The Role of Cohesive Devices in L2 Writing Development: Insights from an Intensive EAP Program

Mahmoud Abdi Tabari[1] & Attila M. Wind[2]

[1]University of Nevada, Reno | USA

[2]Eötvös Loránd University | Hungary

Abstract: Recent research in L2 writing has increasingly focused on the use and development of cohesive devices at local, global, and textual levels in academic texts. However, little is known about how cohesion unfolds in the writing of English for Academic Purposes (EAP) learners enrolled in intensive instructional programs. Moreover, prior studies often rely on static, decontextualized measures that fail to capture the dynamic and task-sensitive nature of cohesion. Grounded in Complex Dynamic Systems Theory (CDST; Larsen-Freeman, 1997), this study tracked 58 multilingual graduate students enrolled in a five-week intensive EAP course at a major university in the US. Using a multi-wave time-series design, weekly writing samples were analyzed through computational text analysis to examine longitudinal patterns in cohesive device use. Findings revealed significant non-linear growth in text cohesion, modest but statistically significant gains in global cohesion, and inconsistent development in local cohesion. Substantial inter- and intra-individual variability was observed across all cohesion types. Additionally, patterns of cohesive device use were significantly influenced by task type and instructional focus. These results highlight the dynamic, individualized, and context-dependent nature of cohesion in L2 academic writing and underscore the value of longitudinal, fine-grained analyses for understanding developmental trajectories.

Keywords: L2 academic writing, cohesion development, CDST, EAP, longitudinal writing analysis

journal of WRITING RESEARCH

Abdi Tabari, M., & Wind, A.M. (2026, accepted for publication. The role of cohesive devices in L2 writing development: Insights from an intensive EAP Program. *Journal of Writing Research, volume(issue),* ##-##. DOI: xx

Contact: Mahmoud Abdi Tabari, Department of English, University of Nevada, 1664 N. Virginia St, Reno, NV 89557 | USA – mabditabari@unr.edu

1. Introduction

Cohesive devices are essential for structuring clear, coherent, and logically connected discourse in second language (L2) writing. However, their development, especially in high-stakes academic settings, remains underexplored (Godfrey, 2024). Intensive English for Academic Purposes (EAP) programs, aimed at accelerating L2 learners' academic writing skills development in non-English disciplines, offer a unique context to investigate how cohesion evolves under time constraints and instructional pressure. While cohesion in first language (L1) writing has been extensively studied (Crossley et al., 2011; Crowhurst, 1987; Myhill, 2008), longitudinal research on L2 learners' cohesion remains scarce (Crossley et al., 2016). Unlike L1 writers, L2 learners face additional linguistic and cognitive challenges, such as lexical limitations, syntactic complexity, and cross-linguistic interference, making cohesion development a complex, dynamic process (Abdi Tabari et al., 2023; Wei, 2023).

Over the past two decades, computational linguistics has provided large-scale analyses of cohesion in L2 writing. However, these studies often rely on static, decontextualized measures that fail to capture the nuanced, evolving use of cohesive devices across diverse academic tasks, especially in intensive EAP contexts (Schmid, 2020). Existing research mainly uses pre-test/post-test (two-wave) designs (Abdi Tabari & Wind, 2025). While useful for measuring broad outcomes, these designs offer only limited snapshots of linguistic development. Language acquisition is non-linear and dynamic, requiring methods that capture variability, adaptation, and individual trajectories over time (Abdi Tabari & Johnson, 2023; Johnson & Abdi Tabari, 2024; Larsen-Freeman, 1997; Verspoor et al., 2012; Wind & Verspoor, 2026). Complex Dynamic Systems Theory (CDST) offers a robust framework for analyzing the interdependent, emergent properties of L2 writing development (Fogal, 2022). However, empirical CDST studies on L2 cohesion often struggle with small samples, limited generalizability, and difficulty distinguishing individual variability from broader trends (Wind, 2021). The ergodicity problem (Lowie & Verspoor, 2019) underscores the risk of assuming that group-level patterns reflect individual learning, highlighting the need for integrating group- and individual-level analyses.

To address these gaps, this study conducts a fine-grained longitudinal investigation of cohesive device development in L2 academic writing within an intensive EAP program. It examines how learners use and develop local, global, and text-level cohesive devices across multiple academic tasks over five weeks. Using a multi-wave time-series design with weekly writing samples and computational text analysis, this research offers a multidimensional perspective on cohesion in the writing of business and engineering students. Combining macro-level trends with micro-level case studies enables a nuanced exploration of individual variability, adaptation, and instructional influences on cohesion growth. By moving beyond traditional two-wave designs, this study provides a detailed account of the development of cohesion over time, contributing to theoretical and pedagogical discussions of cohesive devices in academic writing.

2. Literature review

Research on cohesion distinguishes between 'cohesion' and 'coherence'. Cohesion concerns the linguistic cues that link concepts within a text. A highly cohesive text makes it easier for readers to connect ideas. These cues are usually local but can also be global. Local cohesion refers to cues connecting sentences within a paragraph, while global cohesion connects ideas between paragraphs. Additionally, text cohesion refers to the presence or absence of these cues throughout the entire text.

Local cohesion includes overlapping words and concepts between sentences and the use of connectives such as 'consequently,' 'therefore,' and 'because' (Halliday & Hassan, 1976). Global cohesion involves semantic and word overlaps between paragraphs (Foltz, 2007), where lexical items and concepts are reiterated in successive paragraphs. Cohesion can also be examined at the broader text level; for example, givenness gauges the extent to which the writer assumes specific contextual information is familiar to the reader. The measurement of givenness is based on how many words are new or given in a text. Crossley et al. (2016) point out that linguistic cues related to local cohesion are more explicit, whereas those related to global and text cohesion are more implicit.

Coherence depends on cohesion cues as well as non-linguistic factors such as reading skills and background knowledge (McNamara et al., 1996; O'Reilly & McNamara, 2007). Crossley and McNamara (2010) claim that high-knowledge readers, unlike low-knowledge readers, can make inferences necessary to bridge conceptual gaps in texts lacking cohesion cues. High-knowledge readers may benefit from low-cohesion texts because the need to infer implicit connections prompts deeper processing and more active integration of prior knowledge, which can enhance comprehension and retention (McNamara et al., 1996; O'Reilly & McNamara, 2007). Consequently, cohesion cues can sometimes be redundant or even distract readers familiar with the topic. Several studies have demonstrated that cohesion cues are crucial indicators of text comprehensibility. Texts with more cohesion cues tend to be easier to understand (Abdi Tabari & Johnson, 2023; Crossley et al., 2014b; Crossley & McNamara, 2011; Gernsbacher, 1990). However, McNamara et al. (1996) found that low-knowledge readers benefit more from cohesion cues than high-knowledge readers.

2.1 The development of cohesion in L1 and L2 writing

Research on the development of cohesion in L1 students' writing shows that L1 writers generally use more cohesive devices as they progress, especially in elementary and middle school (Bereiter & Scardamalia, 1987). However, the use of cohesion cues varies across grade levels (Crowhurst, 1987; Fitzgerald & Spiegel, 1986; Yde & Spoelders, 1985). For example, studies have documented growth in local cohesion, such as connectives and referential pronouns, in second-grade L1 writers' texts (King & Rentel, 1979) and increased word repetition from Grades 1 to 4 (Rentel et al., 1983). Research also shows that over time, the distance between cohesive ties decreases, meaning referents are placed closer together, resulting in more cohesive texts (Fitzgerald & Spiegel, 1986; McCutchen & Perfetti, 1982; Yde

& Spoelders, 1985). McCutchen and Perfetti (1982) found that, over the course of schooling, the use of local cohesive devices in young writers' texts increases up to Grade 8.

This upward trend in cohesion use, however, appears to level off by eighth grade. McCutchen (1986) found that eighth graders' writing contains more connectives and greater lexical and conceptual overlap between sentences compared to sixth graders. As L1 writers become more proficient in higher grades, there is a shift from relying on local cohesion to employing more complex syntactic structures, such as embeddings and modifications, that create implicit conceptual links (Crossley et al., 2011; Haswell, 2000; McCutchen & Perfetti, 1982). For instance, Crowhurst (1987) observed a decline in the use of causal and temporal conjunctions in narrative texts from 6th to 12th grade, noting that older students use conjunctions more diversely but not necessarily in greater quantity.

Crossley et al. (2011) found that college freshmen's essays displayed more implicit cohesion than those of 11th graders, evidenced by less overlap of content words and fewer explicit logical connectives. Interestingly, 11th graders' essays were less cohesive than those of ninth graders. These results suggest that more proficient writers rely less on explicit local cohesion. Additionally, Haswell (1990) reported differences in givenness and the use of logical connectives between junior and freshman writers and between undergraduate and graduate writers, indicating continued development of cohesion strategies at higher academic levels.

Research on cohesion has focused more on L1 than L2 students over the past decades. Additionally, studies on L2 students have mostly examined cohesion in oral production rather than writing. For example, Crossley et al. (2010a) tracked the speech development of six L2 students over one year and found increased semantic similarity in their utterances. Another study by Crossley et al. (2010b) reported a decline in misunderstandings between L2 students and L1 speakers over a year, linked to growth in text and global cohesion measures such as causal particles, verbs, and semantic similarity, indicating that as L2 speech became more coherent, students used more global cohesive devices.

Longitudinal research on cohesion development in L2 writing remains limited, and existing findings suggest that cohesion growth is neither uniform nor linear. Although Yang and Sun (2012) did not adopt a longitudinal design, their cross-sectional comparison showed that more advanced Chinese EFL undergraduates used a greater range of local cohesion devices and produced fewer cohesion-related errors than less advanced peers. This pattern implies developmental sensitivity in the use of overt cohesive markers, particularly in argumentative writing.

Semester-long longitudinal studies provide a more direct view of change, yet their findings complicate a simple developmental narrative. For instance, Crossley et al. (2016) documented increases in global cohesion (e.g., noun overlap across paragraphs), broader text cohesion (e.g., content lemma type-token ratio), and semantic similarity across sentences and paragraphs, alongside greater bigram repetition. These results suggest that development may involve a shift toward stronger global integration and more systematic lexical patterning rather than merely an increased use of surface connectives.

