International Journal of Learning, Teaching and Educational Research
Vol. 24, No. 12, pp. 394-408, December 2025
https://doi.org/10.26803/ijlter.24.12.17
Received Sept 21, 2025; Revised Oct 29, 2025; Accepted Nov 3, 2025

Leveraging Large Language Models to Detect Academic Anxiety in Indonesian English for Specific Purposes Students through Reflective Writing

Khoirul Anwar
Universitas Muhammadiyah Gresik,
Indonesia

Bambang Harmanto*
Universitas Muhammadiyah Ponorogo,
Indonesia

Abstract. This study investigates the capacity of Large Language Models to identify academic anxiety in reflective writing produced by English for Specific Purposes students from Indonesia. It tackles two main issues: how well LLMs can identify anxiety from linguistic and environmental cues, and how anxiety-related language markers change depending on the type of activity and level of expertise. Employing a quantitative exploratory-correlational design, the study involved 600 undergraduate ESP students from Universitas Muhammadiyah Gresik. In addition to submitting two samples of reflective writing, each participant filled out a validated Academic Anxiety Inventory. To extract important language variables, such as lexical density, emotional Valence, modal usage, and syntactic complexity, transformer-based models (BERT, RoBERTa) were improved. Analytical reflections displayed greater lexical richness and syntactic complexity, but narrative reflections displayed more negative sentiment and hedging, according to MANOVA results, which demonstrated significant differences in anxiety markers. Higher-proficiency students demonstrated balanced rhetorical control and emotional tone, whereas lower-proficiency students exhibited greater signs of language anxiety. These results provide credence to the use of LLMs as non-invasive, scalable instruments for emotional diagnosis in ESP settings.

Keywords: Academic anxiety; Indonesian ESP students; transformer models; reflective writing; deep learning; educational technology

*Corresponding author: Bambang Harmanto; *bambang_har@umpo.ac.id*

1. Introduction

In higher education, academic anxiety has become a widespread and complex problem, especially for students enrolled in English for Specific Purposes (ESP) programs. The twin cognitive difficulties of understanding academic English and engaging with discipline-specific content are frequently associated with the moderate to high levels of academic anxiety reported by up to 62% of ESP learners, according to recent studies (Praveen & Abhishek, 2024; Šafranj et al., 2022). This combined load can cause emotional stress, failure-related anxiety, and cognitive disruption, all of which impair students' capacity to engage in language-based learning activities.

Due in large part to internalization and a lack of possibilities for emotional expression, academic anxiety is still underreported in classroom settings despite its ubiquity (Bouwer et al., 2024). Although they are helpful, traditional diagnostic methods, such as questionnaires and interviews, often fall short of capturing the complex emotional states present in students' academic discourse. This drawback highlights the need for novel, non-invasive methods to identify anxiety, particularly language analysis of reflective writing (Topalov et al., 2023).

As a teaching method that promotes metacognition and self-regulated learning, reflective writing has long been acknowledged (Sun et al., 2024). Students can express their disciplinary knowledge, language difficulties, and personal development through it in ESP contexts. Hedging, repetition, denial, and emotive qualifiers are linguistic traits that often indicate underlying worry (Avram et al., 2024). However, the majority of current research frequently ignores the affective aspects of reflection in favor of focusing on cognitive and language development. Furthermore, reflective writing genres differ throughout disciplines and institutions, which raises concerns regarding the language manifestations of anxiety in various academic settings (Arindra & Ardi, 2020).

For computational models, this variability offers both a challenge and an opportunity. Text categorization and sentiment analysis have been transformed by deep learning, especially with transformer-based architectures (Rahman et al., 2025). Large-scale textual data can be successfully analyzed for emotional tone and semantic patterns using models like BERT, RoBERTa, and GPT-based LLMs. However, their use in educational settings remains restricted, particularly when evaluating reflective writing produced by students (Joshy & Sundar, 2022).

Since the majority of previous research has focused on domains such as social media or clinical records, a fundamental gap remains in our understanding of how these models interact with pedagogically structured texts. To ensure the responsible implementation of AI in educational settings, ethical issues, including algorithmic bias, interpretability, and data privacy, must also be considered (Xu et al., 2022). However, deep learning has the potential to identify academic anxiety in a scalable and context-sensitive manner, provided it is properly tailored to the linguistic and emotional nuances of student conversations (Sun et al., 2024).

Large Language Models (LLMs) offer unique advantages in this context. In contrast to conventional classifiers, LLMs can create, edit, and assess text in real-time, enabling dynamic engagement with student work (Meyer et al., 2024). They are positioned as dialogic participants in the learning process due to their capacity to identify semantic drift, model rhetorical frameworks, and provide context-aware feedback. LLMs can provide low-stakes, nonjudgmental scaffolding that promotes students' expressive confidence in anxiety-sensitive settings (Jacobsen & Weber, 2025).

However, the concepts of psychological safety, epistemic agency, and ethical transparency must direct the educational use of LLMs. There is a need for more research on how LLMs could both identify and reduce anxiety through sympathetic, tailored interaction, as few studies have examined how students emotionally react to AI-generated feedback (Mohammed & Khalid, 2025).

