Entrance exams and first-semester performance: A multi-cohort visual correlation analysis addressing inter-instructor grading bias

Exámenes de admisión y rendimiento en el primer semestre: Un análisis visual de correlaciones multicohorte que considera el sesgo de calificación entre docentes

Luciana N. Huertas-Condori[1], Israel N. Chaparro-Cruz[2], Silvana B. Cabana-Yupanqui[2], Americo Chaparro-Guerra[3]

[1] Universidad Nacional Jorge Basadre Grohmann. Facultad de Ingeniería. Escuela Profesional de Ingeniería en Informática y Sistemas. Tacna, Perú. E-mail: lhuertascon@unjbg.edu.pe.

[2] Universidad Nacional Jorge Basadre Grohmann. Facultad de Ingeniería. Departamento Académico de Ingeniería en Informática y Sistemas. Tacna, Perú. E-mail: ichaparroc@unjbg.edu.pe scabanay@unjbg.edu.pe.

[3] Universidad Nacional Jorge Basadre Grohmann. Instituto de Investigación. Vicerrectorado de Investigación. Tacna, Perú. E-mail: achaparrog@unjbg.edu.pe.

DOI: 10.64966/ingeniare.v34.04

Abstract

University entrance exams are widely used to select candidates for higher education, yet their predictive validity remains contested, especially given institutional biases such as grading inconsistencies among instructors. This study assesses the extent to which subject-area scores from entrance exams predict first-semester academic performance at the Universidad Nacional Jorge Basadre Grohmann (UNJBG) in Peru, incorporating a novel visual correlation analysis that accounts for inter-instructor grading bias. Using data from three consecutive admission cohorts (2017–2019) and more than 4,500 students across four academic tracks, section-level Spearman correlations were computed between entrance exam scores and course grades, enabling fairer, bias-controlled insight into performance patterns. The results reveal that certain cognitive domains, particularly Verbal Reasoning and Mathematical Reasoning, show consistent, moderate-to-strong associations with early academic success across tracks, while others demonstrate track-specific relevance. Visualization these relationships, disaggregated by Study Track, uncovers nuanced predictors of academic progression and provides empirical evidence to support more equitable and discipline-sensitive admission policies. This approach contributes both methodologically and practically to the field of educational assessment, emphasizing the role of data visualization in informing institutional decision-making.

Keywords: entrance exams, academic performance, visual analysis.

Resumen

Los exámenes de admisión universitaria se utilizan ampliamente para seleccionar candidatos a la educación superior, pero su validez predictiva sigue siendo objeto de debate, especialmente al considerar sesgos institucionales como las inconsistencias de calificación entre docentes. Este estudio evalúa en qué medida los puntajes por áreas de evaluación de los exámenes de admisión predicen el rendimiento académico del primer semestre en la Universidad Nacional Jorge Basadre Grohmann (UNJBG) en el Perú, incorporando un novedoso análisis visual de correlaciones que toma en cuenta el sesgo de calificación entre docentes. A partir de datos de tres cohortes consecutivas de ingreso (2017–2019) y más de 4,500 estudiantes en cuatro áreas académicas, se calcularon correlaciones de Spearman a nivel de sección entre las notas de admisión y las calificaciones de los cursos, lo que permite una visión más justa y controlada del desempeño. Los resultados muestran que ciertos dominios cognitivos, en particular Razonamiento Verbal y Razonamiento Matemático, presentan asociaciones consistentes y de moderadas a fuertes con el éxito

Autor de correspondencia: ichaparroc@unjbg.edu.pe

Received: 5 September 2025 | Accepted: 24 April 2026

académico temprano en las distintas áreas, mientras que otros evidencian relevancia específica según la trayectoria académica. Al visualizar estas relaciones y desagregarlas por área de estudio, el trabajo revela predictores matizados de la progresión académica y aporta evidencia empírica que respalda políticas de admisión más equitativas y sensibles a las disciplinas. Este enfoque contribuye tanto metodológica como prácticamente al campo de la evaluación educativa, resaltando el papel de la visualización de datos en la toma de decisiones institucionales.

Palabras clave: exámenes de admisión, rendimiento académico, análisis visual.

INTRODUCTION

University entrance exams are widely used worldwide as a fundamental criterion for admission to higher education. These assessments are intended not only to filter applicants but also to identify candidates most likely to succeed academically. Yet, the actual efficacy of entrance exams in predicting university performance remains a subject of ongoing investigation, particularly when seeking to understand which components of the exam are most strongly associated with early academic outcomes, and how institutional factors, such as variation in grading across different course sections, may distort these measures.

Though these entrance exams are often standardized to provide equal metrics for assessing students' readiness for a higher-level education, the subsequent evaluation of their academic performance is often less standardized. Sometimes, classes that consist of large groups of students require multiple graders; such practice raises concerns about discrepancies between instructors and their different grading standards and expectations [1]. This inconsistency introduces noise and potential bias in grading that can be easily missed at the moment of evaluating students' actual performance.

At the Universidad Nacional Jorge Basadre Grohmann (UNJBG) in Peru, entrance exams are structured by Study Track, each comprising a unique combination of eight subject areas. Thousands of students apply each year, but only a small proportion (approximately 19%) are admitted. For these students, performance during the first semester is critical. This academic period includes foundational specialty courses aligned with each student's field of study and typically sets the trajectory for their continued progression. Unlike subsequent semesters, which may be influenced by prerequisites, course availability, or changes in enrollment behavior, the first semester provides a standardized and comparable context for assessing academic preparedness. This focus is further justified by the university's annual admission structure, which means that students may not be able to enroll in all required second-semester courses due to prerequisite chains or scheduling constraints. These limitations introduce variability in course enrollment and academic metrics, potentially biasing any longitudinal assessment beyond the first term.

Previous studies at UNJBG have examined the relationship between entrance exam results and students' performance. A recent analysis of the 2017 cohort, for example, found moderate positive correlations between scores in Verbal Reasoning and Language and academic achievement, as well as a strong negative association between overall entrance exam scores and student dropout [2]. While valuable, such findings represent a general interpretability that easily could fall into unintentional biases found in the inter-instructor grading, not letting uncover deeper patterns that exist between their admission entrance exam and their actual academic performance.

The present study introduces a new analysis method that aims to reduce the noise in the data to find deeper relationships between how a student scored in the entrance exam and their performance in a particular class of their first semester, to address these biases, reduce the existing bias of their different instructors across the specialty courses, and find more specific patterns. In particular, different visual correlation analyses were applied to account for potential grading biases accross different course sections, enabling us to find specific specialty courses aligned with the subject areas tested in the entrance exam. This enables a more interpretable understanding of which cognitive domains are most aligned with early academic success.

Data visualization is a promising tool for decision-making, as human cognitive abilities prevent the simultaneous, rapid processing of large volumes of data. Visualization facilitates more efficient interpretation of complex information. In higher education, various key actors use data to inform critical decisions. For example, teachers use it to identify gaps and students' knowledge and adjust their strategies and methodologies accordingly. However, the adoption of visualization tools for such analyses remains limited [3].