At the same time, evidence indicates that cohesion development is genre-sensitive. Abdi Tabari and Johnson (2023), analyzing 270 narrative and argumentative essays written over a semester, found that cohesion was realized differently across genres: narrative writing relied more heavily on connective devices, whereas argumentative writing depended more on global-level lexical repetition. Although developmental patterns were less definitive, the study pointed to an expansion in the repertoire of cohesive resources over time. This suggests that cohesion growth cannot be understood apart from genre-specific rhetorical demands.

Subsequent longitudinal analyses further underscore the multidimensional nature of cohesion development. Abdi Tabari et al. (2024) reported increases in several text cohesion indicators (e.g., overall type-token ratio (TTR), noun TTR, function word TTR, proportion of content words), while Abdi Tabari and Wind (2025), drawing on the CDST framework, observed statistically significant growth across global, local, and text-level cohesion indices within a single semester. Together, these findings challenge the notion that development occurs at only one structural level and instead point to dynamic, interacting changes across multiple cohesion dimensions.

Moreover, cohesion development appears to be mediated by task-related factors. Abdi Tabari and Golparvar (2024) reported that pre-task planning and topic familiarity significantly shaped local and global cohesion patterns. Interestingly, planned writing involved fewer overt connectives but greater lexical overlap, indicating that development may reflect increasing conceptual integration rather than reliance on explicit cohesive markers. This finding complicates traditional assumptions that more connectives necessarily signal more advanced writing.

Overall, while L1 research suggests a gradual move away from overt local markers toward more global integration as writers mature, L2 findings remain mixed. Some studies highlight growth primarily in text-level measures, whereas others report simultaneous changes in global, local, and text cohesion. Importantly, genre, task conditions, and instructional context appear to interact with longitudinal development. Given the relatively small number of sustained longitudinal investigations, further research is needed to clarify how cohesion evolves over time and how developmental trajectories are shaped by genre and task-related variables in L2 academic writing.

2.2 The present study

Building on gaps identified in the literature, the present study seeks to deepen the understanding of the longitudinal development of cohesion in the writing of L2 learners enrolled in intensive EAP programs. Longitudinal research is defined as observations that are collected iteratively from the same learners for the same variable or variables over a specific amount of time (Barkaoui, 2014). Previous research has predominantly used static, two-wave designs that fail to capture the dynamic, non-linear nature of language development, especially within high-stakes academic writing contexts. Furthermore, while computational studies have provided valuable insights into cohesion use, they often neglect the instructional and task-specific demands unique to intensive EAP settings, as well as the fine-grained

changes occurring throughout the instructional period. To address these gaps, this study employs a multi-wave, time-series design that tracks the week-by-week development of cohesive devices in graduate students' writing over a five-week EAP course. Participants were recruited using convenience sampling due to the practical constraints of the program, including limited class size and accessibility to students who were actively enrolled in the course. This approach allowed for efficient data collection while ensuring ecological validity within the instructional context. By collecting and analyzing weekly writing samples, the study offers an ecologically valid and temporally sensitive perspective on cohesion development. Integrating macro-level group trends with micro-level individual analyses, the research aligns with the CDST approach, aiming to capture both shared patterns and individual variability in the use of cohesive devices. Focusing on three types of cohesion, local, global, and text level, the study addresses the following research questions (RQs):

RQ1. To what extent does local cohesion develop in the written production of graduate students throughout the five-week EAP course?

RQ2. To what extent does global cohesion develop in the written production of graduate students throughout the five-week EAP course?

RQ3. To what extent does text cohesion develop in the written production of graduate students throughout the five-week EAP course?

3. Methods

3.1 Participants

The study involved 58 graduate students enrolled in a five-week intensive EAP course during the summer session at a major university in the US. Of these, 30 students were from the business program and 28 from the engineering program. Participant recruitment followed the program's existing enrollment structure, as these two cohorts were the only graduate student groups offered the EAP course. This approach ensured that the study reflected the program's authentic instructional ecology rather than a selective sampling process, thereby enhancing the ecological validity of the findings.

All participants were first-year master's students placed into the course based on their TOEFL iBT and university placement test scores, which positioned them at the Advanced-Mid level on the ACTFL scale (or B2 on the CEFR scale). To maintain instructional consistency, all participants received 23 hours of instruction from the same instructor before data collection.

As shown in Table 1, the two groups were broadly comparable in age, proficiency, and academic experience but differed in disciplinary orientation and linguistic background. This balance between comparability and diversity provided a meaningful basis for examining how learners from distinct academic domains engage with EAP writing tasks while controlling for potential instructional and proficiency-related confounds.

Table 1. Participant demographics

Variable	Business (n = 30)	Engineering (n = 28)	Total (N = 58)
Gender	18 women, 12 men	9 women, 19 men	27 women, 31 men
Age (years)	M = 24.15, SD = 0.92	M = 24.03, SD = 0.88	Range = 23–27
TOEFL iBT score	72–94	72–94	72–94
University placement test (writing)	7–8 (out of 9)	7–8 (out of 9)	7–8 (out of 9)
Proficiency level	Advanced-Mid (ACTFL); B2 (CEFR)	Advanced-Mid (ACTFL); B2 (CEFR)	—
Years of English study	M = 8.76, SD = 2.03 (range 5–12)	M = 9.42, SD = 2.14 (range 6–14)	—
Previous writing courses	M = 3.98, SD = 1.72 (range 1–6)	M = 4.62, SD = 1.85 (range 2–7)	—
L1 distribution	Chinese (7), Spanish (6), Arabic (5), French (3), Other (9)	Chinese (12), Spanish (4), Hindi (3), Korean (2), Other (7)	Diverse multilingual sample
Instructional context	5-week intensive EAP course; 23 instructional hours	5-week intensive EAP course; 23 instructional hours	Same instructor for both groups

Note. All participants were first-year master's students enrolled in Business or Engineering programs at a major university in the US. The study design reflected the authentic enrollment structure of the EAP program rather than selective sampling.

3.2 Writing tasks

As part of the intensive five-week EAP writing program, students followed a structured sequence of discipline-specific tasks designed to develop academic and professional communication skills. Both groups began with analytical summaries of two academic articles, focusing on key idea extraction and rhetorical analysis. Subsequent tasks were tailored to the disciplinary focus of each group, building on this shared foundation (see the full descriptions of the EAP writing tasks in Appendix A). Table 2 summarizes the full sequence and nature of the writing tasks completed by the business and engineering students. This progression was designed to scaffold essential writing competencies, moving from foundational analysis to more complex, discipline-specific genres and communicative purposes.

Table 2. Sequence of writing tasks and targeted skills

Order	Task (Business)	Task (Engineering)	Key Skills Targeted	Learning Objectives
1	Analytical summary 1	Analytical summary 1	Critical reading, summarization	Extract key ideas, identify main arguments, and synthesize information from academic texts.
2	Analytical summary 2	Analytical summary 2	Critical reading, summarization	Develop accuracy and coherence in summarizing academic content
3	Cover letter (part of job application)	Extended definition	Professional correspondence (Business) / Academic writing (Engineering)	Refine formal writing for professional or technical contexts; practice discipline-specific conventions
4	Rhetorical analysis of a popular press business article	Rhetorical analysis of a popular press engineering article	Rhetorical evaluation, audience awareness	Analyze persuasive strategies and evaluate effectiveness for the target audience.
5	Extended definition of key business concept	Technical description and recommendation report	Conceptual explanation, structured argumentation, and visual communication	Communicate complex ideas clearly; integrate research, argumentation, and visual elements appropriate to the discipline.

3.3 Procedure

Participant access and data collection procedures

The course instructor was initially contacted to discuss the study's objectives and procedures. After obtaining approval, permission was granted to attend selected classes during the five-week summer program. Data collection spanned the entire term, ensuring that writing samples reflected students' authentic performance in a real academic setting. To safeguard authenticity and mitigate the potential use of generative AI tools, submission protocols were enforced. Students completed all assignments under firm deadlines within the university's learning management system, which tracked submission timestamps and document revision

histories to detect irregularities. Explicit instructions prohibited AI-based writing assistance, and students were reminded of academic integrity policies emphasizing original work. While external resources such as dictionaries, style manuals, and grammar guides were permitted to reflect genuine academic practices, these measures collectively ensured that the writing samples authentically represented each student's independent abilities.

Participant assessment and writing task sequence

To ensure comparability in English proficiency, participants from the business and engineering disciplines completed both the TOEFL *i*BT and the university placement test as pre-assessments. Their writing profiles were also reviewed to obtain a comprehensive understanding of their linguistic backgrounds, prior writing experience, and exposure to academic discourse. This information informed the interpretation of subsequent writing performance and allowed for meaningful comparisons across participants.

Throughout the program, students engaged in writing tasks organized within a structured weekly sequence that was carefully aligned with the university's core writing curriculum. The sequencing of tasks was designed to progressively scaffold key writing skills, ranging from idea generation and organization to drafting, revising, and editing, thereby supporting the development of both linguistic accuracy and rhetorical competence. Assignments were submitted according to designated deadlines, and participants consistently complied with timely submission, ensuring the collection of authentic, complete writing samples that accurately reflected their independent abilities. This approach not only maintained instructional consistency but also facilitated the analysis of writing development across the duration of the program.

Preprocessing and spelling corrections for cohesion analysis

After data collection, writing samples underwent manual spelling corrections during pre-processing to prepare texts for automated analysis. This step was critical to ensure the accurate computation of cohesion indices, which rely on the precise identification of lexical and grammatical markers signaling textual connectedness (Lu & Hu, 2024). Even minor spelling errors can disrupt the detection of cohesive devices by automated tools, including conjunctions, referential pronouns, and lexical repetitions, potentially causing systematic measurement errors that undermine validity and reliability (Tris & Yuan, 2025). Correcting spelling preserved the integrity of the analyses focused on local, global, and textual cohesion, enabling a valid and nuanced assessment of discourse-level writing development.

Summary of writing sample lengths

Among business students, analytical summaries averaged 323.63 words (*SD* = 134.98), with a range of 169 to 904 words. Cover letters averaged 332.07 words (*SD* = 41.37), ranging from 243 to 403 words. Rhetorical analyses of popular press business articles were longer, averaging 457.5 words (*SD* = 232.49) and ranging from 251 to 1,015 words. Extended

definition assignments were the most extensive, averaging 649.6 words (*SD* = 114.39), with samples ranging from 374 to 896 words. For engineering students, analytical summaries were shorter on average, at 257.71 words (*SD* = 92.5; range = 146–550). Extended definition assignments averaged 441.29 words (*SD* = 120.22), with a range of 274 to 851 words. Rhetorical analyses of popular press engineering articles averaged 351.21 words (*SD* = 146.66), ranging from 210 to 683 words. Technical description and recommendation reports were the longest, averaging 551.18 words (*SD* = 204.24), with a range of 198 to 739 words.

