Open Access Library Journal
2026, Volume 13, e15130
ISSN Online: 2333-9721
ISSN Print: 2333-9705

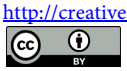
Using Unsupervised Learning to Identify Student Performance Profiles in English Language Education: A Clustering Analysis from the Nigeria Maritime University

Ebitiminipre Mercy Ogbise[1], Akpofure Avwerosuoghene Enughwure[2]

[1]School of General Studies, Nigeria Maritime University, Okerenkoko, Nigeria
[2]Department of Electrical Engineering, Nigeria Maritime University, Okerenkoko, Nigeria
Email: akpofureenughwure@gmail.com

How to cite this paper: Ogbise, E.M. and Enughwure, A.A. (2026) Using Unsupervised Learning to Identify Student Performance Profiles in English Language Education: A Clustering Analysis from the Nigeria Maritime University. *Open Access Library Journal*, **13**: e15130.
https://doi.org/10.4236/oalib.1115130

Received: March 9, 2026
Accepted: April 7, 2026
Published: April 10, 2026



Abstract

The ability to communicate effectively in English is fundamental to academic success in Nigerian universities, yet persistent low achievement in the compulsory "Use of English" course remains a concern, particularly in specialized institutions like maritime universities where proficiency underpins both academic performance and professional competence. This study addresses the limitation of reactive academic support systems by employing unsupervised machine learning to analyze English language learning characteristics among undergraduates. Using K-means clustering on survey data from 248 students at the Nigeria maritime university, the research identified four distinct student performance clusters validated through chi-square analysis ($p = 0.000949$). The clusters revealed multidimensional profiles incorporating academic, psychological, and environmental factors: Cluster 1 (Diligent High-Achievers) demonstrated strong study habits and confidence; Cluster 2 (Steady Performers) represented average students with consistent patterns; Cluster 0 (Quiet Achievers) achieved comparable results through different behavioral pathways; and critically, Cluster 3 (At-Risk Group) exhibited low study hours, diminished confidence, elevated anxiety, and minimal AI tool adoption, with a 36.4% failure rate. The findings demonstrate that clustering techniques enable early identification of learning difficulties before traditional assessment methods detect poor performance. Recommendations include developing differentiated institutional support systems targeting specific cluster needs, from foundational programs for at-risk students to enrichment opportunities for high achievers, advancing data-driven approaches to language education in Nigerian higher education.

Subject Areas

Language Education

Keywords

Educational Data Mining, K-Means Clustering, English Language Proficiency, Academic Performance, Nigeria Maritime University, Early Intervention

1. Introduction

The ability to communicate effectively in English is fundamental to academic participation and success in Nigerian universities, where English serves as the official language of instruction, assessment, and scholarly communication [1]. The compulsory "Use of English" course is therefore designed to equip undergraduates with essential skills in grammar, comprehension, and academic writing. However, despite its foundational role, consistent reports of low achievement and high failure rates persist, especially among first-year students. These challenges are more pronounced in specialized institutions such as maritime universities, where English proficiency underpins not only academic performance but also professional competence and adherence to international standards [2].

A major limitation of existing academic support systems is their reliance on conventional evaluation methods that identify struggling students only after poor performance has occurred [3]. Such reactive approaches limit the effectiveness of remediation and often allow learning difficulties to persist. Recent advances in educational data mining offer alternative, proactive strategies for analyzing student performance data and identifying hidden patterns associated with learning challenges [4]. In particular, unsupervised learning techniques enable exploratory analysis without the need for predefined outcome labels, making them suitable for early-stage academic monitoring [5].

Despite the growing application of data mining techniques in education, limited attention has been given to language-based courses and institution-specific contexts in Nigeria. This study addresses this limitation by proposing a clustering-based approach to analyze "Use of English" examination data from the Nigeria maritime university. By identifying distinct performance groupings, the study highlights the significance of unsupervised learning as a tool for early detection of learning difficulties and informed academic intervention [6].

2. Literature Review

Research on academic performance has long emphasized the importance of early identification of learning difficulties in higher education.

Lestari employed Mixed-Method Action Research design that offers a distinct contribution by repositioning clustering from an analytical tool to an instructional

intervention [7]. It demonstrates that clustering techniques can enhance English writing performance by helping students visually organize their thoughts before writing, thereby improving both engagement and learning outcomes.

Nafuri and his team proposed an unsupervised clustering framework to classify the academic performance of B40 (low-income) undergraduate students across 20 Malaysian public higher education institutions [8]. Using a dataset of 117,069 students with 16 attributes, the research evaluated three clustering algorithms: k-means, BIRCH, and DBSCAN. This study demonstrates the scalability of clustering approaches for identifying vulnerable student populations and validates k-means as an effective algorithm for educational data mining in under-resourced student contexts.

A study conducted by Oguike and his team demonstrates the practical application of machine learning techniques to analyze undergraduate examination result repositories in a Nigerian university context [9]. Using student data from the Department of Computer Science at the University of Nigeria, Nsukka, and the research trained unsupervised learning model in Jupyter Notebook to extract actionable insights from academic records. However, the work didn't provide the implications on their result and the student performance in future intervention.

Rahma and Ulfah employed K-means clustering to analyze academic performance patterns among students by integrating both academic and social-demographic variables [10]. The dataset included mathematics, reading, and writing scores alongside demographic attributes such as gender, ethnicity, parental education level, and lunch type. A limitation of their study is the reliance on only basic academic and demographic variables. Incorporating additional factors—such as emotional well-being and learning preferences—would likely yield more comprehensive educational models.

Cao carried a study that applied clustering techniques to English language education by analyzing skill-specific performance data across listening, reading, writing, translating, and speaking [11]. The four-cluster solution provides a framework for understanding diverse learner profiles and supporting differentiated instruction. However, the exclusion of socio-demographic and emotional variables is a limitation, as prior research confirms these factors influence student performance.