This study contributes to the literature in several key ways. It expands prior analyses by incorporating data from three consecutive admission cohorts (2017–2019); it focuses exclusively on first-semester outcomes to ensure comparability; and it applies an in-depth data analysis to address known institutional biases, such as variation in grading across course sections. Ultimately, this research aims not only to assess how well entrance exams predict early academic performance but also to visualize these correlations and generate insights that support the development of more effective and equitable admission policies in higher education. Given that few changes have been made to the admission process since 2019, the findings underscore the need of revisit and refine current practices based on empirical evidence.

LITERATURE REVIEW

Given that the relationship between university entrance exams and future academic performance is a wide area of research, this literature review focuses on three main branches: i) traditional statistical methods, ii) visual analytics and interpretability techniques, and (iii) instructors' roles in students' performance. These topics represent the baselines of the current research and their influence in the uncovering of nuanced patterns.

Studies using traditional statistical methods

Several studies have evaluated the predictive validity of entrance exams using correlation coefficients, regression models, or descriptive statistics. For example, [2] conducted a longitudinal case study at Jorge Basadre Grohmann National University (UNJBG) using Spearman correlations and regression-based thresholds. They found moderate positive correlations between verbal reasoning and language scores and first-year GPA, and a strong negative correlation between entrance scores and dropout rates.

[4] evaluated the predictive validity of a uniform entrance test across three health science programs in Pakistan. They reported modest correlations (rs = 0,334 in the first semester) and noted a decline in predictive power over time, particularly in later semesters.

Similarly, [5] examined data from Czech universities and identified a statistically significant but modest relationship between entrance exam scores and university GPA using Spearman's rank correlations and regression analysis. In the Peruvian context, [6] reported a low but positive relationship between cognitive entrance test scores and first-year academic achievement at UNMSM.

Other traditional approaches highlighted additional predictors. [7] found that while MDCAT (Medical College Admission Test) scores correlated with pre-clinical performance, a combination of cognitive and non-cognitive measures (such as MMI scores) improved predictive capacity. Likewise, [8] in Portugal used multilevel modeling and showed that entrance scores were significant predictors of GPA, although the magnitude of the effect varied by academic program.

In general, these studies suggest that entrance exams have some degree of predictive validity, particularly in early semesters, but their explanatory power is limited when using linear models alone.

Studies incorporating visual analytics and interpretability

A smaller group of studies emphasizes visualization. [9] described visual data analysis as a blend of statistics, data mining, and graphical exploration. They emphasized its ability to navigate complex educational datasets and to understand conceptual relationships.

For example, [10] employed visual analytics within a clustering approach to identify similarities among failed courses, associate them, and evaluate their impact on student performance. However, the study was limited to data collected after students' enrollment, thereby constraining the analysis to variables related exclusively to their academic trajectory within the institution.

[11] employed visual tools to distinguish patterns between high- and low-performing students, while [12] used t-SNE to clarify GPA cluster separability. [13] utilized Shapley additive explanations to assess variable importance, revealing that social and contextual factors often surpass cognitive ones in explanatory value.

Together, these studies underscore the growing relevance of visual analytics in education, particularly for guiding decision-making and enhancing model transparency.

Studies including the inter-instructor bias

Several studies highlight the variability in grading outcomes when multiple instructors are involved in the assessment process. Grading remains the most common method for assessing student performance, whether through numerical scores or more standardized categorical scales. However, concerns regarding its reliability and consistency have persisted over time. As noted by [14], percentage-based systems (numerical) used in the early twentieth century were pointed as highly unreliable, with different instructors assigning significantly different grades to the same work. Grading practices continue to face criticism for inconsistency and subjectivity, as students may receive varying grades depending on the evaluator and the timing of the assessment.

[15] proposed a method to "bridge the gap" of grading by statistically aligning said grading scales across different instructors for the same class. Arguing that multi-grader teaching environments arise bias due to differences in severity and error, and assigning just one instructor to large classes

is an unrealistic expectation. This approach supports the use of methods to adjust for inter-instructor biases and track and address certain grading patterns.

Prior literature highlights that while entrance exams can predict early academic performance to some extent, this correlation is usually moderate and points to other factors. Visual analytics remains an essential tool for the different education agents to inform and refine their instructional strategies. The inter-instructor bias is mentioned as a significant concern during evaluation, particularly in grading, as it can introduce inequity and data variability. Addressing this bias is essential when analyzing patterns.

RESEARCH METHODOLOGY

Research design

This study adopts a quantitative, observational, and exploratory research design, integrating visual analytics with bias-aware correlation techniques to assess the predictive validity of university entrance exam scores. The research focuses on identifying relationships between subject-area scores from entrance exams and students' academic performance during the first semester, while accounting for instructor-based grading variability.

The design includes the following key features:

- Multi-cohort analysis: The study examines data from three consecutive admission cohorts (2017–2019), allowing for the identification of consistent trends across time.
- Cross-sectional academic focus: Analysis is limited to the first academic semester to ensure comparability and avoid bias introduced by course sequencing or delayed enrollment.
- Bias-controlled correlation analysis: Section-level visual analytics are employed to reduce grading bias and better capture the relationship between cognitive entrance domains and university performance.
- Integrated institutional data: The study links entrance exam records with detailed course-level performance data for admitted students, enabling subject-level and course-level correlation studies.

This research design is structured to support both descriptive insights and methodological contributions, providing a replicable framework for evaluating the effectiveness of exam entrance in other institutional contexts.

Datasets description

This study utilizes two institutional datasets from the Universidad Nacional Jorge Basadre Grohmann (UNJBG) in Peru. These datasets correspond to the university entrance exam and academic performance records. These datasets were collected for three consecutive admission cohorts: 2017, 2018, and 2019. Each dataset serves a distinct yet complementary role in assessing the efficacy of entrance exams as a predictor of first-semester academic success.

Entrance exam dataset

The first dataset contains records of all applicants who participated in the university's annual entrance exam process. The dataset comprises: 7,829 applications in 2017, 8,651 applications in 2018, 9,507 applications in 2019.

Each exam is tailored according to the student's chosen study track. UNJBG organizes its 34 undergraduate programs into four major study tracks (Table 1 in the previous study in [2]):

Each study track features a unique combination of 8 subject areas in the entrance exam, drawn from a broader pool that includes Verbal Reasoning, Mathematical Reasoning, Language, Arithmetic and Algebra, Geometry and Trigonometry, Physics, Chemistry, Biology, Logic, History, Literature, Economy, and Geography. The weight and structure of the exam are determined by the track, and the number of questions per subject area varies accordingly (Table 1 in the previous study [2]).

For each applicant, the dataset contains:

- Year of application.
- Academic track (Study Track).
- Program applied for.
- Subject area scores (as percentage of correct answers).
- Total exam score (out of 600).
- Admission process.

Academic performance dataset

The second dataset consists of university academic records of students who were admitted and enrolled in at least one course during the first academic semester following their admission. The total number of students included in the analysis by cohort is: 1,499 in 2017, 1,529 in 2018, and 1,506 in 2019.

Table 1. Number of students per year and study track.

Year	Study Track 1	Study Track 2	Study Track 3	Study Track 4	Total
2017	317	619	326	237	1499
2018	322	621	333	253	1529
2019	318	665	303	220	1506
Total	957	1905	962	710	4534

These records contain detailed grade reports for each course the student took during their first semester. The selection of only first-semester records was deliberate and is based on two key considerations:

1. Curricular integrity: The first semester includes foundational courses specific to the academic program and provides a standardized academic experience before variations introduced by elective choices or program specializations.