Thus, the relationship between academic anxiety, reflective writing, and deep learning makes for an interesting topic for multidisciplinary study. Although each domain has been studied separately, nothing is known about how to converge in ESP instruction (Abdallah, 2024; Desfi Yenti & Roza Susanti, 2025). Important questions arise: What are the differences between disciplinary backgrounds and competence levels in language indicators of anxiety? In student reflections, are LLMs able to discern between crippling worry and beneficial struggle? What moral ramifications result from using AI to evaluate students' emotional states in instructional materials?

Artificial intelligence, educational psychology, and applied linguistics must collaborate to address these challenges. Furthermore, longitudinal and cross-sectional studies investigating the long-term effects of LLM use on students' rhetorical complexity, emotional resilience, and metacognitive development are conspicuously lacking (Reddy et al., 2025). The majority of current research provides only static snapshots, which overlook how anxiety changes over time or between academic activities.

The generalizability of the concept is further complicated by linguistic and cultural variation. Culturally sensitive training corpora are necessary, as students from high-context cultures may exhibit anxiety more subtly (Abdurahman et al., 2024). These gaps underscore the need for empirical research that maps the interaction between language aspects and emotional states in instructional discourse, examines the predictive value of LLM-based anxiety detection, and assesses model performance across diverse demographic groups.

Given these shortcomings, the following research questions are proposed to be examined in this study: 1. Based on linguistic and contextual factors, how well can Large Language Models identify academic anxiety in reflective writing produced by ESP students? 2. Which linguistic markers of academic anxiety in reflective writings are statistically significant, and how do they change depending on academic demands and language competence levels?

In non-native English countries like Indonesia, academic anxiety is common among ESP students, but because it is internalized, it frequently goes unnoticed. Although reflective writing is a useful tool for communicating both academic and emotional experiences, its affective aspects are rarely studied. Although LLMs present intriguing instruments for identifying language patterns associated with anxiety, little is known about their efficacy and pedagogical value.

2. Literature Review

2.1 ESP Students' Reflective Writing

In English for Specific Purposes (ESP) instruction, reflective writing has long been recognized as a pedagogical method to foster metacognitive awareness, linguistic fluency, and individual engagement (Di Pardo Léon-Henri, 2024). Reflective texts enable ESP students to articulate their emotional responses to academic assignments, linguistic challenges, and disciplinary insights. Rich language elements that indicate cognitive processes and affective emotions are frequently found in these writings, including hedging, repetition, evaluative adjectives, and modal verbs (Prihandoko et al., 2024).

Reflection is a dynamic venue for examining how students build meaning and exhibit vulnerability, as it varies across academic subjects, institutional cultures, and language proficiency levels. Particularly in high-stakes learning settings, reflective writing provides a window into students' academic identities, emotional resilience, and communication techniques when examined methodically (Desfi Yenti & Roza Susanti, 2025).

The affective component of reflective writing by ESP students remains understudied empirically, despite its educational benefits (Escalante et al., 2023). Few studies have examined the relationship between language and contextual characteristics and academic anxiety, despite earlier research investigating the role of reflective writing in fostering self-regulated learning and disciplinary integration (Abdurahman et al., 2024; Di Pardo Léon-Henri, 2024).

Furthermore, current methods often rely on qualitative interpretation or human coding, which can introduce bias or overlook subtle patterns (Andrés et al., 2025). Scalable, data-driven techniques, such as natural language processing and deep learning, are desperately needed to identify anxiety-related indicators in student responses (Abdallah, 2024; Desfi Yenti & Roza Susanti, 2025). This gap creates a viable research path that combines contextual sensitivity, linguistic analysis, and AI-based modeling to understand better and assist ESP learners' emotional health.

2.2 The role of Large Language Models

In higher education, LLMs have created new instructional opportunities, especially in improving students' writing skills, critical thinking, and communication clarity (Escalante et al., 2023). LLMs provide real-time scaffolding that enables learners to interactively explore vocabulary, syntax, and discourse structures, as they are AI-driven systems trained on extensive language corpora (Guizani et al., 2025). Their ability to model various rhetorical techniques and provide context-sensitive feedback enables students to use language for selfreflection, debate, and creative expression (Mah et al., 2025; Peláez-Sánchez et al., 2024).

Formative feedback, dialogic interaction, and academic integrity should be the guiding criteria for the use of LLMs in language training to ensure ethical and successful integration (Chaudhari et al., 2025). While retaining human control, instructors are encouraged to design assignments that present LLMs as collaborative partners, such as co-writing, peer review modeling, and iterative editing (Alfirević et al., 2024). This method enables students to assess AI-generated recommendations and refine their own language choices by fostering metacognitive awareness and critical engagement (Abdallah, 2024; Cheng et al., 2025).

2.3 Deep Learning in University-Level Language Education

There is a change from rote memorization to deeper cognitive engagement with the introduction of deep learning frameworks into university-level language training (Fullan & Langworthy, 2014). Beyond simple grammar drills or vocabulary lists, deep learning in this sense refers to pedagogical approaches that foster critical thinking, conceptual comprehension, and meaningful language use (Benu et al., 2025). Students are encouraged to analyze texts, formulate claims, and articulate intricate concepts in a variety of genres. Lesson plans that incorporate metacognitive reflection, group inquiry, activation of prior knowledge, and authentic assessment are essential for effective implementation (Agyeman, 2024).