Huang contributes a domain-specific improvement to K-means clustering by incorporating multiple behavioral factors (borrowing behavior and course learning behavior) relevant to English education. The resulting three-tier stratification model provides a structured framework for implementing differentiated instruction in language teaching contexts [12].

3. Methodology

3.1. Research Design

This study employed a quantitative, exploratory research design utilizing unsupervised machine learning techniques (K-means clustering) to identify natural groupings among university students based on their English language learning

characteristics and performance indicators.

3.2. Data Collection and Preparation

3.2.1. Sampling and Data Collection Procedure

Data were collected via Google Form from students enrolled in the "Use of English" course at a Nigeria maritime university during the 2024-2025 academic session. The survey was administered over a four-week period (February-March 2025). All enrolled students (N ≈ 1000) were invited to participate through their departmental WhatsApp groups. A total of 248 completed responses were received, yielding a response rate of approximately 24.8%.

a) Eligibility criteria

Students were eligible to participate if they enrolled in the "Use of English" course and had completed the examination at the time of survey administration. The sample includes students from Tarma 1 through Tarma 5 (all undergraduate years), representing multiple cohorts within the institution.

b) Ethical consideration

Participation was voluntary and anonymous. The Google Form included an informed consent statement at the beginning, explaining the purpose of the research, the voluntary nature of participation, and the confidentiality of responses. No personally identifying information (names, matriculation numbers) was collected.

3.2.2. Dataset Description

The complete set of variables collected in the survey is presented in Table 1, including all features subsequently used in the clustering analysis.

Table 1. Dataset column profile.

Column	Data Type	Unique Count	Sample Values (first 5)
Age	Int64	14	26, 17, 21, 20, 18
Gender	Object	2	Male, Female
State of Origin	Object	26	Rivers, Osun, Delta, Cross River, Bayelsa
City of Residence	Object	70	Port Harcourt, Osogbo, Warri, Ughelli, Yenagoa
Department	Object	15	Civil Engineering, Marine Geology, Marine Engineering, Electrical Engineering, Mechanical Engineering
Faculty	Object	3	Engineering, Environmental Management, and Transport
Secondary School Mode		2	Government, Private
Year of Study	Object	5	Tarma 3, Tarma 1, Tarma 2, Tarma 5, Tarma 4
English Exam Grade	Object	3	A, B, C and F
Confidence Level	Object	5	High, Average, Low, Very High, Very Low
AI tools usage	Object	2	Yes, No
Study Hours	Object	3	0 hour, 1 - 3 hours, and 4+ hours
Exam Anxiety	Object	5	Low, Average, Very Low, Very High, High

Continued

Class Attendance	Object	4	Rarely, Sometimes, Often, Always
Assignment Participation	Object	5	Never, Rarely, Sometimes, Often, Always
Motivation Level	Object	5	Very Low, Low, Average, High, Very High
Private Tutoring Experience	Object	2	Yes, No

3.2.3. Data Cleaning

The data cleaning process was carried out to ensure that irrelevant data columns like timestamp were eliminated to ensure anonymity. Categorical variables such as state of origin, department and city of residence were standardized in a bid to remove spelling errors at the data entry stage. Also, the study resolved the inconsistent responses by the candidates in certain columns by merging values such as "Good, Excellent" into "Good".

3.3. Variable Transformation

All variables listed in Table 1 were included in the clustering analysis. Categorical variables without inherent order were transformed using LabelEncoder:

a) Gender (2 categories)
b) State of Origin (26 categories)
c) City of Residence (70 categories)
d) Department (15 categories)
e) Faculty (3 categories)
f) Secondary School Mode (2 categories)
g) Year of Study (5 categories)
h) Private Tutoring Experience (2 categories)

Variables with inherent ordinal relationships were mapped to numerical values as shown in Table 2.

Table 2. Ordinal variable lookup table.

Variable	Ordinal Mapping
English Exam Grade	F = 1, C = 2, B = 3, A = 4
Class Attendance	Rarely = 1, Sometimes = 2, Often = 3, Always = 4
Confidence Level	Very Low = 1, Low = 2, Average = 3, High = 4, Very High = 5
Study Hours	0 hour = 1, 1 - 3 hours = 2, 4+ hours = 3
AI Tool Usage	No = 1, Yes = 2
Assignment Participation	Never = 1, Rarely = 2, Sometimes = 3, Often = 4, Always = 5
Motivation Level	Very Low = 1 to Very High = 5
Exam Anxiety	Very Low = 1 to Very High = 5

3.4. Data Standardization

All features were standardized using StandardScaler to ensure equal contribution

to the clustering algorithm, preventing variables with larger scales from dominating the distance calculations.

3.5. Definition of Performance Outcome

The target outcome variable "Pass/Fail" was derived from the self-reported English Exam Grade as follows:

Pass: Grades A, B or C.

Fail: Grade F.

3.6. Clustering Procedure

K-means clustering was implemented in this study. This technique was employed because it is a popular, simple, and efficient unsupervised machine learning method used to partition a dataset into a user-specified number, k, of distinct groups or clusters. It works by grouping data points such that those in the same cluster are more similar to each other than to those in other clusters, based on a distance measure, typically the Euclidean distance.

The optimal number of clusters used in the study is determined by the use of two complementary methods:

1) Elbow Method: Plotting inertia (within-cluster sum of squares) against cluster numbers 2 - 10 to identify the "elbow point".

2) Sihouette Analysis: Computing silhouette scores to measure cluster cohesion and separation.

3.7. Validation Technique

Cluster validity was assessed through:

1) Cross-tabulation between cluster assignments and actual performance outcomes (Pass/Fail).

2) Chi-square test of independence to determine statistical significance of the relationship.

