Impact and Usage Patterns of Supplemental Online Learning by UCAB Informatics students [*]

Franklin D. Quintana   *Universidad Católica Andrés Bello*
Luis A. Ochoa   *Universidad Católica Andrés Bello*

This report investigates the self-directed learning behaviors of Informatics students at UCAB, specifically examining their utilization of online educational platforms to supplement formal curriculum and acquire job market-relevant skills. The study identifies predominant platforms, content types, and usage frequencies, while also exploring student perceptions of these resources' effectiveness. Furthermore, the analysis addresses how students overcome common obstacles like course pricing and language barriers. The findings reveal a significant reliance on free resources, particularly YouTube, and a preference for external tools over content in Spanish when facing language challenges. The report also highlights the analytical limitations posed by the discrete nature and sample size of the collected data, precluding the development of certain linear regression models to establish definitive relationships between platform usage, academic performance, and professional development.

Keywords: online learning, self-directed learning, informatics education, student skills, external resources, R Markdown

Introduction

The Informatics program at Universidad Católica Andrés Bello (UCAB) provides a strong academic foundation in essential areas such as methodologies, standards, computer science theory, and telematics. This knowledge is delivered through a comprehensive curriculum that covers various aspects of information technology, project management, and software development. However, the rapidly evolving technological landscape and the demands of the modern job market necessitate continuous skill development beyond traditional coursework. Students, particularly during practical courses and community service projects, frequently encounter the need to acquire specialized technical skills that are not always taught within the core curriculum. Furthermore, employers increasingly place high value on certifications and technical knowledge, which pushes students to learn and acquire these skills through alternative resources.

The proliferation of accessible online educational platforms, including Coursera, Udemy, and YouTube, has transformed how individuals acquire knowledge and skills. Recognizing this shift, the Informatics faculty at UCAB seeks a deeper understanding of how its students leverage these external resources to complement their formal education. Investigating these self-directed learning behaviors is essential for understanding student preparedness, identifying emerging skill gaps, and potentially informing future curriculum development to better align with professional demands and student learning strategies. This research is highly relevant as it sheds light on the dynamic interplay between formal education and autonomous learning in a field characterized by rapid innovation.

To address these areas, this study explores the following research questions:

---

[*]Replication files are available on the author's Github account (http://github.com/svmiller). **Current version**: julio 16, 2025; **Corresponding author**: svmille@clemson.edu.

1.How does the quality and price of specialized courses determine the selection of platforms such as Udemy, Coursera, or Youtube among UCAB Informatics students?

2.What measures Informatics students take when facing obstacles, such as payment barriers or difficulties with the English language, that limit their autonomous learning of technologies?

3.What is the relationship between free or paid courses and the contribution to the Informatics students' academic performance and professional development?

Hypotheses

-**Hypothesis 1:** Students of lower semesters (1-4) prioritize free or low cost platforms such as youtube due to economic reasons, even when they recognize limitations in their quality or content structure, compared to upper semesters students (5-8), whose disposition to invest in paid or premium platforms is greater (Udemy, Coursera).

-**Hypothesis 2:** In the latest semesters (7-8), the possibility to obtain certifications valid in the laboral market (Google Cloud Platform, Microsoft Azure, AWS) are a dominant criteria, even over costly prices. This tends to be more prevalent among students who are about to graduate or are seeking professional internships.

-**Hypothesis 3:** Students with limited resources prefer free platforms (Youtube, FreeCodeCamp, official docs), even when paid platforms (Udemy, DevTalles) offer more structured and comprehensive content due to their acquisition prices.

-**Hypothesis 4:** Students that utilize free resources have a bigger difficulty finding quality courses in Spanish, so they are forced to consume content in English with external tools (translators, subtitles).

-**Hypothesis 5:** Structured courses in paid platforms facilitate a better understanding of complex concepts improving their academic performance in comparison to free tutorials.

-**Hypothesis 6:** Students that complete paid or certified courses have a greater chance of including those self-developed projects through these platforms in their GitHub as a portfolio, contributing to their professional development.

Methodology

This section outlines the approach taken to collect, process, and analyze the data relevant to the research questions and hypotheses. It details the characteristics of the dataset, the sampling procedure, and the computational methods employed for data cleaning, visualization, and analysis, along with the rationale for these choices.

Data Collection and Sample Description

The primary data for this study was collected through an online survey administered via Google Forms. This platform was chosen for its accessibility, ease of distribution, and efficiency in gathering structured responses from a student population. The survey was designed to capture information regarding students' self-directed learning habits, preferences for online platforms, challenges faced (e.g., pricing, language barriers), and the perceived impact of these resources on their academic performance and professional development.

The target population for this research comprised students enrolled in the Informatics Engineering program at UCAB.

