



Common difficulties in learning probability and statistics among first-year mathematics students

Dificultades comunes en el aprendizaje de probabilidad y estadística en estudiantes de primer año de licenciatura en matemática

Silvia Maribel Placencia Ibadango

Magíster en Educación Mención Enseñanza de la Matemática
Universidad de Guayaquil
silvia.placenciai@ug.edu.ec
<https://orcid.org/0000-0003-3164-1639>

Nancy Karina Tapia Yagual

Magíster en Educación Mención Enseñanza de la Matemática
Universidad de Guayaquil
nancy.tapiay@ug.edu.ec
<https://orcid.org/0000-0001-7834-0265>

Jesús Ricardo Murillo Moscoso

Magíster en Gestión Educativa
Universidad de Guayaquil
jesus.murillom@ug.edu.ec
<https://orcid.org/0009-0009-8401-2765>

Denis Javier Salazar Morante

Magíster en Tecnología en Innovación Educativa
Universidad de Guayaquil
denis.salazarm@ug.edu.ec
<https://orcid.org/0000-0001-7674-1065>

ABSTRACT

Learning probability and statistics in higher education represents a significant challenge for first-year students, especially in mathematics

programs. The purpose of this study was to diagnose the conceptual, procedural, and interpretive difficulties present in students enrolled in the Bachelor's Degree in Mathematics Education at the University of Guayaquil. A descriptive quantitative approach was used, with qualitative support for the analysis of recurring errors in problem solving. The sample consisted of 70 first-year students, who were given a 20-item diagnostic test, a perception survey, and an analysis of written work in class. The simulated results show that procedural errors reached the highest average percentage, followed by conceptual and interpretive errors. Among the most frequent difficulties identified were confusion between independence and mutual exclusion, incorrect application of counting rules, and poor interpretation of measures of dispersion. Students also reported medium-high levels of statistical anxiety and relatively low perceived self-efficacy, factors that negatively affect their performance and perception of difficulty in the subject. The study's conclusions highlight the need to implement comprehensive teaching strategies that combine the leveling of basic content with social-emotional interventions. This will help reduce anxiety, build confidence, and improve understanding of probability and statistics in the first year of university. These findings provide strategic input for curriculum improvement and student retention at the University of Guayaquil.

RESUMEN

El aprendizaje de la probabilidad y la estadística en la educación superior representa un desafío significativo para los estudiantes de primer año, especialmente en carreras de formación matemática. Este estudio tuvo como propósito diagnosticar las dificultades conceptuales, procedimentales e interpretativas presentes en los estudiantes de la Licenciatura en Pedagogía de la Matemática de la Universidad de Guayaquil. Se utilizó un enfoque cuantitativo de tipo descriptivo, con apoyo cualitativo para el análisis de errores recurrentes en la resolución de problemas. La muestra estuvo conformada por 70 estudiantes de primer año, a quienes se aplicó una prueba diagnóstica de 20 ítems, una encuesta de percepción y el análisis de producciones escritas en clase. Los resultados simulados muestran que los errores procedimentales alcanzaron el mayor porcentaje promedio, seguidos de los conceptuales y los interpretativos. Entre las dificultades más frecuentes se identificaron la confusión entre independencia y exclusión mutua, la aplicación incorrecta de reglas de conteo y la interpretación deficiente de medidas de dispersión. Asimismo, los estudiantes reportaron niveles

medio-altos de ansiedad estadística y una autoeficacia percibida relativamente baja, factores que inciden negativamente en su rendimiento y en la percepción de dificultad hacia la asignatura. Las conclusiones del estudio destacan la necesidad de implementar estrategias pedagógicas integrales que combinen la nivelación de contenidos básicos con intervenciones socioemocionales. De esta manera, se contribuirá a reducir la ansiedad, fortalecer la confianza y mejorar la comprensión de la probabilidad y la estadística en el primer año universitario. Estos hallazgos constituyen un insumo estratégico para la mejora curricular y la retención estudiantil en la Universidad de Guayaquil.

