E-ISSN: 3110-3979, Vol. 2, No. 1, June 2026, pp. 118-137
Journal of Research in English Language Teaching and Linguistics (JRELL)
https://ejournal.earlypublishing.com/index.php/JRELL
Published by Early Publishing Group

THE USE OF CORPUS TOOLS IN TEACHING ACADEMIC WRITING TO EFL STUDENTS

Utami Rosalina[1*] and Davlatova Nargis Dilafruz[3]
[1]Institut Prima Bangsa, Indonesia
[2]Tajik National University, Tajikistan
*Corresponding Author: utamirosalin@ipbcirebon.ac.id
DOI: https://doi.org/10.65431/jrell.v2i1.36

Abstract
The use of corpus tools in teaching academic writing has become increasingly important in supporting EFL learners to produce accurate and contextually appropriate texts. While existing studies highlight the benefits of corpus tools in providing authentic language input and improving linguistic accuracy, limited research has explored how students engage with these tools and how such interaction shapes their writing decisions and awareness. This study aims to examine students' engagement with corpus tools and their influence on academic writing development. Using a qualitative case study design, the study involved 19 EFL students in an advanced academic writing course in Indonesia. Data were collected through students' written texts, reflective surveys, and semi-structured interviews, and analyzed thematically with triangulation to ensure credibility. The findings show that students mainly engage in querying and exploring corpus data to address immediate writing needs, particularly in collocation, vocabulary, and grammar. Corpus use improves vocabulary choice, grammatical accuracy, argument development, and academic tone, while gradually enhancing students' awareness from noticing patterns to reflecting on writing decisions. However, deeper engagement such as application and evaluation remains limited. These results suggest that corpus tools effectively support writing accuracy and awareness, but require structured guidance to promote deeper and more reflective use in academic writing development.

Keywords: Academic writing, corpus tools, EFL students

INTRODUCTION

Academic writing represents a cornerstone of English language learning, particularly for students whose first language is not English. Mastery of formal structures, appropriate vocabulary, and coherent argumentation remains a major challenge for EFL students, often limiting their ability to communicate complex ideas effectively. Advances in technology have introduced various tools to support language development, among which corpus tools have gained prominence. These tools allow learners to analyze authentic language use, identify patterns, and model their writing on real examples, offering a bridge between theory and practice. Using corpus tools in teaching can foster awareness of lexical and syntactic choices, helping students produce writing that is both accurate and contextually appropriate. Despite their growing presence in language instruction, their integration in classroom settings varies widely. This study centers on exploring how corpus tools enhance the teaching and learning of academic writing in EFL contexts.

The importance of equipping students with strong academic writing skills cannot be overstated, as proficiency in this area underpins academic success and professional communication. For EFL learners, challenges include limited exposure to authentic language, insufficient understanding of academic conventions, and difficulty applying theoretical knowledge to practical tasks. Corpus tools can support students in overcoming these difficulties by providing a structured environment for observing patterns, collocations, and frequency of use in scholarly

This work is licensed under CC BY-NC-SA.
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

texts. When students actively engage with corpora, they develop a heightened sense of language norms and stylistic variation, which can strengthen both their confidence and competence. This approach offers opportunities for learners to move beyond prescriptive rules and explore writing as a dynamic process shaped by actual usage. The increasing interest in technology-enhanced language learning underscores the relevance of examining corpus-based teaching strategies.

Recent studies have highlighted the growing role of corpus tools in language instruction, particularly in EFL settings. Hussain (2021) investigated the integration of corpora for improving lexical accuracy and found positive effects on students' vocabulary choices. Kaya (2022) examined the impact of concordance tools on student writing, revealing improved grammatical precision and coherence. Similarly, Götz and Granger, (2024) explored learner engagement with corpora, emphasizing how these tools facilitate self-directed learning and metalinguistic awareness. Research by Alsolami and Alharbi (2020) and Cheng (2021) shows that corpus tools can assist instructors in identifying common learner errors, enabling targeted feedback that aligns with authentic language patterns. These findings indicate a growing acknowledgment of corpus-based methods as valuable pedagogical resources, yet classroom practices remain uneven and underexplored in terms of implementation strategies.

Despite these developments, there remains limited understanding of how EFL students interact with corpus tools in authentic academic writing tasks. Previous research emphasizes vocabulary enhancement or error correction, while deeper insights into how students interpret and apply corpus data to shape their ideas and argumentation are scarce. Few studies have examined the cognitive processes and decision-making students undertake while consulting corpora during writing tasks. The ways in which corpus tools influence rhetorical structure, coherence, and stylistic choices in student-generated texts are still insufficiently examined. This leaves questions about the pedagogical integration of corpora in shaping writing strategies largely unanswered, suggesting the need for closer observation of learners' experiences and practices in classroom contexts.

This study explores the use of corpus tools as a medium for supporting academic writing in EFL classrooms, focusing on how students engage with and interpret corpus information in their written production. Unlike prior research that centers primarily on error correction or vocabulary acquisition, this investigation captures the processes students use to make informed choices in their writing. The study's objectives are to examine students' interaction with corpus tools and understand how this engagement shapes their writing decisions and awareness of academic conventions. By observing and analyzing students' writing practices, the research illuminates the role of corpus tools in bridging the gap between theoretical knowledge and applied language use. The findings provide insights into both student behavior and instructional design, emphasizing the potential of data-driven writing approaches.

The study informs perspectives on language learning and teaching by illustrating how technological resources can reshape writing practices. Insights emerge regarding how learners internalize patterns and structures through corpus consultation, revealing strategies that enhance lexical and syntactic precision. The

findings also highlight instructional techniques that can scaffold students' exploration of authentic texts, thereby strengthening their capacity to produce coherent and persuasive academic writing. Teachers gain visibility into students' decision-making processes, allowing them to tailor guidance and feedback more effectively. This knowledge supports curriculum design that integrates evidence-based tools into writing instruction, fostering active learner participation and critical engagement with texts. The study thus situates corpus use within a broader understanding of student-centered, process-oriented writing development.

This research provides a detailed examination of how EFL students interact with corpus tools and apply findings to their academic writing. It presents evidence on student practices and instructional strategies that demonstrate the practical utility of corpus-based methods. Observations of learners' engagement with authentic language data offer guidance for educators seeking to implement data-driven teaching approaches. The study presents an opportunity to understand both the cognitive and procedural aspects of academic writing development, illustrating methods for enhancing writing proficiency through corpus use.

Research questions:

1. How do EFL students engage with corpus tools while completing academic writing tasks?
2. In what ways does interaction with corpus tools influence students' decisions and awareness in constructing academic texts?