2. Institutional enrollment policy: UNJBG admits students only once per year. Therefore, in later semesters, students may not always enroll in the full set of recommended courses due to prerequisite constraints or academic delays. This can result in an inconsistent courses-taken, potentially biasing average performance measures.
3. Pandemic-related consistency: The decision to limit the analysis to the first semester is also grounded in the need to maintain temporal and contextual consistency in academic delivery. Starting in 2020, due to the global COVID-19 pandemic, UNJBG transitioned to virtual instruction, a modality that remained in place until 2022. This shift introduced significant changes in course structure, assessment formats, and student engagement, potentially affecting grade distributions and academic performance patterns. By focusing exclusively on pre-pandemic semesters, the study ensures that all evaluated academic records correspond to a uniform in-person instructional context, thereby avoiding confounding effects associated with emergency remote education.

The academic dataset includes detailed first-semester information for each enrolled student, with records structured at the course level. For every course taken, the dataset captures:

- Student identifier code.
- Year of admission.
- Faculty and academic program.
- Course name.
- Shift (morning/evening).
- Section identifier (used to control for potential instructor-related grading effects).
- Final grade (0 to 20).

Each student is typically enrolled in approximately seven courses during the first semester. This structure enables both student-level and course-level analyses, ensuring that academic performance can be evaluated consistently across individuals and programs.

The final integrated dataset comprises 4,534 students across three cohorts and four academic study tracks. Specifically, Study Track 1 (Health and Biomedical Sciences) accounts for 957 students, Study Track 2 (Exact Sciences, Engineering, and Architecture) for 1,905 students, Study Track 3 (Social Sciences and Humanities) for 962 students, and Study Track 4 (Actuarial and Business Sciences) for 710 students. Each academic year contributes approximately one-third of the total sample, with 1,499 students in 2017, 1,529 in 2018, and 1,506 in 2019. This balanced distribution across both time and academic domains enhances the reliability of the findings and enables meaningful cross-cohort and cross-disciplinary comparisons in the subsequent analysis. Figure 1 presents a schematic diagram outlining the data flow and analysis process, providing a comprehensive overview of the available data.

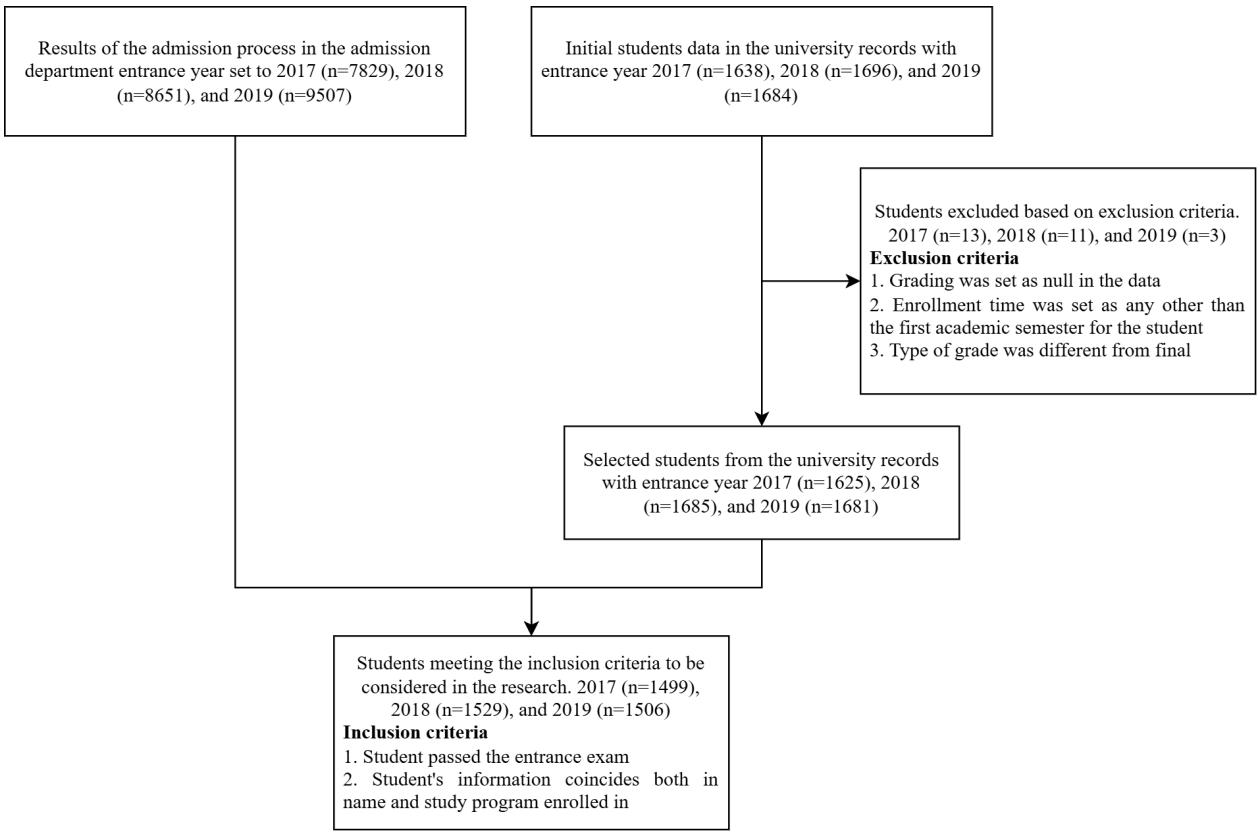

Figure 1. Flow diagram of student data selection and eligibility criteria.

Preprocessing and data cleaning

The data preparation process involved several stages to ensure data integrity, consistency across cohorts, and robustness in the analytical and modeling phases. The preprocessing pipeline integrated both entrance exam records and academic performance data, applying filters and transformations aligned with the study's objectives.

- The two primary data sources, entrance exam scores and academic performance records, were merged by matching student identifiers and the academic program into which each student was admitted. This ensured that the entrance exam subject areas corresponded to the appropriate academic track and course enrollment records.
- Records with null or missing final course grades were excluded to maintain the reliability of academic outcome variables.
- Only official final grade records were retained. Grades from substitute exams, validations, or other non-standard grade sources were removed to avoid distortion of performance metrics.

Following this general preprocessing, the dataset was prepared for the Visual Data Analysis. Requiring tailored treatment of the academic performance variables to account for instructor-related grading variability, commonly known as evaluation bias.

Visual data analysis

This study investigates the predictive validity of entrance exams while accounting for inter-instructor grading bias by employing a section-level correlation methodology (Figure 2 for an overview of the process).

