Extending Academic Analytics to Engineering Education

Anthony M. DeRocchis[1], Ashley Michalenko[2], Laura E. Boucheron[1], and Steven J. Stochaj[1]
[1]Klipsch School of Electrical and Computer Engineering
[2] Department of Computer Science
New Mexico State University
Las Cruces, NM 88003
Email: {tonydero, ash702, lboucher, sstochaj}@nmsu.edu

***Abstract*—This Innovative Practice Category Work In Progress paper presents an application of machine learning and data mining to student performance data in an undergraduate electrical engineering program. We are developing an analytical approach to enhance retention in the program especially among underrepresented groups. Our approach will provide quantitative assessment of student performance in courses. Specifically, by hierarchically mapping the content of assignments to course learning objectives, we can better decipher which concepts a particular student is struggling with and, with the help of peer mentors, create tailored intervention techniques to help the student be successful in the program. These results will also be useful to academic advisors who can work with the student to determine class schedules that promote success in the program. In addition, students can take a proactive approach to their learning. In our approach, data from our learning management system and other available sources will be used to predict several outcomes for individuals such as when a student is beginning to have trouble with the material or if factors outside of the classroom are affecting their success. Here, we present our initial database schema and preliminary results relating number of class re-takes to time-to-graduation.**

I. Introduction

It is increasingly recognized that there are significant retention problems in Science, Technology, Engineering, and Math (STEM) fields [1]–[4] and that these issues appear to be more prevalent among underrepresented minorities (URMs) and women [3], [4]. As a Minority Serving Institution and Hispanic Serving Institution, we are in a unique position to address retention issues and increase completion rates for STEM degrees among URM students, including URM women. In this project, we adopt a data-driven approach to tracking New Mexico State University students' academic performance, determining indicators of potential academic performance problems, and matching academic indicators to appropriate interventions. This will enable and empower all students, including URM students, to take a more active role in their academic career by proactively addressing academic behaviors that put them at risk, and will improve their academic performance and reduce their time-to-degree [5]. In this project, we initially focus on Electrical & Computer Engineering (ECE) students.

In this paper, we discuss our initial progress on our Institutional Review Board approved data-driven analysis of student performance. We present a brief overview of our project in Section II. In Section III we discuss the design of our database management system. In Section IV we present our preliminary work studying the effect re-taking a course has on time-to-graduation. Finally, in Section V we conclude.

II. Project Overview

A. Academic Analytics

The use of data-driven analytics to flag academic performance problems is expected to provide a significant increase in the ability to implement and maintain "intrusive" or "proactive" advising [5], [6] of at-risk students within ECE. We ultimately want to quantitatively track academic performance of students on a weekly basis to identify indicators of academic performance (e.g., poor scores on a homework assignment or quiz). Professors commonly assemble recommendations for students (e.g., high homework scores along with low exam scores may indicate test anxiety), but these are heuristic, incomplete, simplistic, slow to adapt to changing circumstances, and only reach those students whom they advise. Our data-driven approach will generate recommendations that are objective, comprehensive, adaptable, and immediately available to students. This approach will also help with assessment and accreditation of the curriculum [7] by providing quantitative assessment of student learning at a more granular level.

We will begin with manual assignment of interventions (e.g., tutoring, study skills workshops) as indicators are determined from the data analysis, as in [5], [8]–[10]. As interventions are manually linked to academic indicators, we will transition the intervention assignment to machine learning methods. Much of our initial focus will be on implementation of supplemental instruction [11], [12], peer tutoring [13], [14], and peer mentoring [3], [15], [16] (see Section II-B) to help with ECE core courses (required of all ECE majors), which also tend to be "bottleneck" courses. Additionally, we provide a more targeted and individualized approach to tutoring, mentoring and supplemental instruction through the use of course learning objectives (see Section II-C).

B. Peer Tutoring, Peer Mentoring, and Supplemental Instruction

An undergraduate teaching assistant (TA) serves a dual role as a peer tutor and peer mentor for each of the ECE core

978-1-5386-1174-6/18/$31.00 ©2018 IEEE

courses. The undergraduate TA helps formulate the content and structure of the supplemental instruction, attends and participates in the supplemental instruction, holds office hours, is available by appointment, and proactively reaches out to students who are struggling in the class (grade details are withheld due to the peer nature of the relationship). We have a dedicated peer tutor/mentor for each course who have taken the course and shown themselves to be proficient in the course material. These TAs also have open-door office hours in which they are available to tutor or mentor any student.