4. Results and Discussion

4.1. Exploratory Data Analysis

At the end of the data collection phase of the study, the total number of entries was two hundred and forty-eight (248). The age distribution of the students is presented in **Figure 1** with an average age of the respondents is 21.2 years. The sample exhibits strong gender skew (male: 85.9%, female: 14.1%), which may reflect the underlying population characteristics of the study context as presented in **Figure 2**. Private school students (n = 137) slightly outnumber government school students (n = 111) in the dataset, reflecting the significant role of private secondary education in the population studied as shown in **Figure 3**. Students predominantly originate from South-South and South-East states, with Delta (44), Edo (24), Ondo (19), Imo (17), and Bayelsa (16) representing the top five as shown in **Figure 4**. The students in the top 5 states account for 48% of the total respond-

ents in the study. The distribution of students across departments shows Marine Engineering as the most represented (85 students), followed by Electrical Engineering (42), Civil Engineering (41), Mechanical Engineering (32), and Petroleum and Gas Engineering (22). This pattern aligns with broader enrollment trends in engineering education as shown in Figure 5. The students in the top 5 departments account for 90% of the total respondents in the study. The student distribution across faculties is heavily skewed toward Engineering (224 students, 90.3%), with Transport (13, 5.2%) and Environmental Management (11, 4.4%) representing much smaller proportions as displayed in Figure 6. This concentration reflects the institutional emphasis on engineering disciplines, consistent with enrollment patterns observed in technical universities where engineering programs typically dominate student populations.

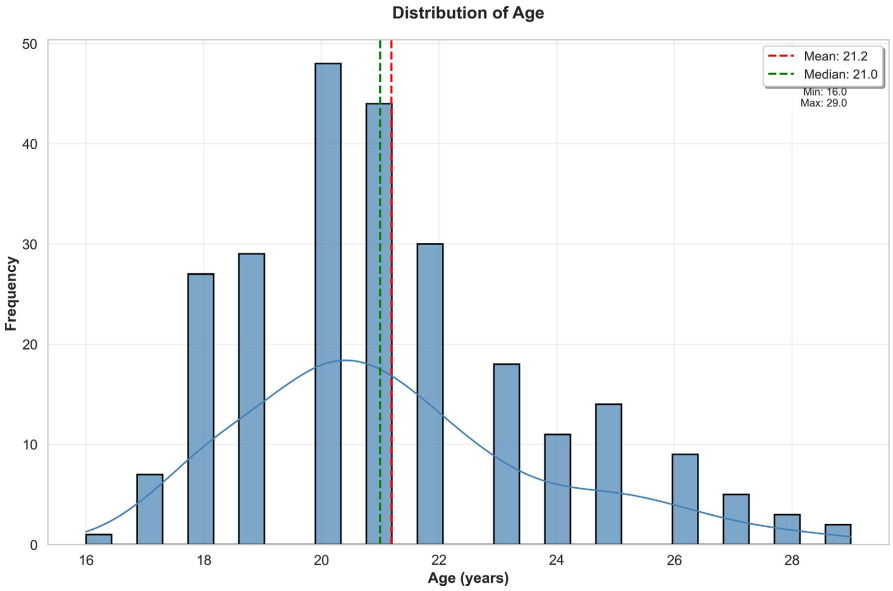


Figure 1. Age distribution of the students.

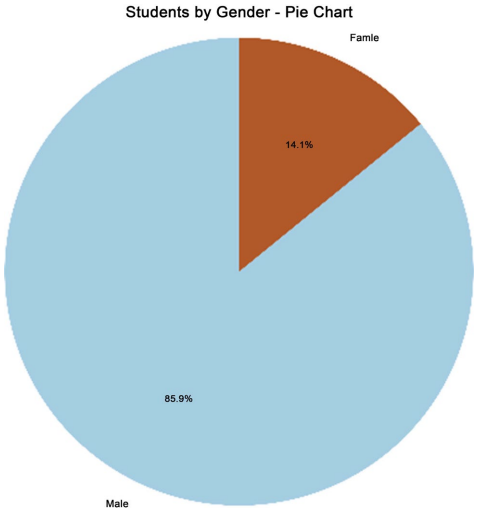


Figure 2. Students by gender.

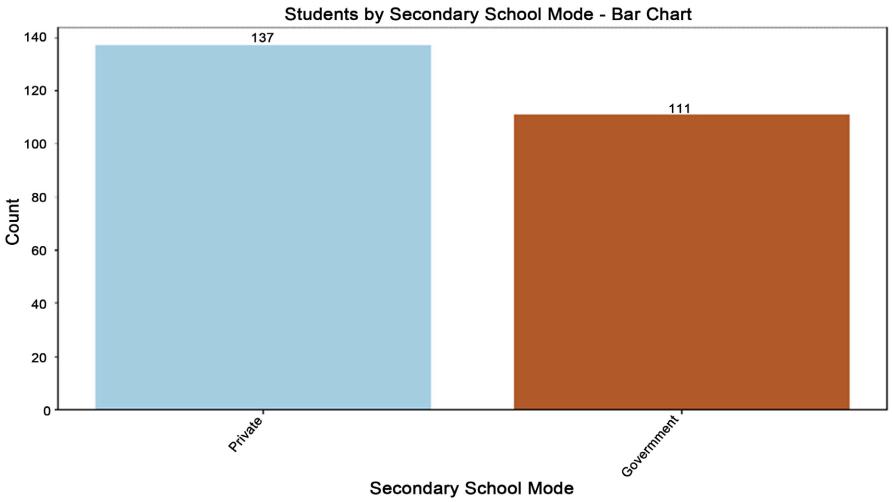


Figure 3. Students by secondary school mode.