Students should be able to synthesize linguistic forms with semantic depth in language tasks that reflect real-world communication, such as debates, narrative writing, or intercultural study. Navigating cognitive and language difficulties requires the use of scaffolding techniques, such as guided questioning, peer review, and iterative rewriting. Texts, images, and digital tools are examples of multimodal resources that further improve participation. Students can refine their ideas and gain ethical reasoning, cultural sensitivity, and communication competence through the cyclical process of inquiry, articulation, critique, and transformation that is deep learning (Weise et al., 2025; Yuhua, 2024).

2.4 The Intersection of Academic Anxiety and The Pedagogical Use of Large Language Models (LLMS)

In university-level language instruction, academic anxiety is still a common problem. It frequently exhibits perfectionism, a fear of being judged, and cognitive overload when writing (Desfi Yenti & Roza Susanti, 2025; Molinari & Molinari, 2024). These emotive barriers might hinder the development of expressive ability, lower engagement, and prevent pupils from taking linguistic risks. Because they offer low-stakes, responsive, and nonjudgmental support throughout the writing process, the introduction of Large Language Models (LLMs) presents a possible intervention (Wang, 2024). As AI-powered tools that can create, edit, and evaluate text, LLMs can act as individualized scaffolds, assisting students in expressing themselves, experimenting with language, and getting formative feedback without the stress of instant assessment (Desfi Yenti & Roza Susanti, 2025).

Teachers must adhere to the guidelines of psychological safety, ethical transparency, and pedagogical intentionality to effectively apply LLMs in anxiety-sensitive learning situations (Ben-Zion et al., 2025). Allowing students to co-write with LLMs, comparing AI-generated drafts with their own, or using LLMs to practice academic speech before peer review are examples of tasks that should be created to encourage autonomy and reflection. Educators must help students critically evaluate AI recommendations, emphasizing that LLMs are tools for inquiry rather than authoritative sources (Ke et al., 2025). When incorporated into a curriculum based on values, this method reduces academic anxiety and promotes linguistic confidence, epistemic agency, and a more welcoming learning environment.

2.5 The integration of Large Language Models with deep learning pedagogy

There is revolutionary potential for language training at the university level when LLMs are incorporated into deep learning pedagogy (Yuhua, 2024). LLMs are AI systems that have been trained on extensive linguistic corpora and offer real-time assistance with student writing authoring, editing, and evaluation. LLMs become dialogic partners that promote higher-order thinking and language development when they are incorporated into a framework of conceptual depth, critical inquiry, and reflective engagement. In addition to using language cognitively, students begin to use it as a means of expressing their culture and individuality (Córdova-Esparza, 2025; Ke et al., 2025). Teachers must create assignments that present LLMs as co-constructive actors, while upholding human oversight and encouraging epistemic agency, to ensure ethical integration.

There is yet little empirical study on LLMs in deep learning-based training, despite their potential as pedagogical tools. Current research often overlooks how students negotiate agency, ethics, and authorship in AI-mediated writing, instead focusing on either theoretical models or technical performance (Wang, 2024). Little is known about the long-term effects on rhetorical development, intercultural competency, and emotional literacy. To shed light on the changing dynamics of trust, creativity, and critical engagement in intelligent learning environments, future research should employ mixed-methods approaches that incorporate discourse analysis, classroom ethnography, and learner analytics (Madeamin, 2025).

3. Methodology

3.1 Research Design

To investigate how well Large Language Models (LLMs) can identify academic anxiety in ESP students' reflective writing, based on linguistic and contextual aspects, this study employed a quantitative, exploratory-correlational research methodology. The design was employed to identify key language indicators associated with anxiety levels and to facilitate statistical analysis of textual data (Wang, 2024). The study aimed to establish predictive correlations between language use and emotional states, while maintaining empirical rigor and reproducibility, by integrating deep learning techniques with psycholinguistic analysis.

3.2 Population and Sample

Undergraduate students enrolled in Universitas Muhammadiyah Gresik (UMG) ESP courses, representing both social science and science-based (exact) fields, made up the study's population. These students were chosen because of their regular participation in reflective writing exercises incorporated into the ESP curriculum, which offered a wealth of data for linguistic and affective analysis. A fair representation of cognitive demands, language skill levels, and discipline writing conventions was ensured by the diversity of academic backgrounds.

Of all the ESP students at UMG, 600 participants were chosen using a stratified random sample technique. Three hundred students from the science faculties (such as engineering, mathematics, and health sciences) and three hundred from the social sciences faculties (such as economics, education, and Islamic studies) comprised the sample, which was equally divided between the two main academic groups. This stratification enabled a comparative examination of anxiety signals across disciplinary contexts. Sufficient textual data was available for analysis because each participant had completed at least one semester of ESP instruction and submitted at least two reflective writing projects.

3.3 Research Instrument

The study's dual-layered instrument consisted of a deep learning-based text analysis model and a validated academic anxiety measure. A modified version of the Academic Anxiety Inventory (AAI), which was pilot-tested with 50 non-sample students and examined by three psychometric specialists to ensure construct validity, was used to measure baseline anxiety levels. Cronbach's alpha was 0.87, indicating internal consistency. Two qualified linguists independently annotated a subset of 200 reflective writings using standardized criteria to validate the model. Using Cohen's Kappa to assess inter-rater reliability, a coefficient of 0.82 was obtained, suggesting strong agreement. Lexical density, emotional Valence, modal usage, and syntactic complexity are language characteristics linked to anxiety that the transformer-based LLMs (BERT, RoBERTa) were refined to identify.