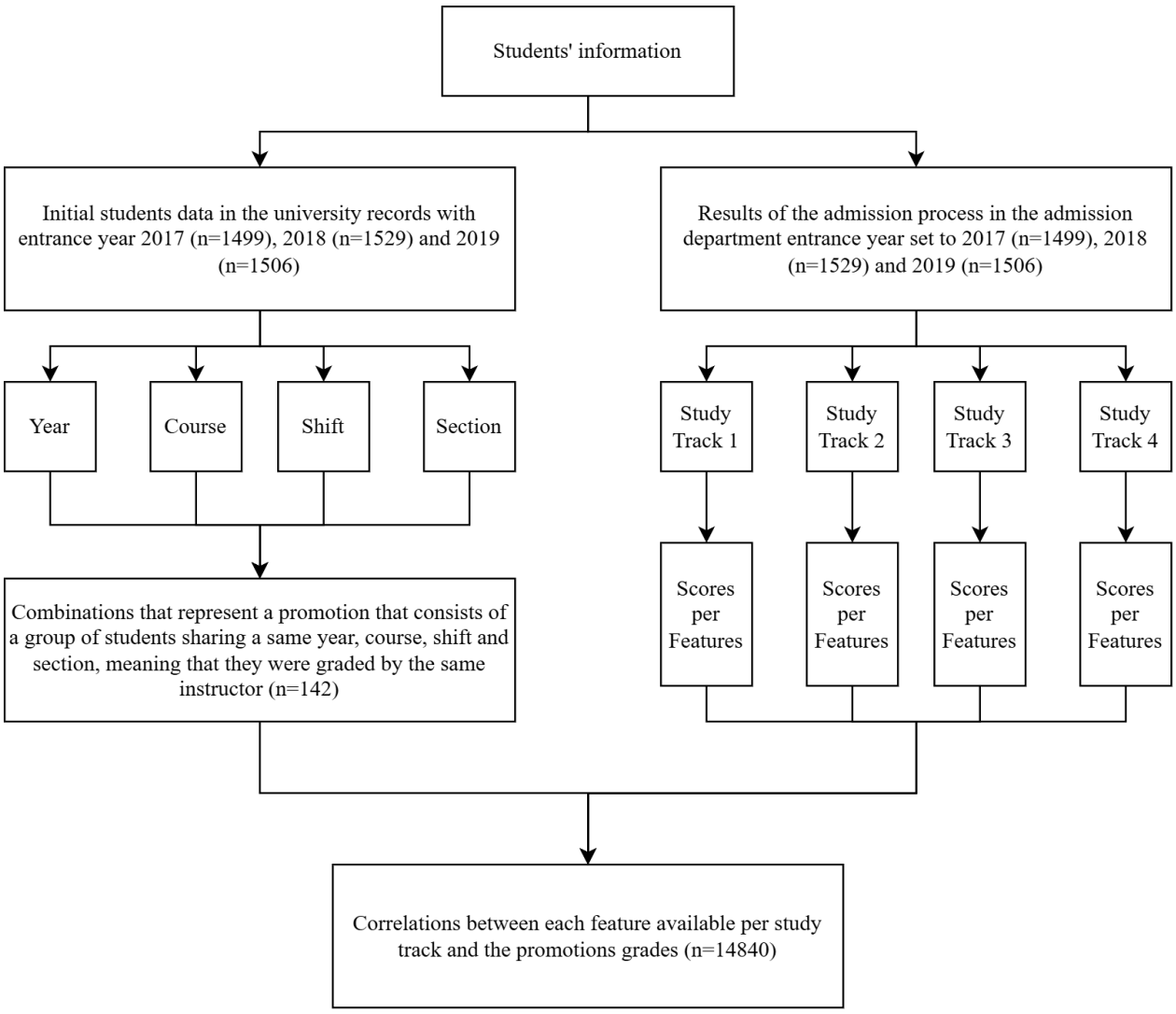

Figure 2. Analytical workflow for section-level correlation analysis of entrance exam scores and first-semester academic performance.

Rationale and problem framing

A major challenge in evaluating the predictive validity of entrance exams is the variability in grading practices across instructors. At UNJBG, as in many universities, a single course may be delivered in multiple sections by different faculty members. Despite a shared syllabus and assessment criteria, grading practices may diverge due to differences in expectations, interpretation of rubrics, or evaluation rigor. This phenomenon, known as inter-instructor grading bias, is documented in a small number of studies in the educational literature [14], [15]. If not addressed, it can obscure genuine patterns between entrance exam performance and subsequent academic achievement.

Conventional correlational analyses that aggregate data across all students, without accounting for the section or instructor, risk capturing grading inconsistencies rather than meaningful cognitive alignments. Therefore, a more granular approach is needed to reduce bias and uncover valid associations between exam subject areas and students' early academic performance.

Section-level correlation computation

This study computes correlation coefficients at the course section level to mitigate the effects of instructor bias. Within each section (i.e., a group of students taught by the same instructor in the same semester), a Spearman correlation is calculated between students' entrance exam scores in each subject area and their final grade in the course.

This approach ensures that every computed correlation reflects a consistent instructional and evaluative context. Because all students in a section are graded by the same instructor under the same conditions, the resulting correlation is less likely to be confounded by grading disparities.

This computation was repeated across all available sections in each course, for each year and study track. As a result, a large set of section-level correlations was generated, covering the full set of first-semester courses offered between 2017 and 2019.

Aggregation and visualization strategy

After computing the section-level correlations, the results were aggregated by course and subject area to facilitate institutional-level analysis. Specifically, boxplots were used to visualize the distribution of correlation coefficients for each subject area across all course sections.

This visual strategy allows for the identification of consistent patterns, such as whether scores in a particular entrance exam subject (e.g., mathematical reasoning) tend to correlate more strongly with student performance in certain courses (e.g., calculus or physics).

The use of boxplots provides a clear summary of the central tendency and dispersion of correlation values, enabling stakeholders to assess not only the strength but also the stability and reliability of the observed relationships. Outliers and variability across cohorts or sections can also be readily identified.

Advantages over pooled analysis

Traditional pooled correlation analyses, while simpler to compute, inherently assume homogeneity in grading across all course sections. This assumption is not valid in practice and can introduce systematic errors, particularly in institutions where multiple instructors are involved in first-year teaching.

By contrast, the section-level disaggregation used in this study offers several advantages:

- Bias reduction: Instructor-related variability is neutralized by isolating correlations within uniform grading environments.
- Higher interpretability: Findings reflect more accurate relationships between entrance exam subject areas and course-specific performance.

- Improved policy relevance: University administrators and curriculum designers can use these insights to refine admission processes and early interventions, grounded in evidence that accounts for institutional realities.
- Reproducibility and transparency: The method is systematic and replicable across departments or future cohorts.

Overall, this visual and bias-aware approach offers a methodologically sound foundation for exploring the academic validity of entrance exams in real-world university settings.

ETHICAL CONSIDERATIONS

This research was conducted in compliance with established ethical guidelines. Ethical clearance was granted by the Ethics Committee of Jorge Basadre Grohmann National University on October 14, 2024, permitting the use of academic records for analytical purposes under strict conditions to protect student confidentiality and privacy.

All datasets were fully de-identified prior to analysis to ensure participant anonymity. Any information that could directly or indirectly reveal the identity of individual students was removed, thereby preventing any potential re-identification. These measures were implemented to uphold data protection principles throughout the entire research process.

RESULTS

An analysis of the relationship between the subject covered in the entrance exam and academic performance in the first-semester was conducted using boxplots. Figure 3 displays the

distribution of Spearman correlation coefficients between each subject area and students' course grades across individual course sections, thereby controlling for instructor-related grading bias.