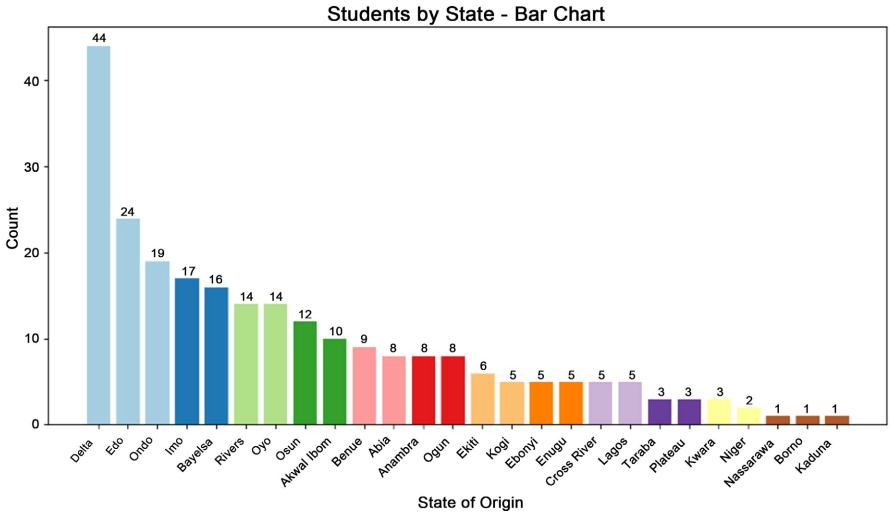


Figure 4. State of origin distribution of the student.

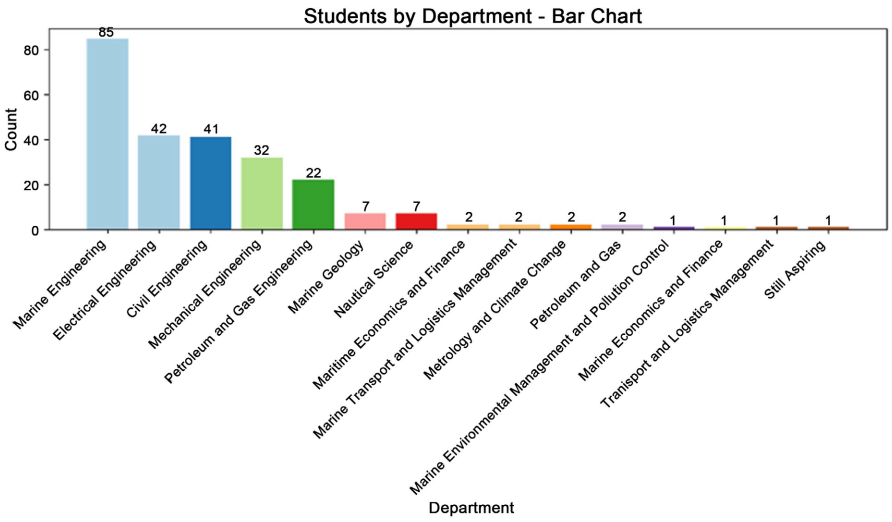


Figure 5. Students by department.

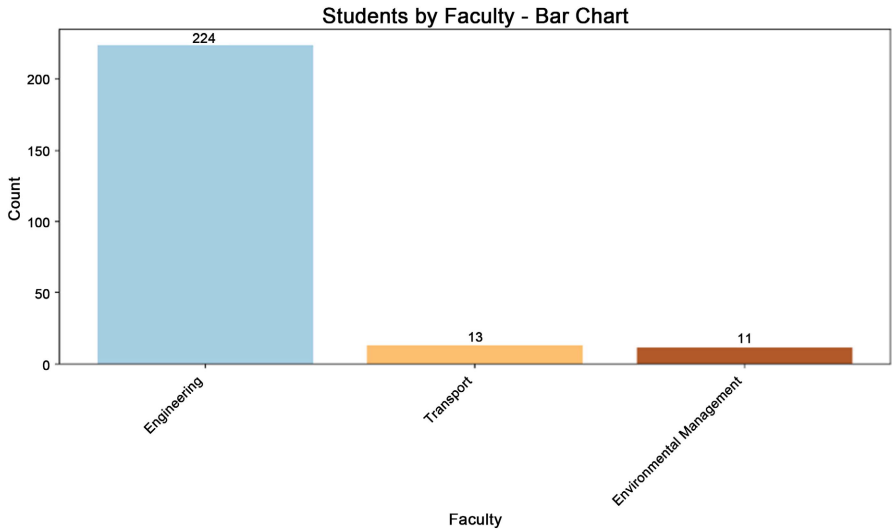


Figure 6. Students by faculty of study.

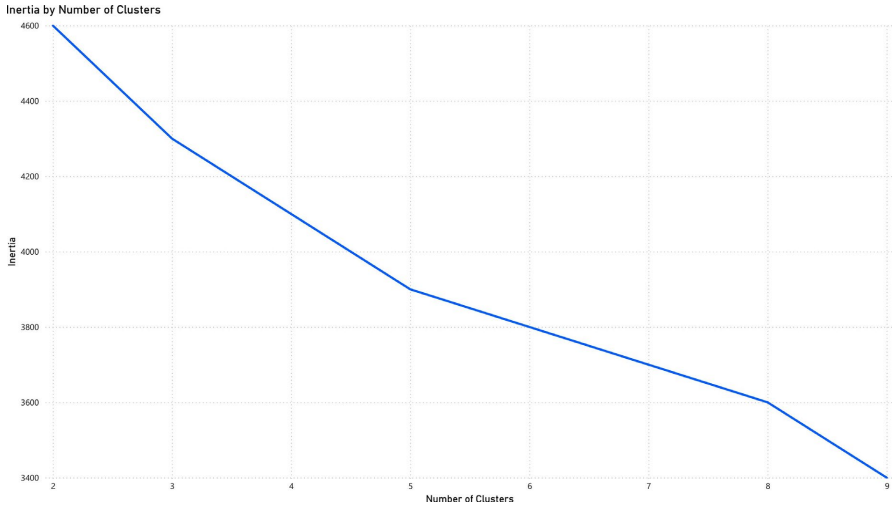
4.2. Optimal Cluster Selection

The optimal cluster selection was determined by the use of elbow method graph and silhouette score analysis. The elbow method graph revealed a gradual decrease in inertia, with a subtle inflection point at k = 4, suggesting diminishing returns in cluster homogeneity beyond this point as shown in Figure 7. The silhouette score analysis supported this selection, showing optimal cluster separation at k = 4 (silhouette score = 0.09). Based on these complementary metrics, four distinct student clusters were identified as the optimal solution.

