

# International Journal of Educational Innovations

ISSN 3078-5677

International Journal of Educational Innovations  
Volume 2, Issue 1, 96-112  
<https://doi.org/10.46451/ije.260108>

Received: 19 May, 2025  
Accepted: 28 October, 2025  
Published: 1 January, 2026

## A Year-long Corpus-based Study of the Development of Academic English Writing Skills among Freshman Students

Anna Poghosyan

Eliza Safaryan\*

Irshat Madyarov

American University of Armenia, Armenia

(\*Corresponding author. Email: elizasafaryan24@gmail.com)

### Abstract

In higher education, where written assignments often serve as the cornerstone of the learning experience and assessment, academic writing skills emerge as a critical aspect of student success. This study explores the progress of freshman students at an English-medium university in Armenia, examining how their academic writing develops during their first year of study. This research employs a corpus-based analysis of 590 pre- and post-year essays to track trends in the use of academic vocabulary, discourse markers, and reporting verbs. The findings reveal a statistically significant increase in the use of academic vocabulary, indicating students' growing familiarity with academic lexis. No significant changes were observed in the use of discourse markers. However, reporting verbs in the *argue* and *show* categories have increased in frequency, while *think* verbs have declined, suggesting a shift from subjective toward more objective and evidence-based writing. These trends provide insights into the developmental potential of academic writing among first-year students in English-medium universities in non-English speaking countries and highlight areas that could be further examined to understand the specific challenges students face in academic writing.

### Keywords

Academic writing, corpus-based approach, discourse markers, reporting verbs, academic vocabulary

### Introduction

This study adds to the current literature by analyzing how freshman students use academic vocabulary, reporting verbs, and discourse markers in their essays before and after their first year at an English-medium university in Armenia. It brings a longitudinal pre-post comparison dimension to existing studies that examine specific aspects of language in student writing (e.g., Al-Khazraji, 2019; Alkhawaldeh et al., 2023; Bakoko & Waluyo, 2021). By focusing on these

specific features—academic vocabulary, reporting verbs, and discourse markers—this study not only contributes to the understanding of novice academic writers' language development but also targets features that have not yet received enough attention in pre-post longitudinal studies. This distinctive approach offers new insights into the evolution of academic writing skills over time, and the results have practical implications for educators and curriculum developers. The insights gained can help shape teaching methods that more effectively support students as they transition into academic discourse during their first year at university.

## **Literature Review**

### **Academic vocabulary in academic writing**

Academic vocabulary and its role in academic writing have been explored by various scholars (e.g. Coxhead, 2000; Coxhead & Byrd, 2007; Gardner & Davies 2014; Nation, 2001). The concept of academic vocabulary encompasses words that are more commonly found in academic writing compared to other types of texts, spanning various academic disciplines. Coxhead (2000) suggests that academic words make up around 10% of academic texts, whereas Gardner and Davies (2014) propose nearly 14% of such texts. This prevalence underscores the significance of academic vocabulary for both understanding and creating academic writing (Coxhead & Byrd, 2007). The majority of studies exploring academic vocabulary in L2 learner writing have relied on Coxhead's (2000) Academic Word List (AWL), which consists of 570 academic word families identified across approximately 3.5 million words from diverse disciplines such as science, arts, commerce, and law.

While the AWL was a foundational tool at the time, prior to the development of larger corpora such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), its relevance in capturing the breadth of academic vocabulary across contemporary disciplines has been questioned. Recent studies have highlighted alternative lists, such as Gardner and Davies's (2014) Academic Vocabulary List (AVL), which comprises 3,015 academic lemmas extracted from a corpus of 120 million words covering nine academic domains, including humanities, social sciences, law, science, technology, medicine, business, philosophy, and religion. Given the broader coverage and larger corpus base of the AVL, the present study relies on the AVL list to capture a more comprehensive representation of academic vocabulary in academic writing. The AWL is still referenced to maintain continuity with previous research and provide a benchmark for comparison.

Some studies in the literature investigated the usage of academic words in students' essays. The research conducted by Lailiyah and Setiyaningsih (2021) explored the most frequently used academic words in English as a foreign language classroom setting. Using Coxhead's 2000 Academic Word List, the researchers identified the frequency of these words in 24 students' argumentative essays. Overall, the corpus included 10,343 tokens, of which only 576 were academic words, comprising less than 6% of the total running words. The research conducted by Vandenhoeck (2022) shows how lexical complexity and vocabulary in first-year student writing changes before and after a 30-week intensive EAP (English for Academic Purposes) course. The study assessed the use of academic vocabulary in both sets of data with Coxhead's 2000 Academic Word List. The results indicated that academic vocabulary usage was approximately 6% in the pre essays and 9% in the post essays corpus. Additionally, it was noted that the pre-essay, written in the fall semester, provided students with more opportunities to use longer structures, whereas the post-essay, written in the spring semester, did not require such constructions, suggesting the need to assign identical tasks for more reliable comparisons. Bakoko and Waluyo (2021) investigated the vocabulary used by high school students in English essay writing. Using Gardner and Davies's (2014) Academic Vocabulary List (AVL),

the researchers manually analyzed 233 essays. The results showed that 48 out of the first 70 most frequently occurring words in academic writing were present, with "important" being the most frequently used word, appearing 119 times. Notably, the study analyzed only 70 words from the 500 list. Durrant's (2016) study is noteworthy because it analyzed AVL coverage in college students' essays across 32 disciplines and 13 text genres at four British universities. While it does not focus on the performance of English language learners in academic contexts, the results could serve as a point of reference for the present study. For example, their findings indicate that the use of AVL increases from the first year of undergraduate study (29%) to the master's level study (34%). The coverage of AVL in the English discipline is 24%, and in the essay genre it is 32%.