The data for this research project was obtained through a survey, which was principally and only directed towards UCAB informatics students. The questionnaire was open from **June 6th to June 22th** for receiving answers. It gathered a total of **30 responses**.

Demographic Characteristics of the sample

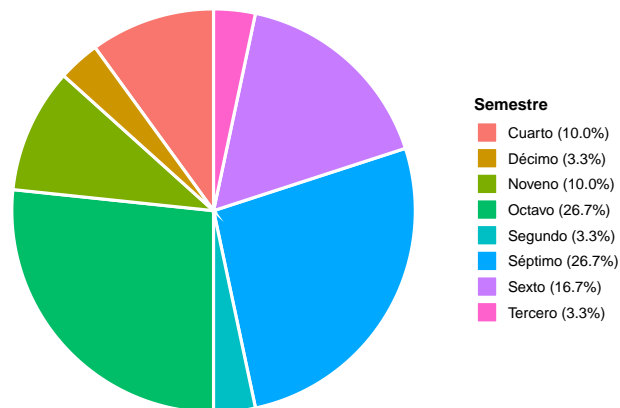**Estudiantes por semestre en la encuesta**

Figure 1: Pie chart showing the distribution of respondents by semester
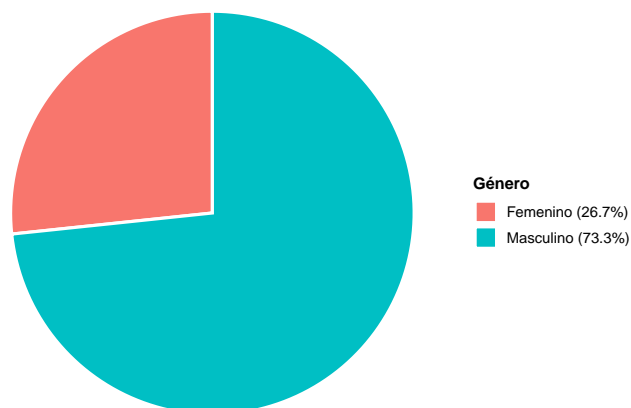
Participación por género en la encuesta

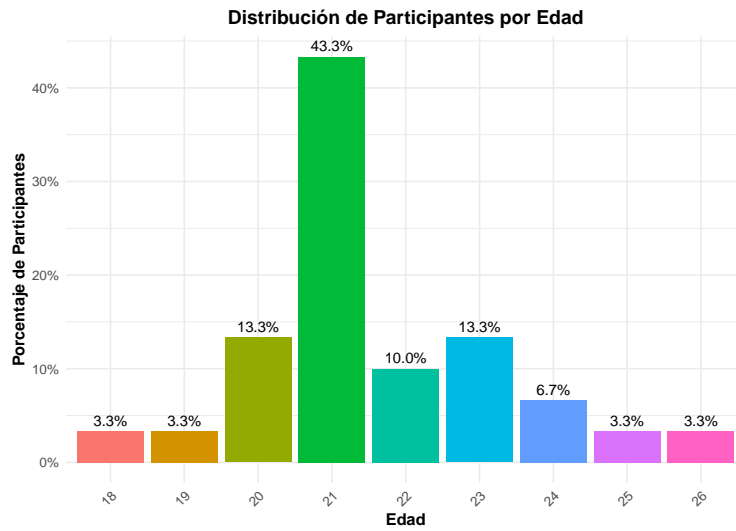Figure 2: Pie chart showing the distribution of respondents by gender
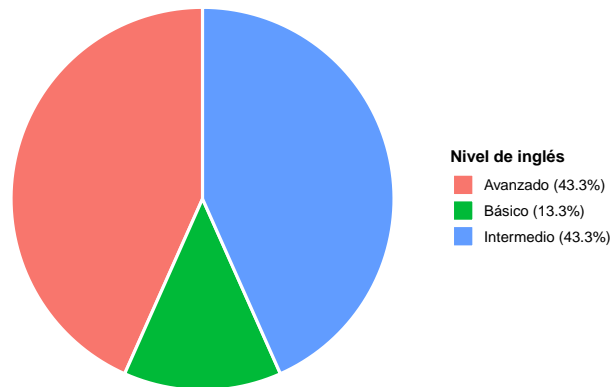
Figure 3: Bar chart showing the distribution of respondents age

Figure 4: Pie chart showing the distribution of respondents' English proficiency level

**Utilización de recursos digitales
por parte de los participantes**

Nivel de inglés
■ Sí (100.0%)

Figure 5: Pie chart showing the distribution of respondents prior experience with online learning resources

These graphs give us information about the sample demographic characteristics such as follow:

- The survey was mostly answered by students in upper semesters (6th to 10th), being this the 83,4% of the total respondents. Notably, both eighth and seventh semesters registered the same highest frequency with 8 respondents each, making them the most prevalent semesters observed.

