2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #38137

Progress Analytics in Support of Engineering Advising and Program Reform

Husain Al Yusuf, The University of Arizona

Husain Al Yusuf is a second year PhD student in the Electrical and Computer Engineering Department at the University of Arizona. He is currently pursuing his PhD with a research focus on higher education analytics, with the goal of improving student outcomes and enhancing the effectiveness of higher education institutions.

Husain Al Yusuf holds a M.Sc in Computer Engineering from the University of New Mexico and has over 10 years of professional working experience as a technology engineer

Prof. Gregory L. Heileman, The University of Arizona

Gregory (Greg) L. Heileman currently serves as the Associate Vice Provost for Academic Administration and Professor of Electrical and Computer Engineering at the University of Arizona, where he is responsible for facilitating collaboration across campus t

Raian Islam, The University of Arizona

Raian Islam is a current Master's student and a Graduate Research Assistant in the Department of Electrical and Computer Engineering at The University of Arizona, Tucson, AZ, USA. She received her BSc. degree in Electrical and Electronic Engineering from Ahsanullah University of Science and Technology, Dhaka, Bangladesh, in 2019. Her current research interests include higher education data analytics, machine learning and photovoltaics.

Yiming Zhang, The University of Arizona

Yiming Zhang is a Ph.D. candidate in the Electrical and Computer Engineering Department at the University of Arizona. His research interest includes machine learning, data analytics and optimization in curricular design.

Tanmay Agrawal

Tanmay is a senior undergraduate student majoring in Computer Science with a minor in Entrepreneurship and Economics and worked on the blockchain-based reward system as well as building dashboards to understand studen outcomes. Tanmay is originally from India, and enjoys learning new things. In addition to his technical pursuits, he has a keen interest in discovering novel ideas, especially related to the social and political dimensions of our everyday existence. He relishes examining the broader context of all things.

Mr. Hayden William Free, Georgia Institute of Technology

Hayden Free is an Master's student studying Computer Science at the Georgia Institute of Technology. His focused area of interests include network science, machine learning, and software design.

Kristina A. Manasil, The University of Arizona

Kristi Manasil is a first-year PhD student in the School of Information at the University of Arizona. She received her bachelor's degree in Computer Science from the University of Arizona. She is interested in data visualization, machine learning, human computer interaction, and learning analytics.

Progress Analytics in Support of Engineering Advising and Program Reform

Husain Al Yusuf,[†] Gregory L. Heileman,[†] Raian Islam,[†] Melika Sharifi,[†]
Roxana Sharifi,[†] Yiming Zhang,[†] Tanmay Agrawal,[‡] Bhavya Sharma,[‡]
Kristi Manasil,[•] Hayden W. Free[⋆]

{halyusuf, heileman, raiani, akbarsharifi, roxanaa, yimingzhang1, tanmayagrawal21, kmanasil, bhavyasharma}@arizona.edu, hayden.free@gatech.edu

[†]Department of Electrical & Computer Engineering
[‡]Department of Computer Science
[•]School of Information
University of Arizona

[⋆]School of Computer Science
Georgia Institute of Technology

Abstract

Students in engineering programs are typically among those having the highest time-to-degree for any of the programs offered on a university campus. Keeping a cohort of students on track towards on-time graduation is extremely difficult given the tightly prescribed nature of engineering programs. Any deviation from the standard degree plan, for any reason, including not passing a class, taking courses out of sequence, etc., often precludes the ability to graduate in four years. In this paper, we describe a cohort tracking analytics platform that can be used by advisors as an aid in keeping students on track, and by program administrators as a tool to better understand the curricular impediments associated with delays in graduation. This cohort analytics platform provides analyses over a population of students, rather than individual students, yielding valuable (often hidden) information regarding the impediments that students face. For instance, this platform makes it easier to determine what courses are most significant in blocking the progress of a cohort, the efficiency of credit hour production within a cohort, where students are losing credit hours (i.e., generating credit hours that do not count towards the satisfaction of any degree requirements), etc.