Supplemental instruction is provided as 1-credit full-semester and mini-semester sections. The supplemental instruction by a graduate TA provides a learning environment complementary to the in-class lecture and laboratory, but in a more peer-oriented small-group structure which is expected to be particularly effective for women and URM students [17]. The student-led structure of the supplemental instruction provides students with more ownership of their learning experience, and helps address generational disconnect in learning styles between faculty and students [18]. Supplemental instruction follows the learning objectives covered in the course assignments each week. In addition to leading the supplemental instruction, the graduate TAs hold office hours, are available by appointment, and proactively reach out to students who are struggling in the class.

C. Learning Objectives

Learning objectives provide measurable behaviors and actions that students should exhibit when demonstrating mastery of course material [19]. These learning outcomes can be used to design course material and pedagogy as well as provide assessment of the curriculum [7], [20]. Since learning objectives are designed to focus on behaviors and skills rather than specific course-related topics, we choose them as an important means to propose student-specific interventions to improve academic performance. The use of learning objectives can help focus student and student-mentor attention to the specific concepts with which students are struggling.

The student mentors work with their assigned core course and map learning objectives to assignments. For those courses that do not have established learning objectives, the student mentors develop learning objectives for each assignment. These learning objectives will be provided to the course instructor for use in subsequent course offerings. Once learning objectives are mapped to assignments, the student mentors collect resources (e.g., additional problem sets, detailed solutions, alternative explanations, links to informative videos) to address each of the learning objectives. These resources are then available for the students to help address any deficiencies in their understanding of course material.

III. Database Management System

There is a wealth of information already gathered for each student at our institution, namely through our student information system (Banner), degree verification system (STARAudit), and learning management system (Canvas). However, this data is not coherently analyzed in relation to students' academic performance and degree progression [5].

In order to efficiently query our dataset, data that are composed of many sources, a database is used to combine data from the various sources. MySQL is used for the database management system software. The database is designed to facilitate queries of interest such as the performance of certain students' homework assignments and their subsequent performance on exams. By organizing the data efficiently we can more effectively pipe data into machine learning algorithms to answer questions of interest. The comprehensiveness and granularity of these data will provide an unprecedented view of student academic progress and degree progression. In this section, we describe the data sources, our database schema, and the potential queries addressable by our database.

A. Data Sources

This project has two main sources of data: (1) grade reports pulled from Canvas, and (2) demographics on enrolled students from Banner. Both sets of data are exported in comma separated value (.csv) form.

Grade reports from the learning management system are reported for each ECE undergraduate core course. The grade reports list all assignments and student scores for the corresponding course. To establish anonymity in our database, we anonymize the student identification number, replacing it with a hashed alphanumeric ID string. The hashed IDs are generated using the Hashids library [21]. A "salt" string (known only to the research team) is used to seed the random generation of the hashed IDs, such that if the same salt is used, the same encoded ID will be generated from an unencoded ID number. Using the same salt also allows the decoder to recover the unencoded ID. Currently the graduate mentors are manually exporting grade reports on a weekly basis and the data is ingested into the database. In the future we would like to have a script automatically pull data from the learning management system to periodically update the database.

The second source of data is the student information system. These data contain demographic and administrative information about the individual student such as incoming SAT/ACT (standardized tests for college admissions in the US) scores, when they first enrolled, high school they attended, ethnic and racial identification, and financial need. These data also contain periodically updated (approximately each semester) information such as cumulative grade point average (GPA) and number of credit hours.

B. Database Organization

Before the data is entered into the database there are some pre-processing steps. As mentioned earlier identifying information (namely the students' identification numbers) are hashed for anonymity. In addition to this security-motivated pre-processing, additional pre-processing is needed to allow for improved performance on queries and to more efficiently pipe data to machine learning algorithms for analysis. In particular, we have found that data from each of the courses

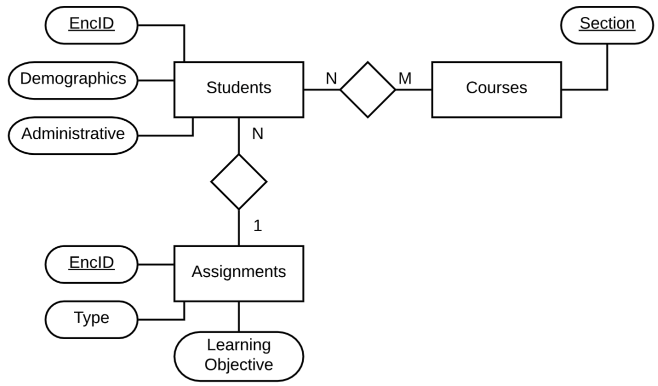

Fig. 1. Diagram of database schema.

should be uniform in how assignments are named. We are developing a list of guidelines in naming assignments to be disseminated to course instructors in the future to streamline this pre-processing step.