- When it comes to gender, the survey was mostly answered by male students, accounting for 73.3% of the sample. The remaining 26,7% corresponds to female students.

- The survey also was answered by respondents in their early 20s (20 to 24 years old), being this the 86,6% of the respondents. The average age of participants was 21.5 years old.

- Most of the respondents had an English proficiency level above basic (intermediate or advanced), accounting for 86.6% of the sample. Among these, both intermediate and advanced levels shared the highest frequency with 13 respondents each, indicating that these are the most common proficiency levels among UCAB informatics students in the sample.

- Finally, 100% of respondents had prior experience with online learning resources as it can be seen on figure 5. This shows that no matter the semester, in this survey every Informatics student had used digital resources for autonomous learning.

Data Management and Cleaning

Upon collection, the raw survey data underwent a rigorous cleaning and preparation process to ensure accuracy and suitability for analysis. All data management and cleaning operations were performed using the **R programming language**, specifically leveraging functionalities from the `tidyverse` suite of

packages. R was selected for its robust statistical capabilities, its extensive ecosystem of data manipulation tools, and its support for reproducible research. The `tidyverse`, with packages such as `dplyr` for data transformation and `tidyr` for data tidying, facilitated efficient handling of missing values, variable recoding, and restructuring of the dataset into a format conducive to analysis. This choice ensures transparency and reproducibility of the data preparation steps.

This is the Project dataset structure After cleaning the Raw Excel dataset:

Table 1: Variables and their Data Types in the Project Dataset

	Data_Type
timestamp	POSIXct, POSIXt
semester	character
gender	character
age	numeric
english_level	character
used_external_resources	character
tech_learning_platforms	character
platform_choice_factors	character
paid_for_online_course	character
online_course_spending	character
alt_high_cost_course	character
is_english_obstacle	character
solution_english_barrier	character
improved_uni_project_performance	character
projects_in_portfolio	character
career_benefits_from_courses	character

Data Visualization

Data visualization played a critical role in exploring patterns, summarizing distributions, and presenting findings from the survey. All graphical representations were generated using the `ggplot2` library in R, a core component of the `tidyverse`. `ggplot2` was chosen for its powerful "grammar of graphics" framework, which allows for the creation of highly customized, aesthetically pleasing, and publication-quality plots. Its declarative syntax promotes consistent visual communication of statistical insights, enhancing the clarity and interpretability of the results presented in this report.

Analytical Approach

The analysis primarily employed descriptive statistics to characterize student behaviors, platform preferences, and challenges encountered. Given the nature of the collected data, which largely consisted of discrete and categorical variables, descriptive summaries (frequencies, percentages, means for applicable numerical scales) were used to answer the research questions and explore the proposed hypotheses.

It is important to note that, as it will be discussed in subsequent sections, the predominance of discrete variables and the modest sample size inhibited the development of advanced statistical models such as linear regression for certain hypotheses. While ideal for establishing specific relationships (e.g., between platform usage and academic performance), the necessary continuous variables (e.g., GPA, detailed

satisfaction scores) were not available for model execution. Therefore, the analysis focuses on identifying patterns, trends, and associations through descriptive means, interpreting observed proportions, and noting qualitative insights if they are applicable.

Results & Analysis

Before delving into the findings, it's important to outline some key observations about the data collected and prepared for analysis:

1. **Data Characteristics and Limitations:** Limited Scope of Insights: While the gathered data addresses the first and second research questions and their corresponding hypotheses, the modest sample size prevents us from drawing definitive or insightful conclusions with certainty.

2. **Predominance of Discrete Variables:** A significant observation during data cleaning and the initial attempts at modeling was that all collected variables are discrete. This characteristic considerably limited the utility of linear regression models, which require continuous variables for robust analysis. For example, including a continuous variable like the GPA of each respondent would have enabled the development of more sophisticated and insightful regression models.

We will provide additional considerations regarding the data and the survey methodology later in the analysis, specifically for the third research question.

First research question

How does the quality and price of specialized courses determine the selection of platforms such as Udemy, Coursera, or Youtube among UCAB Informatics students?

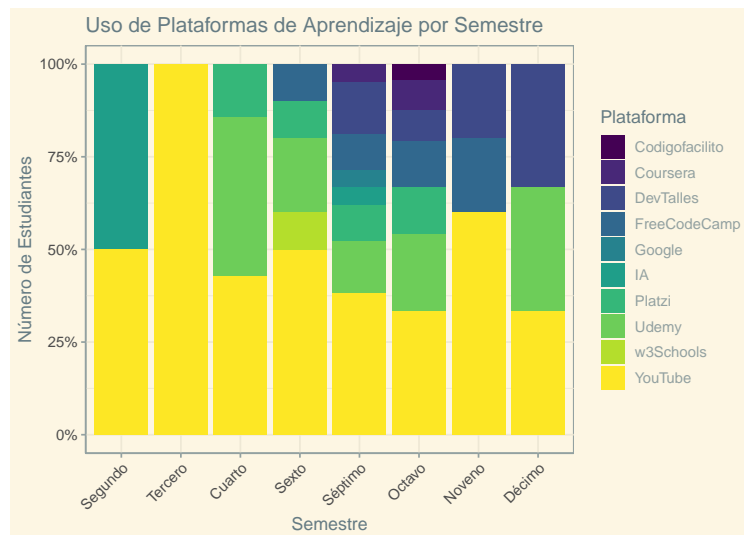To answer this question the following graphs were developed:

Figure 6: Autonomous learning platforms used by semester

Figure 6 illustrates that YouTube is the most frequently used platform across all semesters, likely due to its extensive offering of free technical tutorials covering areas like programming languages, cloud services, and frameworks. Its prevalence is particularly notable among lower-semester students, who also utilize

other free resources like w3schools, alongside paid platforms such as Udemy and Platzi. Additionally, some respondents reported using AI as a learning tool.

On the other hand, upper-semester students show a more diverse range of platform usage, predominantly featuring paid options like DevTalles, Coursera, CodigoFacilito, Udemy, and Platzi. Despite this, free alternatives such as FreeCodeCamp remain in use, and one respondent specifically mentioned Google as a learning platform.
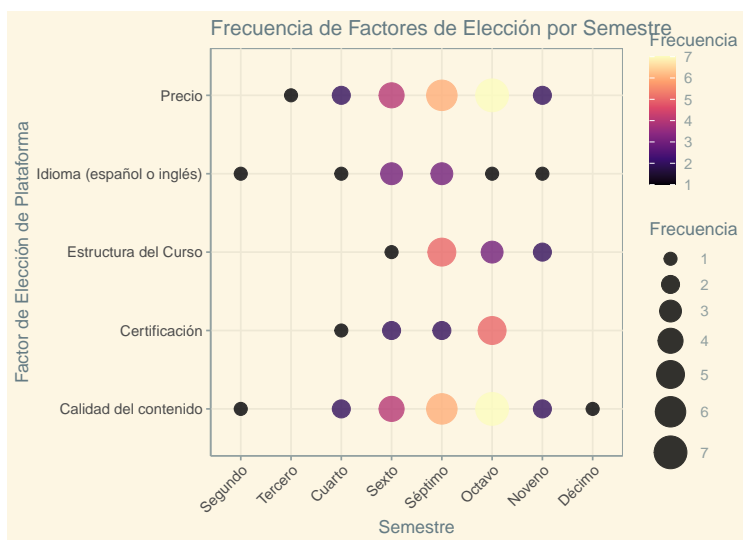

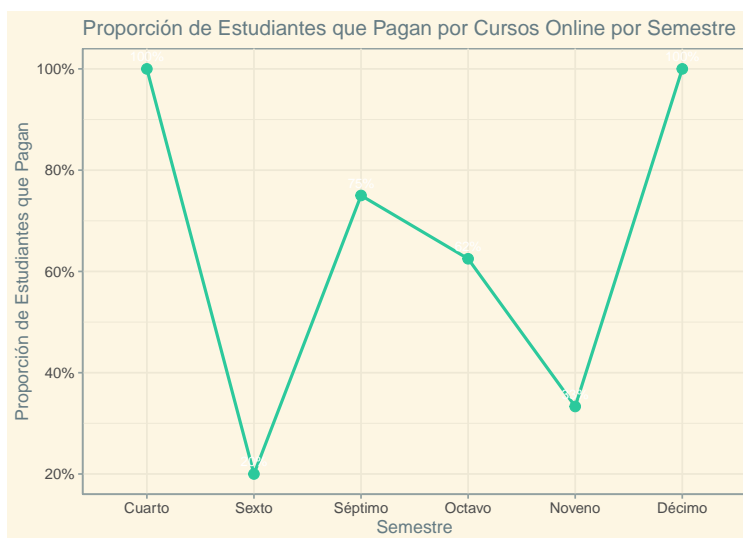Figure 7: Frequency of platforms choice factors by semester

Figure 8: Proportion of students who have spent money on courses by semester

Figure 7 illustrates that in upper semesters (specifically seventh and eighth semesters) Price and the Content Quality are the most influential factors of platform choice. Certifications and Course structure are also noteworthy considerations for some students. Language appears to be an influential factor across

8

most semesters, though its overall frequency is not exceptionally high.

Contrarily, the data collected for lower semesters is insufficient to draw meaningful conclusions.

Finally, Figure 8 indicates a higher proportion of students in upper semesters have paid for at least one online course. However, due to the modest sample size, it is challenging to draw definitive conclusions or identify clear trends from this graph, particularly since some proportions, like that for the tenth semester, are based on a single student. Nevertheless, observations from the eighth and seventh semesters reveal a considerable proportion of students who have not yet paid for an online course, even at this advanced stage of their studies.