Advisors and administrators often suggest programmatic improvements based on anecdotal evi-

dence and experiences related to individual students, not because they are lazy, but because it is inherently difficult to compute cumulative student progress over a cohort. The reason for this is that accurate student progress information is typically difficult to obtain and out of reach for many decision makers, as degree audit capabilities have not been designed with analytics in mind. In an attempt to make this data accessible and actionable, we have developed a platform that can organize student cohorts according to any criteria, and compute progress analytics relative to these cohorts, while also providing useful analytics and visualizations in an appealing and easy-to-understand format. At the core of the platform is a database that stores program degree requirements and student data, as well as a progress reasoner and a curricular analytics engine that can compute cohort-based metrics and display them on an interactive dashboard. The architecture of this platform will be described in this paper, as well as the types of data that must be collected in order to use this platform effectively. We will also discuss the characteristics of cohort-based analytics that have emerged from the study of engineering programs, and how they differ from those generated from non-engineering programs.

Introduction

The revolution in business analytics that has occurred over the past decade, and the concomitant use of predictive analytics to support business planning, was driven by the recognition that transactional data associated with business operations must be restructured in order to make it useful for analytic purposes. Specifically, the gap between transactional processing, the point of origin for most business data, and the analytics-based insights created by manipulating transactional data, is too wide. Data platforms (e.g., data warehouses[1] data lakes,[2] etc.) separate from transactional systems were created to close this gap. These systems align data extracted from transactional systems so as to make it amenable for analytic-based processing, thereby unlocking valuable information contained within business data.

Institutional research offices across higher education have learned much from the business analytics community, and have worked to create their own data platforms in support of institutional analytics.[3–5] The principles are the same: large amounts of transactional data related to conducting the business of higher education are collected in student information systems, learning management systems, registration systems, etc. Many in higher education have created data platforms to mine these data in efforts to improve student success outcomes, as well as other outcomes related to institutional performance. A key missing piece of the higher education analytics puzzle, however, is the ability to easily perform progress-based analyses over cohorts of students, e.g., all of the student currently enrolled in the college of engineering. This type of analysis involves three necessary components, (1) individual student performance data, e.g., courses taken and grades earned; (2) the degree requirements associated with all of the degrees students in a given cohort are pursuing; and, (3) a reasoning algorithm that can reconcile (1) and (2), i.e., the ability to determine how the coursework on student transcripts applies towards the satisfaction of the requirements in the degree programs they are pursuing. Using these three components it is possible to create summary statistics and perform analytics over any defined cohort of students. In doing so, important insights can be gained about the impediments to on-time graduation, and very accurate predictions can be made regarding the expected graduation rate of a cohort.

In this paper we describe a cohort analytics capability we have constructed that can provide key analytical insights to college- and department-level administrators regarding how the students in the programs they offer are progressing towards graduation. In the next section we describe the data requirements of this application, as well as the difficulty of the combinatorial optimization problems involved in this work. Next, we describe the reasoning algorithm we are using to generate our cohort-based analyses. After that, we describe the cohort analytics dashboard that has been built on top of the aforementioned analytics engine. Finally, we provide some useful concluding remarks.

Not Your Father's (or Mother's) Degree Audit

Astute readers may have noticed that parts of what was described above sounds suspiciously similar to traditional degree audit systems. Indeed, degree audit systems, typically buried deep within student information (i.e., transactional) systems, are designed to analyze a student transcript in order to determine the degree requirements it satisfies. This is routinely done in order to certify whether or not a student is eligible to graduate. This process, however, is extremely difficult to repackage for the purposes of more general analytics aimed at better understanding the issues impacting student progress that only come to light when analyzing a larger collection of students. Furthermore, we have found that vendors of degree audit solutions are highly resistant to this idea. In addition, there is a dearth of literature on how to effectively represent degree requirements, and how to reason over them to determine requirements satisfaction. For instance, even the difficultly of this problem is not well understood. Thus, in order to better frame this problem, it is important to formally define and understand the data elements.

Curricula, Degree Plans and Degree Requirements. In order to earn the credential associated with an academic program (e.g., BS in Computer Science), a student must satisfy all of the *degree requirements* associated with the academic program. These degree requirements, almost exclusively, are satisfied by earning sufficient grades in the specific courses that constitute each degree requirement. The particular collection of courses used to satisfy a program's degree requirements is referred to as a *curriculum*. It is generally the case that many different curricula can satisfy a given set of degree requirements. For instance, students are given freedom in how to select the courses needed to satisfy technical electives, general electives, etc., and each choice produces a slightly different curriculum. The notion that a single set of degree requirements can be satisfied by numerous curricula is depicted in Figure 1.