Despite some valuable insights generated by these studies, several gaps remain unaddressed in the literature. Firstly, the size of the corpora analyzed in most studies is relatively small. Secondly, the different essays analyzed for pre and post learning experiences in some studies may have affected the reliability of results. Finally, the use of AVL has been rarely explored among students at English-medium universities in English as foreign language contexts. Addressing these gaps would provide a more comprehensive understanding of academic vocabulary usage in student writing.

### **Discourse markers in academic writing**

Discourse markers in academic writing enhance coherence and guide readers through complex arguments and ideas. As early as 50 years ago, Halliday and Hasan (1976) defined discourse markers as devices "used to relate sentences, clauses, and paragraphs to each other" and expressions that "signal the way the writer wants the reader to relate what is about to be said to what has been said before" (p. 226). For L2 academic writers, discourse markers can present some challenges. Gonzalez et al. (2013) suggest that L2 writers working to achieve smooth and well-structured academic writing in English should prioritize learning a specific set of discourse markers. Li and Schmitt (2009) point out that the lack of using discourse markers in academic writing has always been associated with novice L2 writers. According to them, the problems with discourse markers occur in two ways: a) L2 writers tend to rely heavily on a small set of commonly used phrases, and b) they lack the extensive range needed to use DMs to match the expectations of the target audience.

Several studies explore the use of discourse markers in students' essays. Early research by Francis, Huston, and Manning (1996) provided a foundational list of discourse markers, which has informed subsequent frameworks for analyzing cohesion in student writing. Building on this, Fraser's (2006) list of 107 discourse markers, categorized into contrastive, elaborative, implicative, and temporal markers, offers a more detailed taxonomy for examining the functions of cohesive devices in L2 academic writing. Alkhawaldeh et al. (2023) examined the utilization of discourse markers in argumentative compositions by 120 sophomore and senior students studying English at Hashemite University, Jordan. The findings revealed that both groups employed similar types of discourse markers with seniors using slightly more due to overuse or unnecessary instances. Analysis identified "and", "because", and "but" as the most commonly used, while markers like "although", "yet", "besides", and "furthermore" were rarely used or exclusive to one group. Similarly, Eesa (2021) investigated the utilization of discourse markers in 40 essays of Iraqi English Language Proficiency Test (ELPT) participants, based on Fraser's (2006) 107 discourse markers list. The findings revealed a total of 272 discourse markers used, with elaborative and differential markers being the most prevalent. Topic-related and transition markers were infrequently utilized, comprising only 2% of the total markers. Participants tended to overuse certain markers while neglecting others, possibly

due to a lack of understanding of their proper application or difficulties in essay composition. Further, Al-Khazraji's (2019) study revealed diverse findings, including instances of both misuse and advanced application of discourse markers. However, the scope of the study was constrained by a small sample size of only six essays, and the absence of a predefined list for identifying discourse markers limited the comprehensiveness of the analysis.

Overall, these studies highlight the varying use and challenges associated with discourse markers in student essays, indicating challenges in understanding their proper application and the need for more research in this area. One obvious gap in this literature is a lack of truly longitudinal research to understand developmental potentials of discourse markers among EFL students in English-medium instructional settings.

### **Reporting verbs in academic writing**

Reporting verbs serve as the backbone of academic discourse, facilitating the transmission of knowledge and ideas with precision and authority. According to Kwon et al., (2018) reporting verbs help writers seamlessly integrate and synthesize sources into their arguments and are an important element in writing research papers. In Charles' (2006) words, reporting verbs can be used "to give credit to other researchers and to use their work in the cumulative construction of knowledge" (p. 326).

Some studies aim to investigate the use of reporting verbs in students' essays. Huang (2022) explored the frequency and usage patterns of reporting verbs in undergraduate L2 English major theses compared to L1 students' academic writing. It utilized a list of 48 reporting verbs proposed by Kwon et al. (2018) and found that L2 students used significantly fewer reporting verbs compared to L1 students. Both L2 and L1 students showed a preference for "argue" verbs. Crawford et al., (2021) investigated whether there were variations in reporting verb choices among L2 writers in their argumentative and cause-and-effect essays. Drawing on Kwon et al.'s list of reporting verbs and considering 34 reporting verbs across four semantic categories (argue, think, find, and show), the study compared their occurrence in the two essay types. The findings indicated that students utilized more reporting verbs in argumentative essays compared to cause-and-effect essays. Additionally, argue verbs were found to be the most frequently employed across both types of essays. However, it remains unclear whether it is the essay type or the developmental stage that contributes to these differences. Further, Febriyanti and Yuliawati (2024) examined the usage of reporting verbs in short essays written by undergraduate students majoring in English at the University of Indonesia. The research focused on conducting a frequency analysis of reporting verbs across four semantic categories: *argue*, *find*, *show*, and *think* verbs in narrative, descriptive, causative, and argumentative essays collected over a period of four years. The findings indicated that students predominantly utilized *think* verbs.