3.4 Cohesion indices

The selection of cohesion indices for this study was guided by prior research (Abdi Tabari & Wind, 2025; Crossley et al., 2016) and the goal of capturing distinct aspects of L2 text cohesion without redundancy. Table 3 presents the selected indices, which were calculated using the Tool for the Automatic Analysis of Cohesion 2.0.4 (TAACO; Crossley et al., 2016, 2019) and TAALED 1.4.1 (Kyle et al., 2021) for lexical diversity. Although TAACO 2.0.4 offers a broad range of cohesion metrics, some indices are highly correlated or share conceptual overlap. We, therefore, selected indices that each reflect unique characteristics of cohesion at local, global, and text levels.

Table 3. Selected cohesion indices

Level	Sub-type	Index	Tool	Rationale
Local	Lexical overlap	adjacent_overlap_all_sent	TAACO 2.0.4	Captures repetition of lemmas across adjacent sentences to indicate local continuity.
Local	Synonym overlap	syn_overlap_sent_noun	TAACO 2.0.4	Measures noun synonym continuity across sentences to reflect semantic cohesion.
Local	Semantic similarity	lsa_1_all_sent	TAACO 2.0.4	LSA-based similarity between adjacent sentences reveals underlying conceptual links beyond exact words.
Local	Connectives	all_connective	TAACO 2.0.4	Includes all rhetorical connectives to assess explicit sentence-level signaling of cohesion.
Global	Lexical overlap	adjacent_overlap_all_para	TAACO 2.0.4	Captures lemma overlap across paragraphs to indicate global textual cohesion.
Global	Synonym overlap	syn_overlap_para_noun	TAACO 2.0.4	Measures semantic continuity via noun synonyms between adjacent paragraphs.
Global	Semantic similarity	lsa_1_all_para	TAACO 2.0.4	LSA cosine similarity between paragraphs reflects conceptual cohesion at a higher discourse level.
Text	Type-token ratio & density	mtld_original_aw	TAALED 1.4.1	Measures lexical diversity across the whole text, providing a text-level cohesion indicator.
Text	Givenness	pronoun_density	TAACO 2.0.4	Assesses the degree to which writers assume information is known to the reader, reflecting global referential cohesion.

TAACO 2.0.4 reports potentially redundant indices; therefore, we chose indices that gauged unique cohesion characteristics.

Local cohesion indices

Four local cohesion indices were chosen to capture lexical and semantic continuity across adjacent sentences: lexical overlap, synonym overlap, semantic similarity, and connectives.

Lexical overlaps were measured by adjacent sentence overlap of all lemmas (i.e., the number of lemma types occurring in the next sentence). Synonym overlap reflects sentence-to-sentence overlap of noun synonyms, while semantic similarity is quantified as the average latent semantic analysis (LSA) cosine similarity between adjacent sentences, revealing underlying conceptual relationships across sentences. Connectives were measured using the all-connectives index, which includes all rhetorical connectors recognized by TAACO 2.0.4, reflecting explicit discourse signaling.

Global cohesion

Three indices were selected to assess cohesion at the paragraph level: lexical overlap, synonym overlap, and semantic similarity. Lexical overlap was operationalized as adjacent paragraph overlap of all lemmas, while synonym overlap captures noun synonym continuity across paragraphs. Semantic similarity was measured using LSA cosine similarity between adjacent paragraphs, indicating broader thematic and conceptual cohesion. These indices were chosen to complement local cohesion measures while avoiding redundancy.

Text cohesion

Cohesion at the text level was measured using type-token ratio (TTR) and lexical density, operationalized via the Measure of Textual Lexical Diversity (MTLD) in TAALED 1.4.1, and givenness, operationalized as pronoun density (number of third-person pronouns divided by total words). Givenness reflects the degree to which writers assume that information is already known or accessible to the reader, providing an additional perspective on text-level cohesion that is distinct from sentence- or paragraph-level measures. By selecting indices that target unique aspects of cohesion at multiple linguistic levels, this approach ensures a comprehensive yet non-redundant analysis of L2 writing cohesion.

3.5 Data analysis

We aimed to identify a common L2 developmental tendency in the data set, considering the time spent attending the course. Therefore, the data set was analyzed with individual growth trajectories via several generalized additive mixed models (GAMMs). In this way, insights into the relationship between cohesion and time were obtained using the *mgcv* package of *R* (R Core Team, 2021). A non-linear function of time is permitted by GAMMS during the analysis of nested dependencies in condensed data, as in our case with time intervals within trajectories (Kliesch & Pfenninger, 2021). GAMMs can give a more precise representation of the data structure when there are potential intercorrelations within clusters in the dataset. GAMMS are flexible as opposed to linear models that assume normality of residuals and linearity. The findings were plotted from the GAMM analyses by utilizing the itsadug package in *R* (van Rij et al., 2020).

Smoothing techniques are important in GAMMs to explore underlying trends by fitting smooth forms to the raw data. Smothers facilitate the modeling of complex non-linear

relationships and make it possible to describe data precisely. Therefore, smoothers can convey the direction of change with high effectiveness and provide a general picture of the overall developmental pattern. The effective degrees of freedom (EDF) show the complexity of the smooth term and reflect the non-linearity of the trajectory, whereas the F and p values assess the overall statistical significance, displayed in the output of the mgcv package. The results are unaffected by the property (linear or non-linear) of the relationship with the outcome variable. An EDF value greater than 1.000 indicates a non-linear trajectory, while a value of 1.000 suggests a linear trend. The dependent variables were linguistic measures (e.g., scores on various assessments), while time (data collection points) was treated as the fixed effect, with participants as random intercepts, allowing for a non-linear time function. The model specification for the GAMM analyses was implemented using the mgcv package. The full GAMM syntax is provided in Appendix B.

The mgcv package in *R* is a statistical package for fitting Generalized Additive Models (GAMs) and GAMMs. It is ideal for modeling smooth, non-linear relationships between variables, automatically selecting the degree of smoothness for every predictor. Key characteristics are automatic smoothness selection, a Bayesian approach to confidence intervals, support for various smooths, and functions for fitting models to relatively large datasets (bam).

Random effects, conforming to individual learners, can contribute to unique developmental trajectories due to different initial points. However, GAMMS can discern an overall trend for the development of cohesive devices across students. All analyses are accessible via OSF: https://osf.io/sk5x2/overview?view_only=11577b4d79084a888956df6f4f71d3ad.

4. Results

The extent to which participants developed cohesion was assessed using several cohesion indices. For brevity, only statistically significant results are reported here. Table 4 presents the descriptive statistics for these indices at the first (Time 1) and final (Time 5) measurement points.

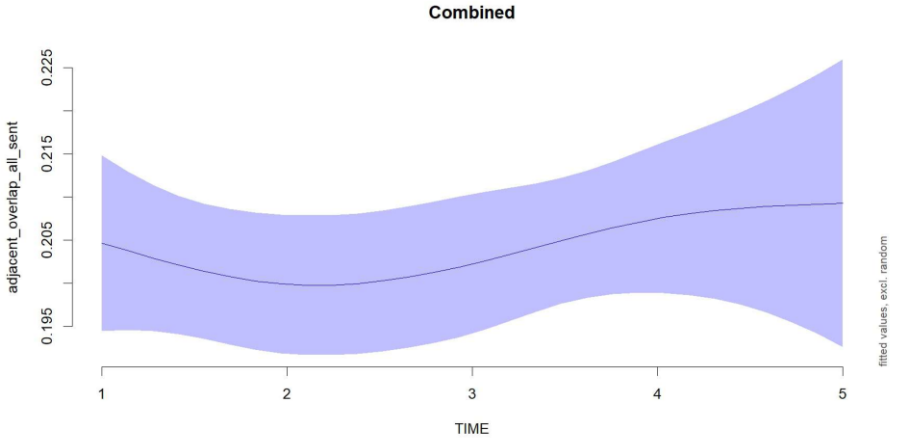
Table 4. Descriptive statistics

Index		Combined		Business		Engineering	
	Time	Mean	SD	Mean	SD	Mean	SD
adjacent_overlap_all_sent	1	0.209	0.039	0.197	0.039	0.221	0.036
	5	0.204	0.034	0.195	0.039	0.213	0.026
syn_overlap_sent_noun	1	1.091	0.612	0.703	0.32	1.507	0.579
	5	1.296	0.548	1.191	0.572	1.409	0.508
lsa_1_all_sent	1	0.342	0.102	0.302	0.072	0.386	0.112
	5	0.396	0.081	0.421	0.088	0.37	0.064
all_connective	1	0.071	0.011	0.07	0.007	0.073	0.013
	5	0.065	0.013	0.066	0.012	0.065	0.013
adjacent_overlap_all_para	1	0.253	0.09	0.252	0.034	0.254	0.126
	5	0.263	0.115	0.257	0.143	0.268	0.078
syn_overlap_para_noun	1	8.287	8.956	4.731	1.674	12.097	11.714
	5	16.9	11.972	15.471	10.458	18.432	13.433
lsa_1_all_para	1	0.513	0.185	0.477	0.112	0.551	0.236
	5	0.601	0.257	0.595	0.315	0.608	0.179
mtld_original_aw	1	74.101	21.552	86.446	15.922	60.874	18.908
	5	73.215	17.75	79.974	19.069	65.974	13.016
pronoun_density	1	0.02	0.013	0.01	0.006	0.031	0.01
	5	0.031	0.017	0.038	0.019	0.023	0.009

4.1 RQ1. To what extent does local cohesion develop in the written production of graduate students throughout the five-week EAP course?

Lexical overlap (Adjacent overlap all sentences)

Statistically significant positive relationships between time and adjacent overlap all sentences (AOAS) were not demonstrated by the GAMM model. The effective degrees of freedom indicated a non-linear relationship for the AOAS index for the combined cohort (EDF = 2.009, Ref.df. = 2.436, F = 0.396, p = 0.653) and for the engineering cohort (EDF = 1.273, Ref.df. = 1.493, F = 0.167, p = .810) but not for the business cohort (EDF = 1.000, Ref.df. = 1.000, F = 0.083, p = .773). In other words, no significant increases were detected in the number of lemma types that occur at least once in the next sentence in participants' essays over five weeks. Figure 1 shows the regression line of the AOAS index, and the summary of the GAMM model is presented in Table 5.