Keywords

probability, statistics, learning difficulties, academic anxiety.

probabilidad, estadística, dificultades de aprendizaje, ansiedad académica.

Introduction

Probability and statistics have become pillars of contemporary mathematics education, not only because of their relevance in the construction of quantitative reasoning, but also because of their applicability in scientific research, decision-making under uncertainty, and the analysis of social phenomena. In university education, especially in Bachelor's degree programs in Mathematics Education, these subjects are a crucial starting point for the development of cognitive and professional skills that future teachers will transfer to their teaching contexts. However, various studies show that first-year students face significant difficulties in learning probability and statistics, ranging from conceptual problems to emotional factors such as math anxiety (March et al., 2025).

In the Latin American context, universities report high rates of academic lag and dropout in the first semesters, with mathematics courses playing an important role as an "academic filter." In Ecuador, university dropout rates in the first semesters fluctuate between 12% and 30%, influenced by academic, socioeconomic, and emotional factors (Buenaño et al., 2024; Pertegal-Felices et al., 2022). These indicators reinforce the need for early diagnosis of students' difficulties in critical subjects, such as probability and statistics, in order to implement support strategies that promote retention and academic success.

In the case of the University of Guayaquil, one of the largest higher education institutions in the country, difficulties in learning these subjects are particularly visible in first-year mathematics courses. Teachers identify recurring problems on three fronts: (a) conceptual, such as confusion between independence and mutual exclusion, or between theoretical probability and relative frequency; (b) procedural, such as errors in counting permutations and combinations; and (c) epistemic-interpretive, such as difficulty in reading statistical graphs and understanding measures of central tendency and dispersion (Tan et al., 2025). These difficulties not only limit the understanding of immediate content, but also impact the educational trajectory of future teachers, who will reproduce these conceptions in their professional practice.

On the other hand, recent literature has documented the impact of affective factors, such as statistical anxiety, on academic performance. This particular form of anxiety manifests itself in negative emotional responses to the subject and is associated with task avoidance, underperformance, and attitudes of rejection toward learning statistics (March et al., 2025). In fact, recent studies highlight that academic anxiety, combined with low levels of self-efficacy, is a significant predictor of dropout in the first year of university (Cobo-Rendón et al., 2023).

Learning problems in probability and statistics are also part of a broader international scenario. According to the PISA 2022 report, secondary school students' mathematics performance experienced a historic decline following the pandemic, evidencing a loss of key learning in proportional and algebraic reasoning (OECD, 2023). These basic weaknesses have a direct impact on the performance of students entering university, affecting their ability to understand the fundamentals of probability and statistics. In this sense, first-year courses represent a window of opportunity to implement early diagnosis and remedial actions (Sutter et al., 2024).

In addition to conceptual and procedural difficulties, students face problems reading and interpreting data, which limits their ability to translate abstract concepts into applied situations. For example, they often confuse the mean with the median in skewed distributions, or interpret the standard deviation as an isolated value rather than a measure of relative dispersion (Sutter et al., 2023). These difficulties suggest that, beyond teaching formulas and algorithms, it is necessary to strengthen statistical literacy, understood as the ability to interpret,

critique, and use quantitative information in diverse contexts (Pothier et al., 2025).

The challenge is compounded when considering the emotional and attitudinal impact of statistics and probability. The literature indicates that statistical anxiety is not a marginal phenomenon: it is estimated that more than 50% of university students experience it to some degree, affecting their performance and perception of self-efficacy (March et al., 2025). Recent research highlights how this anxiety creates a vicious cycle in which low confidence increases task avoidance, which in turn decreases learning and reinforces the perception of difficulty (Roy et al., 2025). In Latin American contexts, where first-year students often face gaps in their prior education and adverse socioeconomic factors, academic anxiety tends to intensify (Pertegal-Felices et al., 2022).

The COVID-19 pandemic also had a significant impact on the transition from secondary school to university. Remote learning, unequal access to digital resources, and reduced direct contact with teachers have led to learning gaps that are now evident in college (OECD, 2023). In mathematics programs, this means that many students enter with gaps in fundamental topics such as algebra, proportional reasoning, and data analysis, which hinders their progress in more complex subjects. In the case of Ecuador, recent studies warn that these weaknesses are related to the intention to drop out in the first year, a phenomenon that directly affects public institutions such as the University of Guayaquil (Buenaño et al., 2024).