The entire set of degree requirements for an academic program can be represented as a Boolean formula, and it should be noted the formula for real degree requirements are exceedingly complicated. A realistic set of degree requirements contains on the order of fifty sub-requirements, and some of these sub-requirements might be satisfied by taking some combination of more than one hundred different courses. Thus, for most programs we can conservatively estimate that tens of thousands of curricula can be constructed to satisfy a set of degree requirements. On the left side of Figure 1 we denote this by showing that a single set of degree requirements can lead to many different curricula depending upon the course options selected. This complexity serves to highlight the importance of academic advisors, who routinely work with students to select a collection of courses, i.e., a curriculum, that will allow them to satisfy degree requirements as efficiently as possible.

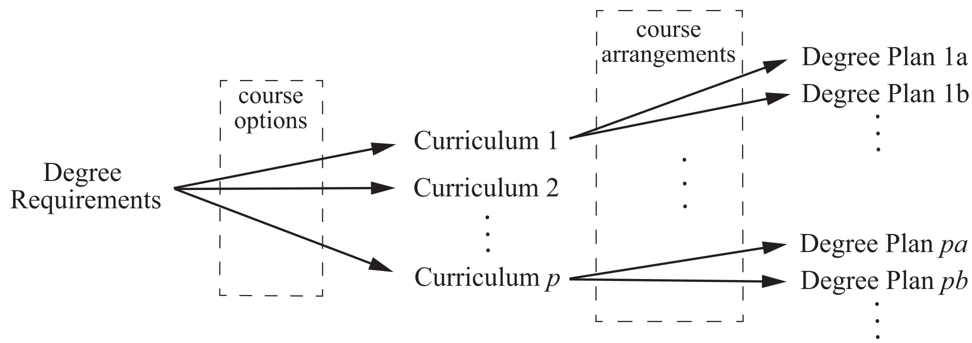

Figure 1: The relationships between degree requirements, curricula and degree plans.

Finally, a *degree plan* is a term-by-term arrangement of the courses in this curriculum constructed so as to satisfy the prerequisite relationships between the courses in the curriculum. The fact that a single curriculum can lead to many different degree plans, depending upon how the courses are arranged, is depicted on the right side of Figure 1. Furthermore, the number of possibilities is much larger than many realize. If we consider a typical bachelor's degree curriculum consisting of forty courses, there more than a million different ways to arrange these courses over eight terms when there are no prerequisites. By accounting for prerequisites, we might reduce the possibilities by a factor of ten or more, but we are still dealing with a massive number of possible degree plans.

Requirements Satisfaction Using the data models described above, degree program requirements satisfaction can be treated as a Boolean formula satisfiability problem. This involves constructing a formal model for representing a set of degree requirements as a Boolean formula, and then treating the grades extracted from a student transcript as Boolean variables. The problem of determining whether or not a particular assignment of grades to these variables satisfies the underlying Boolean formula is a strongly $\mathcal{NP}$-complete problem. We have developed an integer linear programming (ILP) algorithm that solves this assignment problem optimally, and it forms the basis of the underlying analytics engine in our cohort analytics application.

Using the progress analysis results returned by the ILP algorithm, cohort-based statistics can be computed. The term *progress* in this work refers to the extent to which students within a given cohort have satisfied the degree requirements of their respective degree programs. Specifically, for a single student, we define *progress* as the percentage of the requirements completed relative to the complete set of degree requirements that must be satisfied in order for the student to earn a degree in a given program. Summary statistics can then be created by accumulating the progress of individual students within a cohort. Because the various requirements in a degree program may have differing numbers of credit hours associated with them, it does not make sense to measure progress in terms of the number of degree requirements that have been satisfied, rather, a more effective unit of progress will be the *earned credit hour*, as all degree requirements can be "normalized" to the number of credit hours they encompass.

Consider an academic program p, where the minimum number of credit hours required to satisfy the degree requirements in p is denoted $c(p)$. Next consider an individual student x pursing a degree in program p, with a transcript $t(x)$ listing the coursework completed by this student. If we let $c(t(x))^p$ denote the number of earned credit hours on student x's transcript that apply towards the satisfaction of degree requirements in program p, then we can define the progress of student x in degree program p as

$$\gamma_x^p = \frac{c(t(x))^p}{c(p)}. \tag{1}$$

Since $0 \leq c(t(x))^p \leq c(p)$, and $c(p) > 0$, γ_x^p must be in the range $[0, 1]$. Equation (1) can be used to create summary progress statistics over any cohort of students.