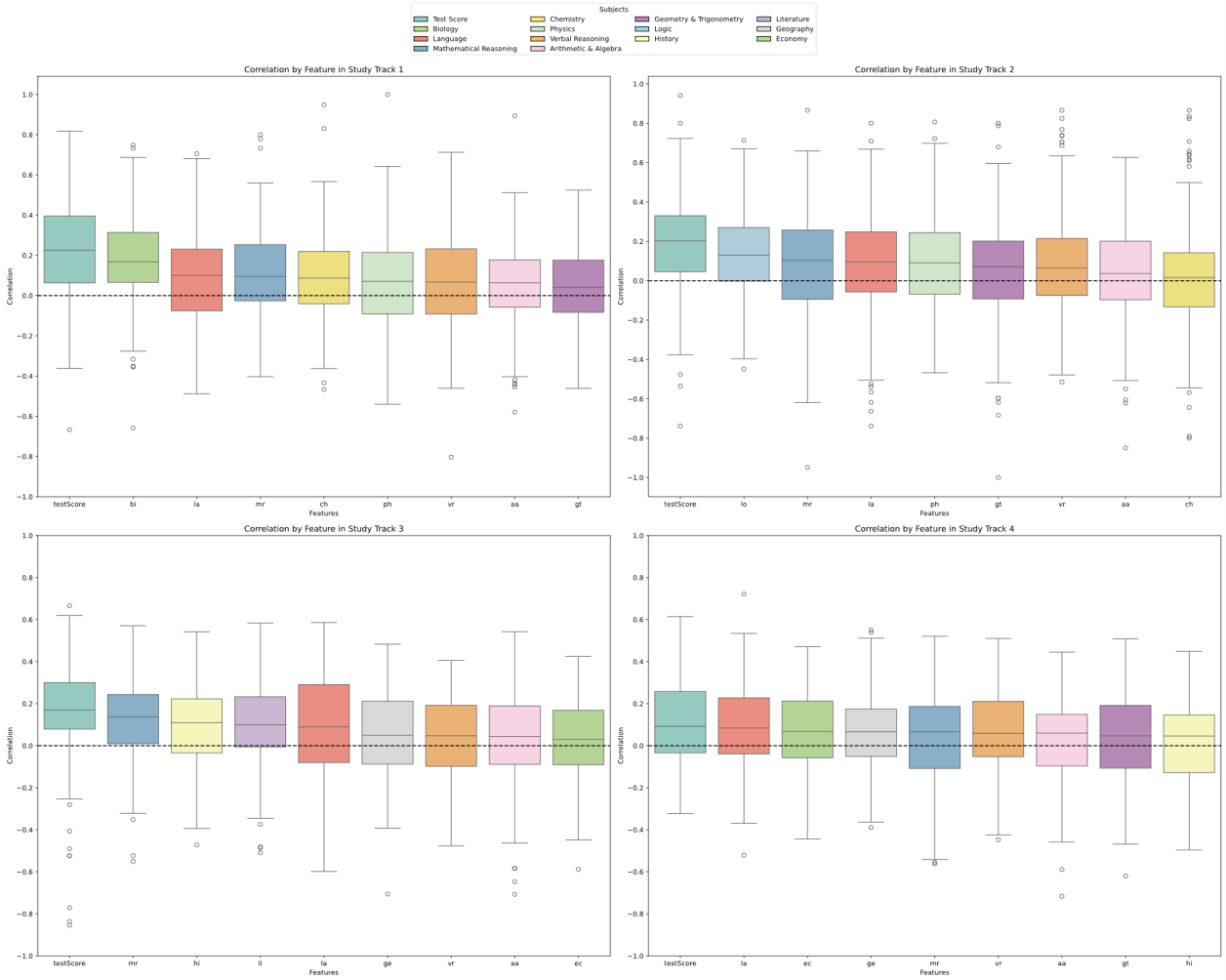

Figure 3. Boxplot of correlation between subject areas and first-semester academic performance along courses.

The visualization reveals a notable asymmetry in predictive strength across subject areas, reflecting how different competencies assessed in the entrance exam align with actual academic success in early university coursework. Among the subject areas, verbal reasoning (VR) and language (LA) consistently exhibit positive median correlations, with relatively narrow interquartile ranges (IQR), indicating a stable, moderately strong association with academic performance across sections. This suggests that reading comprehension and linguistic aptitude may be foundational across disciplines.

In contrast, subject areas such as arithmetic and algebra (AA) and mathematical reasoning (MR) exhibit higher variability and a broader range of correlation values, including some close to zero or slightly negative. This variability may reflect the discipline-specific relevance of these areas, more predictive in quantitative programs and less so in humanities or social sciences.

Other areas, such as history (HI), literature (LI), and economy (EC) generally show weaker, more dispersed correlations, with several sections exhibiting negligible or no relationship. This pattern may be influenced by differences in content alignment between entrance exam questions and first-semester course material, as well as the subjective nature of grading in these disciplines.

Importantly, the boxplot approach allows identification of systematic patterns beyond what a global correlation could provide. Estimating correlations at the section level and visualizing their distribution helps minimize the confounding effect of inter-instructor variability, which often skews aggregate performance measures. This methodological choice ensures that the observed associations are more likely to reflect true content alignment rather than grading idiosyncrasies.

The variability (box width) captures the extent of association across different course sections. When controlling grading bias, the analysis reveals a broader spread of correlation values, suggesting genuine, subject-specific relationships with academic performance. In contrast, pooled analyses without this control typically result in narrower boxplots with values closer to zero, resembling random noise, and obscuring meaningful patterns.