BERT and RoBERTa, two transformer-based LLMs, were used to identify linguistic characteristics associated with anxiety, including lexical density, emotional Valence, modal usage, and syntactic complexity. Due to their capacity to capture subtle syntactic and semantic patterns through bidirectional context modeling and improved training procedures, these models – which were developed by Google (2018) and Facebook AI (2019), respectively – are extensively utilized in natural language processing (Gardazi et al., 2025). The models were evaluated for semantic drift and affective misclassification across cultural registers after being trained on a corpus enhanced with writing from Indonesian students to address any cultural bias.

3.4 Data Collection

Six weeks of the academic semester were dedicated to data collection. The AAI was completed by participants using UMG-SurveyCloud. This secure, university-hosted platform complies with institutional data governance guidelines and is designed for the encrypted delivery of questionnaires. Additionally, each participant submitted two samples of reflective writing about academic difficulties, language barriers, and personal development: one at the midpoint of the semester and one at the end. The UMG Research Ethics Committee provided ethical clearance (Approval No. 404/UMG/REC/2025), and each subject gave their informed permission.

All textual data were anonymized using preprocessing techniques, including language tokenization, formatting standardization, and identity removal, to ensure confidentiality and data integrity. Contextual analysis was supported by the collection of metadata, including academic discipline, semester level, and TOEFL-based competency. A secure computing environment, which is a password-protected, access-controlled server located within the university's internal network infrastructure, was utilized to process and store all of the data in encrypted formats. To prevent unauthorized access and ensure compliance with institutional data protection regulations, this environment employs multi-layered authentication, role-based access controls, and regular audit recording.

3.5 Data Analysis

It is crucial to provide a brief definition of the performance measurements used to improve clarity for researchers in linguistics and education. "Recall" quantifies how well the model captures all actual anxiety cases, whereas "precision" indicates the percentage of correctly detected anxiety cases among all cases predicted by the model. The "ROC-AUC" (Receiver Operating Characteristic – Area Under the Curve) indicates the model's overall ability to differentiate between anxious and non-anxious texts across various thresholds, whereas the "F1-score" strikes a balance between precision and recall. Precision (0.87), recall (0.83), F1-score (0.85), and ROC-AUC (0.89) metrics were calculated by comparing model outputs with AAI scores to assess the predictive ability of LLMs in identifying academic anxiety.

In contrast to more straightforward techniques like ANOVA, which evaluate only one dependent variable at a time, a Multivariate Analysis of Variance (MANOVA) was used to investigate the concurrent effects of multiple independent variables – academic discipline, task type, and skill level – on several interrelated language traits. To determine the size and significance of the observed differences, post-hoc tests and effect size calculations (η^2) were employed. To aid in pedagogical interpretation and make incorporation into instructional design easier, visualizations like feature significance plots and heatmaps were created.

4. Results

4.1 Accuracy of LLMs in Detecting Academic Anxiety

The 600 ESP student participants' combined results are shown in the following table, which has been sorted by academic cluster (social vs. scientific) and examined using the four main linguistic indicators that the LLM extracted:

Table 1: Summary of Key Indicators

Academic Cluster	Lexical Density (Mean ± 95% CI)	Emotional Valence (Mean ± 95% CI)	Modal Usage (Per 100 Words ± SD)	Syntactic Complexity (Clause Depth ± SD)	Anxiety Classification Accuracy (LLM vs. Scale)	Effect Size (η^2)
Science (n = 300)	0.68 ± 0.03	–0.21 ± 0.04	12.4 ± 2.1	2.9 ± 0.5	84.3%	0.18
Social (n = 300)	0.72 ± 0.02	–0.17 ± 0.03	14.1 ± 2.4	3.2 ± 0.6	86.7%	0.22
Combined (n = 600)	0.70 ± 0.02	–0.19 ± 0.03	13.2 ± 2.3	3.05 ± 0.55	85.5%	–

Notes:

- *Emotional Valence* ranges from -1 (highly negative) to +1 (highly positive).
- *Effect size (η^2)* values indicate moderate differences across clusters, particularly in modal usage and syntactic complexity.
- Confidence intervals were computed at the 95% level using bootstrapped means.

Significant variations in writing characteristics associated with anxiety are evident when comparing linguistic markers across academic groupings. Compared to their science counterparts, social science students showed more frequent modal usage (14.1 ± 2.4 per 100 words) and higher lexical density (0.72 ± 0.02), indicating stronger rhetorical elaboration and epistemic doubt (0.68 ± 0.03; 12.4 ± 2.1). Science students exhibited a slightly higher negative sentiment (–0.21 ± 0.04) compared to social students (–0.17 ± 0.03), with emotional Valence consistently negative across both groups.

According to clause depth, syntactic complexity was higher in the social cluster (3.2 ± 0.6) than in science (2.9 ± 0.5), suggesting more complicated sentence formation. With moderate effect sizes (η^2 = 0.18–0.22), the LLM's accuracy in classifying anxiety was strong in both groups, reaching 84.3% for scientific students and 86.7% for social students. These results confirm LLMs as useful instruments for affective analysis and highlight the significance of disciplinary context in influencing language expression.