We are often interested in defining the progress of a student or a cohort of students relative to some schedule *milestone*. For instance, we regularly report four-, five-, and six-year graduation rates, and therefore we would also like to analyze progress relative to these same milestones; that is, progress towards earning a degree in four, five, or six years. For a given milestone, we can categorize the status of each student in terms of their progress towards a milestone as being either: (1) ahead of schedule, (2) on track, or (3) behind schedule. Because students receive grades on a term-by-term basis, the granularity of the milestone specifications should be at the level of a term. We further assume that students should make steady and uniform progress on a term-over-term basis. For instance, a student on track to graduate in four years should have earned roughly 12.5% of the credits required in their degree program after one semester, roughly 25% of the credits required in their degree program after two semesters, etc.

The creation of conditions that facilitate student success has been a focus throughout higher education over the past two decades.[6–10] The value in analyzing progress relative to cohorts is that group-based statistics can reveal systemic issues impacting a large number of students that are often difficult to discern when considering the progress of a single student. Furthermore, group-based analyses naturally lead to improvement-based conversations around the collective impact associated with programmatic changes. For instance, "We have found that 55% of first generation students are not able to complete this course in the first term; if we can improve the pass rate of the course by 15%, we expect a much larger percentage of our first generation students will be retained, and our four-year graduation rate will improve by 5%. Can we think of ways to support this course that would lead to better outcomes?" We have found that faculty are far more likely to respond to improvement opportunities when the outcomes are made tangible.

Cohort Analytics Dashboard

The cohort analytics dashboard, shown in Figure 2, displays the progress of an entire defined cohort as well as a histogram that groups students by completion rate as well as categorizing them into on-track and off-track. It also computes the efficiency of the cohort as described above. Additional information provided includes which courses most commonly do not count towards the satisfaction of a degree requirement (among the programs students in the cohort are pursing), the courses that students have difficulty in completing, and the requirements that are least completed among all of the program requirements students in the cohort must complete. We believe this dashboard is an invaluable tool for faculty and program administrators in making data-informed decisions that can have a positive effect on improving the efficiency of engineering program pathways, and will ultimately lead to better student success outcomes.

The cohort tracking analytics dashboard aims to provide information that is useful in understanding and gaining insights into the performance and progress of engineering students so that fac-