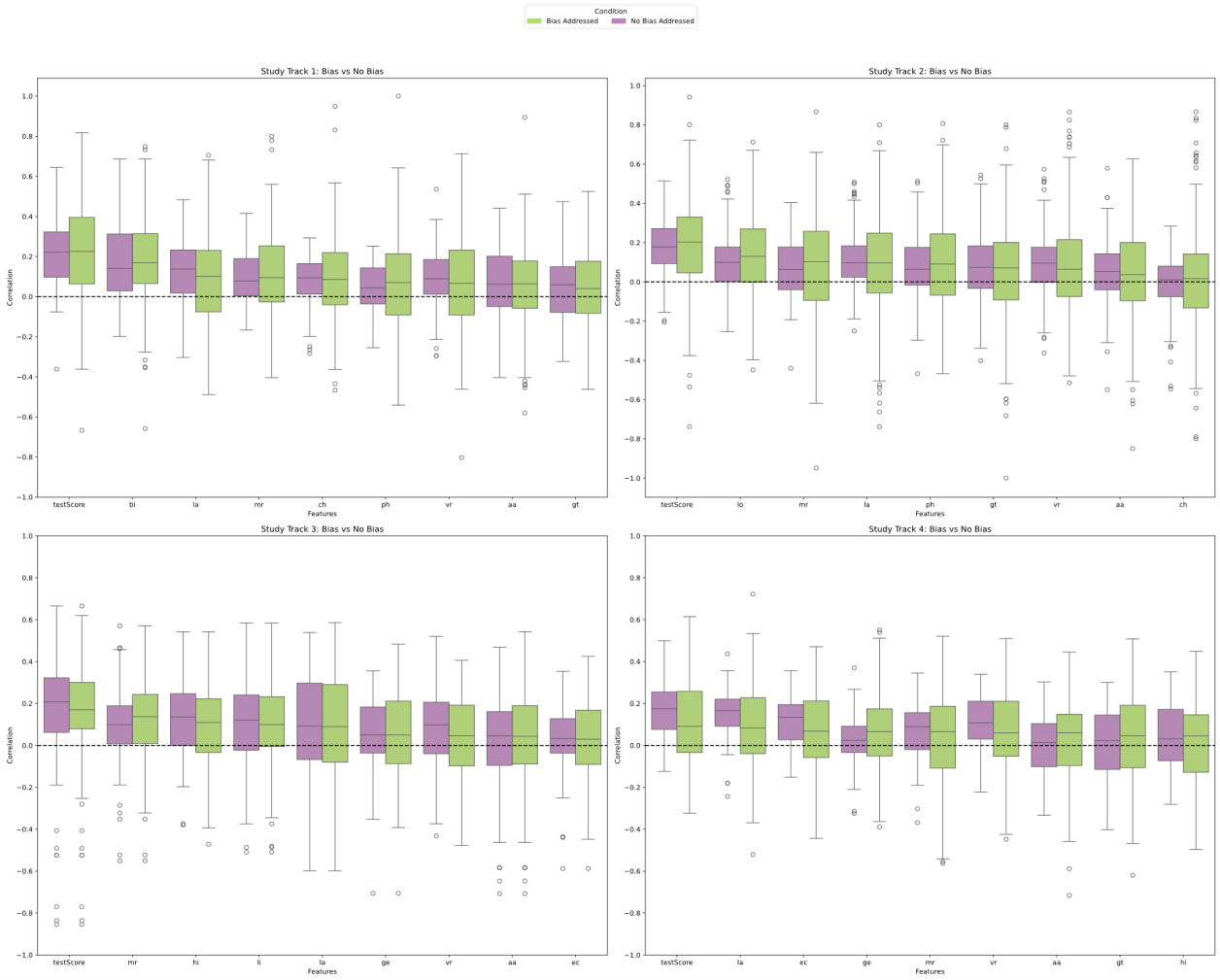

Figure 4. Boxplots comparison of correlation distributions with and without addressing the inter-instructor grading bias.

In contrast, a direct visual comparison is displayed of the two analytical conditions is shown in Figure 4, one addressing inter-instructor grading bias and the other omitting this adjustment, further illustrating the magnitude of the bias's impact on observed relationships. When the bias is explicitly targeted by grouping students according to the previously mentioned characteristics, the distribution of correlation coefficients exhibits both greater dispersion and clearer differentiation

across features, revealing discipline-specific patterns that were previously obscured. The broader interquartile ranges and higher median shifts observed under the "Bias Addressed" condition suggest that once said bias is accounted for, the associations more faithfully represent the intrinsic link between entrance exam competencies and subsequent academic outcomes, disentangling important relationships normally disregarded.

In contrast, the "No Bias Addressed" shows smaller differences in the IQR of the boxes, indicating that the unadjusted data tend to dilute the genuine signal with noise stemming from inconsistent grading practices. Consequently, the contrast between both conditions underscores the methodological importance of isolating inter-instructor effects when studying predictive validity. Without such correction, apparent weak correlations may not signify an absence of association, but rather the masking influence of uneven evaluative processes. Therefore, addressing grading bias enables a more nuanced and authentic interpretation of how entrance exam domains align with academic achievement, supporting the view that academic performance is influenced by previous evaluations and also depends on the instructor's evaluative standards.

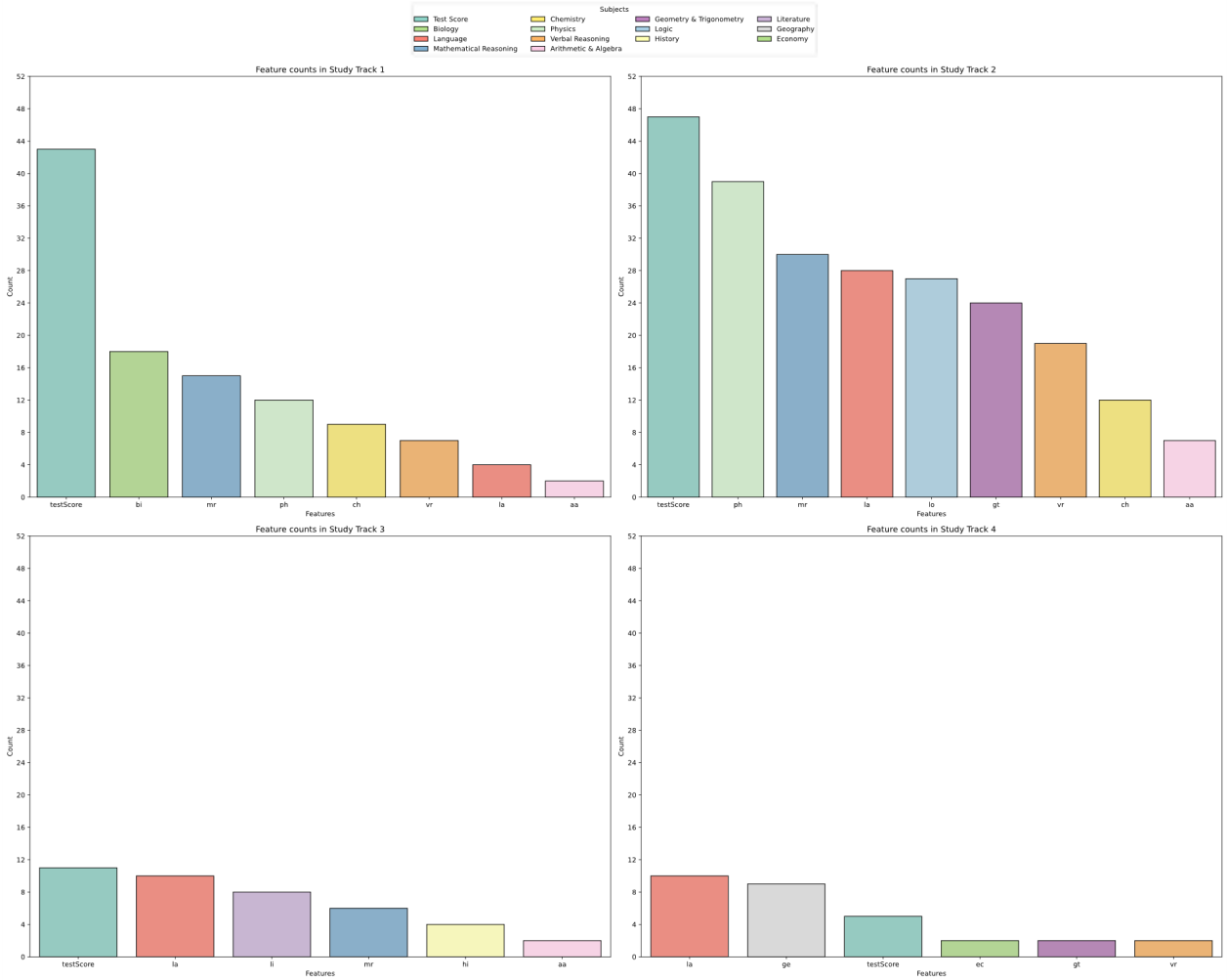

Figure 5. Frequency of entrance exam features with moderate-to-strong correlations (≥ 0,4) to student performance by Study Track.

