth: Make College More Affordable. Restore the Promise of Graduating on Time, 2014. Retrieved from `https://files.eric.ed.gov/fulltext/ED558792.pdf`.

[6] J. P. Cullinane. *The Path to Timely Completion: Supply- and Demand-Side Analyses of Time to Bachelor's Degree Completion*. PhD thesis, University of Texas at Austin, 2014.

[7] Federal Student Aid. Types of Aid. Retrieved from `https://studentaid.gov/resources/types-infographic-accessible`.

[8] J. Fink, D. Jenkins, E. M. Kopko, and F. X. Ran. Using data mining to explore why community college transfer students earn bachelor's degrees with excess credits, February 2018. Retrieved from `https://ccrc.tc.columbia.edu/media/k2/attachments/using-data-mining-explore-why-community-college-transfer-students-earn-bachelors-degrees-excess-credits.pdf`.

[9] S. K. Hargrove and D. Ding. An Analysis of B.S.I.E. Degree Completion Time at Morgan State University. In *International Conference on Engineering Education*. International Network for Engineering Education and Research, October 2004.

[10] D. R. Hush, E. S. Lopez, W. Al-Doroubi, T. Ojha, B. Santos, and K. Warne. Analyzing student credits. 2022. Retrieved from `https://digitalrepository.unm.edu/ece_rpts/55`.

[11] M. Kapur. Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning*, 6:39–56, 2011.

[12] A. Karimi and R. D. Manteufel. Factors influencing student graduation rate. In *2013 ASEE Gulf-Southwest Annual Conference*. American Society for Engineering Education, March 2013.

[13] W. Kilgore, E. Crabtree, and K. Sharp. Excess credit accumulation: An examination of contributing factors for first-time bachelor's degree earners. *Strategic Enrollment Management Quarterly*, 6(4):41–56, 2019.

[14] D. A. Kramer, M. R. Holcomb, and R. Kelchen. The costs and consequences of excess credit hours policies. *Educational Evaluation and Policy Analysis*, 40(1):3–28, 2018. Retrieved from `http://www.jstor.org/stable/44985444`.

[15] M. J. Mahzoon, M. L. Maher, O. Eltayeby, W. Dou, and K. Grace. A sequence data model for analyzing temporal patterns of student data. *Journal of learning analytics*, 5(1):55–74, 2018.

[16] I. Molenaar. Advances in temporal analysis in learning and instruction. *Frontline Learning Research*, 2(4):15–24, 2014.

[17] National Association of Student Financial Aid Administrators (NASFAA). Glossary of Terms for Financial Aid Offers, 2020. Retrieved from `https://www.nasfaa.org/glossary`.

[18] T. Ojha, D. R. Hush, and E. S. Lopez. Credit hour analysis. 2023. Retrieved from `https://digitalrepository.unm.edu/ece_rpts/57`.

[19] J. Saint, Y. Fan, D. Gašević, and A. Pardo. Temporally-focused analytics of self-regulated learning: A systematic review of literature. *Computers and education: Artificial intelligence*, 3:100060, 2022.

[20] J. M. Shedd. The history of the student credit hour. *New Directions for Higher Education*, 2003(122):5–12, 2003. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/he.106`.

[21] Texas Higher Education Coordinating Board. Transfer issues advisory committee report: Identifying and closing the gaps, 2001. Retrieved from `http://www.thecb.state.tx.us/reports/pdf/0427.pdf`.

[22] J. Wellman. The student credit hour counting what counts. *Change: The Magazine of Higher Learning*, 37(4):18–23, 2005. `https://doi.org/10.3200/CHNG.37.4.18-23`.

[23] M. Zeidenberg. Valuable learning or “spinning their wheels”? understanding excess credits earned by community college associate degree completers, April 2012. `https://doi.org/10.7916/D89G5VX0`.

[24] R. E. T. Zeno and R. A. Hansen. An empirical study of degree completion patterns. *NACADA Journal*, 20(1):42–46, November 2000. `https://doi.org/10.12930/0271-9517-20.1.42`.