LITERATURE REVIEW

Corpus tools have become increasingly recognized as essential instruments in supporting EFL students' academic writing, offering insights that traditional textbooks cannot provide. By exposing learners to authentic examples of language in use, these tools allow students to observe how words, phrases, and grammatical structures naturally co-occur (Alsehibany & Abdelhalim, 2023; Kaya, 2022; Luo & Zhou, 2017). Resources such as COCA, BNC, and specialized concordancers like AntConc provide windows into real academic discourse, making abstract rules tangible through patterns that learners can analyze and apply. Such interaction encourages learners to move beyond memorization, fostering intuitive understanding of language structures, which is critical for producing fluent and accurate texts. Empirical studies show that learners who actively engage with corpora refine their writing strategies and develop confidence in constructing academic texts (Alsehibany & Abdelhalim, 2023; Alsolami & Alharbi, 2020; Birhan et al., 2021). These experiences cultivate both precision and ownership in language use. By working with real language data, students begin to internalize patterns naturally, strengthening their ability to write independently while gradually gaining confidence in their own linguistic choices.

Building on this foundation, corpus-based activities demonstrate clear benefits for linguistic precision, particularly in vocabulary, collocations, and grammar. Learners analyzing concordance lines and frequency information make more informed choices and expand their lexical range (Ma et al., 2022; Simpson et al., 2020; Wibowo & Charlotte, 2025). Corpus-informed exercises encourage

learners to analyze patterns, test hypotheses, and engage critically with language, rather than applying rules mechanically (Altun, 2021; Fredrick et al., 2020; Simpson et al., 2020). Repeated exposure to authentic examples strengthens grammatical awareness, enabling learners to notice subtle variations in sentence construction and cohesion (Alzahrani, 2024; Nanyinza, 2023; Solano, 2020). Such practices also enhance reflection, prompting students to actively monitor patterns and adjust their writing accordingly. Observations suggest that corpus use fosters autonomy, as students evaluate and revise texts independently (Fatoni et al., 2025; Huo et al., 2021; Qoura et al., 2018). The integration of corpus activities thus transforms the way learners approach language, combining technical accuracy with analytical thinking, and reinforcing self-directed engagement.

This analytical engagement extends naturally into revision and writing strategies. Corpus consultation during editing allows learners to identify coherence issues, refine lexical choice, and improve rhetorical flow (Birhan & Nurie, 2024; Isbahi & Mendoza, 2025; Yang & Ren, 2025). Students who examine concordance lines for revision are more capable of self-correcting and strengthening argumentation. Analyzing patterns critically supports reflective decision-making, reducing dependence on teacher guidance (Jashari, 2026; Lannin & Townsend, 2020). This process fosters metacognitive awareness, as learners monitor their choices and evaluate outcomes. Over time, these practices cultivate independence and confidence in text production (Karlinda & Srivastava, 2025; Ramadhanti & Yanda, 2021; Rimun & Yumarnamto, 2024). Through corpus engagement, students begin to approach writing as a deliberate process, treating each revision as an opportunity for discovery and refinement. They develop a sense of agency that extends beyond grammar into the strategic organization of ideas.

The type of corpus tool and its pedagogical framing strongly influence learning outcomes, creating a bridge between observation and application. Hands-on tools like AntConc allow learners to explore data directly, whereas hands-off methods rely on curated examples (Kaya, 2022; Liang et al., 2022). Direct interaction fosters experimentation and pattern recognition, deepening engagement with the material (Çyfeku, 2023; Hussain, 2021; Rahayu & Rostika, 2025). Guidance remains crucial, as students require support to interpret outputs accurately and avoid misapplication (Adhikari & Jordan, 2021). Integrating corpus activities into meaningful writing tasks ensures that learners engage cognitively while avoiding frustration. The choice of tools and the manner in which they are embedded in instruction shapes the depth of learning. Well-designed tasks enable students to explore language patterns confidently, enhancing both motivation and understanding.

Learner perception plays a pivotal role in shaping how corpus tools are experienced and applied. Many students find corpus data more authentic and relevant than traditional textbook examples, increasing both engagement and confidence (Falcone, 2025; Yoon & Hirvela, 2004). Initial challenges with

unfamiliar interfaces are common, yet structured support, scaffolded exercises, and gradual exposure help learners overcome these obstacles (Ma et al., 2022; Tosun & Sofu, 2023). As familiarity grows, students move from passive observation to active incorporation of patterns into their own writing (Yunus et al., 2024; Zhou & Preez, 2025). This evolution illustrates that learners' attitudes are closely intertwined with performance; when learners feel capable and supported, they engage more deeply and apply corpus findings meaningfully in their texts. Confidence, skill, and engagement develop hand in hand.

Corpus consultation also strengthens understanding of academic genres and discourse conventions, creating a natural link to writing development. Examining metadiscourse markers, cohesive devices, and academic phraseology through corpus data helps learners internalize structural patterns (Brooks-Gillies et al., 2020; Lannin & Townsend, 2020). Students exposed to corpus-informed tasks demonstrate improved coherence in introductions, transitions, and conclusions. Concordancers reveal frequency and context, guiding learners' lexical and grammatical choices (Birhan et al., 2021; Fredrick et al., 2020). These observations challenge learners to interpret patterns critically and integrate them purposefully. Over time, students internalize genre norms, producing texts that are grammatically accurate and rhetorically coherent (David & Anderson, 2022; Nagao, 2019). Repeated corpus engagement fosters an intuitive understanding of academic conventions. Learners gradually write with greater clarity, organization, and rhetorical awareness, bridging the gap between observation and creation.

Challenges in corpus-based learning emerge as part of the natural learning process. Students with lower proficiency may struggle to interpret concordance outputs, and insufficient guidance can limit meaningful engagement (Ma & Mei, 2021; Youssef, 2020). Time constraints and curriculum design also influence how deeply learners can explore data (Uçar & Zarfsaz, 2022; Youssef, 2020). Long-term, consistent engagement is essential for noticeable improvements, as brief exposure tends to produce only minor gains (Bao, 2021; Steen et al., 2019). Addressing these challenges requires careful preparation and ongoing support. When learners receive appropriate scaffolding, they can navigate difficulties successfully, gradually building both skill and confidence in their writing.

Emerging hybrid approaches, integrating corpus tools with adaptive platforms and automated feedback, further enhance academic writing instruction. These models reduce cognitive load while directing attention to key language patterns and structures (Curry & Riordan, 2021; Khojasteh et al., 2025). Extended corpus use promotes autonomy and self-regulation in writing (Alsolami & Alharbi, 2020; Curry & Riordan, 2021). Teacher development is essential to ensure effective classroom integration and maximize benefits. The evolving research shows that corpus tools, when thoughtfully combined with pedagogical strategies, support sustainable skill development. Students gain both immediate writing improvements and long-term strategies for independent, reflective, and confident academic writing.