Regarding the first hypothesis, observations suggest that students in lower semesters tend to prioritize free platforms for autonomous learning, in contrast to upper semesters where there's a more diverse engagement with both paid and free resources. Also, this shift in upper semesters does not appear to be driven by a preference for better-structured course content as illustrated in Figure 7. While data doesn't definitively explain why lower semester students might avoid paid resources (e.g., economic reasons), Figure 8 provides limited evidence from the fourth semester that students earlier in their careers can indeed afford paid courses.

For the second hypothesis, Price and Content Quality are consistently the most influential factors in platform choice, with this trend appearing particularly prevalent in the later semesters. Nevertheless, the current evidence is insufficient to definitively ascertain that certifications are the primary driving factor for choice over considerations of cost.

Second research question

What measures Informatics students take when facing obstacles, such as payment barriers or difficulties with the English language, that limit their autonomous learning of technologies?

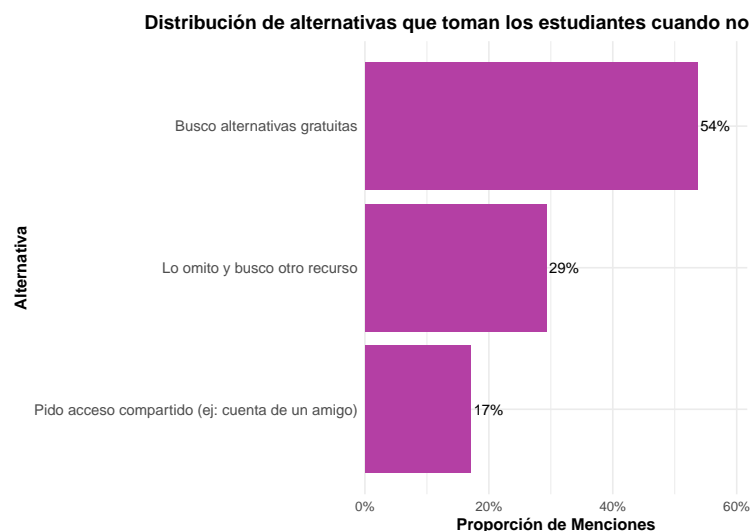To answer this question the following graph were developed:

Figure 9: Proportion of alternatives for not paying a course

Figure 9 illustrates the measures students take when facing pricing obstacles. The most prevalent strategies are searching for free alternatives or seeking other resources. In specific instances, students even share access to courses among friends. This observation aligns with our findings for the first research question, which noted that a significant number of students, even in higher semesters, continue to use free resources.

To understand the types of platforms frequented by students with limited resources (those who haven't paid for a course) and to address the third hypothesis, we developed the following graphs:
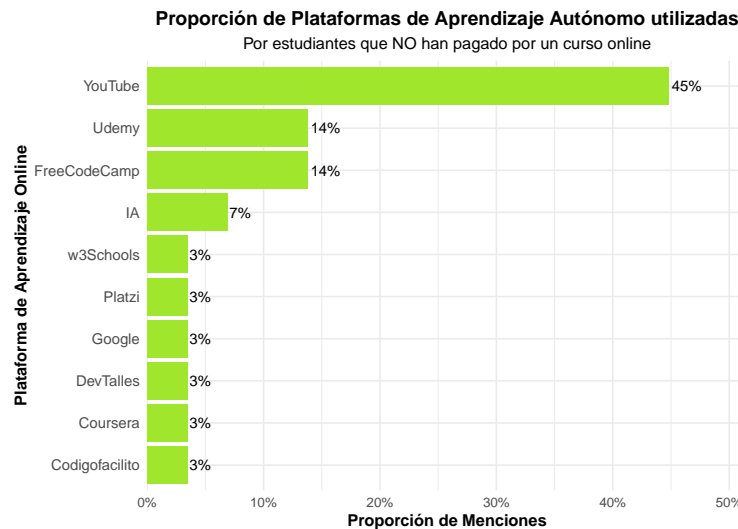
Proporción de Plataformas de Aprendizaje Autónomo utilizadas
Por estudiantes que NO han pagado por un curso online

Plataforma de Aprendizaje Online	Proporción de Menciones
YouTube	45%
Udemy	14%
FreeCodeCamp	14%
IA	7%
w3Schools	3%
Platzi	3%
Google	3%
DevTalles	3%
Coursera	3%
Codigofacilito	3%

Figure 10: Proportion of Autonomous Learning platforms for students who have never paid for an online course