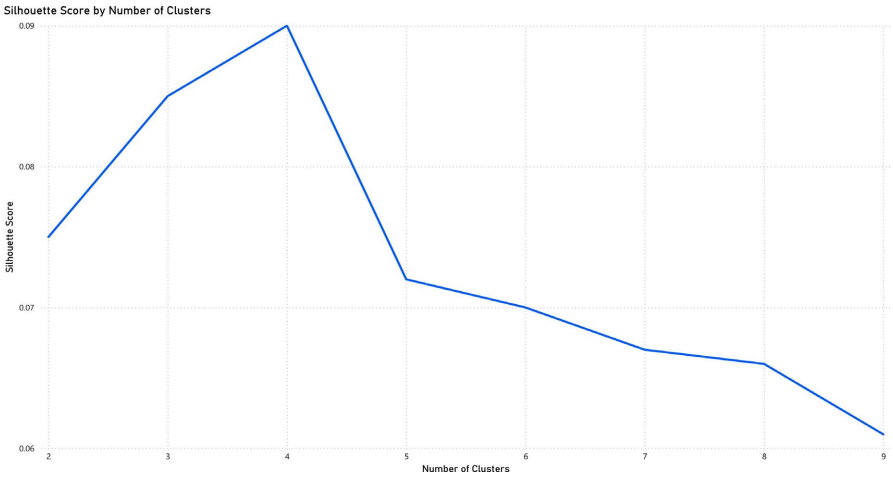
4.3. Cluster Characteristics

Cluster Composition Analysis

The cross-tabulation between cluster assignments and actual performance outcomes revealed meaning patterns are presented in Table 3.



(A)



(B)

Figure 7. Elbow method and silhouette score plots.

Table 3. Cross-Tabulation of cluster assignment and student performance.

Cluster	Pass (%)	Fail (%)	Size (n)
0	84.6	15.4	26
1	94.4	5.6	107
2	83.9	16.1	93
3	63.6	36.4	22

4.4. Statistical Validation

The chi-square test of independence yielded a p-value of 0.000949, indicating a statistically significant relationship between cluster membership and actual performance outcomes at $\alpha = 0.05$. This validates that the identified clusters capture meaningful variance in student performance.

4.5. Discussion of Findings

4.5.1. Cluster Profiling and Behavioral Characteristics

To further validate the clustering solution and develop a holistic profile of each student segment, we analyzed the mean scores of key behavioral and attitudinal variables across the four clusters. **Figure 8** presents a comprehensive comparison of cluster sizes, academic performance, study habits, and psychological factors.

4.5.2. Cluster Distribution and Academic Performance

The cluster solution successfully partitioned the sample into four distinct groups of varying sizes (**Figure 1**, Plot 1). Cluster 1 emerged as the dominant group, containing the majority of students (n = 107), while Cluster 3 represented the smallest segment (n = 22). This distribution suggests that while the high-performing profile is the norm, there exists a specific, smaller subset of the population exhibiting distinct challenges.

