

Academic Performance, Course Load, and Student Persistence: A Longitudinal Modeling of Academic Trajectories in an Engineering Program

Ignacio Joaquín Pérez Chaves¹; Laura Stella Vega Escobar²; Paula Andrea Rodríguez Marín³
Institución Universitaria ITM, Colombia, ignacioperez1155357@correo.itm.edu.co, lauravega@itm.edu.co,
paularodriguez@itm.edu.co

Abstract– Student persistence in engineering programs represents a critical challenge for higher education institutions, particularly during the early semesters, when foundational courses exhibit high repetition and failure rates. In this context, the present study examines the relationship between academic performance, academic load, and student persistence through the construction of longitudinal academic trajectories. The analysis is based on an institutional dataset comprising academic records of students enrolled in foundational courses within a Software Development program between the 2022-1 and 2025-1 academic terms. From these data, a synthetic performance indicator termed the cumulative course pass rate was developed, along with a binary variable representing student persistence. The analytical framework combined descriptive statistics, exploratory visualizations, and logistic regression modeling to assess the relationship between cumulative academic performance and the probability of persistence. The results reveal a strong and statistically significant association between cumulative course pass rate and student persistence, with the proposed model demonstrating high predictive performance. Additionally, individual-level analysis of the relationship between academic load and performance reveals substantial heterogeneity in students' responses, enabling the identification of differentiated academic profiles. These findings highlight the relevance of longitudinal performance indicators for understanding student trajectories and for supporting institutional strategies aimed at strengthening academic persistence.

Keywords– Student persistence, academic performance, academic load, academic trajectories, longitudinal modelling, learning analytics.

I. INTRODUCTION

Higher education constitutes a fundamental human right and a key driver of individual and societal development, as it fosters social mobility, promotes equity, and contributes to economic growth. Despite its recognized importance, Latin America and the Caribbean continue to experience higher student dropout rates than other regions, positioning student attrition as a persistent challenge for the higher education sector [1].

Prior research indicates that dropout is particularly concentrated in the early semesters, especially in foundational courses characterized by high conceptual rigor and a structuring

role in students' professional formation [2]. Within this context, understanding the factors associated with academic performance and their influence on program continuity is essential for designing timely, evidence-based student support strategies.

Traditionally, student persistence has been examined using aggregated indicators at the cohort or semester level. Numerous studies have developed predictive models aimed at identifying critical factors associated with dropout and graduation, leveraging institutional data to estimate attrition risk and support institutional decision-making [3]. These approaches have facilitated earlier identification of at-risk students, particularly through predictive models grounded in historical academic records.

More recently, research has incorporated learning analytics and machine learning techniques to predict academic risk by integrating demographic, academic, and behavioral variables. Evidence suggests that performance- and engagement-based indicators often demonstrate greater explanatory power than traditional sociodemographic variables, and that incorporating temporal dimensions significantly enhances predictive accuracy [4]. Furthermore, analytical frameworks based on minimal institutional data have been proposed not only to identify at-risk students but also to classify them into differentiated profiles, thereby informing targeted institutional policies [5].

From a pedagogical standpoint, the systematic monitoring of academic performance throughout the learning process has also been emphasized. Studies on continuous assessment indicate that frequent performance tracking and timely feedback strengthen student engagement and promote coherence in the learning process, both of which are closely associated with student persistence [6].

Nevertheless, aggregated approaches may obscure the complexity of individual academic trajectories, limiting insight into the processes that lead to persistence or discontinuity. The increasing availability of longitudinal academic data has enabled more nuanced analyses in which academic performance is conceptualized as a dynamic process evolving over time and across multiple courses [7]–[9].

In particular, cumulative academic performance and academic load have emerged as key variables in persistence

research. However, their relationship is neither strictly linear nor homogeneous across students. While some students experience performance decline as academic load increases, others maintain or even improve their outcomes, suggesting the presence of differentiated response profiles.

Within this framework, the present study examines the relationship between academic performance, academic load, and student persistence through the construction of longitudinal academic trajectories. Drawing on institutional records from foundational courses in a Software Development program, the study proposes a modeling approach based on synthetic performance indicators and statistical analysis to characterize performance patterns and their association with persistence. The study aims to contribute empirical evidence toward a deeper understanding of student trajectories and to inform the development of early warning and academic support strategies.