METHOD

Research Design

This study employed a qualitative case study design, which focuses on an in-depth exploration of a bounded system—in this case, the interaction of EFL students with corpus tools during academic writing tasks. The case study was designed to capture the complexity of learners' engagement, decision-making, and evolving awareness in an authentic educational context. Specifically, the study treated each student's engagement with corpus tools as a unit of analysis, allowing for detailed examination of strategies, reflections, and writing outcomes. Multiple data sources were integrated to enrich understanding, including students' written texts, reflective surveys, and semi-structured interviews. This approach enables a comprehensive, context-sensitive investigation, revealing patterns and processes that are often overlooked in purely quantitative designs. The design also emphasizes exploration of the "how" and "why" of learners' behaviors, providing insight into not only what students do, but how corpus tool interaction shapes their academic writing practices. By systematically analyzing these processes, the case study allows researchers to construct a holistic and nuanced understanding of learners' engagement within a bounded academic setting.

Participants

The study involved 19 EFL students enrolled in an advanced academic writing course at a higher education institution in Indonesia. Participants were selected based on their prior exposure to academic writing and their willingness to engage fully with corpus tools during the study. The students represented diverse academic disciplines, which allowed the investigation of corpus use across varied writing contexts while maintaining comparable proficiency levels. Demographic information, such as age, study background, and prior experience with digital resources, was recorded to provide context for interpreting the findings. All participants were involved in every stage of the study, ensuring a complete and consistent dataset for analysis.

Instruments

Data collection employed instruments that captured both observable behaviors and reflective insights. Participants used corpus tools, including AntConc and COCA, to analyze authentic language data, identify collocations, and examine grammatical structures while completing academic writing tasks. Reflective surveys were administered after each task to gather immediate responses regarding engagement, strategies, and awareness development. Additionally, semi-structured interviews were conducted to explore participants' reasoning, decision-making processes, and perceptions of corpus-assisted writing in greater depth. This combination of instruments enabled triangulation, providing a detailed and credible understanding of learners' interactions with corpus tools.

Data Collection

Data collection followed a structured sequence designed to capture rich information on students' engagement. Initially, participants received guidance on corpus tool functionalities and instructions for completing the writing tasks. They

then completed three academic writing tasks, each requiring active consultation of corpus resources, with opportunities to verbalize or document their reasoning during the process. After each task, reflective surveys were completed to capture participants' immediate thoughts on strategies and language choices. Finally, semi-structured interviews were conducted to explore patterns of engagement, challenges, and cognitive processes. All interviews were audio-recorded and transcribed for detailed analysis, ensuring that each of the 19 participants provided a complete set of behavioral and reflective data.

Data Analysis

Data analysis involved examining patterns of corpus engagement, decision-making, and awareness development. Students' written texts were analyzed for corpus-informed choices, such as lexical selection, collocations, and structural revisions. Survey and interview data were coded thematically, with multiple iterations refining initial codes into categories that captured strategies, challenges, and cognitive processes. Triangulation across writing tasks, surveys, and interviews ensured that observed behaviors aligned with participants' reported reasoning. The analysis focused on understanding how corpus interactions shaped learners' awareness and influenced their writing decisions. By integrating multiple data sources, the study provided a holistic view of learners' engagement with corpus tools, highlighting both the practical and cognitive dimensions of academic writing development.

FINDINGS

RQ 1: How do EFL students engage with corpus tools while completing academic writing tasks?

Figure 1 illustrates a clear progression of students' engagement with corpus tools, starting from query (31%), which represents the most dominant activity, indicating that students primarily rely on entering keywords or phrases to address immediate linguistic needs. This is followed by exploration (27%), where learners examine concordance lines, showing a strong tendency to observe authentic language use before making decisions. The analysis stage (21%) reflects a moderate level of deeper engagement, as students begin identifying patterns and language features, suggesting that not all learners move beyond surface-level consultation. A noticeable decline appears in the apply stage (15%), indicating that fewer students consistently transfer corpus findings into their own writing. The lowest proportion is seen in evaluation (6%), which shows that only a small number of students revisit the corpus to check and revise their work. This distribution highlights that while students engage in initial stages of corpus use, fewer progress toward higher-order processes such as application and evaluation, suggesting the need for stronger pedagogical support to encourage deeper and more sustained engagement.

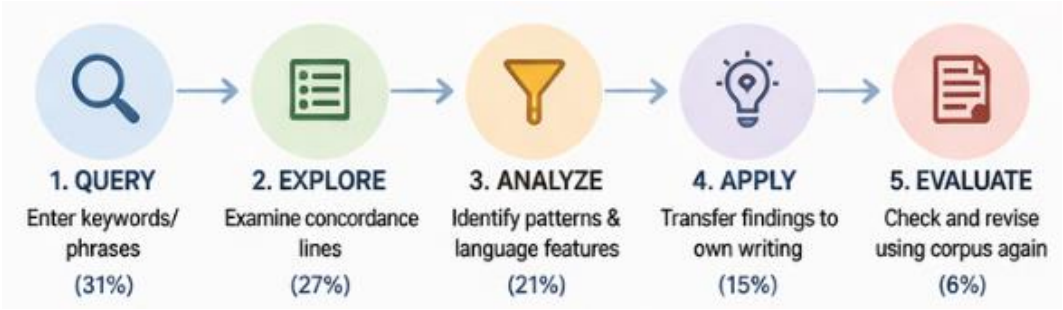


Figure 1. Engagement phases in using corpus tools

Figure 2 illustrates the range of students' engagement with specific corpus tool features, with viewing concordance lines (92%) emerging as the most dominant activity, indicating that nearly all students depend on direct observation of authentic language examples as their primary strategy when completing writing tasks. This is followed by examining multiple examples (76%), showing that a large proportion of learners attempt to verify patterns by comparing several instances before making decisions. A moderate level of technical interaction is reflected in using filters such as part of speech, genre, or collocation (58%), suggesting that more than half of the students begin to explore advanced corpus functions. However, deeper analytical practices appear less frequent, as seen in comparing word or phrase alternatives (47%) and checking frequency information (41%), indicating that fewer students engage in critical evaluation of language data. The least common activity is saving or noting results (35%), which points to limited engagement in reflective or documentation practices. This overall distribution suggests that students use corpus tools, but their engagement is largely concentrated on surface-level observation and immediate problem-solving rather than systematic analysis and reflection. It also reveals a gradual decline in participation as tasks require higher-order thinking, indicating that while students are comfortable accessing corpus data, they are less inclined to process it deeply or reuse it strategically in their writing.