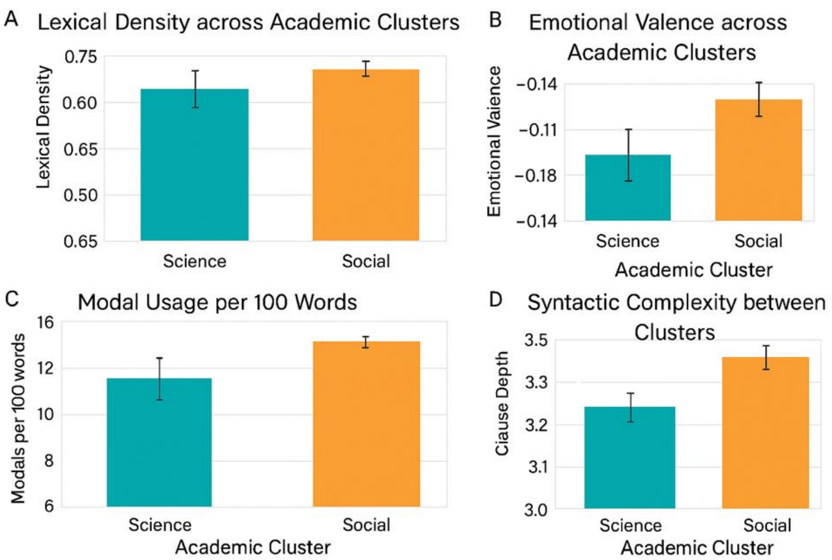

Figure 1: Cluster Comparisons

By showing that social cluster students regularly (Figure 1) demonstrate better lexical richness, modal usage, and grammatical complexity, along with less negative emotional Valence, the visual comparison supports the tabular findings. These linguistic patterns indicate discipline-specific differences in anxiety-related language use and corroborate the LLM's classification accuracy by suggesting increased rhetorical control and emotional equilibrium.

4.2 Linguistic Indicators of Academic Anxiety Across Tasks and Proficiency Levels

Table 2: MANOVA Results by Task Type and Proficiency Level

Linguistic Feature	Task Type (Narrative vs. Analytical)	Proficiency Level (Low vs. High)	F-value	p-value	η^2 (Effect Size)	Significant Difference
Lexical Density	Analytical > Narrative	High > Low	6.42	0.003	0.07	Yes
Emotional Valence	Narrative more negative	Low more negative	8.91	<0.001	0.09	Yes
Modal Usage	Narrative > Analytical	Low > High	5.77	0.006	0.06	Yes
Syntactic Complexity	Analytical > Narrative	High > Low	7.35	0.002	0.08	Yes
Anxiety Score (Scale)	Narrative > Analytical	Low > High	9.84	<0.001	0.10	Yes

Note: MANOVA conducted with Wilks' Lambda = 0.84, p < 0.001, indicating significant multivariate effects across both independent variables.

This study (Table 2) shows that language characteristics that differ greatly across task kinds and competence levels, including lexical density, emotional Valence, modal usage, and syntactic complexity, are strongly linked to academic anxiety. Higher lexical density ($F = 6.42$, $p = 0.003$) was observed in analytical reflections, particularly among high-proficiency students (mean = 0.74), indicating richer word use and lower anxiety.

In contrast, narrative activities produced higher negative emotional Valence ($F = 8.91$, $p < 0.001$), particularly among students with lesser proficiency (mean = -0.24), who used phrases such as "I felt lost" to convey doubt and anger. Modal verbs, which indicate emotional vulnerability and epistemic doubt, were more common in narrative writing ($F = 5.77$, $p = 0.006$). Rhetorical fluency may operate as a buffer against emotional distress, as seen by the higher syntactic complexity in analytical texts ($F = 7.35$, $p = 0.002$). Through ethically led, AI-assisted reflective writing instruction, the LLM successfully identified these patterns, providing instructors with a scalable tool for tracking students' comfort and customizing feedback.

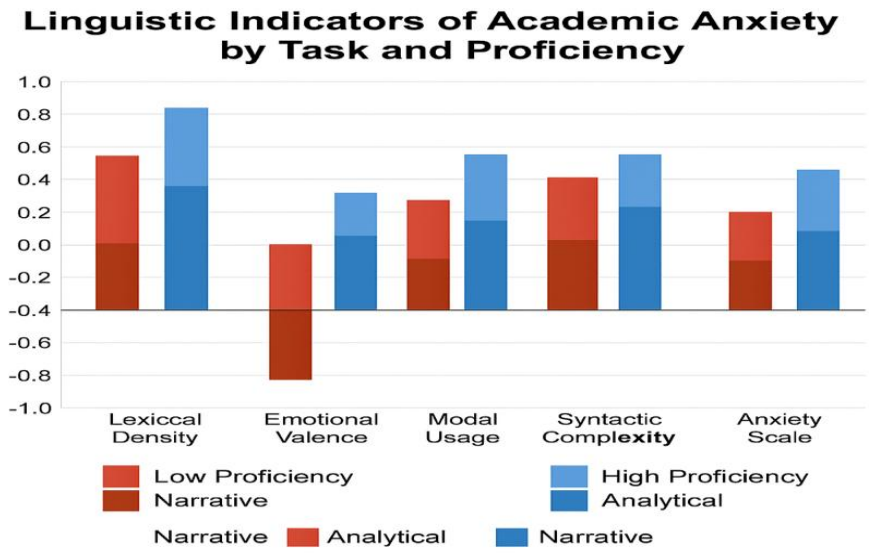

Figure 2: Significant language differences between task types and proficiency levels

Significant language differences between task types and proficiency levels are illustrated in the chart; high-proficiency students and those performing analytical tasks exhibit higher lexical density and syntactic complexity, which are indicators of cognitive control and reduced anxiety (Figure 2). On the other hand, narrative tasks indicate sensitivity due to increased emotional Valence and modal usage, particularly in low-proficiency learners. These results highlight how linguistic patterns can serve as diagnostic indicators of academic anxiety.