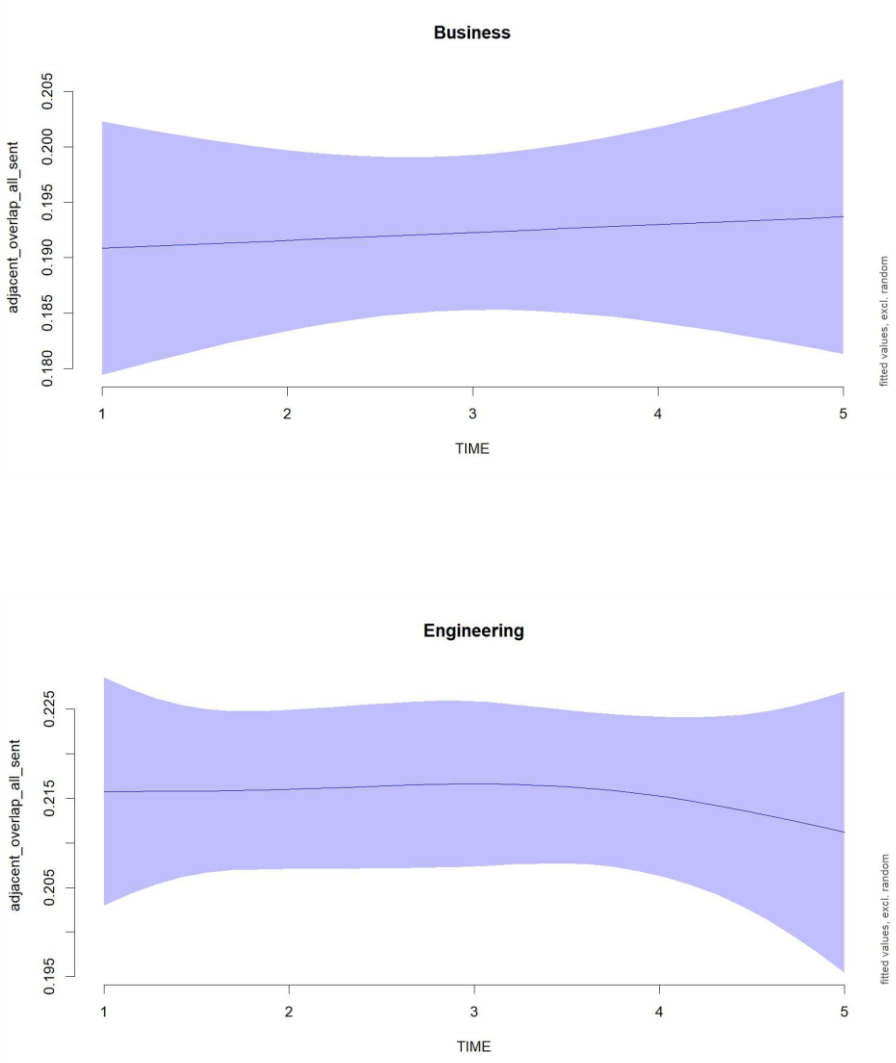


Figure 1: Group Tendency (AOAS)

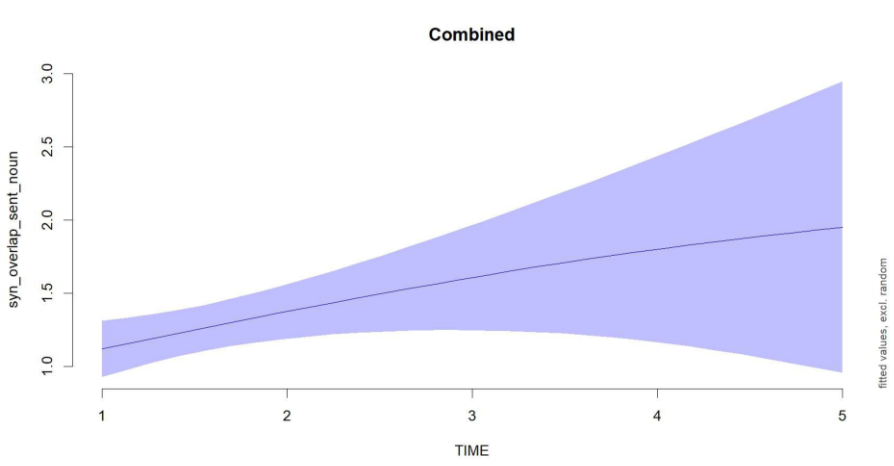
Table 5. Summary of the GAMM regarding AOAS

Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	0.205	0.003	70.53	<.001				
	Smooth terms								
	s(TIME)					2.009	2.436	0.396	0.653
	s(TIME, PAR)					1.000	1.000	0.331	0.565
Business	(Intercept)	0.192	0.003	55.32	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	0.083	0.773
	s(TIME, PAR)					1.000	1.000	0.673	0.413
Engineering	(Intercept)	0.215	0.003	64.28	<.001				
	Smooth terms								
	s(TIME)					1.273	1.493	0.167	0.810
	s(TIME, PAR)					1.000	1.000	0.703	0.403

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

Synonym overlap

Statistically significant positive relationships between time and synonym overlap sentence noun (SOSN) were demonstrated by the GAMM model for the business cohort, but not for the combined and engineering groups. The effective degrees of freedom indicated a non-linear relationship for the SOSN index for the combined cohort (EDF = 1.286, Ref.df. = 1.359, F = 4.665, p = 0.138) and for the engineering cohort (EDF = 1.653, Ref.df. = 2.028, F = 0.343, p = .734), but not for the business cohort (EDF = 1.000, Ref.df. = 1.000, F = 0.16.52, p < .001). In other words, a significant increase was detected in the average sentence-to-sentence overlap of noun synonyms in the participants' essays over five weeks. Figure 2 shows the regression line of the SOSN index, and the summary of the GAMM model is presented in Table 6.



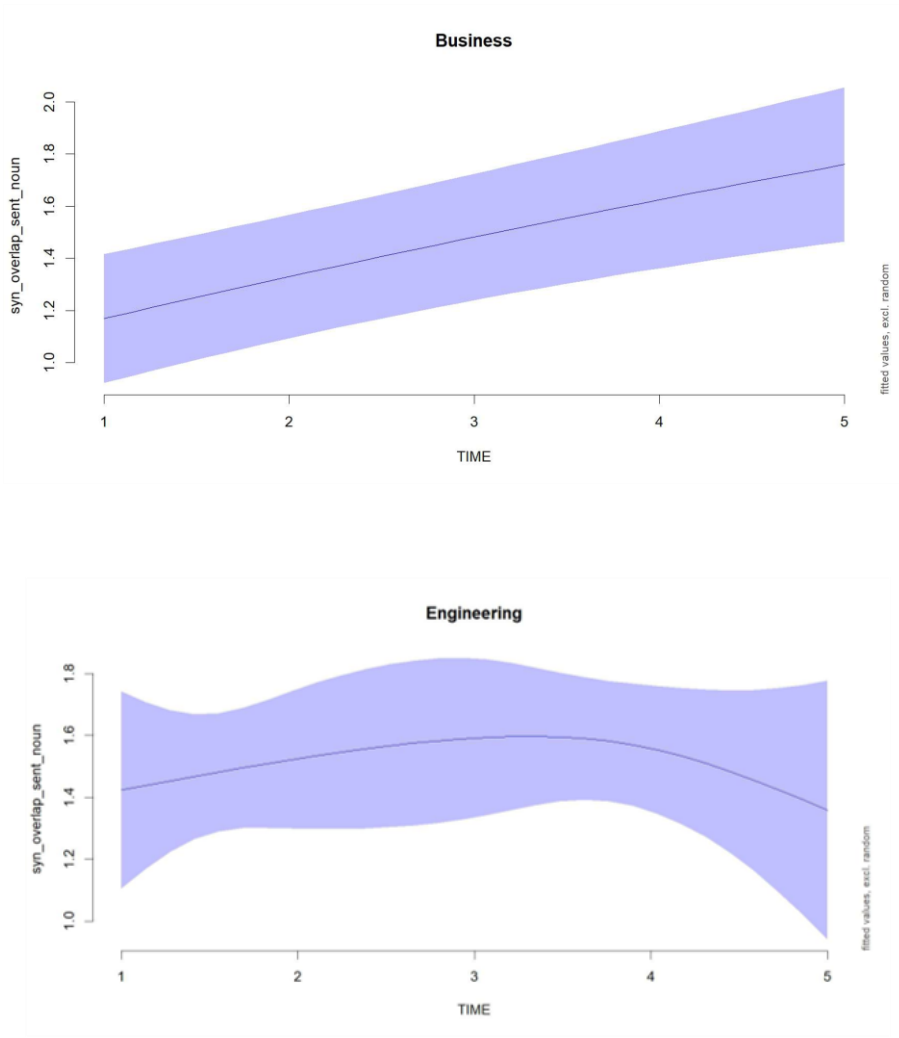


Figure 2: Group Tendency (SOSN)

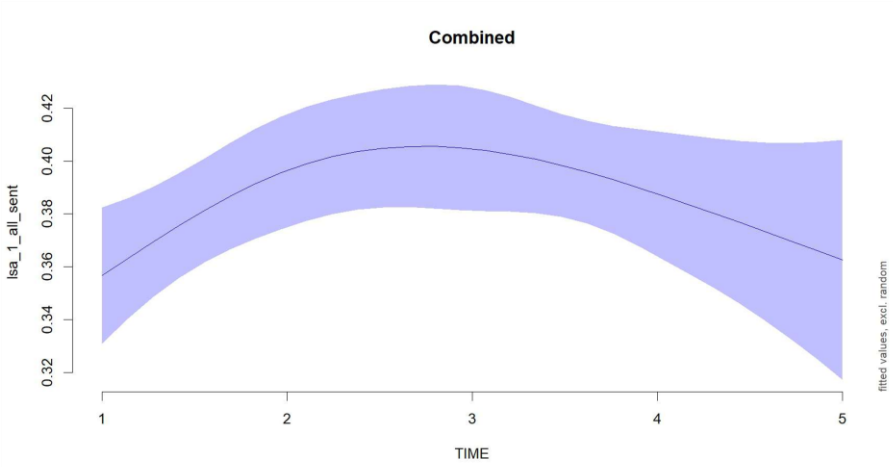
Table 6. Summary of the GAMM regarding SOSN

Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	1.289	0.044	29.14	<.001				
	Smooth terms								
	s(TIME)					1.286	1.359	4.665	0.138
	s(TIME, PAR)					2.121	2.442	1.644	0.135
Business	(Intercept)	1.1	0.061	17.92	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	16.52	<.001
	s(TIME, PAR)					2.534	2.869	2.972	0.02
Engineering	(Intercept)	1.515	0.065	23.32	<.001				
	Smooth terms								
	s(TIME)					1.653	2.028	0.343	0.734
	s(TIME, PAR)					1.000	1.000	0.439	0.509

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

Semantic similarity (LSA all sentences)

Statistically significant positive relationships between time and LSA in all sentences were demonstrated by the GAMM model for the combined and business cohorts, but not for the engineering cohorts. The effective degrees of freedom indicated a linear relationship for the LSA index for the business cohort (EDF = 1.000, Ref.df. = 1.000, F = 23.492, $p < .001$) and for the engineering cohort (EDF = 1.000, Ref.df. = 1.000, $F = 0.35$, $p = .555$), but not for the combined cohort (EDF = 2.371, Ref.df. = 2.75, $F = 2.797$, $p = .03$). In other words, significant increase was detected in the average latent semantic analysis cosine similarity between all adjacent sentences in the participants' essays over five weeks. Figure 3 shows the regression line of the LSA All sentences index, and the summary of the GAMM model is presented in Table 7.



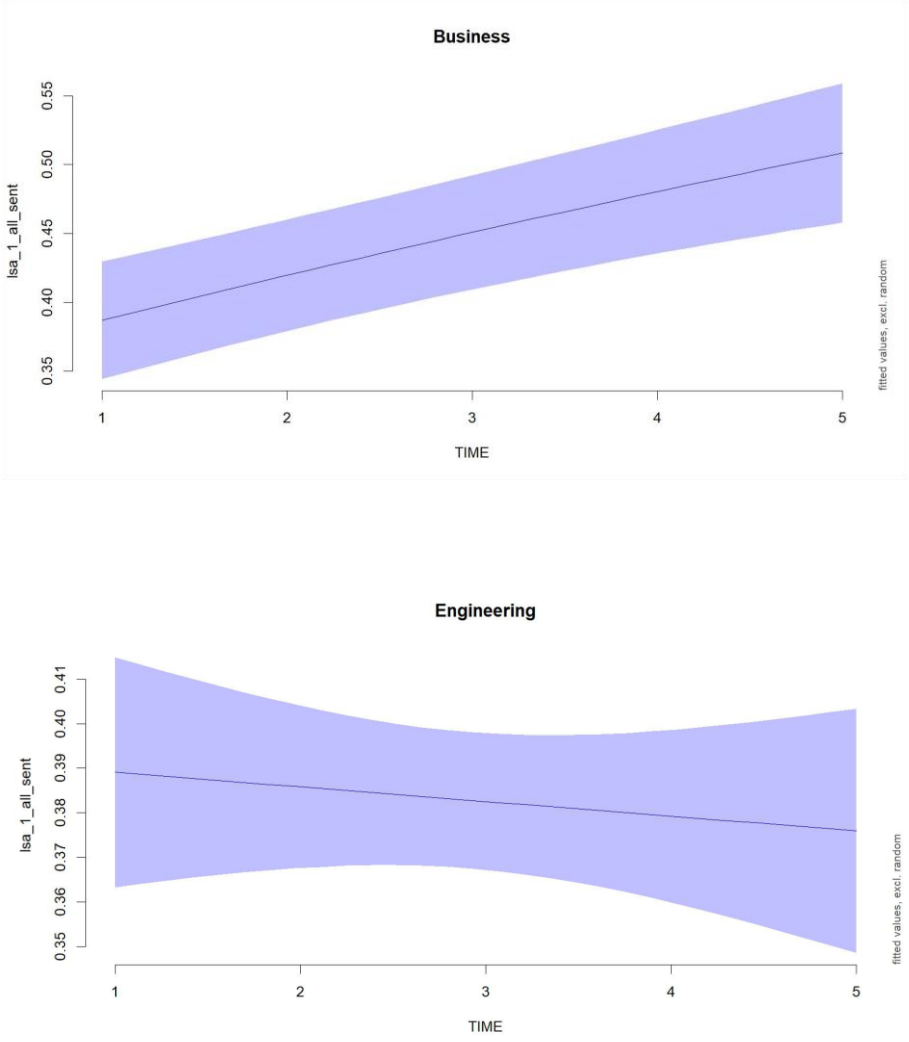


Figure 3: Group Tendency (LSA 1, all sentences)

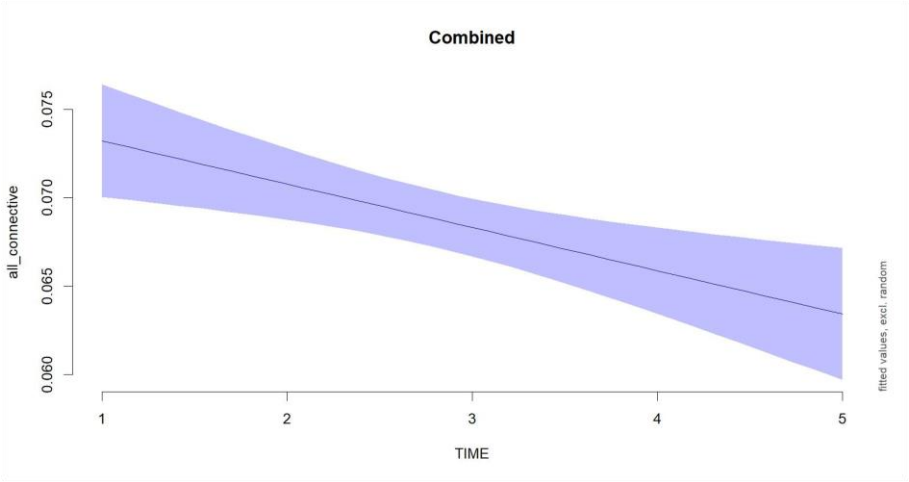
Table 7. Summary of the GAMM regarding LSA all sentences

Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	0.378	0.008	49.93	<.001				
	Smooth terms								
	s(TIME)					2.371	2.75	2.797	0.03
	s(TIME, PAR)					1.000	1.000	3.338	0.069
Business	(Intercept)	0.379	0.01	37.13	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	23.492	<.001
	s(TIME, PAR)					2.682	2.937	4.121	0.005
Engineering	(Intercept)	0.383	0.008	48.51	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	0.35	0.555
	s(TIME, PAR)					1.000	1.000	0.078	0.780

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

Connectives

A statistically significant negative relationship between time and all connective indices was demonstrated by the GAMM model for the combined but not for the business and engineering cohorts. The effective degrees of freedom indicated non-linear relationships for the all-connective index for the business cohort (EDF = 2.45, Ref.df. = 2.643, $F = 4.753$, $p = .055$) and for the engineering cohort (EDF = 1.031, Ref.df. = 1.062, $F = 3.198$, $p = .074$), but not for the combined cohort (EDF = 1.000, Ref.df. = 1.000, $F = 9.836$, $p = .002$). In other words, a significant decrease was detected in the number of all connectives in the participants' essays over five weeks. Figure 4 shows the regression line of the All connectives index, and the summary of the GAMM model is presented in Table 8.



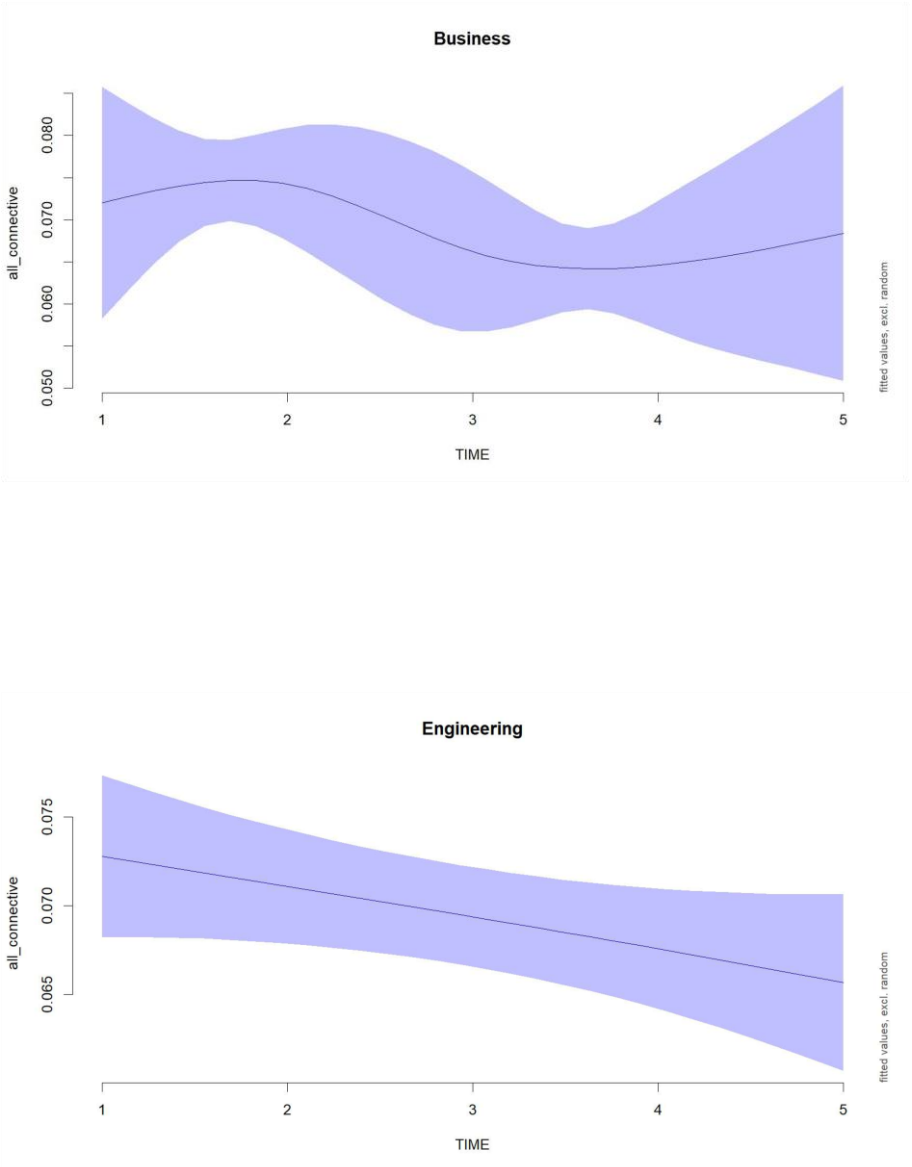


Figure 4: Group Tendency (All connectives)