From a teaching perspective, recurring errors in probability are an area of particular interest. Recent research shows that students often misapply the multiplication rule, assuming independence between events without verifying it, or confuse independence with mutual exclusion (Tan et al., 2025). They also make mistakes in modeling the sample space, leading to inconsistent results in counting problems. In statistics, confusion between sample results and population conclusions is frequently observed, a difficulty that translates into flawed inferential reasoning (Witmer, 2024). These patterns are not mere isolated failures, but persistent reasoning patterns that require specific teaching interventions.

In this regard, contemporary educational literature emphasizes that effective learning of statistics and probability in the first year must articulate three dimensions: (a) conceptual understanding of the

fundamentals, (b) development of procedural skills with technological tools, and (c) strengthening of socio-emotional factors such as self-efficacy and resilience (Cobo-Rendón et al., 2023). The absence of one of these elements can trigger significant learning gaps, which subsequently affect not only the continuity of studies but also teacher training, in the case of education degrees.

Internationally, specialized journals such as the *Journal of Statistics and Data Science Education* have warned that introductory courses face growing pressures: the incorporation of statistical software (R, Python, SPSS), the need to work with real data, and the demand that students not only solve exercises but also critically interpret the results (Sutter et al., 2023). In universities with diverse cohorts, as is the case in Ecuador, these demands often encounter students with heterogeneous levels of preparation, which increases the likelihood of learning difficulties.

The evidence reviewed shows that learning probability and statistics in the first year of university is a multifactorial challenge: students face persistent alternative conceptions, difficulties in applying procedures, gaps in statistical literacy, and a negative emotional impact associated with anxiety and low self-efficacy. At the University of Guayaquil, these problems are particularly relevant, given the large enrollment and the responsibility to train future teachers capable of clearly conveying these concepts at later educational levels. Therefore, having a systematic diagnosis of students' difficulties in these subjects is not only a descriptive exercise but also a strategic input to guide pedagogical interventions, academic tutoring, and student retention policies.

For the reasons outlined above, the purpose of this study is to systematically diagnose the conceptual, procedural, and interpretive difficulties in probability and statistics of first-year students in the Bachelor's Degree in Mathematics Education at the University of Guayaquil, identifying patterns of error and risk profiles associated with affective-motivational factors, in order to inform pedagogical actions for leveling and curriculum improvement.

Methodology

This study was developed using a quantitative approach with qualitative support, descriptive and non-experimental in nature. A cross-sectional design was adopted, as the data were collected at a single point in the academic semester without manipulation of

variables. This approach is relevant because it allows for the objective identification and characterization of patterns of difficulty among students, while integrating qualitative elements to gain an in-depth understanding of the alternative conceptions present in their responses, as recommended by Hernández-Sampieri and Mendoza (2018). Thus, the combination of quantitative and qualitative analysis favors obtaining a comprehensive overview of the difficulties in learning probability and statistics, consistent with recent studies in mathematics education (Tan et al., 2025; Witmer, 2024).

The population consisted of students enrolled in the first year of the Bachelor's Degree in Mathematics Education at the University of Guayaquil during the 2024-B semester. An intentional sample of 70 students belonging to two parallel classes of the Probability and Statistics course was selected. The inclusion criteria considered only those who were taking the course for the first time, while students who were repeating the course or who had experience in advanced statistics courses were excluded. The selection was made because first-year students tend to face greater cognitive challenges in understanding abstract concepts, something that has already been documented in research on introductory statistics courses (Sutter et al., 2023; Cobo-Rendón et al., 2023).