A second relevant stage of visual analysis focuses on identifying which features exhibit moderate to very strong correlations with student promotion outcomes. Following the guidelines proposed by [16], correlations equal to or greater than 0.4 are considered meaningful, as they indicate a moderate to strong relationship. Figure 5 presents a bar plot summarizing the frequency with which each feature, within each study track, meets or exceeds this correlation threshold. This approach provides an aggregated perspective of feature relevance, helping to identify consistent predictors of academic progression across diverse academic paths.

The visualization reveals that certain features, particularly those derived from standardized entrance exam components, demonstrate recurrent associations with promotion. Notably, Mathematical Reasoning (MR) emerges as one of the most consistent and frequently correlated features across all three study tracks. Its persistent presence suggests that quantitative reasoning skills are broadly predictive of academic success, regardless of the student's chosen specialization. This finding aligns with prior research emphasizing the role of foundational mathematical competencies in university performance.

Similarly, Language (LA) appears in all three study tracks, though with varying frequencies. While its influence may not be as dominant as MR, its consistent presence reinforces the idea that verbal and reading comprehension skills are fundamental across disciplines. The variation in frequency, however, suggests that the strength of this association may depend on the specific academic demands of each track, possibly reflecting differences in curricular emphasis or assessment formats.

Interestingly, the plot also highlights specialized subject areas such as Biology (BI), Logic (LO), Physics (PH), Literature (LI), and Geography (GE), which appear frequently despite being exclusive to only one or two study tracks. The recurrence of these features suggests they be of particular importance within their respective academic domains. For example, the prominence of Biology in a track oriented toward the life sciences, or the appearance of Logic in a humanities-focused track, may signal that success in these disciplines is closely tied to performance in track-specific entrance exam components. This underscores the contextual nature of academic predictors: while some competencies are universally valuable (e.g., MR and LA), others serve as strong indicators only within more narrowly defined academic trajectories.

Taken together, the frequency-based analysis supports a nuanced understanding of how different components of the entrance exam relate to academic success. It reveals both generalizable patterns, such as the broad importance of mathematical, and language skills and track-specific associations that reflect the alignment between entrance requirements and academic demands. Such findings not only inform the design of future admission criteria but also highlight the importance of tailoring educational support to students' strengths and needs as entering diverse academic fields.

In summary, the results of the visual analysis suggest that certain subject areas, especially those related to verbal and linguistic reasoning, serve as more reliable predictors of first-semester success across programs. Meanwhile, the predictive power of quantitative and discipline-specific areas appears to be more context-dependent, motivating further investigation through supervised modeling.

The correlation analysis was broken down further by academic track to provide a more detailed picture of how the subjects on the entrance exams relate to academic performance. This allows us to identify which subject areas are most relevant within the disciplinary context of each program group.

Study Track 1- Health and biomedical sciences

In this track, the highest median correlations are observed for Biology (BI), Language (LA), and Verbal Reasoning (VR). These areas exhibit consistent, positive relationships with academic performance, suggesting that both scientific content knowledge and verbal competencies are essential for success in health-related programs. In contrast, areas such as Mathematical Reasoning (MR) and Physics (PH) exhibit lower and more variable correlations, reflecting their secondary importance in the initial curriculum.

Study Track 2- Exact sciences, engineering, and architecture

As expected, quantitative subject areas such as Mathematical Reasoning (MR), Arithmetic and Algebra (AA), and Physics (PH) show stronger median correlations in this track. Also, in the frequency, it can be seen that Logic (LO), a subject present only in this track, has a significant number of appearances with moderate to strong correlations, suggesting its importance. These findings align with the content demands of engineering and scientific programs, where numerical proficiency plays a central role. Notably, Verbal Reasoning (VR) also shows moderate and consistent correlations with Language (LA) regarding frequency, indicating that reading comprehension remains a relevant skill even in technical disciplines.

Study Track 3- Social sciences and humanities

In this track, Language (LA), Verbal Reasoning (VR), and History (HI) emerge as the most predictive subject areas, with moderate median correlations and relatively compact interquartile ranges. These results reflect the alignment between the entrance exam and the humanities-oriented curriculum. In contrast, Mathematical areas tend to show low or near-zero correlations, which is expected given the lower emphasis on formal quantitative training in early humanities coursework.

Study Track 4- Actuarial and business sciences

This track presents a mixed profile. Mathematical Reasoning (MR) shows the highest correlation, consistent with the analytical demands of actuarial and economic studies. However, Economy (EC) and Verbal Reasoning (VR) also contribute meaningfully, suggesting that both numerical and communicative competencies are valued in the business domain. The variability in correlations for Arithmetic and Algebra (AA) and Language (LA) may indicate differing levels of alignment across specific programs (e.g., Accounting vs. Commercial Engineering).

This disaggregated analysis confirms that predictive patterns are context-dependent: subject areas do not contribute equally across all disciplines. However, its predictive power varies across academic tracks, depending on their cognitive demands. This underscores the importance of maintaining track-specific admission criteria and tailoring analytical models to capture disci-

plinary nuances.

The findings also validate the methodological choice to analyze correlations at the section level, as it allows for fine-grained insight into the academic relevance of entrance exam components without the confounding influence of heterogeneous grading practices.

An important global observation from the boxplots is that all subject areas exhibit positive mean correlation values across the dataset, regardless of academic track or cohort. While the strength and dispersion of the correlations vary, the fact that no subject area shows a negative or null average association underscores a strong and consistent pattern: the entrance exam, as a whole, exerts a non-trivial positive influence on first-semester academic performance. This finding validates the exam's predictive value beyond anecdotal or program-specific effects and confirms that its subject areas capture relevant competencies that generalize across disciplines.

DISCUSSION

This study applied a data visual analytics approach to represent the relationship between students' performance and the entrance exam subject areas across three admission cohorts and four academic tracks at UNJBG. The in-depth approach of grading inequities among instructors allowed us to uncover nuanced patterns in the data, contributing to the debate over the validity and utility of university admission tests, as well as the need for more standardized grading practices to ensure fairness in the students' performance qualification.

The results confirm what was reported in the prior literature about the existence of moderate relationships between entrance score and academic performance [4], [5], [6], although the values were more variable given the approach of this study, giving some higher correlations like they can be seen in the Figure 3, the boxplots show an amplitude in the range of these values outside of the quartiles.

When disaggregated by study track, some patterns are even more evident, with subject areas designed specifically for the tracks showing higher correlations (e.g., Biology in Health and Biomedical Sciences, Logic in Exact Sciences, Engineering, and Architecture, Literature in Social Sciences and Humanities, and Geography in Actuarial and Business Sciences). It has been evident that general reading comprehension skills, along with the language domain (Verbal Reasoning and Language), consistently present high values, as shown in Figure 3 (boxplots) and Figure 5 (frequency of appearance).

Furthermore, when comparing the outcomes of the correlation analysis conducted with and without controlling for inter-instructor grading bias, a notable difference emerges. Specifically, the number of correlations equal to or greater than 0.4, a threshold indicating moderate to strong association per [16], is 42% lower when the bias is not accounted for. This substantial reduction highlights the extent to which aggregated, uncorrected analyses can obscure meaningful patterns between entrance exam subject areas and academic performance. The result reinforces the methodological importance of disaggregating by course section to control for grading variability, ensuring that observed associations more accurately reflect students' cognitive alignment with curricular demands rather than inconsistencies in instructor evaluation.