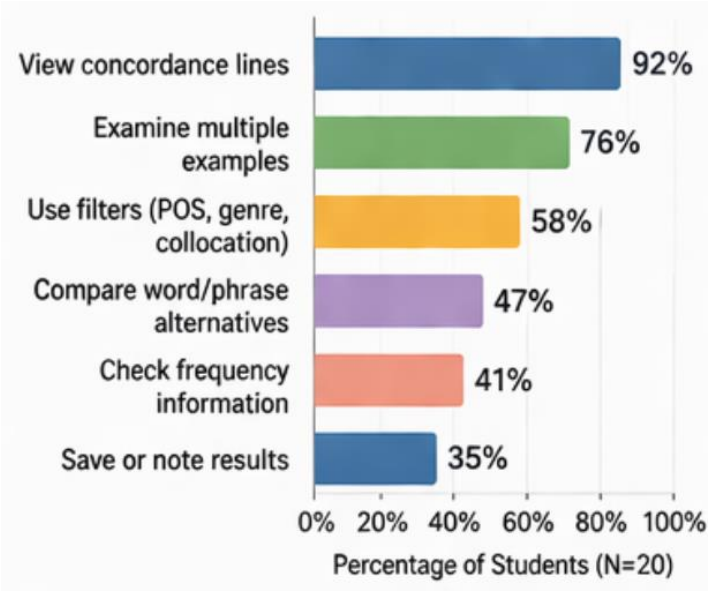


Figure 2. Frequency of actions performed

Figure 3 presents the distribution of corpus tools used by students, with AntConc accounting for 45%, making it the most frequently utilized tool among participants. This indicates that nearly half of the students prefer AntConc, likely due to its accessibility, flexibility, and ease of use in analyzing concordance lines. The second most commonly used tool is COCA (30%), showing that a substantial proportion of students rely on large, structured corpora to explore authentic academic language patterns. Meanwhile, BNC represents 15% of usage, suggesting a more limited but still relevant role in supporting students' writing, possibly due to differences in interface familiarity or corpus scope. The remaining 10% falls under other tools, indicating that only a small number of students experiment with alternative corpus resources. This distribution reveals that students tend to concentrate on a limited range of familiar tools rather than diversifying their corpus use. It also suggests that accessibility and usability strongly influence tool preference, with AntConc emerging as the dominant choice for hands-on interaction. At the same time, the lower percentages for BNC and other tools indicate that students may not fully explore the variety of corpus resources available, which could limit their exposure to different language datasets and analytical possibilities.

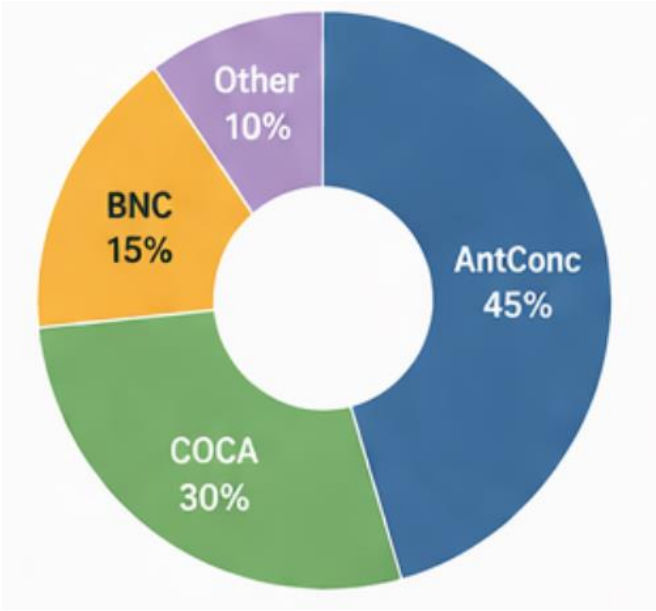


Figure 3. Corpus tools most frequently used

Figure 4 illustrates how students use corpus tools to support specific academic writing purposes, with checking collocations or word partnerships (85%) emerging as the most dominant activity. This indicates that a large majority of students rely on corpus data to ensure that their word combinations are natural and contextually appropriate. This is followed by finding appropriate academic phrases (70%), suggesting that many learners use corpus tools to enhance the formality and precision of their writing. Additionally, verifying grammar patterns (60%) shows that more than half of the students consult corpus resources to confirm correct grammatical usage. In contrast, more analytical and comparative activities such as comparing synonyms or alternatives (45%) and understanding register or usage

(40%) are less frequent, indicating that fewer students engage with deeper aspects of language variation and contextual appropriateness.

This distribution suggests that students primarily use corpus tools for immediate and practical writing needs, particularly in ensuring lexical accuracy and appropriateness. The high percentage for collocation checking (85%) reflects a strong awareness of the importance of natural language use in academic writing, while the relatively lower engagement in synonym comparison and register understanding indicates that higher-level linguistic sensitivity develops more gradually. These patterns also reveal that students tend to focus on correctness and phrase selection rather than exploring subtle differences in meaning or stylistic variation. As a result, corpus tools function mainly as a support for refining language accuracy, while their potential to enhance deeper rhetorical and stylistic awareness is not yet fully utilized by a significant portion of learners.

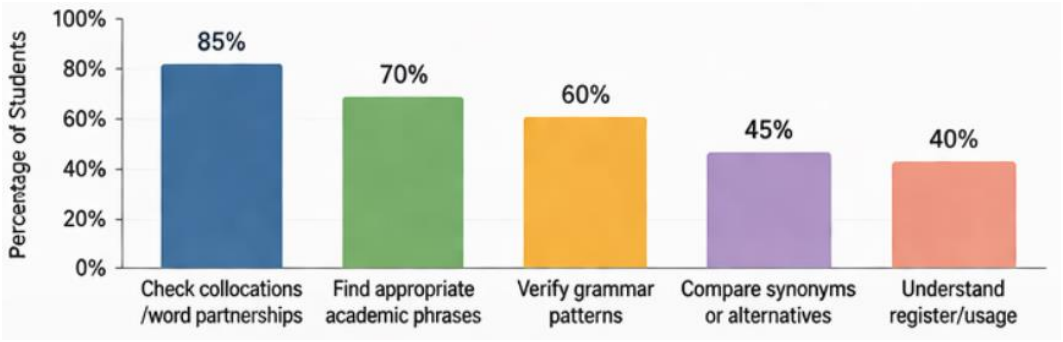


Figure 4. Purpose for using corpus tools

Across the four figures, EFL students' engagement with corpus tools can be characterized as active, frequent, and primarily oriented toward immediate linguistic needs during academic writing tasks. The process begins with a strong emphasis on initial consultation, where 31% of activities involve query formulation, followed by 27% exploration and 21% analysis, indicating that most students engage at the early and middle stages of corpus interaction. This pattern is reinforced by tool usage data, where AntConc (45%) and COCA (30%) dominate, suggesting a preference for accessible and familiar platforms. In terms of specific behaviors, 92% of students view concordance lines and 76% examine multiple examples, showing that observation and pattern exposure are central to their engagement. Furthermore, task-oriented use is evident in writing purposes, where 85% check collocations and 70% search for appropriate academic phrases, confirming that corpus tools are mainly used to support vocabulary accuracy and natural expression during writing.