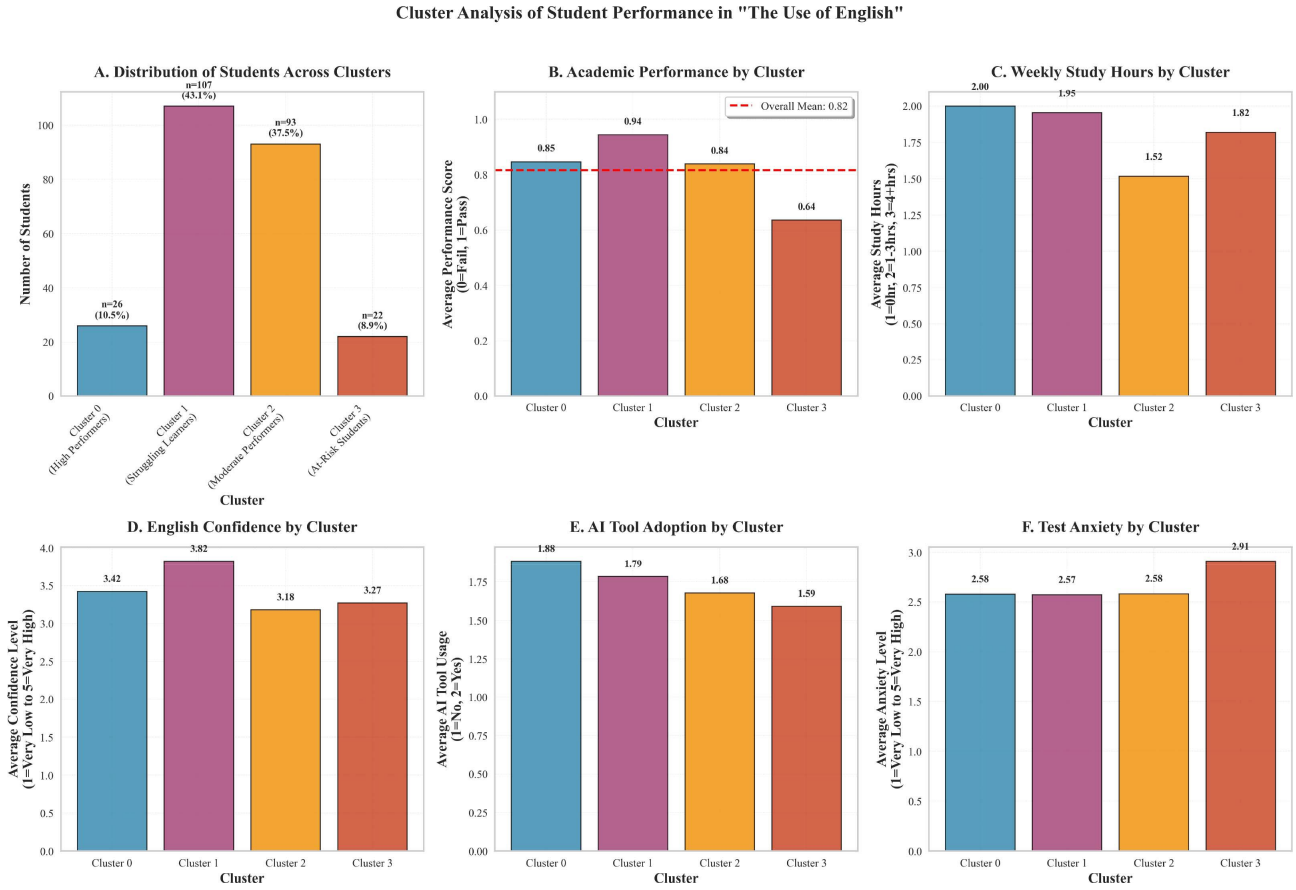


Figure 8. Cluster analysis of student performance in the "Use of English".

As illustrated in the performance metric plot (**Figure 1**, Plot 2), the clusters align hierarchically with the target outcome (final pass/fail). Cluster 1 demonstrates the highest average performance score, significantly exceeding the overall mean. Any intervention for this group should focus on enrichment or maintaining their high standards, as they require the least amount of remedial support. Clusters 0 and 2 exhibit near-identical performance levels, hovering just above the global average. The similarity in pass rates between Clusters 0 and 2 suggests that the clustering algorithms distinguished them based on other features (study habits, demographic data, prior grades), even though their final outcome is similar. Consistent with the cross-tabulation, Cluster 3 shows the lowest average performance, reinforcing its classification as the "At-Risk" segment.

4.5.3. Behavioral and Attitudinal Profiles

The analysis of non-performance variables reveals the underlying drivers of the cluster assignments, painting a rich picture of each student persona.

a) **Study Habits:** A clear positive correlation exists between cluster performance and study hours (**Figure 1**, Plot 3). The high-achieving Cluster 1 reports a high average study time. Conversely, the at-risk Cluster 3 reports the lowest investment in external study, suggesting a potential lack of engagement or time management as a key factor in their academic struggle.

b) **Psychological Factors (Confidence and Anxiety):** The confidence metric (Figure 1, Plot 4) mirrors the performance hierarchy almost perfectly, with high-performing students reporting the highest self-confidence. Interestingly, the anxiety profile (Figure 1, Plot 6) provides a nuanced insight. While the lowest-performing Cluster 3 reports high anxiety, the highest-performing Cluster 1 also reports elevated anxiety levels. This could indicate that while some anxiety is a performance driver (facilitative anxiety), excessive anxiety in Cluster 3 becomes debilitating.

c) **Technology Adoption**: The analysis of AI tool usage (Figure 1, Plot 5) reveals a distinct pattern. Clusters 1 and 2, the larger and higher-performing groups, show moderate adoption of tools like ChatGPT and Grammarly. However, the struggling Cluster 3 reports the lowest average usage of AI tools. Therefore, this data presents a paradox: the students who arguably need the most external support (Cluster 3) are the least likely to utilize available AI resources, while moderately high achievers (Clusters 1 & 2) are leveraging them to potentially enhance their already solid performance.

4.5.4. Summary of Cluster Personas

Synthesizing the data from the cross-tabulation and the behavioral plots, we can define the following student personas:

a) Cluster 1 (The Diligent High-Achievers): The largest group (n = 107, 43.1%). They study the most (4+ hours weekly), possess high confidence, and achieve the highest pass rates (94.4%). Their anxiety is moderate, potentially serving as a motivator.

b) Cluster 2 (The Steady Performers): A large group of solid, average students (n = 93, 37.5%). They have good study habits (1 - 3 hours weekly), and moderate confidence, resulting in consistently average pass rates (83.9%). They represent the "typical" student profile.

c) Cluster 0 (The Quiet Achievers): A small group (n = 26, 8.9%) that achieves above-average results (84.6% pass) despite reporting lower confidence and moderate study hours. They may achieve outcomes through different pathways not fully captured in measured variables.

d) Cluster 3 (The At-Risk Group): The smallest, most vulnerable group (n = 22, 8.9%). They study the least (predominantly 0 - 1 hour weekly), have low confidence, experience high anxiety, and do not leverage AI tools. This group requires the most urgent and targeted pedagogical intervention.

5. Conclusion and Recommendations

5.1. Summary of Findings

This study successfully employed K-means clustering to identify four distinct student profiles based on comprehensive English language learning indicators. The statistical validation through chi-square analysis confirmed a significant relationship between cluster membership and actual examination performance, support-

ing the predictive validity of the clustering solution.

The optimal number of clusters (k = 4) was empirically determined through both the elbow method and silhouette score analysis, ensuring methodological rigor in the segmentation approach.

5.2. Theoretical Implications

The findings contribute to educational psychology literature by demonstrating that student performance clusters are multidimensional, incorporating:

a) Academic factors (study hours)

b) Psychological factors (confidence, anxiety, motivation)

c) Environmental factors (resource availability, AI tool usage)

5.3. Practical Recommendations

The identification of distinct student performance clusters offers valuable insights for educational practice. Building upon both empirical findings and established literature, the following recommendations are proposed for stakeholders at various levels.