II. METHODOLOGY

A. Context and Data Source

This study draws upon an institutional dataset provided by the Academic Persistence Office. The dataset includes partial academic performance records of students enrolled in foundational courses within the early-stage curriculum of the Software Development program Differential Calculus, Linear Algebra, Mathematics for Computer Science, and Programming Logic covering the 2022-1 to 2025-1 academic terms.

These data constitute the basis for analyzing academic performance in key components of the curriculum and enable the subsequent reconstruction of longitudinal academic trajectories, the identification of performance patterns, and the examination of their association with student persistence in the program.

B. Data Processing and Integration

Data processing was conducted in the *Google Colab* environment using *Python*-based analytical libraries. The original information was distributed across seven Excel workbooks corresponding to different academic terms, each containing multiple sheets with sociodemographic information and course-level academic records.

An initial exploratory phase identified redundant attributes across sheets, where multiple variables described the same characteristic. These attributes were removed, along with non-analytical variables such as names, alternative identification numbers, phone numbers, email addresses, and other personal information. The “Documento” field was defined as the unique integration key, given its consistent presence across all records.

The datasets were consolidated into a single structured file, preserving student identifiers across course

enrollments to maintain course-level performance granularity. Variations in the number of partial assessments across courses were detected, which complicated direct comparisons. To address this heterogeneity, a standardization procedure was implemented to harmonize grading structures. The resulting dataset was clean, internally consistent, and suitable for longitudinal analysis.

C. Construction of Academic Trajectories and Persistence Variable

Based on the integrated dataset, student’s academic trajectories were modeled longitudinally. The core variable, “Resultados,” recorded at the course–semester level, included categories such as *Passed*, *Failed*, *Withdrawn*, and *Failed due to Absence*. These categories were recoded into a binary indicator (*resultados_num*), where a value of 1 denotes course approval and 0 represents all other outcomes.

The records from the seven semesters were merged into a longitudinal dataframe using “Documento” as the primary key, enabling student-level trajectory reconstruction on a course-by-course basis. Two structural components were defined: (i) chronological order, representing the temporal sequence in which each course was taken, and (ii) results sequence, defined as an ordered binary vector summarizing performance in a specific course over time.

Based on this longitudinal representation, a binary Persistence variable was defined to capture whether students successfully passed the foundational courses of the early-stage curriculum. Students whose sequences consisted exclusively of non-passing outcomes and showed no subsequent academic records were classified as non-persistent (0), whereas those who achieved course approval were classified as persistent (1).

D. Analysis of Academic Performance and Its Relationship with Student Persistence

The final stage of the study focused on analyzing the quantitative relationship between academic performance and student persistence. Descriptive statistics and exploratory visualizations were first employed to assess the association between academic load measured as the number of courses enrolled per semester and the number of courses passed.

A cumulative performance indicator, termed approval rate, was then defined as the proportion of total courses passed to total courses attempted by each student during the observation period. This indicator served as the primary predictor in a logistic regression model, with the binary Persistence variable as the outcome.

At the individual level, the association between academic load and performance was further examined by computing the correlation between semester-level enrollment load and the semester approval rate. Based on these correlation coefficients, students were classified into distinct response profiles, which

were subsequently contrasted with the Persistence variable to identify differentiated patterns of academic performance.

E. Ethical Considerations

The handling of student data in this study was conducted in accordance with institutional confidentiality standards. Prior to accessing the dataset, a formal confidentiality agreement was signed with the Academic Persistence Dependence, establishing the conditions under which the information could be used, stored, and reported.

The dataset provided contained personally identifiable information that was not relevant to the analytical objectives of the study. Accordingly, variables such as full names, contact details, residential addresses, and alternative identification numbers were removed during the data processing stage, as described in Section B. The "Document" field was retained exclusively as a technical identifier to enable longitudinal trajectory reconstruction across academic terms and was not used to identify, profile, or reference any individual student.

All results are reported in aggregated form, ensuring that no individual student can be identified from the findings presented in this study. The information was used solely for academic research purposes within the scope agreed upon with the issuing institution.

III. RESULTS

A. Joint Distribution of Course Enrollment and Course Completion

The initial stage of the results analysis examined the association between the number of courses enrolled and the

number of courses passed by students in each semester. To visualize this relationship, heatmaps were generated, with the vertical axis representing semester-level academic load and the horizontal axis indicating the number of courses passed.

La Fig 1. illustrates this distribution for the 2022-1 academic term.

Overall, the heatmaps reveal increasing heterogeneity in academic performance as academic load rises. Under higher academic loads, students who pass most of their courses coexist with others who experience substantial academic difficulties.