However, engagement tends to decline as tasks require deeper cognitive processing and reflective practices. While 60% of students verify grammar patterns, only 47% compare alternatives, 41% check frequency information, and 35% document findings, indicating limited movement toward analytical and reflective use. Similarly, higher-order stages such as application (15%) and evaluation (6%) are significantly less frequent, suggesting that fewer students fully integrate corpus insights into revision and critical review processes. This trend reveals that students predominantly engage with corpus tools as problem-solving resources during drafting rather than as tools for systematic refinement and reflection. Overall, EFL

students demonstrate strong initial engagement and practical use of corpus tools, yet their interaction remains largely surface-level, with fewer learners progressing toward deeper, more strategic, and sustained engagement throughout the entire writing process.

RQ2: In what ways does interaction with corpus tools influence students' decisions and awareness in constructing academic texts?

Figure 5 illustrates the influence of corpus tools on different aspects of students' academic writing, with improved vocabulary and collocation choices (85%) emerging as the most prominent outcome. This indicates that a large majority of students benefit from corpus use in selecting more precise and contextually appropriate words. This is followed by more accurate grammatical structures (70%), showing that corpus consultation also supports learners in refining sentence correctness. Additionally, 55% of students demonstrate stronger support for argumentation, suggesting that corpus tools contribute not only to language accuracy but also to the development of ideas and content. Improvements in academic tone (50%) further indicate that students become more aware of formal writing conventions, while greater variety in sentence structure (40%) appears as the least developed aspect, reflecting a more gradual development of structural flexibility.

This distribution suggests that corpus tools have the strongest impact on lower-level linguistic features, particularly vocabulary and grammar, which are more immediately accessible through concordance analysis. The moderate percentages in argumentation (55%) and academic tone (50%) indicate that students begin to extend corpus insights beyond surface-level corrections into broader aspects of writing, although this development is not yet consistent across all learners. The relatively lower figure for sentence variety (40%) suggests that more complex structural changes require deeper engagement and sustained practice. Overall, the pattern shows that while corpus tools effectively support accuracy and appropriateness in writing, their potential to enhance higher-level writing skills, such as stylistic variation and rhetorical development, is still developing among students.

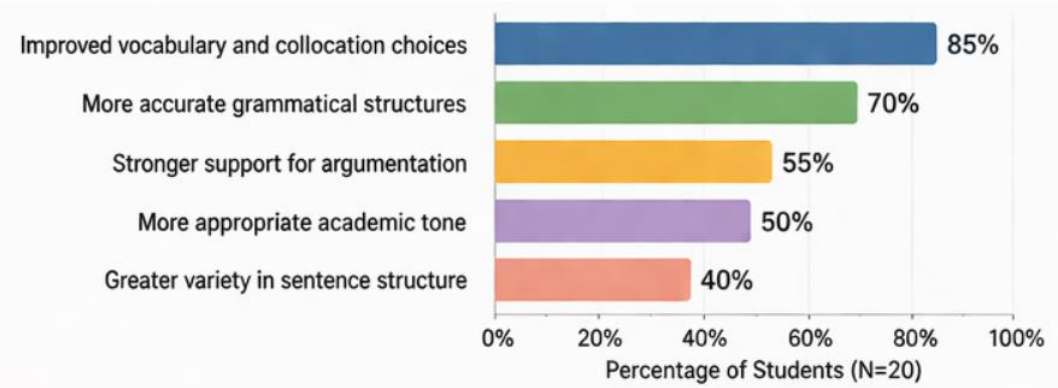


Figure 5. How corpus use influences writing decisions

Figure 6 captures how students' awareness gradually unfolds as they engage with corpus tools, beginning from noticing patterns and moving toward deeper

reflection. The highest proportion appears in awareness of usage patterns (90%), showing that most students become highly attentive to how words and structures function in real academic texts. This sensitivity then develops into awareness of context (75%), where learners begin to understand that meaning is not fixed but shifts depending on how and where language is used. A further step is seen in awareness of appropriateness (65%), indicating that many students start to judge whether certain expressions fit academic writing conventions. The progression reaches its most complex level in metacognitive awareness (55%), where just over half of the students actively reflect on and monitor their own writing decisions. The steady decrease across these stages suggests that while noticing is widely achieved, deeper reflection requires more effort and experience.

Looking more closely, this pattern feels less like a gap and more like a natural learning journey. Students first become observers of language, then gradually turn into users who make more informed choices, and eventually some begin to question and refine their own decisions. The drop from 90% to 55% reflects how challenging it is to move from simply recognizing patterns to actually thinking critically about one's own writing. The fact that 65% reach the stage of appropriateness shows promising progress, as students begin to align their writing with academic expectations. However, the lower level of metacognitive awareness suggests that many learners are still developing confidence in evaluating their own work independently. In this sense, corpus tools do more than provide answers—they slowly reshape how students see language, encouraging them to move from noticing to understanding, and eventually to self-awareness in writing.

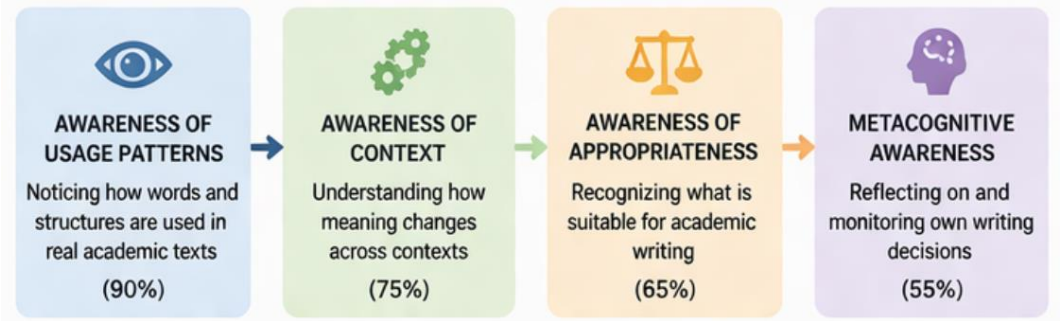


Figure 6. Development of students' awareness

Table 1 below offers a close look at how students' writing begins to change in subtle but meaningful ways after working with corpus tools. In collocation, what starts as a single, somewhat rigid choice like *"make a decision"* expands into more natural and flexible options such as *"make a decision / reach a decision,"* showing that students are no longer settling for the first phrase that comes to mind. In terms of academic phrases, expressions like *"many important factors"* shift to *"a number of important factors,"* which carries a more formal and polished tone. The change in grammar may look small on the surface, but correcting forms like *"Researchers have found that…"* reflects a growing attention to accuracy and detail. Perhaps the most noticeable transformation appears in tone, where students move from personal statements like *"I think it is good because…"* to more academically appropriate forms such as *"It can be argued that…"*. These shifts show that students are not

just editing their sentences, but rethinking how their ideas should sound in an academic context.