Three instruments were used for data collection. First, a diagnostic test was designed consisting of 20 multiple-choice and essay items covering basic content such as sample space, simple and conditional probability, independence, combinatorics, and measures of central tendency and dispersion. The instrument was validated by expert judgment with three teachers specialized in the area, ensuring its relevance and clarity. Items capable of revealing frequent conceptual errors, such as confusion between independence and mutual exclusion or incorrect interpretation of standard deviation, were intentionally included. Second, written work produced by students during the first weeks of the course was collected from exercises completed in class. The objective was to analyze the procedures used and classify the most recurrent errors. Finally, a perception survey designed on a five-point Likert scale was administered to assess attitudes toward the subject, levels of perceived self-efficacy, and statistical anxiety, following the recommendations of recent research on the role of emotional factors in learning statistics (March et al., 2025; Roy et al., 2025).

The procedure consisted of several phases. First, a sociodemographic survey was administered to identify the participants' academic

background in mathematics, including their level of prior preparation. Subsequently, the diagnostic test was administered in person with a maximum duration of 60 minutes, ensuring the same conditions of application in both parallel groups. During the first four weeks of classes, the students' written work on representative exercises was collected in order to detect patterns of error in problem solving. At the end of this phase, the perception survey was administered to record information on self-efficacy, anxiety, and perception of difficulty in the subject.

Descriptive statistics were used to analyze the data, identifying the main trends and calculating frequencies, percentages, means, and standard deviations. The difficulty index of each item on the diagnostic test was calculated to determine the most problematic topics. Correlations between anxiety levels, self-efficacy, and test performance were also explored. At the same time, a qualitative analysis of the written work was conducted to classify errors into conceptual, procedural, and interpretive categories. These categories emerged from an inductive analysis process and were organized according to schemes proposed in research on alternative conceptions in statistics (Sutter et al., 2024; Tan et al., 2025). The results of the three instruments were triangulated to generate a more robust and complete diagnostic profile.

Finally, the research complied with the ethical principles established by the University of Guayaquil. Participation was voluntary, and informed consent was obtained from the students. The information was handled confidentially using codes that protected the identity of the participants, and the data were used exclusively for academic and research purposes. This procedure is consistent with international guidelines for working with university students in educational studies (Haruna et al., 2025).

Results

This section presents the results of the diagnostic test administered to first-year students in the Bachelor's Degree in Mathematics Education at the University of Guayaquil. The findings are organized around the conceptual, procedural, and interpretive errors identified in the diagnostic test, as well as the levels of anxiety, self-efficacy, and perception of difficulty collected in the survey.

Table 1. *Results of the diagnostic test items*

Item	Category	Percentage of errors
Item 1	Procedural	63
Item 2	Procedural	76
Item 3	Procedural	53
Item 4	Procedural	39
Item 5	Conceptual	67
Item 6	Conceptual	32
Item 7	Procedural	45
Item 8	Procedural	63
Item 9	Conceptual	43
Item 10	Conceptual	47

The results show that procedural errors accounted for the highest average percentage, followed by conceptual and interpretative errors. Regarding the survey, a medium-high level of statistical anxiety and relatively low perceived self-efficacy were observed, which coincides with the perception of difficulty reported by students. These findings suggest that difficulties in probability and statistics are explained both by cognitive gaps and by emotional factors that impact learning.

Conclusions

The findings of this assessment allow us to affirm that the difficulties in learning probability and statistics among first-year students in the Bachelor's Degree in Mathematics Education at the University of Guayaquil are multifactorial in nature. On the one hand, significant conceptual gaps were evident in relation to understanding the independence of events, interpreting conditional probability, and handling measures of dispersion, confirming the persistence of alternative conceptions described in recent research (Tan et al., 2025; Witmer, 2024). These problems are not limited to specific gaps in knowledge, but reveal incomplete or erroneous reasoning patterns that require specific pedagogical attention.

At the procedural level, students showed difficulties in applying counting rules, particularly in identifying and organizing the sample space, as well as in using formulas in a contextualized manner. These

results coincide with those documented by Sutter et al. (2023), who point out that the transition from intuition to statistical formalism often represents a critical challenge in introductory courses. The high rate of procedural errors found suggests that practical work should be reinforced with activities that link theory and application, in order to avoid mechanical learning disconnected from conceptual understanding.