We show a diagram of the schema for our database in Fig. 1. The schema shows the relationship between the attributes of our database. We create a table labeled `Students` that holds a student identifier (our encrypted alphanumeric ID string `encID`), as well as the demographics (e.g., ethnicity) and administrative (e.g., GPA) data available for the student. A second table labeled `Courses` contains an N to M relationship to the `Students` table; this denotes that there are N students in a course and an individual student can belong to M courses. The `Courses` table contains attributes such as section number and instructor. A third table labeled `Assignments` holds information on the assignments and has an N to M relationship to the `Students` table. The `Assignments` table contains attributes such as the assignment type and the associated learning objectives (Section II-C). The learning objectives attribute is a key part of the database: by mapping the learning objectives to the different assignments and assignment types, we can better assess a student's performance and quantify where a student is having problems in their studies.

C. Possible Queries

Using the database we expect to be able to answer various questions related to a student's academic progress in a course and their degree progression. For example, if we are interested in the learning objective(s) that a student is having trouble grasping, we can query which assignments a student performed poorly on and retrieve the learning objectives associated with those assignments. Such a query may look like 'select learning objective where assignment_grade <75.' Further, we can see if that lack of competency was overcome by the time of an exam by querying the exam and associated learning objectives and develop intervention methods for the future. With the incorporation of data from our university's degree verification system and aggregating data over many semesters, we can perform the query again to see if students have the same issues over time and either apply intervention methods or assess the intervention methods that were applied.

IV. Effect of Unsuccessful Course Attempts on Completion Rate

To assist students most effectively we need to determine which factors play the largest role in successfully and expediently obtaining their degree. One such factor that we study here is the number of unsuccessful attempts at courses the students accrue during the time they are working toward a degree [22]–[24]. Unsuccessful attempts include failing the course by earning a 'D' or 'F' final grade, or withdrawing from the course after the third Friday of the semester, thus earning a 'W' (withdraw) grade; we refer to these collectively as DFWs. Here, we study the relationship between DFWs in ECE courses and time-to-degree. Future work will expand this analysis to other courses, with specific interest in math courses which also tend to be bottleneck courses for our ECE students.

For the ECE core courses, either a failure or withdrawal will result in the need to retake the course. Since all core courses are offered each semester, this will most likely be the following semester. It is important to note, however, that it is not only that one class that is affected, but also any subsequent courses which require that class as a prerequisite. Intuitively we would expect that fewer DFWs would correlate to a faster completion rate or time to graduation; other studies have demonstrated this, e.g., [23], [24].

A. 2017 Cohort

Our initial study of the effect of DFWs on degree completion rate is performed on the cohort of students who earned degrees in 2017, including the Spring, Summer, and Fall semesters. It is interesting to note, however, that while these students completed their ECE *degrees* (BSEE) in 2017, these students did not necessarily complete their ECE *courses* in 2017. Specifically, there were $N = 63$ students who graduated with the BSEE in the 2017 cohort, but only $N = 22$ who enrolled in an ECE course in 2017. This could be due to other outstanding degree requirements (e.g., general education courses) or courses required for minors or second majors.

The number of DFWs accrued for each student in the cohort was determined by searching each semester's `Courses` table for every instance of ECE course enrollment. If a student was enrolled in a course, the string in the final grade field was examined to see if it included the characters 'D', 'F', or 'W'. If it did, the DFW count for that student was incremented by one. Searching only for 'D' and 'F' allowed the inclusion of, for example, 'D+' grades. The counts for semesters enrolled in ECE courses were calculated similarly: he value was incremented for each semester in which the student was found to be enrolled in at least one course.

The BSEE degree at our university was revised in 2016 to consist of a total of 120 credits, allowing for 8-semester graduation assuming 15 credits per semester. The 2017 cohort largely fall under the previous curriculum, requiring a total of 132 credits and a nominal graduation time of 8-9 semesters

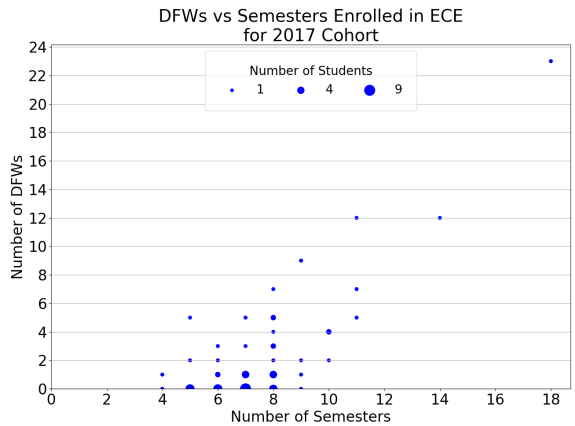

Fig. 2. Number of counted DFWs versus the number of counted semesters enrolled in ECE courses for the 63 students in the 2017 cohort. Mark size is proportional to the number of students (larger marks represent more students).