There are still notable gaps in the current research on reporting verbs. A small sample size is one such drawback. Additionally, these studies often employed different pre- and post-essays, which could affect the consistency and reliability of the results.

### **Methodology**

This study tracks academic writing gains of Armenian freshmen students' by analyzing the trends in academic vocabulary, reporting verbs, and discourse markers in their entrance and exit essays.

The study addressed the following research questions:

1. How does the use of academic vocabulary change in freshman students' pre- and post-essays?
2. How does the use of discourse markers change in the students' pre- and post-essays?
3. How does the use of reporting verbs change in the pre- and post-essays?

### **Setting and participants**

The study took place in an English-medium university in Armenia, accredited by the Western Association of Schools and Colleges (WASCUC), a US accrediting organization. The participants of the study were freshman students enrolled in the 2022-2023 academic year. Applicants to the university are expected to demonstrate academic proficiency in English via TOEFL iBT (score of 79 or above) or IELTS (6.5 or above). These students represented various undergraduate degree programs at the university from four colleges: College of Humanities and Social Sciences, College of Science and Engineering, College of Business and Economics, and College of Health Sciences. Between the pre and post essays, the students took two Freshman Seminar courses of 3 credits each (3 hours of classwork per week) within two 15-week semesters (Fall and Spring), where they frequently wrote different types of essays and received feedback on their drafts. In addition, the students took 3-4 other courses per semester taught in English, some major-specific and some broader general education electives.

### **Ethical considerations**

Informed consent was not applicable in this study because the data was secondary. The university where the study took place had collected the data for administrative purposes. The research team then applied and received approval for the use of anonymized secondary data from the Ethics Committee of the International TESOL Union. To protect students' identity, their essays were coded.

### **Data collection**

The students hand-wrote their essays in groups of 15-25 under the same proctored conditions in classrooms without any resources such as dictionaries or access to the internet. The entrance and exit essays were two semesters apart (over eight months). For this reason, the students followed the same essay prompt (Appendix A). The prompt asked the students to read two short passages and write their own opinion on the topic with reference to the passages. Initially, 440 students wrote the entrance essay. Due to attrition at the exit essay and illegible handwriting, the final corpus comprised the essays of 295 students, resulting in a set of 590 essays (295 entrance and 295 exit essays). The entrance essays contained 60,116 tokens and 5,089 types, while the exit essays included 62,230 tokens and 5,925 types. The average entrance and exit essays contained 203 and 211 tokens, and 17 and 20 types, respectively.

To digitize the handwritten essays, optical character recognition (OCR) Pen-to-Print software was employed. To ensure accuracy of character recognition, two researchers from the team checked a random sample of 64 essays (roughly 10%). In analyzing 64 papers totaling approximately 13,116 tokens, various errors were identified and quantified. Among these, false negative spelling errors, where an actual spelling mistake was not detected, accounted for 171 tokens, representing 1.30% of the total. False positive spelling errors, where a correct word was mistakenly flagged as incorrect, were found in 72 tokens, making up 0.55%. Crossed out text errors, where text was mistakenly marked for deletion, were present in 85 tokens, or 0.65%, and punctuation errors appeared in 81 tokens, comprising 0.62% of the overall count. These figures suggest a relatively small proportion of inaccuracies made by the OCR software in the



examined texts. Further, when querying students' essays for vocabulary, reporting verbs, and discourse markers in the RStudio software, the RegEx function was applied. This function enables the search for all versions of the words queried regardless of spelling mistakes made by the student or misrecognized by the OCR software.

### Data analysis

Most of the analysis of raw essays was conducted in RStudio (2025.09.1 Build 401). AntConc (4.2.4) software was employed to query selected discourse markers and reporting verbs within the essays to verify accuracy of counts in RStudio and to add an element of qualitative analysis. RStudio produced frequencies for all three measures of academic writing: 1) Academic Vocabulary List (Gardner & Davies, 2014), 2) Fraser's (2006) list of 107 discourse markers in four categories (CDMs, EDMs, IDMs, and TDMs) (Appendix B), and 3) Kwon et al. 's (2018) 53 reporting verbs categorized as argue, show, find, and think verbs (Appendix C). As part of the analysis, RStudio also lemmatized the lists to account for different grammatical forms of the words.

The final descriptive and inferential statistical analyses were produced both in RStudio and JASP (Version 0.18.3) to verify the results. The initial assumption checks showed that all measures (vocabulary, discourse markers, and reporting verbs) did not meet the normality assumption based on the Shapiro-Wilk test, QQ plots, and histograms. This was expected because of the frequency nature of the data, which sometimes had very small counts per essay e.g., 0, 1, or 2. For this reason, the non-parametric equivalent of the paired *t*-test (Wilcoxon signed-rank test) was used. All analyses are accompanied with descriptive statistics.