Table 8. Summary of the GAMM regarding All connectives

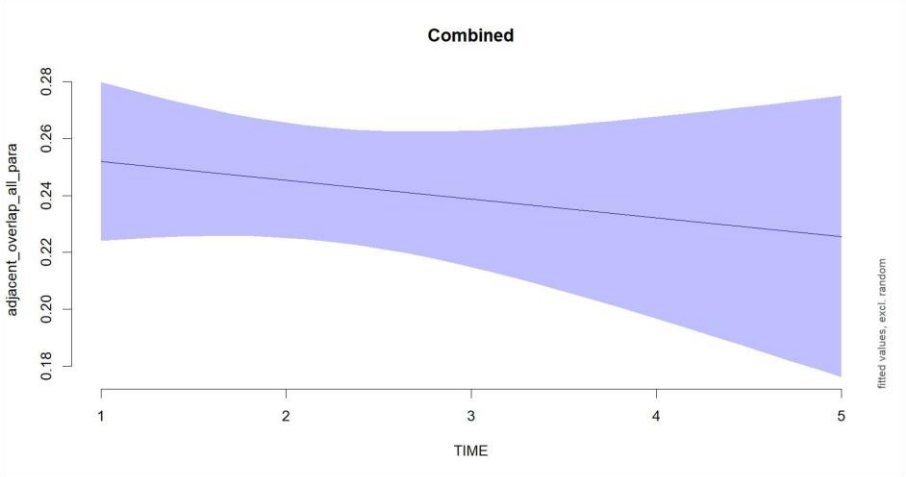
Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	0.068	0.001	74.89	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	9.836	0.002
	s(TIME, PAR)					1.000	1.000	1.514	0.22
Business	(Intercept)	0.068	0.001	54.54	<.001				
	Smooth terms								
	s(TIME)					2.45	2.643	4.753	0.055
	s(TIME, PAR)					1.251	1.297	0.012	0.957
Engineering	(Intercept)	0.07	0.001	49.74	<.001				
	Smooth terms								
	s(TIME)					1.031	1.062	3.198	0.074
	s(TIME, PAR)					1.000	1.000	0.071	0.791

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

4.2 RQ2. To what extent does global cohesion develop in the written production of graduate students throughout the five-week EAP course?

Lexical overlap

Statistically significant positive relationships between time and adjacent overlap all paragraphs (AOAP) were demonstrated by the GAMM model for the business and engineering, but not for the combined cohorts. The effective degrees of freedom indicated a non-linear relationship for the AOAP index for the business (EDF = 2.38, Ref.df. = 2.753, F = 4.758, p = 0.034) and engineering cohorts (EDF = 2.437, Ref.df. = 2.788, F = 5.183, p = .016) but not for the combined cohort (EDF = 1.000, Ref.df. = 1.000, F = 0.623, p = .431). In other words, no significant increases were detected in the number of lemma types that occur at least once in the next sentence in participants' essays over five weeks. Figure 5 shows the regression line of the AOAP index, and the summary of the GAMM model is presented in Table 9.



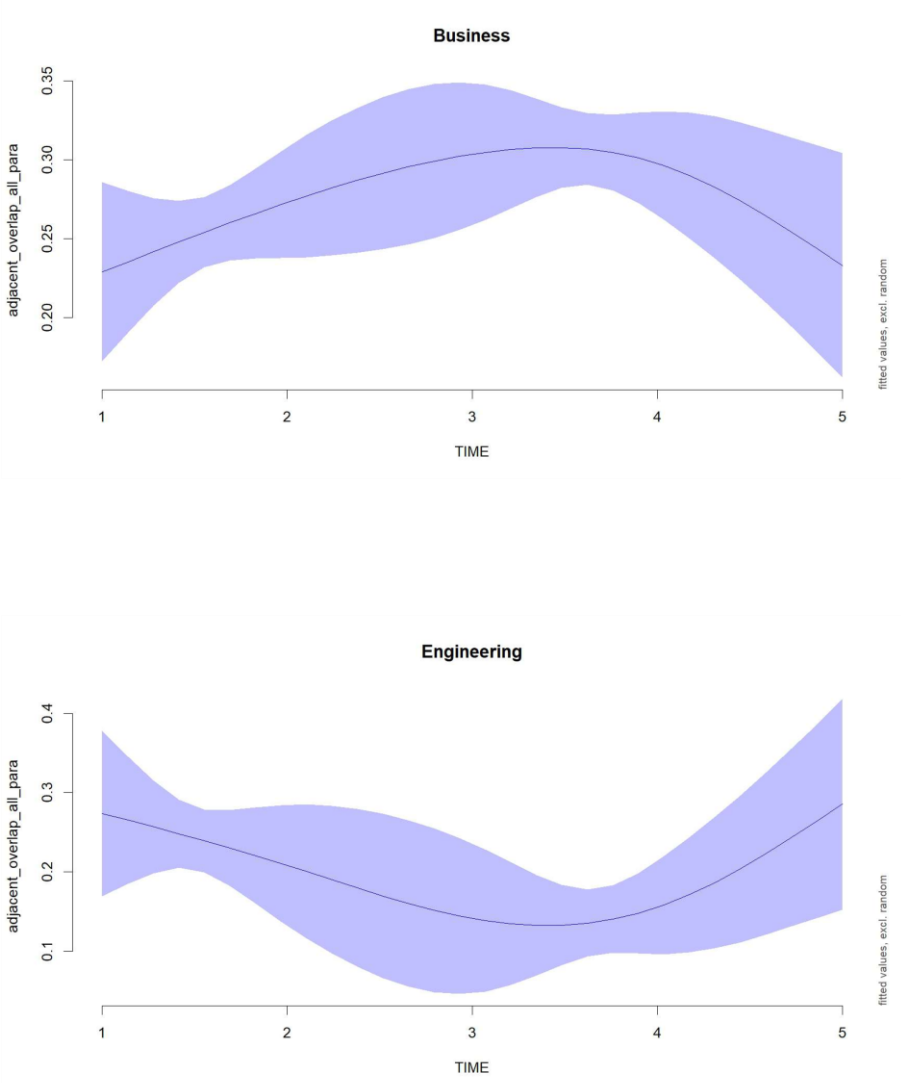


Figure 5: Group Tendency (AOAP)

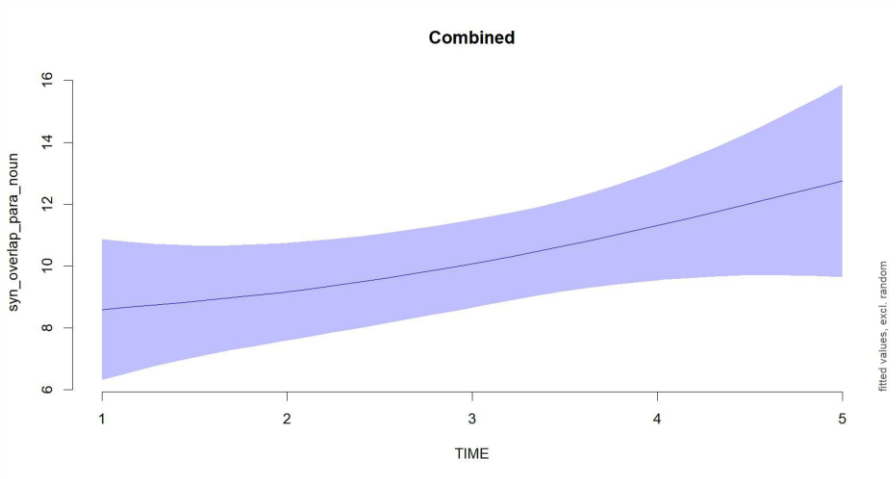
Table 9. Summary of the GAMM concerning AOAP

Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	0.237	0.008	29.96	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	0.623	0.431
	s(TIME, PAR)					1.425	1.709	1.674	0.286
Business	(Intercept)	0.274	0.007	41.2	<.001				
	Smooth terms								
	s(TIME)					2.38	2.753	4.758	0.034
	s(TIME, PAR)					1.000	1.000	0.768	0.382
Engineering	(Intercept)	0.209	0.016	13.06	<.001				
	Smooth terms								
	s(TIME)					2.437	2.788	5.183	0.016
	s(TIME, PAR)					1.000	1.000	0.144	0.705

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

Synonym overlap

Statistically significant positive relationships between time and average paragraph-to-paragraph overlap of noun synonyms (SOPN) were demonstrated by the GAMM model for the business and engineering disciplines, but not for the combined cohorts. The effective degrees of freedom indicated a non-linear relationship for the SOPN index for the combined (EDF = 1.335, Ref.df. = 1.594, $F = 18.13$, $p < .001$) and engineering cohorts (EDF = 2.602, Ref.df. = 2.89, $F = 5.883$, $p = .001$), but not for the business cohort (EDF = 1.000, Ref.df. = 1.000, $F = 49.111$, $p < .001$). In other words, significant increases were detected in the average paragraph-to-paragraph overlap of the noun synonyms index in the participants' essays over five weeks. Figure 6 shows the regression line of the SOPN index, and the summary of the GAMM model is presented in Table 10.