assuming typical full-time courseload. Fig. 2 shows the relationship between number of DFWs and number of semesters to graduation. We see the majority of the 63 students in the 2017 cohort graduated within the expected 8-9 semesters. Using the scikit-learn [25] library's implementation of ordinary least squares linear regression, the first 31 students in the cohort were used to fit a linear regression model. The last 32 students were then used to test the model. The resulting fit line had a slope of 0.956, with a mean squared error of 8.20 and variance of 0.61. From this we can tentatively infer that, on average, 0.956 (or very nearly 1) DFW increases the number of semesters a student remains enrolled in ECE courses by 1. While a causal relationship cannot be inferred from these results, this does agree with the expectation that the student will need to retake courses from which they withdrew or failed, which then adds a semester to the time required to finish the ECE requirements.

We note that these results could be misleading for those students with significant transfer credits. Hypothetically, a student that transfers 66 credits toward a degree should finish in half as many semesters. If this student finished in 8 semesters with many DFWs, they would seem to be completing the program at an appropriate rate with this data, even though they did not. We will need a better metric for completion rate than the number of semesters a student was enrolled in ECE courses.

B. Additional Factors

As discussed above, the data used here contain only those ECE courses the students completed. This means we cannot see if the student is in other degree programs, and whether that has impacted their overall time to graduation. Furthermore, using this dataset, we cannot tell what was happening outside the semesters they were not enrolled in any ECE courses. If a student began taking ECE courses in 2014, for example, they might have started that semester, or they might have started in 2013 or even earlier. This would affect their true time to graduation. Additionally, if a student finished their ECE coursework in 2016, and didn't obtain their degree until 2017, we do not know if that was from working toward another degree, personal difficulties, or other factors. This, again, affects their true time to graduation. We hope to enhance our (and students') understanding of where and why students are (and are not) progressing with further work.

V. Conclusions and Future Work

We have presented here the initial steps taken to apply academic analytics to the Electrical & Computer Engineering program at New Mexico State University. We compiled the initial MySQL database which contains the data to be used for data-driven analytics. The database will continue to expand and evolve as we collect more data and new sources.

This database was used to determine the number of unsuccessful ECE course attempts per student in the 2017 cohort, as well as the number of semesters the student was enrolled in ECE courses. From this we have been able to make an initial observation that the change to a student's completion rate is approximately one additional semester per DFW, given our current data.

We would like to complete a more comprehensive and quantitative analysis of the effect of DFWs on time to graduation, including all courses students are taking to get a more complete picture. Additionally, we would like to determine if a DFW for a specific course is most detrimental to successful completion (e.g., some courses are prerequisites for more courses than others), as well as to understand why students are withdrawing from these courses (e.g., whether it is the specific time the course is offered).

Additionally, our plan of mapping learning objectives to assignments will allow peer mentors to provide effective feedback to students, as they will be able to tell which subjects the student is struggling with based on assignment scores. We would also like to examine mid-semester grades to use in combination with learning objectives as an early warning indicator of probable outcomes for a particular student. This will aid our efforts to supply timely advice and help the student successfully complete the learning objectives required to improve their academic performance. Seeing where they are struggling might help students appreciate which key topics they need to spend more time on long before it becomes a more difficult problem from which to recover.

While we do not currently have an automatic feedback method, nor has the presented work provided the level of insight necessary to realize such a method yet, our future work will allow the possibility. Once implemented, these additional tools and analyses will allow mentors, advisors, and the students themselves to provide deep insights into the work necessary for the student to expediently and successfully complete the Electrical & Computer Engineering program at our university.

Acknowledgment

The authors gratefully acknowledge support from the US Department of Education Minority in Science and Education Improvement Program (MSEIP) Grant P120A170064.

REFERENCES

[1] X. Chen, "STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001," *National Center for Education Statistics*, 2013.

[2] M. J. Graham, J. Frederick, A. Byars-Winston, A.-B. Hunter, and J. Handelsman, "Increasing persistence of college students in STEM," *Science*, vol. 341, no. 6153, pp. 1455–1456, 2013.