Table 1. Influence on writing decisions

Aspect	Before Using Corpus	After Using Corpus	Impact
Collocation	make a decision	make a decision / reach a decision	Chose more natural collocation
Academic Phrase	many important factors	a number of important factors	Used more formal academic phrase
Grammar	*Researchers have found that...	Researchers have found that...	Corrected subject-verb agreement
Tone	I think it is good because...	It can be argued that...	Adopted more academic tone

What makes these changes interesting is how they reflect a quiet shift in mindset rather than just technical improvement. Before using corpus tools, students often rely on habit, writing in ways that feel familiar but not always suitable for academic purposes. After interacting with real language data, their choices begin to feel more intentional, as if they are weighing options rather than guessing. The movement from informal to formal tone suggests that students are starting to see writing not just as expression, but as positioning their ideas within a wider academic conversation. Even small adjustments, like choosing a different phrase or restructuring a sentence, show growing awareness of audience and purpose. In this sense, corpus tools do not simply correct language—they gently guide students toward thinking more carefully about how their writing is perceived, helping them develop a more confident and mature academic voice.

DISCUSSION

Students' tendency to remain at the stages of querying and exploring reflects how they approach corpus tools primarily as immediate problem-solving resources during writing. At this stage, learners are often concerned with finding quick answers to lexical or grammatical uncertainties rather than engaging in deeper analysis of language patterns. This behavior aligns with the notion that exposure to authentic data initially supports noticing, as emphasized by Alsehibany and Abdelhalim (2023) and Luo and Zhou (2017), where learners begin by observing how language is used in real contexts. However, the limited transition toward application and evaluation indicates that engagement does not automatically extend into higher-order thinking. This extends earlier findings by suggesting that access to authentic input alone is not sufficient to promote deeper learning. Learners appear to require structured opportunities to move beyond observation and into reflection. The pattern also reflects how writing tasks often encourage efficiency, leading students to prioritize immediate solutions over extended analysis. As a result, corpus tools are used more as references than as spaces for deeper exploration.

This concentration on early-stage engagement can also be linked to learners' developing analytical abilities when interacting with complex linguistic data.

Corpus tools present dense and often unfamiliar information, which can be cognitively demanding for learners who are still building their linguistic competence. In such situations, students tend to rely on surface-level strategies that feel more manageable, such as identifying familiar patterns. While studies like Altun (2021) and Simpson et al. (2020) emphasize that corpus-based learning promotes analytical thinking, the present findings suggest that this potential is not consistently realized. Instead, analytical engagement appears to depend on learners' readiness and experience. Without sufficient guidance, students may not feel confident in interpreting patterns or drawing conclusions from data. This creates a gap between the theoretical benefits of corpus tools and their practical use in classrooms. The findings therefore extend previous research by showing that analytical skills develop gradually through repeated exposure. Over time, learners may begin to engage more critically, though this process requires sustained interaction and instructional support.

The strong reliance on viewing concordance lines and examining multiple examples highlights the central role of noticing in corpus-based learning. Learners appear to depend heavily on visual exposure to authentic language as a way of building understanding. This aligns with research suggesting that repeated encounters with patterns help learners internalize vocabulary and grammatical structures (Kaya, 2022; Luo & Zhou, 2017). Through this process, students begin to recognize how words function in context, even if they are not yet analyzing them deeply. However, the limited use of features such as frequency analysis and alternative comparison suggests that learners are not fully engaging in hypothesis testing. This contrasts with studies by Simpson et al. (2020), which describe corpus use as a more investigative and evaluative activity. The difference may stem from variations in training or task design, where learners are not always encouraged to explore beyond observation. As a result, corpus interaction remains largely input-driven rather than inquiry-driven. This indicates that noticing serves as a foundation, though not the endpoint, of corpus-based learning.

Students' preference for a small number of familiar corpus tools reflects the role of usability and accessibility in shaping engagement. Tools such as AntConc and COCA provide straightforward interfaces and quick access to examples, which makes them appealing to learners who seek efficiency. This supports the perspective of Kaya (2022) and Liang et al. (2022), who highlight the importance of hands-on tools in facilitating interaction. The reliance on familiar platforms suggests that learners may prioritize comfort over exploration. This behavior extends previous findings by showing that accessibility can both support and limit learning opportunities. When students repeatedly use the same tools, they may not encounter diverse types of language data. This can restrict their exposure to different registers, genres, and linguistic variations. The findings also suggest that learners' choices are influenced not only by functionality but also by confidence and familiarity. In this sense, tool selection becomes part of the learning process itself.

The strong focus on collocations, academic phrases, and grammar demonstrates that students use corpus tools primarily to address immediate linguistic concerns in writing. This pattern is consistent with studies by Ma et al. (2022) and Wibowo and Charlotte (2025), which show that corpus tools effectively support lexical and grammatical accuracy. Learners appear to turn to corpora when

they need confirmation about how language should be used correctly. This reflects a practical orientation toward writing, where accuracy is prioritized over exploration. The lower engagement with register and variation suggests that learners are less focused on stylistic and contextual nuances. This adds depth to previous research by indicating that not all aspects of language benefit equally from corpus use. While accuracy improves, sensitivity to variation develops more slowly. The findings therefore highlight a layered process of learning, where basic features are addressed first before more complex aspects emerge.

The improvements in vocabulary choice and grammatical accuracy reflect how repeated exposure to authentic data shapes learners' linguistic development. Through interaction with corpus examples, students begin to notice patterns that guide their language use. This aligns with findings by Alzahrani (2024) and Nanyinza (2023), who emphasize the role of corpus tools in enhancing precision. Over time, learners move from uncertainty toward more confident decision-making in writing. This process illustrates how input gradually becomes internalized knowledge. Corpus tools act as mediators between observation and production. They provide learners with evidence that supports their choices, reducing reliance on guesswork. The findings also suggest that improvement is not immediate but accumulates through repeated use. As learners engage more frequently, their writing becomes more accurate and controlled.

The development of argumentation and academic tone indicates that corpus tools also influence higher-level writing features. Learners begin to adopt more formal expressions and structure their ideas in ways that align with academic conventions. This supports research by Birhan and Nurie (2024) and Yang and Ren (2025), which highlights the role of corpus consultation in shaping rhetorical awareness. However, the progression in these areas appears less consistent than improvements in vocabulary and grammar. This suggests that higher-level skills require more sustained engagement and deeper processing. While corpus tools provide useful input, learners must actively interpret and apply patterns to their own writing. The findings extend previous studies by showing that rhetorical development is gradual and uneven. Some learners begin to integrate these features more effectively than others. This variation reflects differences in experience, practice, and confidence.

Students' awareness develops in a gradual and layered manner, moving from noticing patterns to understanding context and appropriateness. This progression reflects the learning trajectory described by Yoon and Hirvela (2004) and Falcone (2025), where exposure leads to increased sensitivity to language use. At the initial stage, learners focus on recognizing recurring forms and structures. As they gain experience, they begin to consider how meaning changes depending on context. This shift marks an important step toward more informed language use. The emergence of metacognitive awareness, although less widespread, suggests that some learners begin to reflect on their own writing decisions. This indicates a move toward greater independence in learning. Reaching this level appears challenging and requires continued practice and support. The findings highlight that awareness evolves progressively rather than appearing instantly.