Representación de si el inglés es un obstáculo para el aprendizaje autónomo de los estudiantes
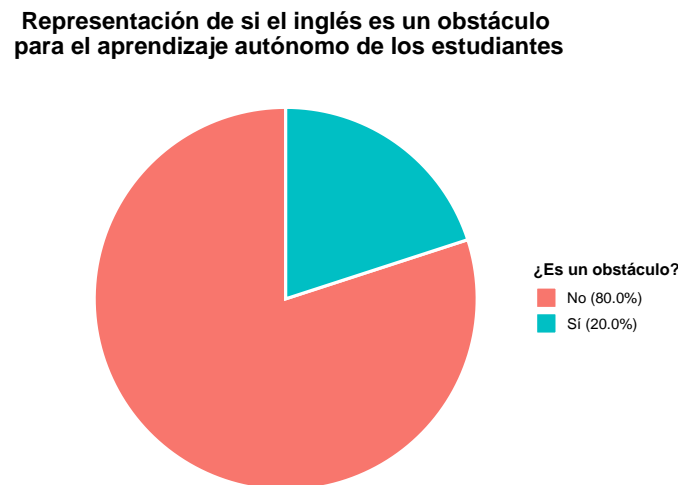
¿Es un obstáculo?
- No (80.0%)
- Sí (20.0%)

Figure 11: Proportion of students for whom English has been a barrier for learning technologies

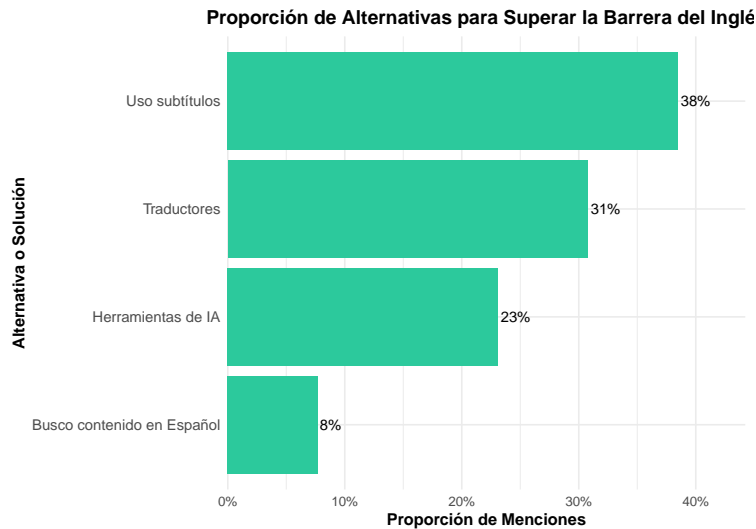**Proporción de Alternativas para Superar la Barrera del Inglé**

Figure 12: Proportion of Alternatives to surpass the English Language barrier

Figure 10 further reinforces that YouTube is the most frequently used platform for free learning resources, followed by Udemy, FreeCodeCamp, and AI tools. It's worth noting that while some typically paid platforms like Platzi, DevTalles, and Coursera are listed, most of these offer free courses. In certain instances, such as with CodigoFacilito, free course access is specifically provided to UCAB informatics students via referral codes.

Figure 11 notably demonstrates that for the majority of the sample, the English language does not present a barrier. However, to address the fourth hypothesis, we specifically analyzed the smaller percentage of the sample who do face language difficulties.

Figure 12 illustrates that these students primarily use subtitles and translators. In other cases, they rely on AI tools, and less frequently, they search for content in Spanish.

This analysis considers all students, regardless of whether they pay for courses or not. To further narrow the population, the following graph was elaborated:
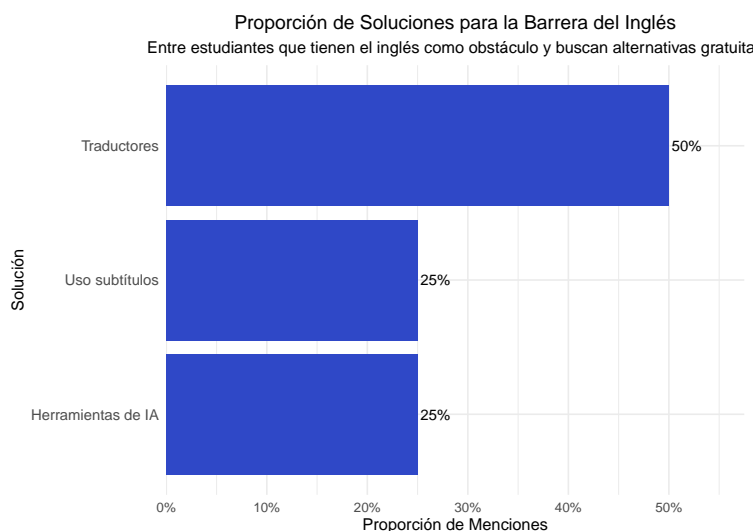

Figure 13: Proportion of Alternatives to surpass the English Language barrier by students who search for free alternatives

The most remarkable analysis from this graph reveals that students who search for free alternatives and also experience difficulties with the English language do not tend to search for content in Spanish.