Conversely, at lower levels of academic load, a noticeable concentration of students pass only a subset or none of their enrolled courses. This pattern suggests that a reduced academic load does not, by itself, guarantee improved performance and may instead reflect course repetition dynamics or persistent academic difficulties in foundational courses of the early-stage curriculum.

B. Relationship between approval rate and student persistence

To synthesize individual academic performance over time, a cumulative course pass rate was constructed for each student, defined as the proportion of the total number of courses successfully completed relative to the total number of courses attempted throughout the student's academic trajectory, as shown in (1). In this formulation, the term $rijk$ represents the outcome obtained by student i in course j during attempt k , taking the value 1 if the course was successfully completed and 0 otherwise. The set S_i comprises all courses attempted by student i , while n_{ij} denotes the number of attempts made by the student to successfully complete course j .

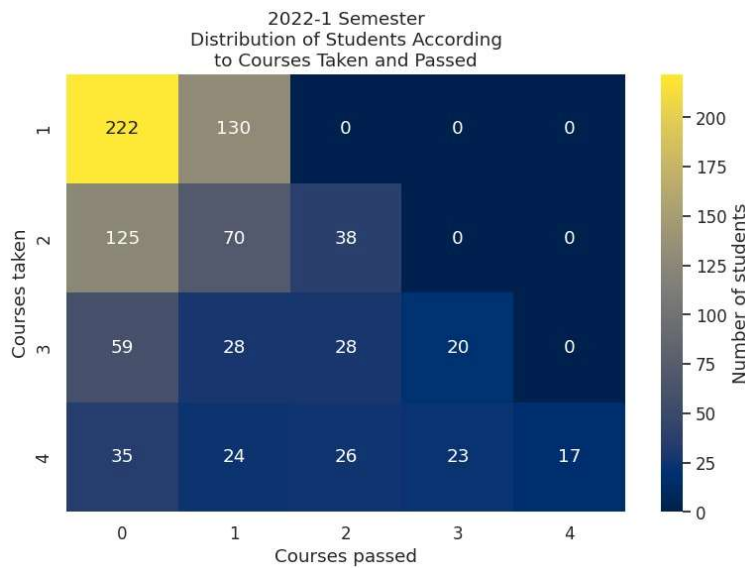


Fig 1. Heatmap illustrating the joint distribution between the number of courses enrolled and the number of courses successfully completed during the 2022-1 academic term.

$$PassRate_i = \frac{\sum_{j \in S_i} \sum_{k=1}^{n_{ij}} r_{ijk}}{\sum_{j \in S_i} n_{ij}} \quad (1)$$

The relationship between this index and student persistence was subsequently examined using a logistic regression model, with persistence specified as a binary dependent variable. The dataset was partitioned into 70 % for model training and 30 % for testing to evaluate predictive performance.

The results demonstrate strong predictive capability, with an overall accuracy of approximately 92 %. These findings indicate that the approval rate is a statistically meaningful predictor of student persistence. Fig. 2 illustrates the estimated probability of persistence as a function of the approval rate, revealing a positive and non-linear relationship. Notably, the probability of persistence increases more sharply at intermediate and higher index values, suggesting the presence of academic performance thresholds associated with sustained enrollment in the program.

C. Relations between academic course load and student performance

To further examine the relationship between academic load and performance, an individual-level analysis was conducted to assess the association between the number of courses enrolled per semester and the semester approval

rate. For each student, a correlation coefficient between these two variables was calculated, considering only cases with sufficient longitudinal information.

The distribution of these correlations (Fig. 3) reveals substantial heterogeneity in students' responses to changes in academic load. Strong negative correlations are observed, along with positive correlations and values close to zero, indicating that the effect of academic load on performance is not uniform across the analyzed population.

Based on these results, students were classified into three main profiles: (i) overload students, characterized by negative correlations; (ii) students whose performance improves with higher academic load; and (iii) independent students, showing no clear relationship between load and performance. Table I summarizes the distribution of these profiles and their relationship with overall academic performance.

TABLE I.
Distribution of students by academic load–performance profile and proportion of low academic performance in each group.