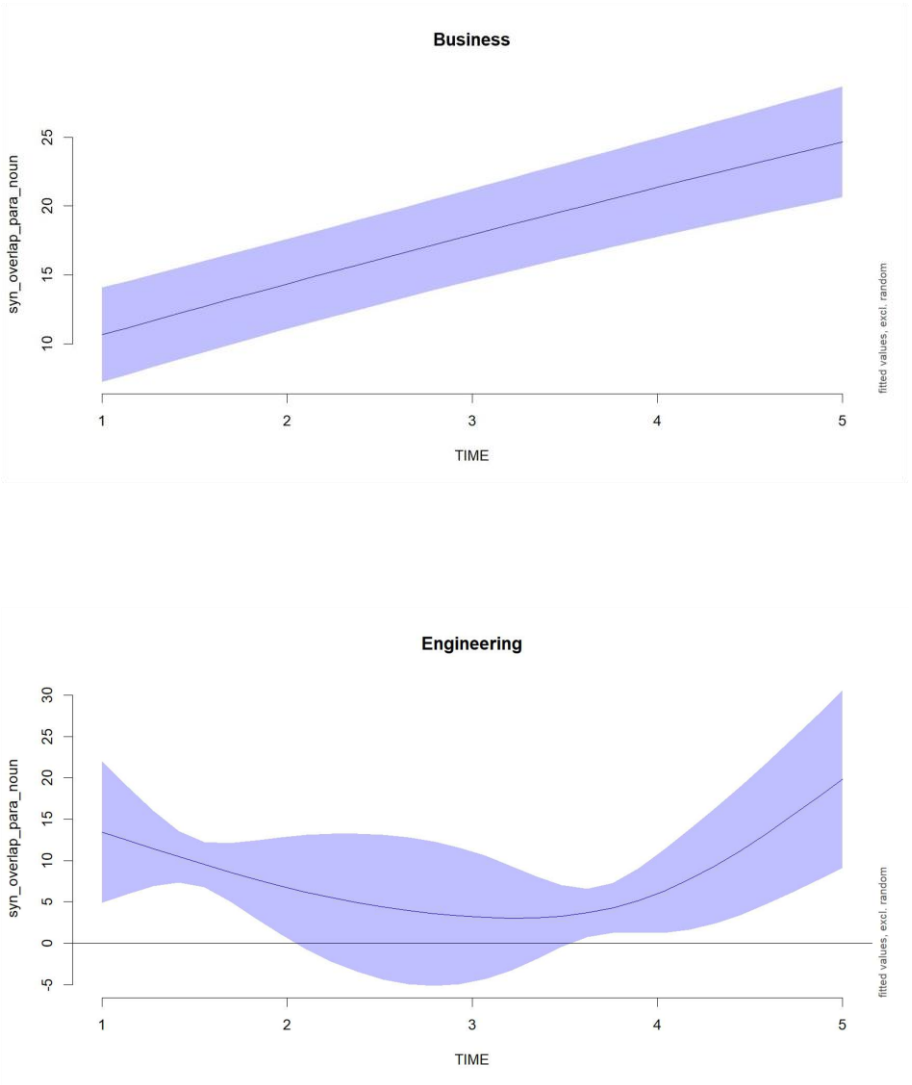


Figure 6: Group Tendency (SOPN)

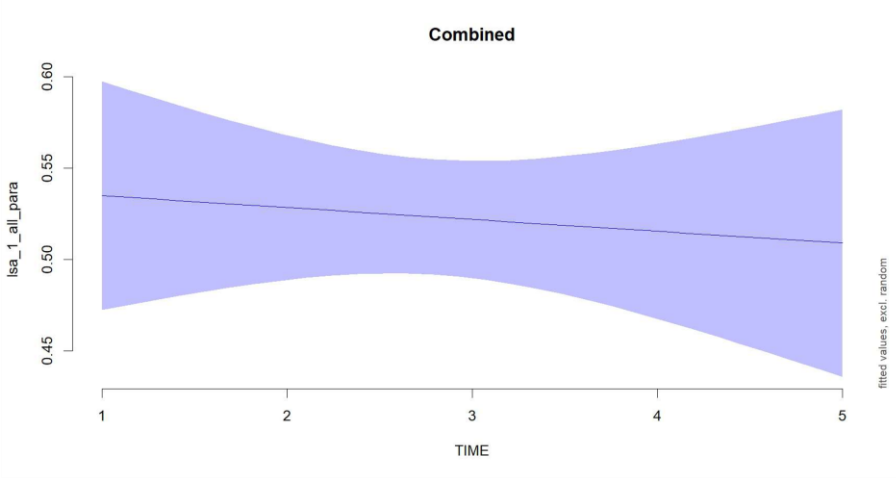
Table 10. Summary of the GAMM concerning SOPN

Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	10.407	0.574	18.13	<.001				
	Smooth terms								
	s(TIME)					1.335	1.594	2.107	0.128
	s(TIME, PAR)					1.000	1.000	7.263	0.007
Business	(Intercept)	12.113	0.829	14.62	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	49.111	<.001
	s(TIME, PAR)					2.675	2.933	7.355	<.001
Engineering	(Intercept)	9.729	1.158	8.399	<.001				
	Smooth terms								
	s(TIME)					2.602	2.89	5.883	0.001
	s(TIME, PAR)					1.000	1.000	0.117	0.733

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

Semantic similarity

Statistically significant positive relationships between time and LSA in all paragraphs were demonstrated by the GAMM model for the business and engineering disciplines, but not for the combined cohorts. The effective degrees of freedom indicated a linear relationship for the LSA index for the combined cohort (EDF = 1.000, Ref.df. = 1.000, F = 0.179, p = .672) and for the business cohort (EDF = 1.000, Ref.df. = 1.000, F = 18.82, p < .001), but not for the engineering cohort (EDF = 2.581, Ref.df. = 2.878, F = 7.575, p = .607). In other words, a significant increase was detected in the average latent semantic analysis cosine similarity between all adjacent paragraphs in the participants' essays over five weeks. Figure 7 shows the regression line of the LSA 1, all paragraphs index, and the summary of the GAMM model is presented in Table 11.



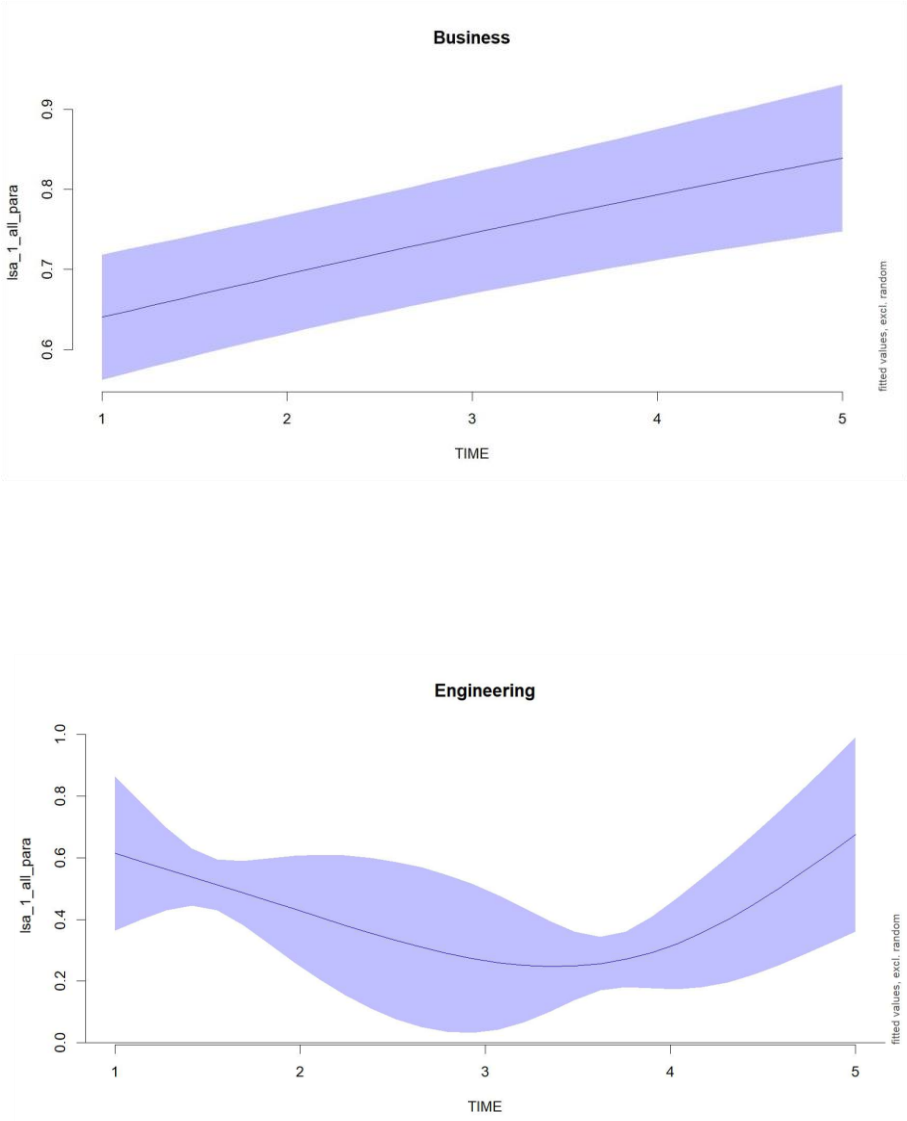


Figure 7: Group Tendency (LSA 1, all paragraphs index)

Table 11. Summary of the GAMM concerning LSA 1, all paragraphs

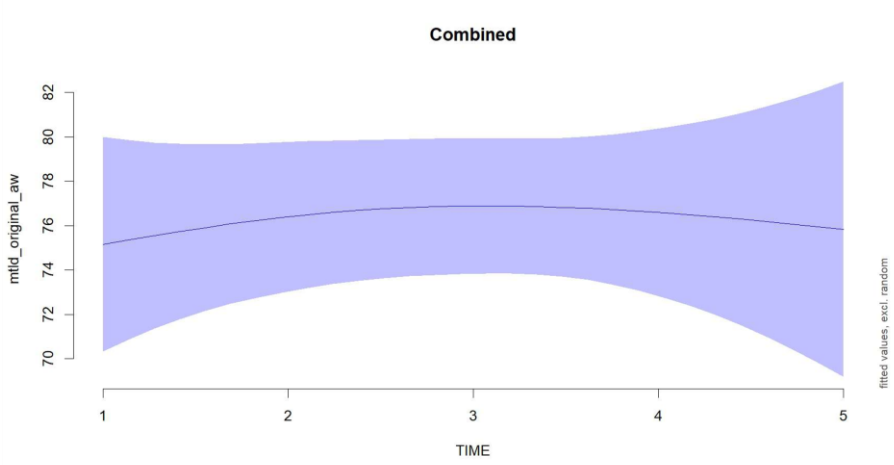
Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	0.518	0.017	30.14	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	0.179	0.672
	s(TIME, PAR)					1.000	1.000	9.518	0.039
Business	(Intercept)	0.62	0.018	34.34	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	18.82	<.001
	s(TIME, PAR)					2.656	2.928	6.73	<.001
Engineering	(Intercept)	0.433	0.029	15.15	<.001				
	Smooth terms								
	s(TIME)					2.581	2.878	7.575	.001
	s(TIME, PAR)					1.000	1.000	0.266	0.607

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

4.3 RQ3. To what extent does text cohesion develop in the written production of graduate students throughout the five-week EAP course?