### Results

***Research Question 1: How does the use of academic vocabulary change in freshman students' pre and post essays?***

Table 1

*Descriptive Statistics Summarizing the Pre- and Post-AVL Results (N=295)*

	<i>Median</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Tokens %</i>
Pre-AVL	23	24.4	9.54	3	62	12%
Post-AVL	28	31.1	12.95	5	74	14.6%

As shown in Table 1, both the mean and median values increased from pre- to post-essays, suggesting an overall improvement in students' use of academic vocabulary. Similarly, the percentage of academic tokens increased from 12% to 14.6%, reflecting a noticeable shift toward more academic language use. The Wilcoxon signed-rank test confirmed that this increase was statistically significant ( $W = 9146$ ,  $z = -8.07$ ,  $p < .001$ ). This finding demonstrates that students incorporated significantly more academic vocabulary in their writing after a year of English-medium instruction. The observed improvement points to students' growing familiarity with formal academic language and suggests progress in adopting more discipline-appropriate vocabulary for academic writing.

***Research Question 2: How does the use of discourse markers change in the students' pre and post essays?***

To answer the second research question, the usage of discourse markers in students' pre- and post-essays is categorized into four types based on Fraser's (2006) classification: Contrastive Discourse Markers, Elaborative Discourse Markers, Implicative Discourse Markers, and Temporal Discourse Markers (Appendix B).

Table 2

*Descriptive Statistics Summarizing the Pre- and Post-Results for Discourse Markers (N=295)*

	<i>Median</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Pre DM contrastive	1	1.20	1.25	0	7
Post DM contrastive	1	1.20	1.21	0	7
Pre DM elaborative	9	9.13	4.14	1	24
Post DM elaborative	9	9.36	4.77	0	29
Pre DM implicative	1	1.89	1.72	0	8
Post DM implicative	1	1.75	1.61	0	9
Pre DM temporal	3	3.35	2.24	0	15
Post DM temporal	3	3.2	1.93	0	10

As seen in Table 2, the mean scores for most discourse marker types remained almost unchanged from pre- to post-essays. Contrastive and temporal markers showed identical mean and median values, while elaborative markers increased only slightly (from 9.13 to 9.36). Implicative markers slightly decreased. These minimal changes suggest that, unlike academic vocabulary, the use of discourse markers did not show notable development after one year of instruction. Students appeared to maintain similar patterns of cohesive device use across both writing samples. Table 3 shows that indeed none of the differences were statistically significant. Appendices D and E show a sampling of 15 discourse markers that have shown the highest increase between entrance and exit essays, and 15 discourse markers that have shown the highest decrease across this time frame. The discussion section below elaborates on possible interpretations of these findings.

**Table 3***Wilcoxon Signed-rank Results for Discourse Markers in Pre- and Post-essays (N=295)*

	<i>W</i>	<i>z</i>	<i>p</i>
DM contrastive	11271	-0.14	0.888
DM elaborative	17455	-0.45	0.654
DM implicative	13638	1.07	0.279
DM temporal	16833	0.89	0.371

**Research Question 3: How does the use of reporting verbs change in the pre and post essays?**

To answer the third research question, the usage of reporting verbs in students' pre- and post-essays was categorized into four groups (argue, show, find, and think) based on Kwon et al.'s (2018) list of reporting verbs (Appendix C).

**Table 4***Descriptive Statistics Summarizing the Pre- and Post-Results for Reporting Verbs (N=295)*

	<i>Median</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Pre RV argue	1	1.495	1.428	0	9
Post RV argue	1	1.661	1.616	0	7
Pre RV show	0	0.481	0.689	0	3
Post RV show	1	0.695	0.834	0	4
Pre RV find	0	0.163	0.429	0	3
Post RV find	0	0.163	0.388	0	2
Pre RV think	1	1.264	1.319	0	6
Post RV think	0	0.827	1.057	0	5

According to Table 4, there was a slight increase in the mean frequency of *argue* and *show* verbs, while *find* verbs remained unchanged and *think* verbs decreased. Table 5 confirms that the increase in *show* verbs was statistically significant ( $p < .001$ ), while the decrease in *think*



verbs was also significant ( $p < .001$ ). No significant change was observed for *argue* or *find* verbs. These patterns suggest that students gradually shifted from informal or subjective expressions of stance (e.g., *think*) toward more objective and academic forms (e.g., *show* or *argue*), reflecting a growing awareness of formal reporting conventions. Appendices F and G illustrate 15 specific reporting verbs that have increased and reduced in use between entrance and exit essays.

Table 5

*Wilcoxon Signed-rank Results for Reporting Verbs in Pre- and Post-essays (N=295)*

	<i>W</i>	<i>z</i>	<i>p</i>
Pre RV argue	11601	-1.56	0.111
Pre RV show	4967	-3.27	< .001
Pre RV find	1409	-0.08	0.927
Pre RV think	13793	4.57	< .001

## Discussion

The findings of this year-long corpus-based study reveal patterns in the development of freshman students' academic writing skills across three key linguistic dimensions: academic vocabulary, discourse markers, and reporting verbs. The results show clear lexical advancement, no changes in the use of discourse markers, and refinement in the use of reporting verbs.