Type	Total	Low Performance	Proportion
Overload	604	62	44 %
Improves with load	465	124	34 %
Independent	317	118	23 %

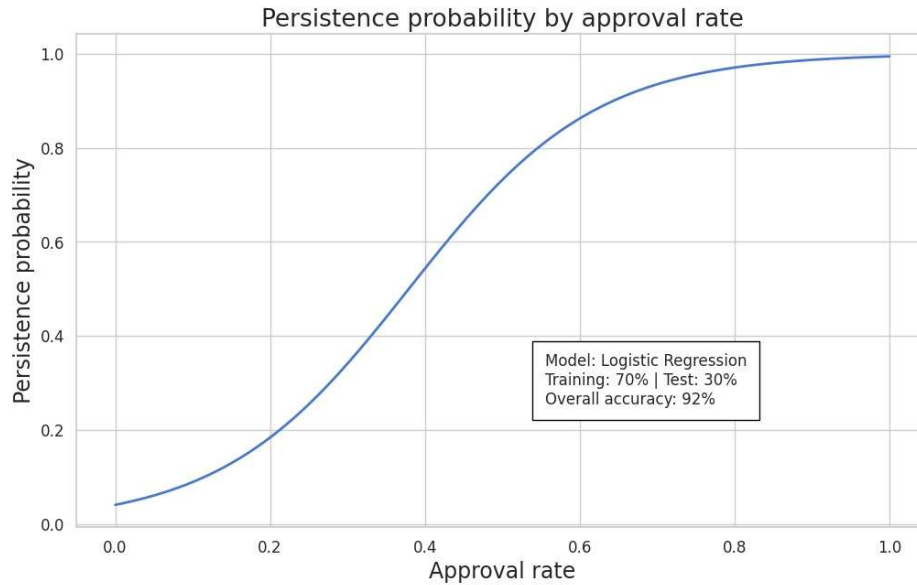


Fig 2. Estimated relationship between the cumulative course pass rate and the probability of student persistence.

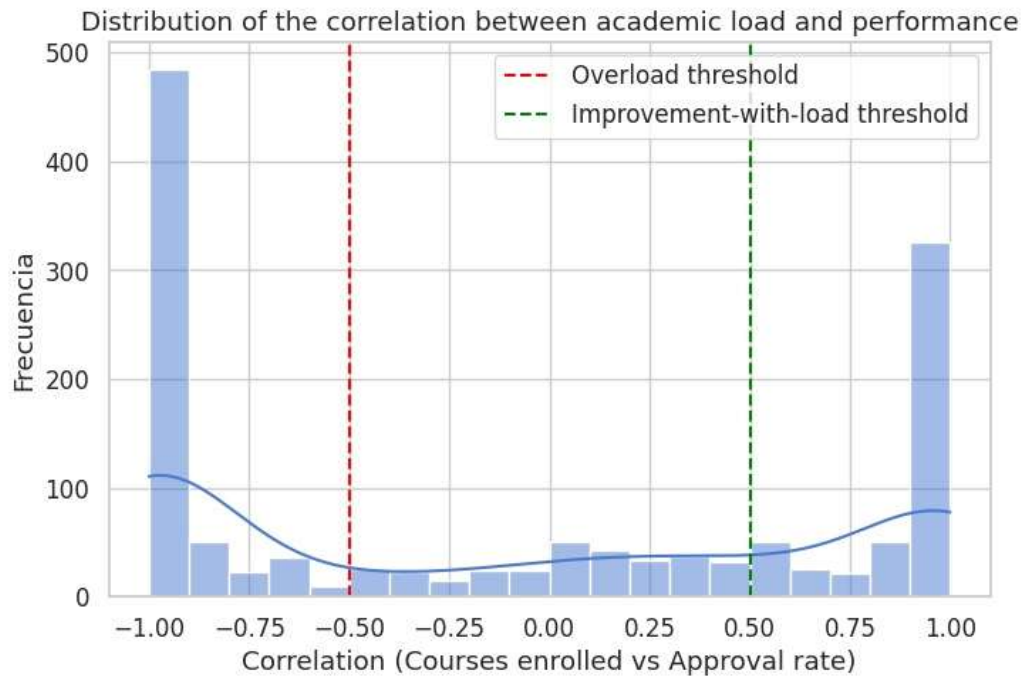


Fig 3. Distribution of students according to the type of relationship between academic load and academic performance, determined based on the correlation between the number of courses enrolled and the semester pass rate.

When these profiles are contrasted with overall performance levels (Table I), a particularly relevant result emerges. Although the overload group exhibits strong negative correlations indicating greater sensitivity of performance to increases in course load—this same group paradoxically concentrates a lower proportion of students with low overall performance.