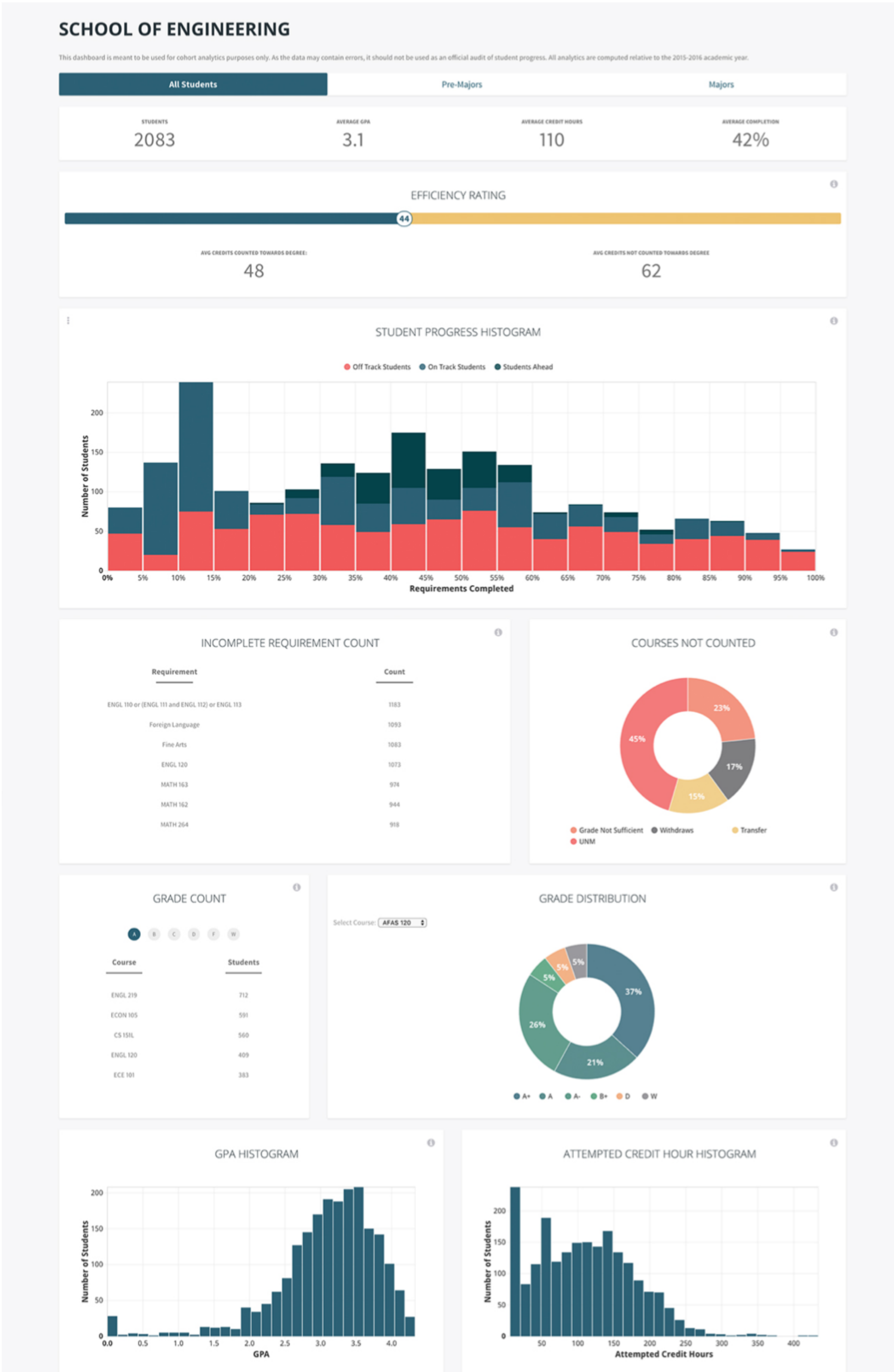

Figure 2: Summary cohort statistics provided by the cohort analytics dashboard.

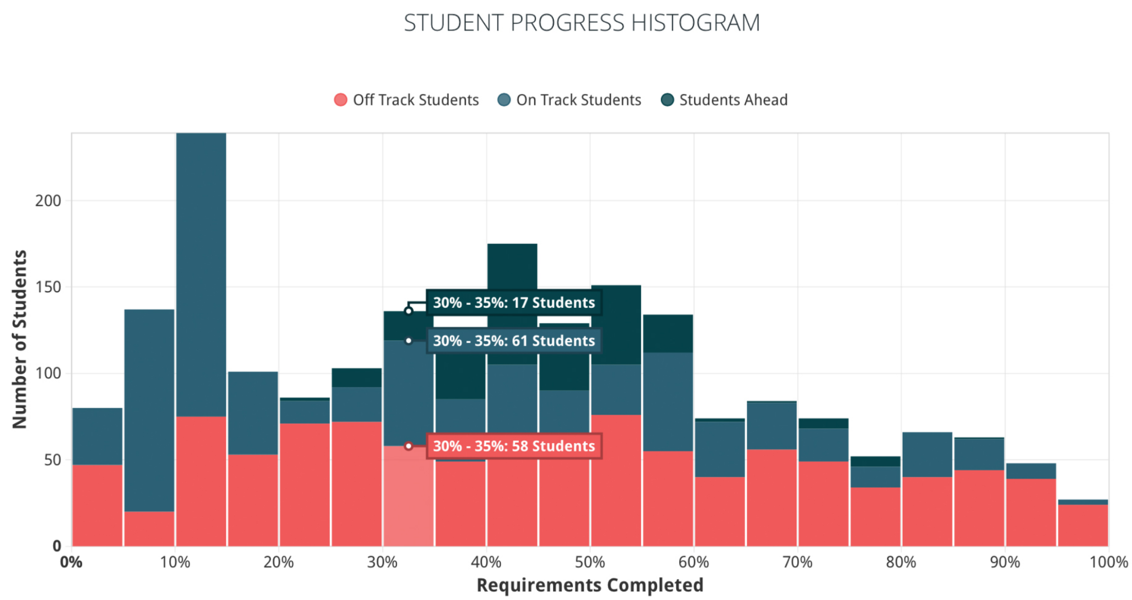

Figure 3: The cohort tracking dashboard's progress histogram showing the progress and status of students in the School of Engineering. The highlighted bar shows that 136 students are between 30% and 35% complete, and of these, 58 are off-track, 61 are on-track, and 17 are ahead of schedule in graduating within four years.