The statistically significant improvement in students' use of academic vocabulary (AVL) from 12% to 14.6% demonstrates a meaningful rise in students' ability to incorporate academic vocabulary after a year of instruction in an EFL college-level English-medium instructional setting. This growth suggests that students are moving toward a more formal and scholarly style, showing greater lexical sophistication and reduced reliance on general, non-academic language. These findings align with previous research (e.g., Vandenhoeck, 2022), which emphasizes the positive effect of sustained academic exposure on lexical development. The resulting 14.6% is also consistent with Gardner and Davies's (2014) estimate that approximately 14% of AVL is present in academic materials in the BNC and COCA corpora. One might be surprised that these results fall quite short of the AVL coverage in students' papers found in university students's papers in the UK (Durrant, 2016). In their study, the essay genre - the same genre used in this study - showed an AVL coverage of 32%. The results of this study fall noticeably short even of the lowest coverage of 21% found by Durrant among students studying classics. One likely explanation is that the corpus in Durrant's study consisted of students' papers whose "had received at least an 'upper-second class' grade, and so can be deemed examples of 'successful' student writing" (p. 52). Second, the students' papers collected between 2004 and 2007 in Durrant's study were likely take-home assignments - carefully planned, proof-read, and typed on computers. Therefore, the nature of their corpus must excel in quality compared to the corpus in this study, compiled of responses written in proctored conditions within 30 minutes.

The results of discourse markers are harder to interpret. The mean frequency for all four types—contrastive, elaborative, implicative, and temporal—remained nearly unchanged. None of these differences were statistically significant, indicating that discourse marker usage remained relatively stable throughout the year. While this lack of increase might initially appear as stagnation, previous studies suggest that advanced writers tend to use discourse markers more selectively to maintain coherence without redundancy (e.g., Altenberg & Tapper, 1998; Yang & Sun, 2012). Thus, the relatively consistent or reduced use of discourse markers in students' post-essays may signal an emerging awareness of the functional rather than purely quantitative use of cohesive devices—an important step toward writing precision and rhetorical control.

The findings for reporting verbs displayed a nuanced developmental trend. The mean frequency of *agree* and *show* verbs increased over time, while *show* verbs also demonstrated statistically significant growth. In general, *argue* verbs are the most frequently used among the four categories of reporting verbs. This pattern is found in previous research where *argue* verbs tend to be predominant in students' writing, and it increases with longer exposure to academic writing and from L2 to L1 students (Briguglio, 2021; Crawford et al., 2021; Huang, 2022). This shift towards more objective verbs suggests that students are developing a more analytical and evidence-based approach to expressing stance and attribution. Another positive finding in this study is that *think* verbs have decreased significantly. The reduced reliance on *think* verbs reflects a move away from personal opinion and toward academic conventions of reporting. Compared with studies that report a persistence of subjective stance markers among novice writers (e.g., Febriyanti & Yuliawati, 2024), the current results indicate promising progress in rhetorical maturity and verb choice sophistication.

Taken together, these results illustrate a multidimensional trajectory of academic writing development. The statistically significant rise in academic vocabulary use demonstrates clear lexical growth, while the refined employment of reporting verbs and stable discourse marker usage point to a developing awareness of rhetorical appropriateness and linguistic economy. Rather than a uniform increase across all features, students' writing shows signs of selective improvement—an indication that they are not merely producing longer or denser texts but are making more deliberate linguistic choices. This pattern reflects an evolving balance between linguistic expansion and strategic control, marking an important stage in their transition from novice to more proficient academic writers.

### Limitations and Directions for Future Studies

One limitation, albeit minor based on the analysis, pertains to possible tokenization and tagging errors introduced by the OCR software, particularly in spelling and punctuation accuracy. Similarly, a number of papers had to be removed from the analysis because of illegible handwriting, hence reducing the total sample size in the data.

Moving forward, a few insights can be gained from this study. Methodologically, how well OCR software is able to interpret less legible handwriting impacts the quality of learner-based corpus studies. It is likely that these software solutions will increase in quality, and it is important to assess continuously the performance of available OCR software. Further, future research could replicate this study to confirm the longitudinal trends in similar English-medium instructional contexts in other countries where English is considered a foreign language. Specifically, it was puzzling that AVL coverage in this study is quite short of those found in British universities. It will be instructive to see findings in other similar educational settings. It may be that one year in this study was not sufficient to see the longer-term growth; therefore,

a study encompassing the full college experience could address many remaining questions. Finally, this study did not perform qualitative analysis of specific vocabulary, discourse markers, and reporting verbs. More contextual discourse analysis will shed more light on the patterns revealed in this study. For example, it would be valuable to identify developmental patterns of L2 students when it comes to grammatical forms of reporting verbs.