5.3.1. Educational Institutions

a) Developing Differentiated Support Systems

The heterogeneity across clusters underscores the limitations of uniform intervention strategies. For the "At-Risk" cluster (Cluster 3), characterized by low confidence, limited study hours, minimal AI tool usage, and restricted textbook access, institutions should establish foundational support programs addressing both academic and affective domains [13]. These might include mandatory writing workshops, peer-assisted study sessions, and access to learning specialists. Conversely, the "High Performers" cluster (Cluster 1) would benefit from enrichment opportunities beyond the standard curriculum—honors modules, undergraduate research, and advanced workshops that nurture existing strengths.

b) Evidence-Based Resource Allocation

Students in the "At-Risk" cluster reported significantly lower access to both traditional materials (textbooks) and emerging technologies (AI writing tools). This disparity within the same institution represents an equity concern demanding administrative attention [14]. Following principles of targeted universalism, universities might provide supplemental resource packages specifically to students in resource-limited clusters while maintaining baseline access for all [15]. A "technology lending library" targeting students in Cluster 3 could democratize access to essential learning supports.

c) Developing Predictive Early Warning Systems

The significant relationship between cluster membership and subsequent performance ($\chi^2 = 0.000949$, $p < 0.001$) suggests cluster analysis can serve as a component of early monitoring systems. Rather than waiting for examinations, institutions could use cluster assignments—available early in the academic year—as indicators of potential support needs. By integrating cluster membership data into

student information systems, academic advisors could identify students from high-risk clusters for proactive outreach [16].

5.3.2. Instructors

a) Responsive Pedagogical Practices

Classroom instructors might consider differentiated instructional strategies that acknowledge students' varied starting points [17]. In courses serving predominantly "Moderate Performers" (Cluster 2), instructors might incorporate more frequent low-stakes assessments that provide regular feedback without overwhelming test anxiety. For classes with significant "High Performers" (Cluster 1), seminar-style discussions and project-based learning could capitalize on these students' confidence and motivation. Instructors can offer multiple pathways to mastery—varied readings, assignment options, and participation modes—that allow students to engage with content in ways aligned with their current confidence and skill levels [18].

b) Cultivating Peer Learning Communities

The identification of distinct clusters creates opportunities for intentional peer learning arrangements. Rather than allowing friendship-based groupings that may reinforce existing academic divisions, instructors could strategically compose small groups including students from different clusters [19]. Students from high-performing clusters might serve as peer mentors within courses—a role benefiting both mentors and mentees [20].

5.3.3. Curriculum Development

a) Integrating Affective and Cognitive Support

The strong association between psychological factors—confidence, anxiety, motivation—and cluster membership challenges purely cognitive models of curriculum design [21]. Students in the at-risk cluster reported significantly lower confidence (mean = 3.27 vs. 3.82 for cluster 1) and higher test anxiety (mean = 2.91 vs. 2.57 for cluster), suggesting curriculum reforms must address both skill development and emotional regulation. Following universal design for learning principles, curriculum developers might embed affective supports directly into course materials: metacognitive prompts helping students recognize progress, explicit test-taking strategies to reduce anxiety, and normalizing statements about language acquisition challenges that encourage help-seeking [22].

b) Purposeful Technology Integration

Differential patterns of AI tool adoption across clusters reveal both opportunity and concern. While high-performing students appear to leverage AI writing tools effectively (Cluster 1 mean = 1.79), those in the at-risk cluster either lack access or have not integrated these tools into their learning practices (mean = 1.59) [23]. Curriculum developers should consider explicit instruction in AI literacy as a core component of English courses [24]. Rather than treating tools like Grammarly as optional supplements, course designs could include structured activities teaching

students to use these technologies critically. At the same time, curriculum must address ethical dimensions—appropriate attribution, boundaries between legitimate assistance and academic dishonesty, and developing personal voice in an age of automated text generation [25].

5.4. Limitation and Future Research

Several limitations should be considered when interpreting these findings:

a) Single-Institution Design: Data were collected from one university, limiting generalizability to other institutional contexts. The unique characteristics of this institution—particularly its engineering focus and regional student composition—may influence the cluster profiles identified.

b) Sample Skew: The sample exhibits strong male skew (85.9%) and engineering faculty concentration (90.3%). While this reflects the institutional population, it limits the applicability of findings to more gender-balanced or multidisciplinary settings.

c) Self-Reported Data: All variables except the final grade classification were self-reported through Google Forms, introducing potential response biases including social desirability bias and recall inaccuracies. Students may over report positive behaviors (study hours, attendance) and underreport negative ones (anxiety, limited resource access).

d) Cross-Sectional Design: Data were collected at a single time point, capturing associations rather than causal relationships. The cluster solution demonstrates concurrent validity with performance outcomes but has not been tested for predictive validity on future cohorts.

e) Methodological Limitations of K-means with Mixed Data: The application of K-means clustering to label-encoded nominal variables (e.g., gender, state of origin) imposes an artificial ordinal structure and Euclidean distance metric that may not be appropriate for all variable types. Future research should employ clustering algorithms designed for mixed data types, such as K-prototypes or hierarchical clustering with Gower's distance, to validate these findings.

f) Variable Coverage: While this study incorporated a broader range of variables than many previous educational data mining studies, unmeasured factors—including prior academic achievement, English proficiency at entry, and detailed socioeconomic status—may provide additional explanatory power.

5.5. Conclusion

Underlying all these recommendations is recognition that students are more than their cluster assignments. The statistical patterns describe tendencies, not destinies. The most effective practices will use these insights to inform—not replace—the human work of teaching. Cluster analysis provides a map of the territory, but the journey of learning remains a fundamentally human endeavor, requiring presence, relationship, and the courage to see each student as a whole person with unique strengths, challenges, and aspirations.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Jegede, O.O. (2026) English Medium of Instruction and Its Effects on Higher Education Outcomes in Nigeria. *International Journal of Multilingualism*, **23**, 697-716. https://doi.org/10.1080/14790718.2025.2519949

[2] Pasyah, A.C. and Anggraini, L.P. (2025) Empowering Lecturers' Professionalism in Improving the English Language Skills of Maritime Institute Cadets: A Socio-Linguistic Perspective. *Edelweiss Applied Science and Technology*, **9**, 1016-1029. https://doi.org/10.55214/25768484.v9i6.8024