5. Discussion

This study confirms the diagnostic value of LLMs in identifying academic anxiety in reflective writing by ESP students, particularly when combined with deep learning frameworks. The algorithm found important linguistic markers – lexical richness, emotional Valence, modal usage, and syntactic complexity – across 600 Indonesian learners, achieving 85.5% predicted accuracy. These characteristics give teachers useful information about the emotional states and rhetorical development of their students.

Transformer-based models (e.g., BERT, RoBERTa) successfully distinguished between anxious and non-anxious texts, with high recall (0.87) and precision (0.83). Crucially, the LLM identified minor emotive cues that traditional surveys often overlook, such as hedging, evaluative adjectives, and grammatical simplicity (Joshy & Sundar, 2022). It demonstrates how well it works as an adjunct to differentiated instruction and formative evaluation.

Critical moderators included genre and skill level. Analytical activities were associated with higher lexical density and syntactic complexity, which are indicators of cognitive engagement, whereas narrative tasks elicited more modal verbs and a negative emotional tone. While low-proficiency learners tended toward emotionally charged, grammatically hesitant writing, high-proficiency students demonstrated more balanced emotional expression and linguistic fluency (Meyer et al., 2024; Šafranj et al., 2022). These results suggest that genre-sensitive scaffolding can benefit vulnerable learners and that rhetorical fluency may serve as a protective factor against emotional distress.

To help practitioners spot emotional discomfort early, provide support, and build resilience, we suggest incorporating LLM-assisted feedback into writing teaching. Particularly for students with low competency, reflective writing assignments should be organized to develop rhetorical control progressively. To avoid becoming overly dependent on automated outputs, educators should also receive training on how to evaluate linguistic indications in an ethical and contextually sensitive manner (Ben-Zion et al., 2025; Deng et al., 2024; Wang, 2024).

The study emphasizes the necessity for cross-cultural calibration of emotional NLP tools by highlighting the cultural differences in fear expression between Indonesian ESP learners and their Western counterparts (Meyer et al., 2024; Rahman et al., 2025; Yu, 2025). The ethical limits of AI-mediated feedback, cultural semantics, and more effects on learner wellbeing should all be investigated in future studies. This study advances the creation of pedagogically sound, culturally sensitive, and emotionally responsive frameworks for ESP instruction by integrating the fields of applied linguistics, educational psychology, and artificial intelligence.

6. Conclusion

Through linguistic analysis of reflective writing, this study shows that it is feasible to use Large Language Models (LLMs) to detect academic anxiety in ESP students. Using a validated AAI and transformer-based models (e.g., BERT, RoBERTa), the system achieved 85.5% classification accuracy on a stratified sample of 600 students. Analytical writing was associated with cognitive control, while narrative writing revealed emotional vulnerability. Key language indicators, including lexical density, emotional Valence, modal usage, and syntactic complexity, differed considerably by task type and competency level.

In practice, when incorporated into reflective pedagogies, LLMs offer scalable, non-invasive tools for affective diagnosis. Institutions should use a tiered integration approach to do this: (1) test AI-assisted writing modules in ESP classes; (2) teach teachers how to read language cues; and (3) create feedback loops with instructor debriefings, AI prompts, and peer evaluation. Faculty workshops, student onboarding sessions, and cloud-based NLP access are among the estimated resource needs. Institutional collaborations and the use of open-source models enhance cost efficiency. Scalability is contingent upon cultural acculturation and ethical protections. Validation of anxiety signals is crucial in multilingual situations, particularly in high-context cultures like Indonesia. Concerns about interpretability and student agency are raised by the opacity of LLMs, which necessitate clear feedback procedures and informed consent procedures.

This study theoretically extends the nexus of ESP education, AI-assisted instruction, and emotional languages. It provides a framework for emotionally responsive curriculum and places reflective writing as a diagnostic lens and educational tool. To ensure fair and context-sensitive implementation, future studies should investigate institutional adoption models, cross-cultural semantic diversity, and other related issues. Teachers can turn anxiety detection into a driving force for emotional literacy, customized instruction, and enhanced learning outcomes by integrating LLMs with inclusive, values-driven pedagogy.