## References

- Alkhawaldeh A., Huneety A., Mashqba, B., & Zeidan Z. (2023). The use of discourse markers in argumentative compositions by Jordanian EFL learners. *Humanities and Social Sciences Communications*, 10(1), 1-8. DOI: [10.1057/s41599-023-01525-0](https://doi.org/10.1057/s41599-023-01525-0)
- Al-Khazraji, A. (2019). Analysis of discourse markers in Essays Writing in ESL classroom. *International Journal of Instruction*, 12(2), 1308-1470. DOI: [10.29333/iji.2019.12235a](https://doi.org/10.29333/iji.2019.12235a)
- Anditasari, A. W., & Sitompul, S. K. (2022). Challenges in writing academic research: An exploration of master's students' experiences. *Getsempena English Education Journal*, 9(2), 136-148. <https://doi.org/10.46244/geej.v9i2.1805>
- Aulbert-Siepelmeier, A., Boltzmann, M., Gerth, I., Menkhaus, K., & Russeler, J. (2012). Alpha Plus: An innovative training program for reading and writing education of functionally illiterate adults. *Creative Education*, 3(3), 357-361. <http://dx.doi.org/10.4236/ce.2012.33056>
- Badley G. (2019). Post-academic writing: Human writing for human readers. *Qualitative Inquiry*, 25(2), 180-191. <https://doi.org/10.1177/107780041773633>
- Bakoko, R. & Waluyo, B. (2021). Vocabulary list learning supported by gamification: Classroom action research using a quizlet. *The Journal of AsiaTEFL*, 18(1), 289-299. DOI: [10.18823/asiatefl.2021.18.1.20.289](https://doi.org/10.18823/asiatefl.2021.18.1.20.289)
- Byrd, P., & Coxhead, A. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16(3), 129-147. DOI: [10.1016/j.jslw.2007.07.002](https://doi.org/10.1016/j.jslw.2007.07.002)
- Bowker, N. (2007) Academic writing: A guide to tertiary level writing. Palmerston North: Massey University. 1-112. <http://www.massey.ac.nz>
- Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(3), 310-331. DOI: [10.1016/j.esp.2005.05.003](https://doi.org/10.1016/j.esp.2005.05.003)
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. <https://doi.org/10.2307/3587951>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge Falmer
- Crawford, W. J., Uludag, P., & McDonough, K. (2021). English for academic purposes writers' use of reporting verbs in argumentative and cause-and-effect essay exams. *BC TEAL Journal*, 6(1), 42-58. <https://doi.org/10.14288/bctj.v6i1.388>
- Davies, M., & Gardner, D. (2013). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327. <https://doi.org/10.1093/applin/amt015>
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49-61.
- Eesa, R. M. (2021). The discourse markers. A corpus-based study of written texts by Iraqi ELPT participants. *LARQ Journal of Philosophy, Linguistics and Social Sciences*, 2(41), 1241-1258, <https://doi.org/10.31185/lark.Vol2.Iss41.1793>
- Febriyanti, N. D., & Yuliawati, S. (2024). A corpus-based study of reporting verbs in short essays. *Journal of Linguistic Phenomena*, 2(2), 42-48. DOI: [10.24198/jlp.v2i2.51533](https://doi.org/10.24198/jlp.v2i2.51533)
- Feak, C. B., & Swales, J. M. (2012). Academic writing for graduate students: Essential tasks and skills (3rd ed.). Ann Arbor, MI: The University of Michigan Press.

- <https://www.press.umich.edu/pdf/9780472034758-unit1.pdf>
- Francis, G., Huston, S. & Manning. (1996). Collins COBUILD grammar patterns 1: Verbs. London: HarperCollins. <https://arts-ccr-002.cal.bham.ac.uk/>
- Fraser, B. (2006). Towards a theory of discourse markers, in K. Fischer (ed.), *Approaches to discourse particles*. 189–204. Amsterdam. Elsevier. [https://doi.org/10.1163/9780080461588\\_012](https://doi.org/10.1163/9780080461588_012)
- Ganobcsik-Williams, L. (2004). *A report on the teaching of academic writing in the UK higher education*. Royal Literary Fund
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied linguistics*, 35(3), 305-327.
- Gibbon, S., & Lillis, T. (2012). "The development of academic writing in English-medium instruction: Linguistic challenges for students and teachers." *Journal of Second Language Writing*, 21(3), 262-275.
- Gonzalez-Rodriguez, E., & Hernandez A. T. (2013). Impact of instruction on the use of L2 discourse markers. *Journal of Second Language Teaching & Research*, 2(1), 3-32. <https://pops.uclan.ac.uk/index.php/jsltr/article/view/508>
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Pearson Longman
- Harris, D. (2019). *Literature review and research design: A guide to effective research practice* (1st ed.). Routledge. <https://doi.org/10.4324/9780429285660>
- Hellekjaer, G. (2010). "English as a medium of instruction in higher education: A case study of EMI at a Norwegian university." *Journal of English for Academic Purposes*, 9(3), 198-211.
- Huang, Y. (2022). A corpus-based study on the semantic use of reporting verbs in English majors' undergraduate thesis writing. *Journal of Language Teaching and Research*, 13(6), 1287-1295. <https://doi.org/10.17507/jltr.1306.17>
- Jaroongkhongdach, W. (2015). Idea sharing: reporting verbs in research writing: Lessons from experts. *PASAA*, 50(0), 142-153, DOI:[10.24198/jlp.v2i2.51533](https://doi.org/10.24198/jlp.v2i2.51533)
- Kapranov, O. (2020). The use of discourse markers in academic writing by in-service primary school teachers of English. *Prague Journal of English Studies*, 9(1), 197-229. <https://doi.org/10.2478/pjes-2020-0009>
- Kwon, M. H., Partridge, R. S., & Staples, S. (2018). Source work in the first-year L2 writing classroom: Undergraduate L2 writers' use of reporting verbs. *Journal of English for Academic Purposes*, 34(0), 86-96. <https://doi.org/10.1016/j.jeap.2018.04.001>
- Lailiyah, M., & Setiyaningsih, L. A. (2021). Academic vocabulary of EFL students' writing: A corpus-based study. In *Proceedings* 5, 28-32. <https://ojs.pnb.ac.id/index.php/Proceedings/>
- Lestari, D. (2020). An analysis of the students' difficulties in writing undergraduate thesis at the English education study program of the Muhammadiyah University of Bengkulu. *Premise: Journal of English Education and Applied Linguistics*. 9(1). 17-29. <http://dx.doi.org/10.24127/pj.v9i1.2588>
- Li J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18(2), 85-102. <https://doi.org/10.1016/j.jslw.2009.02.001>
- Lillis, T., & Curry, M. J. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. Routledge
- McEnery, T., Tono, Y., & Xiao, R. (2006). *Corpus-based language studies: An advanced resource book*. Routledge
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press
- Paltridge, B., & Phakiti, A. (2015) *Research methods in applied linguistics: a practical resource*. Bloomsbury

- Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics*, 29(6), 1139-1172. <https://doi.org/10.1515/ling.1991.29.6.1139>
- Sultan, N. (2003). British students' academic writing: Can academia help improve the writing skills of tomorrow's professionals?, *Sage Journals*, 27(2), 138-446. <https://doi.org/10.5367/ihe.2013.014>
- Tadayyon, M., & Farahani, V. M. (2017). Exploring discourse markers used in academic papers: A comparative corpus-based inquiry of Iranian and English native writers. *Iranian EFL Journal*, 13(2), 130-150.
- Thaiss, C., & Zawacki T. M. (2022). *Engaged writers and dynamic disciplines: Research on the academic writing life*. The WAC Clearinghouse. <https://wac.colostate.edu/books/landmarks/engaged>
- Tuononen, T., & Parpala, A. (2021). The role of academic competencies and learning processes in predicting bachelor's and master's thesis grades. *Studies in Educational Evaluation*, 70(0), <https://doi.org/10.1016/j.stueduc.2021.101001>
- Vandenhoeck, T. (2022). Lexical and syntactic changes in first-year EAP student writing. *Language research bulletin*. 37(0), 94-106. DOI: 10.34577/00005200

## Appendix A

The essay prompt

### Directions

**Read the texts below. In a brief essay, explain the possible relationship between mood and learning. Refer to both texts as you support your explanation. Before you start writing, take a few moments to organize your thoughts. You will have 30 minutes to complete this activity.**

(1) "...a positive mood has a specific influence on learning. It affects your ability to learn things that require some amount of flexibility and creativity. It does not influence learning where flexibility is not required.

What does this mean for you? If you know that you are going to have to exhibit some degree of mental flexibility, then you should do what you can to put yourself in a positive mood. Listen to some happy music. Talk to colleagues and friends you enjoy. Spend a few minutes checking out the latest viral YouTube video. There is real value to feeling good." (Markman, 2010)

(2) "Recent research into neuroscience is confirming that the Chinese philosophers are correct: Brain scans reveal that our unconscious awareness of emotions and phenomena around us are actually what drive the decisions we believe we are making with such logical rationality. According to Marianne LaFrance, a psychology professor at Yale, if we see a happy face for just a fraction of a second (4 milliseconds to be exact), that's long enough to elicit a mini emotional high. In one study viewers who flashed a smile — even though it was shown too quickly for them to even realize they had seen it — perceived the things around them more positively." (Gross-Loh, 2013)



## Appendix B

Fraser's (2006) list of discourse markers

Discourse Markers	Definition	Examples
<b>Contrastive Discourse Markers (CDMs)</b>	CDMs signal that the explicit interpretation of S2 contrasts with an interpretation of S1.	alternatively, although, but, contrariwise, contrary to this/that, conversely, despite (doing) this/that, even so, except, however, in comparison (with/to this/that), in contrast (with/to this/that), in spite of (this/that), instead of (this/that), nevertheless, nonetheless, notwithstanding, on the contrary, on the other hand, rather (than (do) this/that), regardless (of this/that), still, though, whereas, yet ...
<b>Elaborative Discourse Markers (EDMs)</b>	EDMs signal a quasi-parallel relationship between S2 and S1 so that S2 constitutes an elaboration of S1.	above all, also, alternatively, analogously, and, as well as, besides, better yet, by the same token, correspondingly, equally, for another thing, for example, for instance, further(more), in addition, in any event, indeed, in fact, in other words, in particular, likewise, more accurately, more importantly, more precisely, more to the point, moreover, namely, on that basis, on top of it all, or, otherwise, similarly, that is (to say), to cap it all off, what is more ...
<b>Implicative Discourse Markers (IDMs)</b>	IDMs signal that S2 conveys a message which is, in some sense, consequential to some aspect of S1.	accordingly, after all, all things considered, as a (logical) consequence/conclusion (of this/that), as a result (of this/that), because (of this/that), consequently, due to, for this/that reason, hence, in this/that/any case, in view of, it can be concluded that it follows that, of course, on this/that condition, on these/those grounds, overall, since, so, then, therefore, thus ...
<b>Temporal Discourse Markers (TDMs)</b>	TDMs signal that the event in S2 is temporally related to some occurrence in S1.	after, already, as long as, as soon as, before, earlier, eventually, finally, first, following, immediately afterward, meantime, meanwhile, originally, previously, recently, second, since, still, subsequently, then, when, while, until ...