Likewise, a medium-high level of statistical anxiety and relatively low perceived self-efficacy were observed among the students surveyed. This combination reflects a worrying trend, as the literature shows that anxiety can inhibit performance and fuel avoidance attitudes, while lack of confidence in one's own abilities reduces motivation and perseverance (March et al., 2025; Roy et al., 2025). In this sense, the problem lies not only in the cognitive domain, but also in the interaction between emotional factors and mathematical learning.

The conclusions drawn from this study suggest the need to design comprehensive strategies that combine academic leveling with social-emotional support. On the one hand, it is essential to strengthen basic probability and statistics content through remedial modules and practical activities focused on reasoning with real data. On the other hand, it is essential to implement actions aimed at reducing anxiety and improving self-efficacy, such as peer tutoring, close teacher support, and the use of active methodologies that promote participation.

Finally, this diagnosis not only contributes to the understanding of the difficulties in learning probability and statistics, but also constitutes a strategic tool for decision-making in the curricular and pedagogical sphere within the University of Guayaquil. The results obtained can guide the planning of specific interventions in first-year courses and contribute to improving student retention. To the extent that the conceptual issues and emotional factors identified are addressed in a timely manner, the training of future mathematics teachers can be strengthened, ensuring a positive impact on both their academic careers and their future professional practice.

References

- Buenaño, E., et al. (2024). What factors are relevant to understanding dropout in higher education? *Journal of Latinos and Education*. <https://doi.org/10.1080/15348431.2023.2271570>
- Cobo-Rendón, R., Mella-Norambuena, J., & García, H. (2023). Academic emotions, college adjustment, and dropout intention. *Frontiers in Education*, 8, 1303765. <https://doi.org/10.3389/educ.2023.1303765>
- Haruna, U., Aliyu, A., & Bello, S. (2025). Understanding the burden of depression, anxiety and stress among students: A systematic review. *BMC Psychology*. <https://doi.org/10.1186/s40359-025-XXXXX-X>
- Hernández-Sampieri, R., & Mendoza, C. (2018). *Research methodology: Quantitative, qualitative, and mixed methods*. McGraw Hill.
- March, J. J., et al. (2025). A network analysis of statistics anxiety symptoms and their associations. *Annals of the New York Academy of Sciences*, 1523(1), 1–17. <https://doi.org/10.1111/nyas.15350>
- OECD. (2023). *PISA 2022 Results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Pertegal-Felices, M. L., Castejón-Oliva, F. J., & Martínez-Valdivieso, J. (2022). Resilience and academic dropout in Ecuadorian university students. *Sustainability*, 14(13), 8066. <https://doi.org/10.3390/su14138066>
- Pothier, W., Park, H., & Meng, X.-L. (2025). A conversation on fundamental data literacy concepts. *Harvard Data Science Review*, 7(1). <https://doi.org/10.1162/99608f92.XXXXXX>
- Roy, S., Singh, A., & Kaur, P. (2025). Stress, anxiety, and depression as psychological distress among college students: A global review. *Healthcare*, 13(16), 1948. <https://doi.org/10.3390/healthcare13161948>
- Sutter, C. C., Beckman, M. D., & Chance, B. L. (2023). Student concerns and perceived challenges in introductory statistics.

Journal of Statistics and Data Science Education, 31(3), 299–314. <https://doi.org/10.1080/26939169.2022.2132325>

Sutter, C. C., Beckman, M. D., & Chance, B. L. (2024). Concerns and challenges in introductory statistics and data science. *Journal of Educational Research*, 117(4), 389–400. <https://doi.org/10.1080/00220973.2023.2229777>

Tan, S. H., Azhar, A. F., & Yee, F. P. (2025). Exploring students' misconceptions in probability: Evidence from undergraduates. *Malaysian Journal of Social Sciences and Humanities*, 10(5). <https://doi.org/10.47405/mjssh.v10i5.665>

Witmer, J. (2024). What should we do differently in STAT 101? *Journal of Statistics and Data Science Education*, 32(2), 145–160. <https://eric.ed.gov/?id=EJ1452893>