Regarding the third hypothesis, it's evident that students with presumably limited resources overwhelmingly prefer free platforms or free courses offered by typically paid platforms. However, our data doesn't definitively explain why students choose these free options, even when they might recognize that paid platforms often provide more structured content. As Figure 7 indicates, course structure is only a noteworthy consideration for some students in semesters where they are already paying for courses. Furthermore, the collected data also lacks insight into whether students avoid paid courses due to their acquisition prices.

For the fourth hypothesis, it's evident that even among the small subset of students facing language difficulties, particularly those who haven't paid for a course, there's a strong preference for using external tools like subtitles, translators, or AI, rather than seeking out content in Spanish. This preference likely stems from a perception of higher quality in the English-language resources. However, it's important to note that our collected data doesn't directly support this last assertion regarding perceived quality.

Third research question

What is the relationship between free or paid courses and the contribution to the Informatics students' academic performance and professional development?

The third research question posed a particular analytical difficulty due to the exclusively discrete nature of the collected variables. Ideally, a linear regression model would have been employed, using academic

performance or professional development as dependent variables against the survey's dummy (yes/no) independent variables. However, **the absence of continuous data (e.g., GPA for academic performance, satisfaction grades for professional development) rendered these desired models impossible to develop.**

Even so, given that the survey recorded respondents' opinions on these binary questions, providing the corresponding graphs may help to extract any observable patterns.
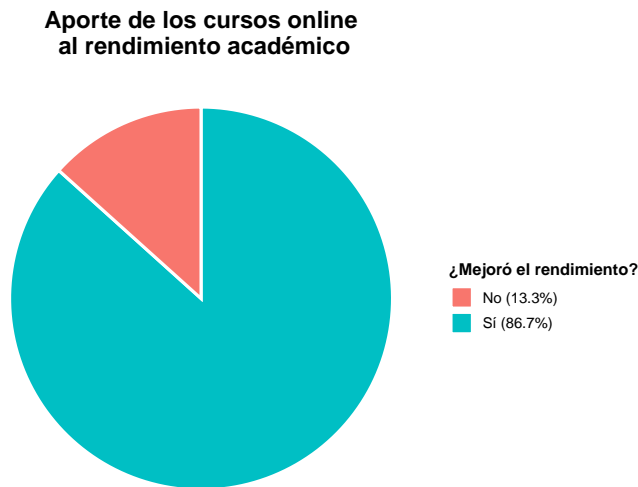
**Aporte de los cursos online
al rendimiento académico**

Figure 14: Proportion of students who have improved their performance in college projects.

**Inclusión de proyectos hechos en cursos online
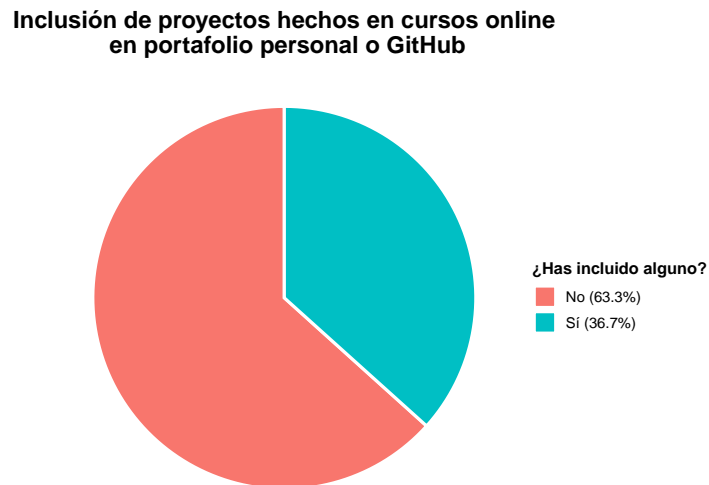en portafolio personal o GitHub**

Figure 15: Proportion of students who have included courses' projects in their portfolio.

Obtención de certificaciones o habilidades relevantes para el trabajo o pasantías gracias a los cursos online
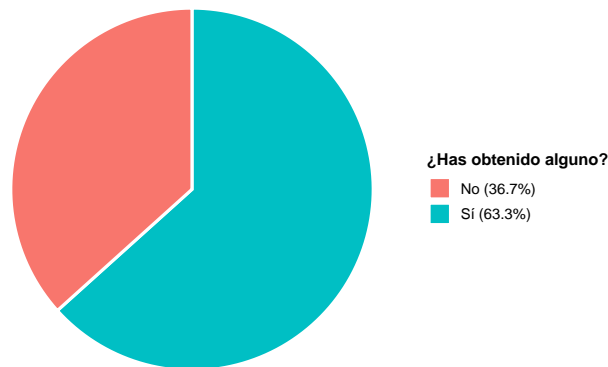
¿Has obtenido alguno?
No (36.7%)
Sí (63.3%)

Figure 16: Proportion of students who have obtained certificates or abilities useful for work or internships.