### Appendix C

Kwon et al.'s (2018) list of reporting verbs

Argue Category ( $n = 32$ ): argue, suggest, assert, predict, write, explain, conclude, mention, admit, observe, accept, imply, add, complain, hypothesize, insist, propose, remark, reply, speculate, stress, contend, report, postulate, posit, claim, point out, maintain, say, state, talk about, acknowledge

Show Category ( $n = 7$ ): demonstrate, illustrate, indicate, confirm, reveal, mean, show

Find Category ( $n = 8$ ): realize, find out, discover, establish, infer, recognize, note, identify

Think Category ( $n = 6$ ): know, assume, think, hold, feel, hope

### Appendix D

Top 15 discourse markers (Fraser, 2006) with the highest increase in use from entrance to exit essays. The table shows the means of 15 discourse markers used in the essays of 264 students.

Category_dm	Mean entrance	Mean exit	Mean difference
contrastive_on_the_other_hand	0.057	0.080	0.023
elaborative_further	0.019	0.042	0.023
implicative_since	0.102	0.133	0.030
implicative_therefore	0.102	0.133	0.030
temporal_since	0.102	0.133	0.030
contrastive_in_contrast	0.000	0.030	0.030
elaborative_indeed	0.030	0.061	0.030
implicative_thus	0.129	0.167	0.038
temporal_while	0.265	0.307	0.042
implicative_consequently	0.008	0.053	0.045
contrastive_however	0.288	0.371	0.083
elaborative_moreover	0.110	0.197	0.087
temporal_first	1.083	1.174	0.091
implicative_overall	0.038	0.152	0.114
elaborative_and	7.470	7.917	0.447

### Appendix E

Bottom 15 discourse markers (Fraser, 2006) with the highest decrease in use from entrance to exit essays. The table shows the means of 15 discourse markers used in the essays of 264 students.

Category_dm	Mean entrance	Mean exit	Mean difference
implicative_so	0.686	0.492	-0.193
contrastive_but	0.648	0.504	-0.144
implicative_because	0.360	0.216	-0.144

elaborative_or	1.133	0.996	-0.136
temporal_when	1.106	0.970	-0.136
elaborative_for_example	0.269	0.170	-0.098
elaborative_for_instance	0.201	0.117	-0.083
temporal_after	0.205	0.140	-0.064
elaborative_that_is	0.174	0.117	-0.057
temporal_before	0.223	0.182	-0.042
implicative_then	0.174	0.140	-0.034
implicative_of_course	0.080	0.045	-0.034
temporal_eventually	0.042	0.011	-0.030
temporal_second	0.333	0.307	-0.027
temporal_as_soon_as	0.015	0.004	-0.011

## Appendix F

Top 15 reporting verbs (Kwon et al., 2018) with the highest increase in use from entrance to exit essays. The table shows the means of 15 reporting verbs used in the essays of 295 students.

Category_verb	Mean entrance	Mean exit	Mean difference
argue_report	0	0.01	0.01
show_indicat	0.003	0.014	0.01
argue_argue	0.007	0.02	0.014
show_demonstrate	0.01	0.024	0.014
argue_stress	0.092	0.108	0.017
argue_maintain	0.014	0.031	0.017
find_identify	0	0.017	0.017
find_note	0.01	0.034	0.024
argue_conclude	0.102	0.132	0.031
argue_argue	0.003	0.034	0.031
show_reveal	0.037	0.081	0.044
show_confirm	0.044	0.112	0.068
argue_state	0.129	0.241	0.112
show_show	0.231	0.346	0.115
argue_suggest	0.061	0.237	0.176

## Appendix G

Bottom 15 reporting verbs (Kwon et al., 2018) with the highest decrease in use from entrance to exit essays. The table shows the means of 15 reporting verbs used in the essays of 295 students.

Category_verb	Mean entrance	Mean exit	Mean difference
think_think	0.359	0.576	-0.217
think_feel	0.288	0.461	-0.173
argue_mention	0.186	0.295	-0.108
argue_say	0.237	0.339	-0.102
think_know	0.132	0.180	-0.047
argue_explain	0.041	0.088	-0.047
show_mean	0.102	0.129	-0.027
find_realize	0.034	0.054	-0.020
argue_talk_about	0.024	0.041	-0.017
show_illustrate	0.000	0.017	-0.017
find_realiz	0.010	0.020	-0.010
argue_claim	0.068	0.075	-0.007
find_find_out	0.027	0.034	-0.007
argue_point_out	0.010	0.014	-0.003
find_discover	0.017	0.020	-0.003