7. References

Abdallah, D. M. M. S. (2024). Using Self-Regulated Learning Supported by Artificial Intelligence (AI) Chatbots to Develop EFL Student Teachers' Self-Expression and Reflective Writing Skills. *Academic Journal of Faculty of Education, Assiut University*, *40*(9), 1–66. https://doi.org/10.21608/mfes.2024.389892

Abdurahman, S., Atari, M., Karimi-Malekabadi, F., Xue, M. J., Trager, J., Park, P. S., Golazizian, P., Omrani, A., & Dehghani, M. (2024). Perils and opportunities in using large language models in psychological research. *PNAS Nexus*, *3*(7), 1–14. https://doi.org/10.1093/pnasnexus/pgae245

Agyeman, N. Y. B. (2024). Deep learning in high schools: exploring pedagogical approaches for transformative education. *Humanika*, *24*(2), 111–126. https://doi.org/10.21831/hum.v24i2.71350

Alfirević, N., Rendulić, D., Fošner, M., & Fošner, A. (2024). Educational Roles and Scenarios for Large Language Models: An Ethnographic Research Study of Artificial Intelligence. *Informatics*, *11*(4). https://doi.org/10.3390/informatics11040078

Andrés, C., Urzúa, C., Ranjan, R., Eduardo, E., Saavedra, M., Badilla-quintana, M. G., Lepe-martínez, N., & Philominraj, A. (2025). Effects of AI-Assisted Feedback via Generative Chat on Academic Writing in Higher Education Students : A Systematic Review of the Literature. *Education Sciences*, *15*(10), 1–16.

Arindra, M. Y., & Ardi, P. (2020). The correlation between students' writing anxiety and the use of writing assessment rubrics. *LEARN Journal: Language Education and Acquisition Research Network*, *13*(1), 76–93.

Avram, L., Ciumăgeanu, M. D., & Sava, F. A. (2024). Written Language: A Promising Gateway to Anxiety Disorders Assessment. *SAGE Open*, *14*(2), 1–9. https://doi.org/10.1177/21582440241241420

Ben-Zion, Z., Witte, K., Jagadish, A. K., Duek, O., Harpaz-Rotem, I., Khorsandian, M. C., Burrer, A., Seifritz, E., Homan, P., Schulz, E., & Spiller, T. R. (2025). Assessing and alleviating state anxiety in large language models. *Npj Digital Medicine*, *8*(1), 1–6. https://doi.org/10.1038/s41746-025-01512-6

Benu, N. N., Beeh, N., & Nenotek, S. A. (2025). Implementing Deep Learning in The Efl Classroom: Strategies For Fostering Mindful, Meaningful. *Journal on Language, Education, Literature and Culture*, *3*(1), 70–85. https://doi.org/10.33323/l.v3i1.64

Bouwer, R., van Braak, M., & van der Veen, C. (2024). Dialogic writing in the upper grades of primary school: How to support peer feedback conversations that promote meaningful revisions. *Learning and Instruction*, *93*(July), 101965. https://doi.org/10.1016/j.learninstruc.2024.101965

Chaudhari, Y. S., Chaudhari, M. Y., & Pawar, H. A. (2025). Generative AI in Scholarly Writing: Opportunities and Ethical Dilemmas. *Trends in Scholarly Publishing*, *4*(1), 22–28. https://doi.org/10.21124/tsp.2025.22.28

Cheng, A., Calhoun, A., & Reedy, G. (2025). Artificial intelligence-assisted academic writing: recommendations for ethical use. *Advances in Simulation*, *10*(1), 1–9. https://doi.org/10.1186/s41077-025-00350-6

Córdova-Esparza, D. M. (2025). AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges. *Information (Switzerland)*, *16*(6). https://doi.org/10.3390/info16060469

Deng, C., Duan, Y., Jin, X., Chang, H., Tian, Y., Liu, H., Wang, Y., Gao, K., Zou, H. P., Jin, Y., Xiao, Y., Wu, S., Xie, Z., Lyu, W., He, S., Cheng, L., Wang, H., & Zhuang, J. (2024). Deconstructing The Ethics of Large Language Models from Long-standing Issues to New-emerging Dilemmas: A Survey. *CsCL*, 1–52. http://arxiv.org/abs/2406.05392

Joshy, A., & Sundar, S. (2022). Analyzing the Performance of Sentiment Analysis using BERT, DistilBERT, and RoBERTa. *2022 IEEE International Power and Renewable Energy Conference, IPRECON 2022*, *December 2022*, 1–6. https://doi.org/10.1109/IPRECON55716.2022.10059542

Ke, L., Tong, S., Cheng, P., & Peng, K. (2025). Exploring the frontiers of LLMs in psychological applications: a comprehensive review. *Artificial Intelligence Review*, *58*(10). https://doi.org/10.1007/s10462-025-11297-5

Madeamin, R. (2025). Transformation of Indonesian language learning : Integration of Desfi Yenti, & Roza Susanti. (2025). Students' Writing Anxiety in Academic Writing Practice. *ELP (Journal of English Language Pedagogy)*, *10*(1), 36–44. https://doi.org/10.36665/elp.v10i1.988

Di Pardo Léon-Henri, D. (2024). Multimodal reflective journals and life writing: A didactic approach to enhanced learning. *Innovative Language Teaching Practices in Higher Education in a Post-Covid Era*, 49–72. https://doi.org/10.5334/bdd.d

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00425-2

Fullan, M., & Langworthy, M. (2014). New Pedagogies for Deep Learning Getting Started Guide 2023. *Education in Motion (New Pedagogies for Deep Learning*. https://deep-learning.global/wp-content/uploads/2023/05/New-Pedagogies-for-Deep-Learning-Getting-Started-Guide-2023.pdf

Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: a review. *Artificial Intelligence Review*, *58*(6). https://doi.org/10.1007/s10462-025-11162-5