- **Figure 14** demonstrates that a significant majority of students report improved performance in college projects directly attributable to these external learning resources.
- **Figure 15**, however, reveals that 63% of students, more than half the respondents, have not included any projects from these courses in their portfolios. This outcome is likely influenced by YouTube being the most frequently used learning platform, as many of its tutorials don't involve the development of substantial, portfolio-worthy projects.
- In contrast to the findings in Figure 15, **Figure 16** indicates that 63% of students state they've gained abilities useful for work or internships through these courses. This suggests that, for a considerable portion of students, these external learning resources are effective in imparting valuable, practical skills.

Regarding the fifth hypothesis, a definitive answer is not feasible due to the absence of suitable data for linear regression modeling. This model would ideally assess the relationship between students' engagement with paid or free courses and their academic performance.

Specifically, two separate regression models were envisioned: one to explore the relationship between GPA and the usage of paid platforms, and another for GPA and the usage of free platforms. For these models, GPA would serve as the dependent variable, while the dummy variable indicating platform usage (as illustrated in Figure 14) would be the independent variable. Data would first be filtered to separate cases based on whether students had paid for a course or not.

This same modeling approach could be extended to assess the perceived level of satisfaction with professional development. In this scenario, the level of satisfaction would be the dependent variable, with the independent variable being the dummy variable from Figure 16.

For the sixth hypothesis, similar linear regression models could be developed. Here, the grade of satisfaction with professional development would remain the dependent variable, but the independent variable would be the dummy variable illustrated in Figure 15. This analysis would also involve filtering the data between students who have paid for courses and those who have not.

Without the capacity to develop linear regression models to analyze these relationships, it is not feasible to definitively answer the third research question.

Conclusion

First, This research aimed to investigate the autonomous learning behaviors of UCAB Informatics students, focusing on their engagement with online educational platforms, the factors influencing their choices, and the strategies employed to overcome learning obstacles. Despite certain analytical limitations, the study could grasp some key insights into student practices.

Regarding the first research question, it was observed that **students in lower semesters tend to prioritize free platforms for their autonomous learning, whereas in upper semesters where there's a more diverse engagement with both paid and free resources**. Additionally, **price and content quality consistently emerged as the most influential factors in platform choice**, with this trend appearing particularly prevalent in the later semesters.

For the second research question, findings indicated that **students with presumably limited resources overwhelmingly prefer free platforms or free courses offered by typically paid platforms**. Also, course structure is only a noteworthy consideration for some students in semesters where they are already paying for courses. Furthermore, it became evident that even among the small subset of students facing language difficulties, particularly those who haven't paid for a course, **there's a strong preference for using external tools like subtitles, translators, or AI, rather than seeking out content in Spanish**.

Second, the analysis of the dataset presented several limitations. First of all, the predominance of discrete variables and the modest sample size inhibited the depth of insights and the ability to draw definitive conclusions. This was particularly evident for the third research question, which could not be definitively answered due to the absence of necessary quantitative variables (e.g., student GPA, satisfaction grades for professional development) that would enable the development of linear regression models to analyze some relationships.

Additionally, the limited number of responses from lower semester students led to a concentration of conclusions around upper semesters, which may not fully reflect the experiences or learning strategies of students earlier in their careers.

For future research, in order to get more comprehensive and conclusive insights into student autonomous learning behaviors, the following adjustments to survey design are recommended.

For the first hypothesis, it would be needed a specific questions exploring students' reasons for not paying for courses (e.g., "What factors have influenced your decision to never pay for an online course?"). This same question could prove to be helpful for getting some insight into the third hypothesis, alongside a graded scale measuring students' preference for English content over Spanish content which would provide valuable quantitative insight for the fourth hypothesis.

For the third research question, future surveys should collect continuous quantitative variables such as student GPA and a graded level of satisfaction regarding the impact of online courses on professional

development. This would enable the application of linear regression models to analyze the relationships associated with academic performance and professional development.

Also, some efforts should be made to secure a larger and more balanced sample, especially from lower semesters, so that the research could provide a better understanding of student experiences across all academic stages.

This research, despite its data limitations, has provided some valuable insights into the autonomous learning practices of UCAB Informatics students. It has highlighted the significant role of external online platforms for students and their preferences, influential factors over their decisions, and their coping mechanisms for common barriers. While certain relationships could not be established, the discernible patterns and trends suggest that further investigation is justified. With a bigger dataset and the inclusion of key quantitative variables, future research could uncover more reliable insights, providing the UCAB Informatics Engineering faculty with the necessary information to support and enhance student learning in the educational landscape.