[3] Z. S. Wilson, L. Holmes, M. R. Sylvain, L. Batiste, M. Johnson, S. Y. McGuire, S. S. Pang, I. M. Warner *et al.*, "Hierarchical mentoring: A transformative strategy for improving diversity and retention in undergraduate stem disciplines," *Journal of Science Education and Technology*, vol. 21, no. 1, pp. 148–156, 2012.

[4] G. Lichtenstein, H. L. Chen, K. A. Smith, and T. A. Maldonado, "Retention and persistence of women and minorities along the engineering pathway in the United States," *Cambridge Handbook of Engineering Education Research*, pp. 311–334, 2014.

[5] (2013, Jan) Predictive Analytics in Higher Education: Data-Driven Decision-Making for the Student Life Cycle. [Online]. Available: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=YTL03202USEN

[6] W. R. Earl, "Intrusive advising of freshmen in academic difficulty," *NACADA journal*, vol. 8, no. 2, pp. 27–33, 1988.

[7] C. W. Starr, B. Manaris, and R. H. Stalvey, "Bloom's taxonomy revisited: specifying assessable learning objectives in computer science," *ACM SIGCSE Bulletin*, vol. 40, no. 1, pp. 261–265, 2008.

[8] A. Kamenetz. (2016, Oct. 30) "How One University Used Big Data to Boost Graduation Rates". National Public Radio. [Online]. Available: http://www.npr.org/sections/ed/2016/10/30/499200614/how-one-university-used-big-data-to-boost-graduation-rates

[9] D. L. Kirp. (2016, Apr. 30) "What Can Stop Kids From Dropping Out". The New York Times. [Online]. Available: http://www.nytimes.com/2016/05/01/opinion/sunday/what-can-stop-kids-from-dropping-out.html

[10] M. Ekowo and I. Palmer. (2016, Oct) The Promise and Peril of Predictive Analytics in Higher Education: A Landscape Analyis. [Online]. Available: https://na-production.s3.amazonaws.com/documents/Promise-and-Peril_4.pdf

[11] R. A. Blanc, L. E. DeBuhr, and D. C. Martin, "Breaking the attrition cycle: The effects of supplemental instruction on undergraduate performance and attrition," *The Journal of Higher Education*, pp. 80–90, 1983.

[12] N. A. Pendergrass, R. E. Kowalczyk, J. P. Dowd, R. N. Laoulache, W. Nelles, J. A. Golen, and E. Fowler, "Improving first-year engineering education," in *Frontiers in Education Conference, 1999. FIE'99. 29th Annual*, vol. 3. IEEE, 1999, pp. 13C2–6.

[13] R. D. Roscoe and M. T. Chi, "Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors explanations and questions," *Review of Educational Research*, vol. 77, no. 4, pp. 534–574, 2007.

[14] K. J. Topping, "The effectiveness of peer tutoring in further and higher education: A typology and review of the literature," *Higher education*, vol. 32, no. 3, pp. 321–345, 1996.

[15] J. W. Colvin and M. Ashman, "Roles, risks, and benefits of peer mentoring relationships in higher education," *Mentoring & Tutoring: Partnership in Learning*, vol. 18, no. 2, pp. 121–134, 2010.

[16] J. M. Holland, D. A. Major, and K. A. Orvis, "Understanding how peer mentoring and capitalization link STEM students to their majors," *The Career Development Quarterly*, vol. 60, no. 4, pp. 343–354, 2012.

[17] H. K. Ro and D. B. Knight, "Gender differences in learning outcomes from the college experiences of engineering students," *Journal of Engineering Education*, vol. 105, no. 3, pp. 478–507, 2016.

[18] Are You Ready to Support 4 Generations of Learners? [Online]. Available: https://www.panopto.com/blog/are-you-ready-to-support-4-generations-of-learners/

[19] R. M. Felder and R. Brent, "Objectively speaking," *Chemical Engineering Education*, vol. 31, pp. 178–179, 1997.

[20] ——, "Designing and teaching courses to satisfy the abet engineering criteria," *Journal of Engineering Education*, vol. 92, no. 1, pp. 7–25, 2003.

[21] hashids: generate short unique ids from integers. [Online]. Available: http://hashids.org/

[22] C. Adelman, "Answers in the Tool Box," Office of Educational Research and Improvement U.S. Department of Education, Tech. Rep., 1999.

[23] J. D. Armstrong and J. Winn, "The Impact of Withdrawals and Failures on Graduation Rates," Florida Department of Education, Tech. Rep., 2005.

[24] G. M. Nicholls and R. K. Gaede, "Exploring the effects of student course withdrawals on time to graduation." *Proceedings of the ASEE Annual Conference & Exposition*, vol. 35899, no. 1, pp. 1–12, 2014.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.