Guizani, S., Mazhar, T., Shahzad, T., Ahmad, W., Bibi, A., & Hamam, H. (2025). A systematic literature review to implement large language model in higher education: issues and solutions. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00424-7

Jacobsen, L. J., & Weber, K. E. (2025). The Promises and Pitfalls of Large Language Models as Feedback Providers: A Study of Prompt Engineering and the Quality of AI-Driven Feedback. *AI (Switzerland)*, *6*(2). https://doi.org/10.3390/ai6020035

deep learning strategy with character education in the era of independent learning. *IJNI: International Journal of Nusantara Islam*, *13*(2), 229–244. https://doi.org/10.15575/ijni.v13i2.46155

Mah, C., Tan, M., Phalen, L., Sparks, A., & Demszky, D. (2025). From Sentence-Corrections to Deeper Dialogue: Qualitative Insights from LLM and Teacher Feedback on Student Writing. EdWorkingPaper No. 25-1193. *Annenberg Institute for School Reform at Brown University*, *25*. https://www.proquest.com/reports/sentence-corrections-deeper-dialogue-qualitative/docview/3237401553/se-2?accountid=10113%0Ahttps://resolver.ebscohost.com/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/ProQ%3Aeric&rft_val_fmt=info:

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, *6*(January 2024), 100199. https://doi.org/10.1016/j.caeai.2023.100199

Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, *15*(1). https://doi.org/10.1186/s40468-025-00343-2

Molinari, A., & Molinari, E. (2024). the Added Value of Academic Writing Instruction in the Age of Large Language Models: a Critical Analysis. *Iadis International Journal on Www/Internet*, *22*(1), 44–58. https://doi.org/10.33965/ijwi_2024220104

Peláez-Sánchez, I. C., Velarde-Camaqui, D., & Glasserman-Morales, L. D. (2024). The impact of large language models on higher education: exploring the connection between AI and Education 4.0. *Frontiers in Education*, *9*(June). https://doi.org/10.3389/feduc.2024.1392091

Praveen, B., & Abhishek, N. (2024). A Systematic Review on Indian Education System & NEP 2020. *International Journal of Research Publication and Reviews*, *5*(1), 1858–1865. https://doi.org/10.55248/gengpi.5.0124.0228

Prihandoko, L. A., Djatmika, D., Nurkamto, J., Morganna, R., Amalia, S. N., & Ardi, P. (2024). Paving the Road to Writing Success: Self-efficacy, Metacognition, a Growth Mindset, and Academic Writing Performance among Indonesian EFL Students. *Pasaa*, *68*(June), 165–202. https://doi.org/10.58837/chula.pasaa.68.1.5

Rahman, M. M., Shiplu, A. I., Watanobe, Y., & Alam, M. A. (2025). RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Sentiment Analysis. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *14*(8), 1–18. https://doi.org/10.1109/TETCI.2025.3572150

Reddy, G. S. L., Kankhara, M. H., Maheshwari, M., Bansal, S., Kapoor, R., M, H. R., & Kumar, B. (2025). Simulating Emotional Intelligence in LLMs through Behavioral Conditioning and Analogical Retrieval. *Proceedings of the 2nd Workshop on Analogical Abstraction in Cognition, Perception, and Language (Analogy-Angle II)*, 81–91. https://doi.org/10.18653/v1/2025.analogyangle-1.7

Šafranj, J., Gak, D., & Bulatović, V. (2022). Anxiety and self-efficacy in ESP learning: Effects on academic success. *Research in Pedagogy*, *12*(2), 447–460. https://doi.org/10.5937/istrped2202447s

Sun, J., Motevalli, S., Chan, N. N., & Khodi, A. (2024). Implementation of Self-regulated Learning Writing Module: Amplifying Motivation and Mitigating Anxiety among EFL Learners. *Forum for Linguistic Studies*, *6*(3), 89–109. https://doi.org/10.30564/fls.v6i3.6600

Topalov, J., Knežević, L., & Halupka-Rešetar, S. (2023). How Anxious Are Online Esp Learners? Exploring Students' Anxiety in Video, Audio and Text-Based Communication in an Online Classroom. *ESP Today*, *11*(2), 395–416. https://doi.org/10.18485/ESPTODAY.2023.11.2.10

Wang, Q. (2024). EFL learners' motivation and acceptance of using large language models in English academic writing: an extension of the UTAUT model. *Frontiers in Psychology*, *15*(January), 1–12. https://doi.org/10.3389/fpsyg.2024.1514545

Weise, T., Demir, K. C., Pérez-Toro, P. A., Arias-Vergara, T., Maier, A., Nöth, E., Schuster, M., Heismann, B., & Yang, S. H. (2025). Towards End-to-End Speech Articulation and Spoken Language Analysis Using Deep Learning. *Human-Centric Intelligent Systems*, *5*(1), 103–122. https://doi.org/10.1007/s44230-025-00094-6

Xu, Q. A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, *3*(June), 100073. https://doi.org/10.1016/j.dajour.2022.100073

Yu, T. (2025). How to Use LLMs Ethically in Academic Writing? *Education Insights*, *2*(5), 25–39. https://doi.org/10.70088/50wkze06

Yuhua, D. (2024). Integrating Deep Learning into English Language Teaching Within the Digital Cultural Framework. *Computer-Aided Design and Applications*, *21*(S16), 71–84. https://doi.org/10.14733/cadaps.2024.S16.71-84