This finding suggests that sensitivity to academic load does not necessarily reflect academic fragility. Instead, it may be associated with high-performance trajectories, in which even moderate variations in course load produce detectable changes in approval rates. In fact, higher-performing students appear to respond in a more structured and predictable manner to changes in their academic load, consistent with prior evidence linking higher approval rates to a greater probability of institutional persistence.

These results open a line of future research aimed at better understanding the mechanisms underlying this academic sensitivity. In particular, it would be relevant to explore whether factors such as self-regulation, semester planning, or strategic course selection help explain why higher-performing students also exhibit a stronger relationship between academic load and performance.

IV. CONCLUSIONS

This study analyzed the relationship between academic performance, academic load, and student persistence through the construction of longitudinal academic trajectories based on institutional records. The findings demonstrate that the

cumulative course pass rate constitutes a robust and highly informative indicator for explaining student persistence in foundational courses.

The logistic regression model confirmed the strong association between sustained academic performance and the probability of continued enrollment. Additionally, the classification of differentiated load performance response profiles revealed substantial heterogeneity across students and showed that negative load sensitivity does not necessarily imply low academic performance.

These findings underscore the importance of adopting longitudinal and individualized analytical approaches that move beyond aggregated perspectives. From an institutional standpoint, cumulative performance indicators and load-response profiles may serve as valuable inputs for academic advising, enrollment planning, and early warning systems.

Future research should incorporate additional sociodemographic and academic variables and extend the analysis to other programs and cohorts to strengthen generalizability and advance toward more comprehensive predictive models of student persistence.

ACKNOWLEDGMENT

Acknowledgment is given to the DATA+STEM Research Seedbed team, affiliated with the “Máquinas inteligentes y reconocimiento de patrones” research group at ITM, within the “Sistemas de recomendación y recuperación de información” research line.

REFERENCES

- [1] N. Chiarino *et al.*, “Abandono y permanencia estudiantil en universidades de Latinoamérica y el Caribe: Una revisión sistemática mixta,” *Actualidades Investigativas en Educación*, vol. 24, no. 2, pp. 1–37, May 2024, doi: 10.15517/aie.v24i2.57306.
- [2] M. Castillo-Sánchez, R. Gamboa-Araya, and R. Hidalgo-Mora, “Factores que influyen en la deserción y reprobación de estudiantes de un curso universitario de matemáticas,” *Uniciencia*, vol. 34, no. 1, pp. 219–245, Jan. 2020, doi: 10.15359/ru.34-1.13.
- [3] L. Addison and D. Williams, “Predicting student retention in higher education institutions (HEIs),” *Higher Education, Skills and Work-Based Learning*, vol. 13, no. 5, pp. 865–885, Oct. 2023, doi: 10.1108/HESWBL-12-2022-0257.
- [4] S. Alhazbi, “Pedagogically-Informed Behavioural Learning Analytics: An Expert Approach to Predicting at-Risk Students,” *Expert Syst.*, vol. 43, no. 3, Mar. 2026, doi: 10.1111/exsy.70210.
- [5] S. Hoca and N. Dimililer, “A Machine Learning Framework for Student Retention Policy Development: A Case Study,” *Applied Sciences*, vol. 15, no. 6, p. 2989, Mar. 2025, doi: 10.3390/app15062989.
- [6] A. Suárez *et al.*, “Multi-Year Evaluation and Enhancement of a Questionnaire-Based Assessment Approach in Engineering Laboratory Instruction,” *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, vol. 21, pp. 37–42, 2026, doi: 10.1109/RITA.2026.3654759.
- [7] M. Saqr, S. López-Pernas, S. Helske, and S. Hrastinski, “The longitudinal association between engagement and achievement varies by time, students’ profiles, and achievement state: A full program study,” *Comput. Educ.*, vol. 199, p. 104787, Jul. 2023, doi: 10.1016/j.compedu.2023.104787.
- [8] K. Jwa and C. G. P. Berdanier, “Witnessing ‘De-Socialization’: Investigating unexpected longitudinal trends in engineering doctoral socialization,” *Journal of Engineering Education*, vol. 115, no. 1, Jan. 2026, doi: 10.1002/jee.70040.
- [9] A. Igualde-Sáez *et al.*, “University Student Dropout: A Longitudinal Dataset of Demographic, Socioeconomic, and Academic Indicators,” *Data (Basel)*, vol. 10, no. 10, p. 162, Oct. 2025, doi: 10.3390/data10100162.