The findings of this study offer actionable insights for refining admission policy. Instead of abolishing standardized tests, as some critics propose, these results support a strategic reweighting of subject areas to better match each program's curricular demands, as also suggested by [5] and [8].

For example, Actuarial programs could emphasize Economics and Algebra, while Humanities could shift focus toward Language and History. These refinements, grounded in evidence, could better align admission expectations with academic realities, thereby enhancing equity and academic success.

Moreover, aligning with [17], this study supports the idea that identifying patterns early allows institutions to develop preemptive academic support systems, potentially lowering dropout rates and improving student retention.

CONCLUSION

This study assessed the efficacy of university entrance exams and identified academic performance patterns using visual data analysis techniques across three admission cohorts (2017–2019) and four academic tracks at UNJBG. Application of visual data analysis demonstrated that the entrance exam is highly predictive of first-semester academic success and that its subject-specific components closely align with the requirements of each academic field. These findings validate the entrance exam as a relevant diagnostic tool and suggest that admission policies could benefit from strategic reweighting of subject areas to improve alignment with program curricula. It is also notable, given the constant presence of test scores as a significant feature, that the ability to learn in general carries strong weight in students' future performance, likely related to more personal factors. Overall, this paper offers an evidence-based framework for improving entrance exam design and evaluating its coherence with academic success, with implications for equitable admissions and early academic support strategies. Future work could expand this approach by integrating additional data sources, such as socioeconomic or motivational variables, and by exploring causal inference techniques to further substantiate observed patterns.

ACKNOWLEDGEMENTS

This study was conducted with the support of the research project "Optimización del rendimiento académico universitario aplicando ciencia de datos al examen de admisión en la Universidad Nacional Jorge Basadre Grohmann", approved by Rectoral Resolution No. 13627-2024-UNJBG. The data used in this study were obtained within the framework of this project. The authors acknowledge Universidad Nacional Jorge Basadre Grohmann for its institutional support.

Funding

The authors declare that they received no funding for this work.

REFERENCIAS

[1] J. Norman, “A method for normalizing students' scores when employing multiple graders, ” *ACM SIGCSE Bulletin*, vol. 33, no. 4, pp. 35–38, 2001, https://doi.org/10.1145/572139.572166 [↑]

[2] I.N. Chaparro-Cruz, L.N. Huertas-Condori, S. B. Cabana-Yupanqui, and A. Chaparro-Guerra, “Relationship between entrance exam scores, academic performance, and student dropout rates: A longitudinal case study” *International Journal of Learning, Teaching and Educational Research*, vol. 24, no. 3, pp. 216–243, 2025, https://doi.org/10.26803/ijlter.24.3.11 [↑]

[3] O. Llaha and A. Aliu, “Students' performance evaluation in higher education using data visualization techniques, ” *Annals of Philosophy, Social & Human Disciplines*, vol. 2, 2022. [↑]

[4] A. Rahila, S. K. Ali, and A. Afzal, “Predictive validity of a uniform entrance test for the health professionals, ” *Pakistan Journal of Medical Sciences*, vol. 35, no. 2, pp. 330–334, 2019, https://doi.org/10.12669/pjms.35.2.334 [↑]

[5] J. Poláčková and S. Libuše, “Relationship of success in university study and admission exam results, ” *Journal on Efficiency and Responsibility in Education and Science*, vol. 6, no. 4, pp. 281–293, 2013, https://doi.org/10.7160/eriesj.2013.060406 [↑]

[6] M. G. Ilave, R. M. B. Tomé, K. L. M. Calderón, J. S. Camayo y E. O. Á. Díaz, “Modelo DECO en procesos de admisión y rendimiento académico en estudiantes al primer año”, *Universidad y Sociedad*, vol. 14, no. S2, pp. 65–75, 2022. [En línea]. Disponible: https://www.scopus.com/pages/publications/85131545333?inward [↑]

[7] S. Iftikhar, K. M. Cheema, and I. Aziz, “MBBS admission criteria as predictor of academic performance in a medical college of Pakistan, ” *Pakistan Journal of Medical Sciences*, vol. 39, no. 6, pp. 1706–1711, 2023, https://doi.org/10.12669/pjms.39.6.6839 [↑]

[8] M. Ferrão and L. Almeida, “Differential effect of university entrance score on first-year students’ academic performance in Portugal, ” *Assessment & Evaluation in Higher Education*, vol. 44, no. 4, pp. 610–622, 2019, https://doi.org/10.1080/02602938.2018.1525602 [↑]

[9] M. Bienkowski, M. Feng, and B. Means, “*Enhancing teaching and learning through educational data mining and learning analytics: An Issue Brief*, ” Office of Educational Technology, U.S. Department of Education, 2012. [Online]. Available: https://eric.ed.gov/?id=ED611199 [↑]

[10] U.G. Inyang, I. J. Eyoh, S. A. Robinson, and E. N. Udo, “Visual association analytics approach to predictive modelling of students’ academic performance, ” *International Journal of Modern Education & Computer Science*, vol. 11, no. 12, pp. 1–13, 2019, https://doi.org/10.5815/ijmecs.2019.12.01 [↑]

[11] Y. Zhao, Q. Hu, M. Chen, and G. M. Weiss, “Predicting student performance in a master of data science program using admissions data, ” in *The 13th International Conference on Educational Data Mining (EDM 2020)*, 2020. [Online]. Available: https://eric.ed.gov/?id=ED607995 [↑]

[12] E. Alhazmi and A. Sheneamer, “Early predicting of students performance in higher education, ” *IEEE Access*, vol. 11, pp. 27579–27589, 2023, https://doi.org/10.1109/ACCESS.2023.3250702 [↑]

[13] M. Suaza-Medina, R. Peñabaena-Niebles, and M. Jubiz-Diaz, “A model for predicting academic performance on standardised tests for lagging regions based on machine learning and Shapley additive expla-

nations, ” *Scientific Reports*, vol. 14, no. 1, Art. no. 25306, 2024, https://doi.org/10.1038/s41598-024-76596-3 [↑]

[14] J. Schinske and K. Tanner, “Teaching more by grading less (or differently), ” *CBE—Life Sciences Education*, vol. 13, no. 2, pp. 159–166, 2014, https://doi.org/10.1187/cbe.cbe-14-03-0054 [↑]

[15] S. Kates, T. Paulsen, S. Yntiso, and J. A. Tucker, “Bridging the grade gap: Reducing assessment bias in a multi-grader class, ” *Political Analysis*, vol. 31, no. 4, pp. 642–650, 2023, https://doi.org/10.1017/pan.2022.27 [↑]

[16] P. Schober, C. Boer, and L. A. Schwarte, “Correlation coefficients: appropriate use and interpretation, ” *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018, https://doi.org/10.1213/ANE.0000000000002864 [↑]

[17] L. Pan et al., “Gathering intelligence on student information behavior using data mining, ” *Library Trends*, vol. 68, no. 4, pp. 636–658, 2020, https://doi.org/10.1353/lib.2020.0015 [↑]