The observable shifts in writing, such as more natural collocations and more formal expressions, illustrate how corpus interaction reshapes learners' approach to writing. Students begin to move away from relying on intuition and instead make

choices based on evidence. This aligns with research on genre awareness and discourse competence (Brooks-Gillies et al., 2020; David & Anderson, 2022), which emphasizes the importance of exposure to authentic academic language. Writing becomes a more deliberate and reflective process. Learners start to consider how their language choices align with academic expectations. These changes may appear subtle, but they signal a deeper transformation in how students perceive writing. The findings suggest that corpus tools contribute not only to accuracy but also to a shift in mindset. Over time, learners develop a stronger sense of control over their writing.

The findings suggest that corpus tools play a significant role in shaping students' writing development, though their impact is closely tied to how they are integrated into learning contexts. Previous studies emphasize autonomy and self-directed learning (Alsolami & Alharbi, 2020; Curry & Riordan, 2021), yet the present results indicate that such outcomes emerge gradually rather than immediately. Learners require guidance, practice, and time to fully benefit from corpus interaction. Without these elements, engagement may remain at a surface level. The findings extend existing literature by highlighting the importance of pedagogical support in maximizing the potential of corpus tools. They also show that learning is not a linear process but involves stages of development. Corpus tools function both as practical resources and as catalysts for deeper learning. Their value lies not only in providing answers but in shaping how learners engage with language over time.

CONCLUSION

The findings show that EFL students mainly engage with corpus tools at the stages of querying and exploring, focusing on observing concordance lines and reviewing examples to solve immediate writing problems. Their use is largely practical, with limited involvement in applying and evaluating corpus results and less frequent use of advanced features. Students rely on a few familiar tools to check collocations, select academic phrases, and verify grammar, while giving less attention to register and variation. Interaction with corpus tools improves vocabulary choice, grammatical accuracy, argument development, and academic tone, and supports a gradual development of awareness from noticing patterns to reflecting on writing decisions. These results indicate that corpus tools effectively support accuracy and awareness, though deeper analytical engagement requires more sustained practice and guidance.

This study has several limitations related to its design and scope. The small sample size and focus on one institution limit the generalizability of the findings. The use of self-reported data may introduce subjectivity, and the short duration of the study may not capture long-term development. In addition, the focus on specific corpus tools narrows the range of possible outcomes. Future research should involve larger and more diverse samples, adopt longitudinal designs, and combine qualitative and quantitative methods. Further studies could also examine different instructional approaches to support deeper engagement with corpus tools.

References

Adhikari, A., & Jordan, M. I. (2021). Interleaving computational and inferential thinking in an undergraduate data science curriculum. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.cb0fa8d2

Alsehibany, R. A., & Abdelhalim, S. M. (2023). Enhancing academic writing vocabulary use through direct corpus consultation: Saudi English majors' perceptions and experiences. *Arab World English Journal*, *14*(1), 166–182. https://doi.org/10.24093/awej/vol14no1.11

Alsolami, T., & Alharbi, A. (2020). Saudi EFL learners' perceptions of the use of corpora in academic writing teaching. *Studies in English Language Teaching*, *8*(4). https://doi.org/10.22158/selt.v8n4p94

Altun, H. (2021). The learning effect of corpora on strong and weak collocations: Implications for corpus-based assessment of collocation competence. *International Journal of Assessment Tools in Education*, *8*(3), 509–526. https://doi.org/10.21449/ijate.845051

Alzahrani, I. H. (2024). Exploring learners' beliefs on grammar learning: Importance and preferred methods. *Theory and Practice in Language Studies*, *14*(5), 1475–1485. https://doi.org/10.17507/tpls.1405.20

Bao, K. (2021). Should we use it in our classrooms: An analysis of data-driven learning research. *English Linguistics Research*, *10*(3), 66–66. https://doi.org/10.5430/elr.v10n3p66

Birhan, A. T., & Nurie, Y. (2024). Developing engineering students' engagement in academic writing classes using corpus-based instruction. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1). https://doi.org/10.1186/s40862-023-00232-2

Birhan, A. T., Teka, M., & Asrade, N. (2021). Effects of using corpus-based instructional mediation on EFL students' academic writing skills improvement. *Theory and Practice of Second Language Acquisition*, *7*(2), 133–153. https://doi.org/10.31261/tapsla.9993

Brooks-Gillies, M., García, E. G.-B., & Kim, S. H. (2020). Introduction: Graduate writing across the disciplines. In *The WAC Clearinghouse; University Press of Colorado eBooks* (pp. 3–19). https://doi.org/10.37514/atd-b.2020.0407.1.3

Cheng, Y.-H. (2021). EFL college students' concordance for error correction. *English Teaching & Learning*, *45*(4), 431–460. https://doi.org/10.1007/s42321-021-00075-5

Curry, N., & Riordan, E. (2021). Intelligent CALL systems for writing development. In *Advances in educational technologies and instructional design book series* (pp. 252–273). IGI Global. https://doi.org/10.4018/978-1-7998-6609-1.ch011

Çyfeku, J. (2023). Alternative flipped learning approach scaffolding student engagement and collaboration: A case study with ELT master students in "Corpora in Applied Linguistics" Subject. *International Journal of Management Knowledge and Learning*, *12*. https://doi.org/10.53615/2232-5697.12.s89-100

David, R. D., & Anderson, C. E. (2022). The universal genre sphere: A curricular model integrating GBA and UDL to promote equitable academic writing

instruction for EAL university students. *Center for Educational Policy Studies Journal*, *12*(4), 35–52. https://doi.org/10.26529/cepsj.1442

Falcone, M. (2025). Integrating corpus linguistics and extended reality for ESP learning material design. *International Journal of English Linguistics*, *15*(5), 1–1. https://doi.org/10.5539/ijel.v15n5p1

Fatoni, M., Wardhono, A., Min-ho, K., & Liu, M. C. (2025). The effectiveness of using Quizizz in improving students' grammar accuracy. *Journal of Research in English Language Teaching and Linguistics*, *1*(1), 136–151. https://doi.org/10.65431/jrell.v1i1.13

Fredrick, T., Stravelli, K., May, S., & Brookman-Smith, J. (2020). The space between MA Students enculturate to graduate reading and writing. In *The WAC Clearinghouse; University Press of Colorado eBooks* (pp. 143–165). https://doi.org/10.37514/atd-b.2020.0407.2.06

Götz, S., & Granger, S. (2024). Learner corpus research for pedagogical purposes. *International Journal of Learner Corpus Research*, *10*(1), 1–38. https://doi.org/10.1075/ijlcr.00039.got

Huo, C., Jiao, L., & Wang, W. (2021). An empirical study of corpus application in the teaching of English writing. *Creative Education*, *12*(10), 2464–2470. https://doi.org/10.4236/ce.2021.1210184