ulty/administrators might discover ways to better serve them. We believe the charts presented in the dashboard provide information useful in making these discoveries. At the heart of these analytics is the measure of progress of a cohort of students towards satisfying their degree requirements (described above); however, other information that could be useful to faculty/administrators in evaluating their engineering programs is also provided. In all, there are ten main components to the dashboard: (1) cohort progress efficiency, (2) cohort credit hour efficiency, (3) student progress histogram, (4) top requirements students have yet to complete, (5) breakdown of credit hours not counted, (6) most frequent course sorted by grades earned, (7) grade distributions per course, (8) GPA histogram, (9) credit hour histogram, and (10) individual student audits.

Cohort View. The centerpiece of the dashboard is a histogram that displays the progress and status of the students within a cohort. An example can be seen in in Figure 3, where a user has hovered over the histogram bar of those students who are between 30%–35% complete. The cohort in this case is the set of all student enrolled in a particular program; however, it is possible to track the progress of any cohort a user may wish to define, e.g., first generation students, full-time students who started in a particular year, all of the students who participated in a summer bridge program, etc. The tool tip highlighted in this figure shows that breakdown of how many students are in each category within this bin. This is visualized in the histogram by separating each bar (representing students within a progress-range) into color-coded segments that each correlate to one of the aforementioned statuses.

A important factor in graduating students on time is making sure their credit hour generation is efficient. The dashboard sheds light on this issue by introducing a metric that quantifies a cohort's efficiency; this metric is positioned at the top of the dashboard shown in Figure 2. Engineering programs tend to have strict curricula that are unforgiving when not precisely followed. As a result, we have found that it is common for engineering students to graduate with far more credit hours than they actually need, as compared to students in other colleges. This indicates that engineering students are often not as efficient as they could be in generating credits that count towards their degree, which could lead to longer average times-to-degree. We define it as the ratio of credit hours that count towards a degree over the total number of attempted credit hours. This measure is computed for every student and then averaged to give the efficiency of the cohort as a whole.

Although it is very useful to know how efficient a cohort of students is in taking courses that count towards their degree, it is equally useful to know where the inefficiencies lie. Students might not be generating useful credits for a variety of reasons and knowing these reasons might help inform interventions aimed at reducing time-to-degree. A by-product of the reasoning engine determining which courses satisfy a set of requirements is knowing the courses that were taken that do not contribute towards earning a degree. These courses can then be sorted into four categories: (1) courses that simply do not satisfy any degree requirements, (2) transfer courses that do not satisfy any degree requirements, (3) courses that a student withdrew from, and (4) courses in which the student did not not make a sufficient grade. To visualize the frequency in which these cases occur, the dashboard provides a donut chart (see Figure 2) showing the percentage in which each category accounts for credits that do not apply towards degree requirement satisfaction. In addition, by selecting a category, a list of course attempts that fall into the selected category are shown, ranked by frequency. Therefore faculty/administrators can easily see why students are generating unusable credit hours and the courses that contribute the most to this inefficiency.

Rounding out the dashboard shown in Figure 2 are several other charts and graphs that provide basic information about the cohort and the performance of students within it. Two of these charts are GPA and attempted credit hour histograms. These simply visualize two metrics that are often used to measure student performance. To see how students perform at the course level two other charts are given. One displays the top five courses in which students most commonly receive a given grade, e.g., a user can see which courses students most commonly receive a “C” in. Also a donut chart shows grade distributions for a selected course in the program. Although these charts might not present novel information, we believe that they can still be useful and the dashboard provides a convenient way to obtain this information.

Individual Student View. Although the dashboard’s main purpose is to display information over a cohort of students, the dashboard also provides views showing individual students and the individual’s degree audit. An example is provided in Figure 4. These views are provided so that users can see which students make up a set of students represented in charts within the graph. For example, by selecting one of the bars in the progress histogram shown in Figure 3, a list of students corresponding to the selected progress range and status will be displayed. This behavior is also present in the GPA and credit hour histograms, as well as in the grade distribution donut chart.