[3] Sajja, R., Sermet, Y., Cwiertny, D. and Demir, I. (2025) Integrating AI and Learning Analytics for Data-Driven Pedagogical Decisions and Personalized Interventions in Education. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-025-09897-9

[4] Sharma, R., Shrivastava, S.S. and Sharma, A. (2023) Predicting Student Performance Using Educational Data Mining and Learning Analytics Technique. *Journal of Intelligent Systems and Internet of Things*, **10**, 24-37. https://doi.org/10.54216/jisiot.100203

[5] Devkar, V. and Mantri, S. (2026) Machine Learning Algorithms in Early Detection of Chronic Diseases Applications of Supervised and Unsupervised Learning for Early Diagnosis and Risk Prediction. *Artificial Intelligence and Machine Learning in Neurology*, **2**, 711-739. https://doi.org/10.1002/9781394389131.ch26

[6] Hu, A. (2024) Developing an AI-Based Psychometric System for Assessing Learning Difficulties and Adaptive System to Overcome: A Qualitative and Conceptual Framework. arXiv:2403.06284.

[7] Lestari, E.A., Budiarti, B. and Juhansar, J. (2022) Utilizing Clustering Technique to Enhance Students' English Writing Performance. *English Review: Journal of English Education*, **10**, 439-452. https://doi.org/10.25134/erjee.v10i2.6245

[8] Mohamed Nafuri, A.F., Sani, N.S., Zainudin, N.F.A., Rahman, A.H.A. and Aliff, M. (2022) Clustering Analysis for Classifying Student Academic Performance in Higher Education. *Applied Sciences*, **12**, Article 9467. https://doi.org/10.3390/app12199467

[9] Oguike, O.E., Ukekwe, E.C. and Elufidodo, G.A. (2024) Using Supervised and Unsupervised Machine Learning Models to Analyze Students Academic Performance. *International Journal of Soft Computing and Engineering*, **14**, 1-6. https://doi.org/10.35940/ijsce.d3640.14040924

[10] Rahma, F. and Ulfah, S.Z. (2025) Clustering Students Based on Academic Performance and Social Factors: An Unsupervised Learning Approach to Identify Student Patterns. *International Journal for Applied Information Management*, **5**, 139-154. https://doi.org/10.47738/ijaim.v5i3.109

[11] Cao, X. (2025) Exploration of College English Test Scores Utilizing the K-Means Clustering Algorithm. *Proceedings of the* 2025 2*nd International Conference on Informatics Education and Computer Technology Applications*, Kuala Lumpur, 17-19 January 2025, 138-142. https://doi.org/10.1145/3732801.3732829

[12] Huang, W. (2025) A Study on Optimizing Student Stratification Management in English Teaching Based on K-Mean Clustering Algorithm. *Journal of Combinatorial Mathematics and Combinatorial Computing*, **127**, 859-873. https://doi.org/10.61091/jcmcc127b-048

[13] Kuh, G.D., O'Donnell, K. and Reed, S. (2017) Ensuring Quality and Taking High-Impact Practices to Scale. *Peer Review*, **19**, 25-28.

[14] Imran, A. (2022) Why Addressing Digital Inequality Should Be a Priority. *The Electronic Journal of Information Systems in Developing Countries*, **89**, e12255. https://doi.org/10.1002/isd2.12255

[15] Powell, J.A., Menendian, S. and Ake, W. (2019) Targeted Universalism. Haas Institute for a Fair and Inclusive Society.

[16] Chang, Y., Chen, F. and Lee, C. (2025) Developing an Early Warning System with Personalized Interventions to Enhance Academic Outcomes for At-Risk Students in Higher Education in Taiwan Region. *Education Sciences*, **15**, Article 1321. https://doi.org/10.3390/educsci15101321

[17] Tomlinson, C.A. and Jarvis, J.M. (2023) Differentiation: Making Curriculum Work for All Students through Responsive Planning & Instruction. In: Renzulli, J.S., Jean Gubbins, E., McMillen, K.S., Eckert, R.D. and Little, C.A., Eds., *Systems and Models for Developing Programs for the Gifted and Talented*, Routledge, 599-628. https://doi.org/10.4324/9781003419426-22

[18] Tomlinson, C.A. (2014) The Differentiated Classroom. 2nd Edition, ASCD.

[19] Cai, L., Msafiri, M.M. and Kangwa, D. (2025) Exploring the Impact of Integrating AI Tools in Higher Education Using the Zone of Proximal Development. *Education and Information Technologies*, **30**, 7191-7264. https://doi.org/10.1007/s10639-024-13112-0

[20] Colvin, J.W. and Ashman, M. (2010) Roles, Risks, and Benefits of Peer Mentoring Relationships in Higher Education. *Mentoring & Tutoring: Partnership in Learning*, **18**, 121-134. https://doi.org/10.1080/13611261003678879

[21] Perkins, N.A. and Bains, R. (2025) From Uncertainty to Competence: A Longitudinal Study of Confidence Development in Occupational Therapy Education. *Occupational Therapy International*, **2025**, Article ID: 1797008. https://doi.org/10.1155/oti/1797008

[22] Mais-Thompson, E., Brown, B. and Paul, N. (2024) Unique Practices in Teaching Affective Learning in a Higher Education Applied Curriculum. *The Curriculum Journal*, **36**, 180-199. https://doi.org/10.1002/curj.285

[23] Dalal, P., Beniwal, G., Sharma, V., Garg, P. and Ahmed, K. (2025) Predicting Student Motivation and Engagement through Machine Learning Models. *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, **32**, 393-411.

[24] Willner, L.S. (2025) AI-Powered Instructional Planning for Integrated Content, Literacy, and English Language Development for K-12 Multilingual Learners. *GATESOL Journal*, **34**, 17-34.

[25] Nwozor, A. (2025) Artificial Intelligence (AI) and Academic Honesty-Dishonesty Nexus: Trends and Preventive Measures. *AI and Ethics, Academic Integrity and the Future of Quality Assurance in Higher Education*, 27.