Hussain, Z. D. (2021). Incorporating digital corpus analysis technology into english for academic purposes (EAP) lesson design for the Sri Lankan tertiary context: A self-experimentation by a 'Novice.' *Advances in Technology*, 483–508. https://doi.org/10.31357/ait.v1i3.5381

Isbahi, B., & Mendoza, E. (2025). A comparative study of grammar achievement between male and female students. *Journal of Research in English Language Teaching and Linguistics*, *1*(1), 119–135. https://doi.org/10.65431/jrell.v1i1.8

Jashari, A. (2026). Enhancing coherence and persuasiveness in academic discourse through lexical cohesion and critical thinking. *Open Journal of Social Sciences*, *14*(1), 471–483. https://doi.org/10.4236/jss.2026.141028

Karlinda, O., & Srivastava, R. A. (2025). The impact of blended learning on English language test scores. *Journal of Research in English Language Teaching and Linguistics*, *1*(1), 100–118. https://doi.org/10.65431/jrell.v1i1.9

Kaya, Ö. F. (2022). Using corpora for language teaching and assessment in L2 writing: A narrative review [Review of *Using corpora for language teaching and assessment in L2 writing: A narrative review*]. *Focus on ELT Journal*, 46–62. https://doi.org/10.14744/felt.2022.4.3.4

Khojasteh, L., Soori, A., & Javed, F. (2025). Comparing teacher E-feedback, AI feedback, and hybrid feedback in enhancing EFL writing skills. *Technology in Language Teaching & Learning*, *7*(3), 102626–102626. https://doi.org/10.29140/tltl.v7n3.102626

Lannin, A., & Townsend, M. A. (2020). Graduate student perspectives: Career development through serving as writing-intensive GTAs. In *The WAC Clearinghouse; University Press of Colorado eBooks* (pp. 95–119). https://doi.org/10.37514/atd-b.2020.0407.2.04

Liang, L., Kai-ying, C., Huang, S., & Guo, Z. (2022). A study on the application of a corpus-based data-driven learning method utilizing an online and offline blended teaching model in a college English reading course. In *Proceedings*

of the 2022 3rd International Conference on Big Data and Informatization Education (ICBDIE 2022) (pp. 81–101). https://doi.org/10.2991/978-94-6463-034-3_11

Luo, Q., & Zhou, J. (2017). Data-driven learning in second language writing class: A survey of empirical studies. *International Journal of Emerging Technologies in Learning (iJET)*, *12*(3), 182–182. https://doi.org/10.3991/ijet.v12i03.6523

Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, *1*(1), 177–190. https://doi.org/10.1515/jccall-2021-2008

Ma, Q., Chiu, M. M., Lin, S., & Mendoza, N. B. (2022). Teachers' perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study. *ReCALL*, *35*(1), 19–39. https://doi.org/10.1017/s0958344022000180

Nagao, A. (2019). The SFL genre-based approach to writing in EFL contexts. *Asian-Pacific Journal of Second and Foreign Language Education*, *4*(1). https://doi.org/10.1186/s40862-019-0069-3

Nanyinza, E. (2023). Theory into practice: Exploring pre-service teachers' application of elaboration strategies to teach english grammar. *Journal of Education & Social Sciences*, *11*(2), 1–18. https://doi.org/10.20547/jess1122311201

Qoura, Y., Hassan, B. A., & Mostafa, A. (2018). The impact of corpus-based program on enhancing the EFL student teachers' writing skills and self-autonomy. *Journal of Research in Curriculum Instruction and Educational Technology*, *4*(1), 11–53. https://doi.org/10.21608/jrciet.2018.24488

Rahayu, Y. S., & Rostika, E. (2025). The effect of using flashcards on vocabulary mastery of elementary students. *Journal of Research in English Language Teaching and Linguistics*, *1*(1), 86–99. https://doi.org/10.65431/jrell.v1i1.10

Ramadhanti, D., & Yanda, D. P. (2021). Students' metacognitive awareness and its impact on writing skill. *International Journal of Language Education*, *5*(3), 193–193. https://doi.org/10.26858/ijole.v5i3.18978

Rimun, J. C., & Yumarnamto, M. (2024). Metacognition in ELT writing: Teacher's facilitation and students' strategies. *Englisia Journal of Language Education and Humanities*, *11*(2), 91–91. https://doi.org/10.22373/ej.v11i2.20563

Simpson, S., Clemens, R., Killingsworth, D. R., Priest, J., & Ford, J. D. (2020). Creating a culture of communication: A graduate-level STEM communication fellows program at a science and engineering university. In *The WAC Clearinghouse; University Press of Colorado eBooks* (pp. 167–187). https://doi.org/10.37514/atd-b.2020.0407.2.07

Solano, D. A. (2020). Implementing guided-discovery grammar learning. *LETRAS*, *68*, 143–173. https://doi.org/10.15359/rl.2-68.6

Steen, S. van der, Steenbeek, H., Hartigh, R. J. R. D., & Geert, P. L. C. V. (2019). The link between microdevelopment and long-term learning trajectories in science learning. *Human Development*, *63*(1), 4–32. https://doi.org/10.1159/000501431

Tosun, S., & Sofu, H. (2023). The effectiveness data-driven vocabulary learning: Hands-on concordance through a pedagogical corpus. *Journal of Language and Education*, *9*(3), 176–190. https://doi.org/10.17323/jle.2023.12426

Uçar, S., & Zarfsaz, E. (2022). A corpus-based study: The employment of lexical bundles in Turkish students' academic writing. *DergiPark (Istanbul University)*. https://dergipark.org.tr/tr/pub/tubad/issue/74645/1178978

Wibowo, R. A., & Charlotte, D. W. (2025). Students' perception and achievement on online English learning platforms. *Journal of Research in English Language Teaching and Linguistics*, *1*(1), 152–168. https://doi.org/10.65431/jrell.v1i1.11

Yang, Y. F., & Ren, H. (2025). The efficacy of the corpus-based error correction method on revision in writing classrooms. *PLoS ONE*, *20*(3). https://doi.org/10.1371/journal.pone.0317574

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, *13*(4), 257–283. https://doi.org/10.1016/j.jslw.2004.06.002

Youssef, A. F. F.-A. (2020). The effectiveness of corpus-based approach on vocabulary learning gains and retention in Saudi tertiary EFL context. *Journal of English Language Teaching and Applied Linguistics*, *2*(3), 1–15. https://doi.org/10.32996/jeltal.2020.2.3.1

Yunus, M., Dalle, M. B., Kaharuddin, K., Hudriati, A., & Strid, J. E. (2024). Enhancing paragraph writing proficiency: A study of students' performance post global English textbooks exposure. *International Journal of Language Education*, *8*(1). https://doi.org/10.26858/ijole.v8i1.60913

Zhou, S., & Preez, G. D. (2025). Bringing Cinderella to spotlight? Developing grammar through ChatGPT in writing. *Language Learning & Technology*, *29*(1), 1–25. https://doi.org/10.64152/10125/73639