Selecting a student from the list brings up an individualized dashboard that shows general information about the student, their credit hour efficiency, and a degree audit. The audit is divided into four segments: satisfied requirements, partially satisfied requirements, unsatisfied requirements,

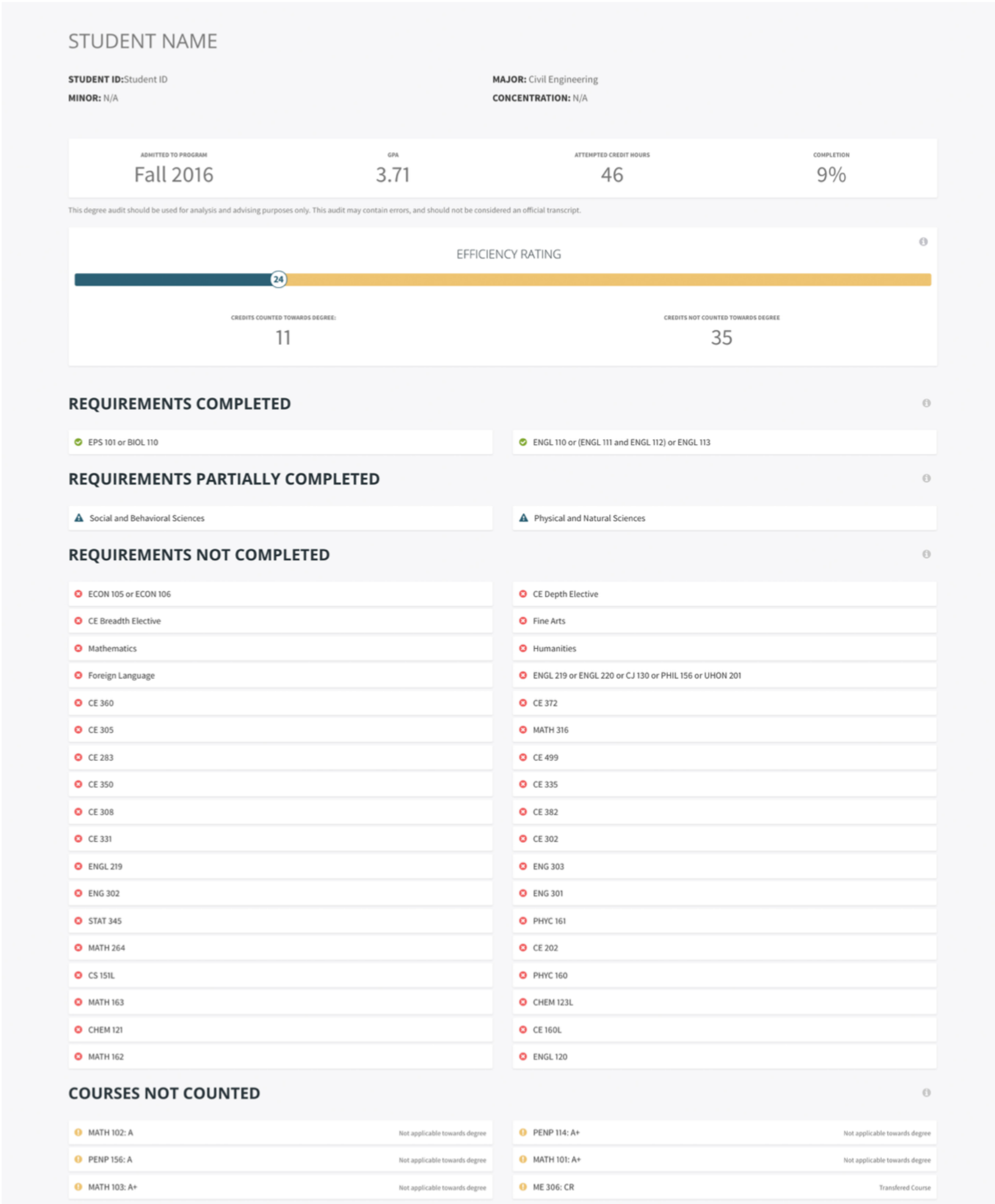

Figure 4: An individual student audit.

and courses that a student took that do not contribute to their major. In the list of satisfied requirements, the course that satisfied a particular requirement is listed and in the list of courses that do not count towards a degree, the reason for the course not counting is given. We believe that these individual student views allow for deeper investigation and will be particularly useful to advisors in giving students informed advice as well as diagnosing inefficient behavior early.

Discussion

The type of analytics supported by the cohort tracking dashboard can be referred to as *action analytics*; that is, analytics conducted for the purpose of making decisions, rather than simply generating reports.[11,12] Even in the short amount of time the cohort analytics dashboard has been deployed, it has proven itself immensely valuable to this end. One area that it has proved useful is simply identifying which students are within a semester or two of graduating. Traditionally this is done based on completed credit hours, followed by manual degree audits. The dashboard shows students are not always efficient in accumulating credits that count towards their degree. While some students might seem to have sufficient credits to put them close to graduation, many of their credit hours may not count towards their degree and therefore they are several semesters away from finishing. Being able to accurately determine the students close to graduating saves time going through individual student records and gives university personnel the ability to provide these students with the classes and other resources needed to complete their degree within the target time.

The dashboard was also useful in dispelling myths regarding where engineering students are generating unusable credit hours. For instance, we have often heard that students in an engineering program are off track due to transfer work that was accepted, but not counting towards the degree. While this is not uncommon, the dashboard did not reveal transfer courses to be the largest source of unusable credits. Instead, it was learned that the majority of non-counting credits were due to courses taken at the home (not transfer) institution, but not applicable for an engineering degree. Upon further inspection the courses that topped this category were math courses that serve as prerequisites to Calculus I—the math course students are expected to start at in order to graduate in four years. As a result of starting in these lower-level math courses, students will be delayed at least one semester. This has begun discussions on how engineering curricula can be redesigned to facilitate students that do not begin in Calculus I, but allowing them a path to graduate within four years.

As more administrators use the application we plan to collect feedback to provide more analytics that might prove helpful. A logical next step is to use the dashboard for predictions. Knowing how far students have progressed in a program is already helpful in getting a feel for when students might graduate, and incorporating this data into formal prediction models could yield better graduation rate predictions. In addition, knowing which requirements student have yet to complete could help administrators plan class offerings to ensure that students are able to take the courses they need to remain on track. However, even in its current state, we believe that the current dashboard lays the groundwork for better data-driven decision making and plan to continue adding useful cohort-based analytics. We believe faculty and administrators will continue to find insights using the dashboard that can improve student success outcomes in engineering programs.

References

[1] Ralph Kimball and Margy Ross. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. John Wiley & Sons, 3rd edition, 2013.

[2] Philipp Wieder and Hendrik Nolte. Toward data lakes as central building blocks for data management and analysis. *Frontiers in Big Data*, 5, 2022.

[3] P. Baepler and C. J. Murdoch. Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, 4(2):17, 2010.

[4] A. G. Picciano. The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, 16(3):9–20, 2012.

[5] M. D. Pistilli and K. E. Arnold. Purdue signals: Mining real-time academic data to enhance student success. *About Campus*, 15(3):22–24, 2010.

[6] Tommaso Agasisti and Alex J. Bowers. Data analytics and decision making in education: Towards the educational data scientist as a key actor in schools and higher education institutions. In *Handbook of Contemporary Education Economics*. Edward Elgar Publishing, Cheltenham, UK, 2017.

[7] Beth Dietz-Uhler and Janet E. Hurn. Using learning analytics to predict (and improve) student success: A faculty perspective. 12(1):17–26, 2013.

[8] Nate Johnson. Three policies to reduce time to degree. Technical report, Complete College America, Washington, DC, February 2011.

[9] George D. Kuh, Jillian Kinzie, John H. Schuh, and Elizabeth J. Whitt. *Student Success in College: Creating Conditions That Matter*. Jossey-Bass, San Francisco, CA, 2010.

[10] Vincent Tinto. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. University of Chicago Press, Chicago, IL, 2nd edition, 1987.

[11] D. Norris, L. Baer, J. Leonard, L. Pugliese, and P. Lefrere. Action analytics: Measuring and improving performance that matters in higher education. *EDUCAUSE Review*, 43(1):42, 2008.

[12] Angela van Barneveld, Kimberly E. Arnold, and John P. Campell. Analytics in higher education: Establishing a common language. In *Educause Learning Initiative*, January 2012. ELI Paper 1.