Measure of textual lexical diversity

Statistically significant positive relationships between time and the measure of textual lexical diversity (MTLD) were not demonstrated by the GAMM model. However, statistically significant interaction effects were detected for the engineering cohort. The effective degrees of freedom indicated a linear relationship for the MTLD index for the business (EDF = 1.000, Ref.df. = 1.000, F = 0.99, p = .321) and engineering cohorts (EDF = 1.000, Ref.df. = 1.000, F = 1.55, p = 215), but not for the combined cohort (EDF = 1.361, Ref.df. = 1.635, F = 0.252, p = .773). In other words, a significant increase was not detected in the measure of textual lexical diversity in the participants' essays over five weeks. Figure 8 shows the regression line of the MTLD index, and the summary of the GAMM model is presented in Table 12.



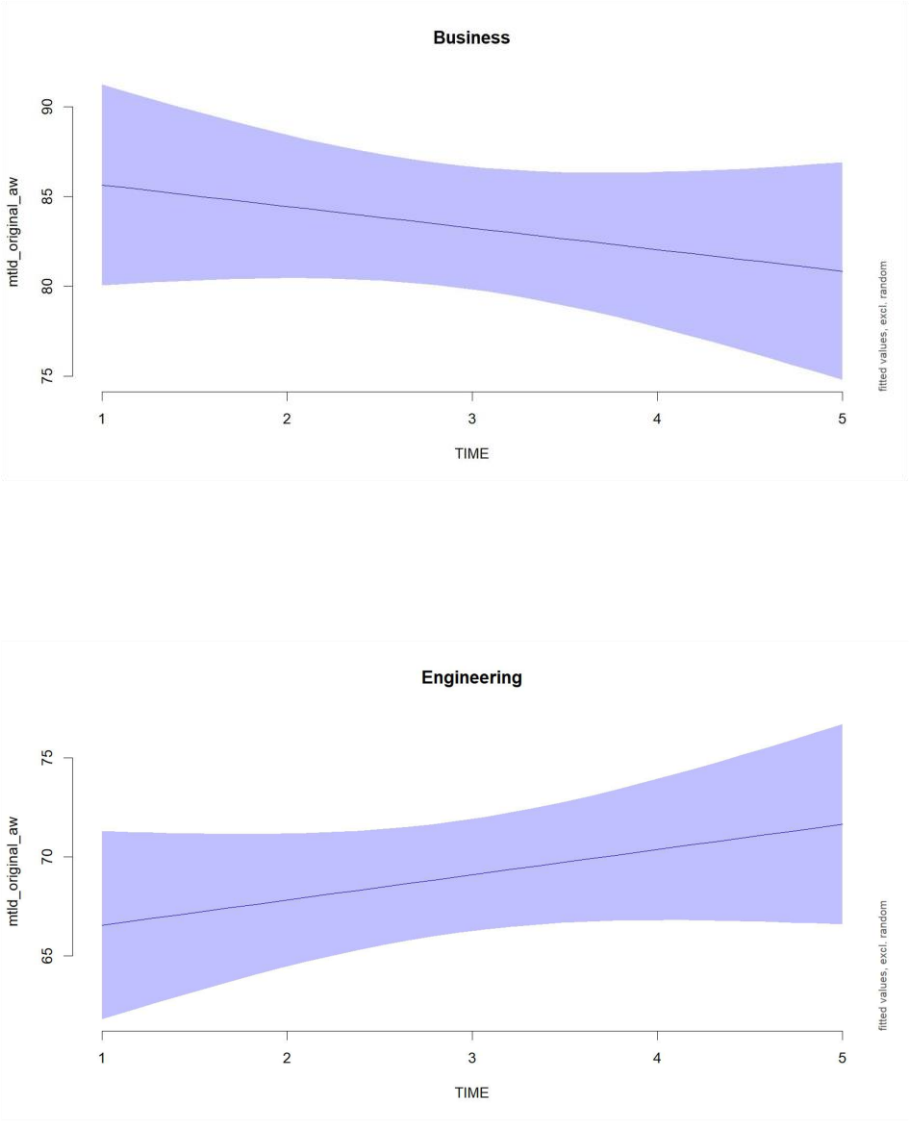


Figure 8: MTLD index regression line

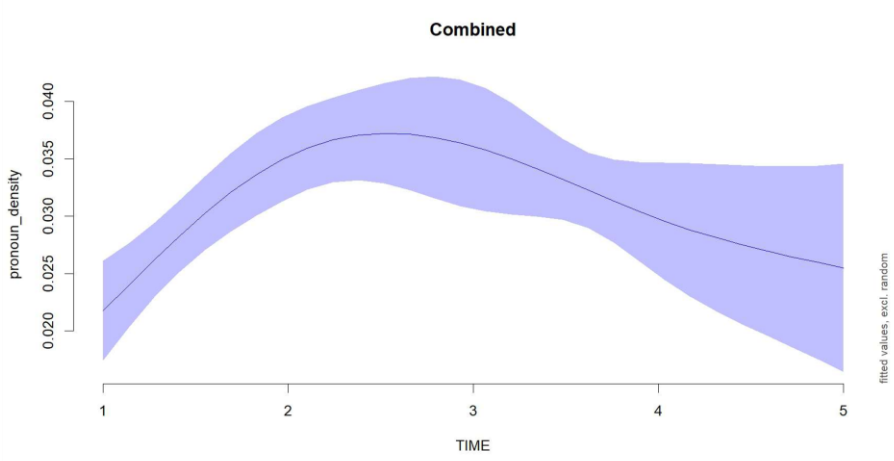
Table 12. Summary of the GAMM concerning MTLD

Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	76.251	1.368	55.75	<.001				
	Smooth terms								
	s(TIME)					1.361	1.635	0.252	0.773
	s(TIME, PAR)					1.000	1.000	0.367	0.545
Business	(Intercept)	83.413	1.698	49.13	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	0.99	0.321
	s(TIME, PAR)					1.000	1.000	0.088	0.767
Engineering	(Intercept)	69.042	1.455	47.44	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	1.55	0.215
	s(TIME, PAR)					1.000	1.000	0.126	0.002

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

Givenness

A statistically significant positive relationship between time and pronoun density was demonstrated by the GAMM model. The effective degrees of freedom indicated a non-linear relationship for the pronoun density index for the combined (EDF = 2.792, Ref.df. = 2.97, F = 9.686, p < .001) and engineering cohorts (EDF = 1.162, Ref.df. = 1.251, F = 4.126, p = .147), but not for the business cohort (EDF = 1.000, Ref.df. = 1.000, F = 77.21, p < .001). In other words, there was a significant increase in the number of third-person pronouns/number of words index in the essays over five weeks. Figure 9 shows the regression line of the pronoun density index, and the summary of the GAMM model is presented in Table 13.



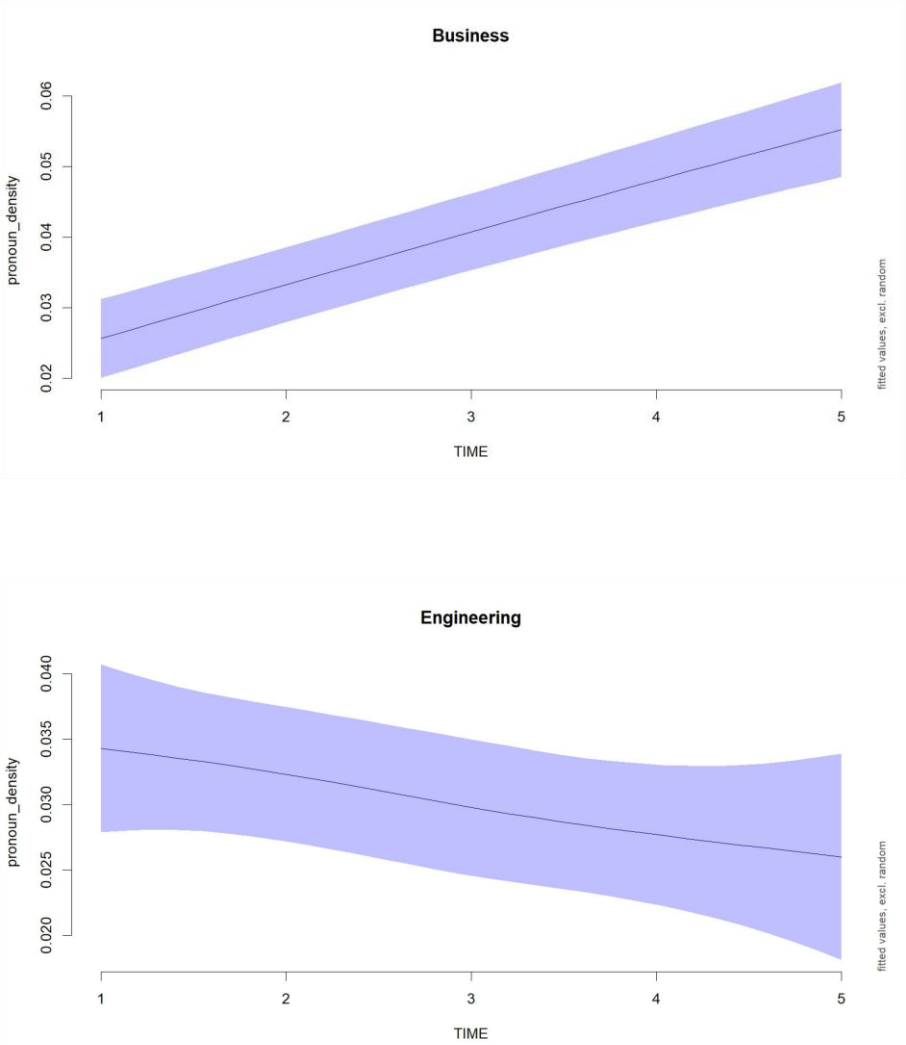


Figure 9: Group Tendency (pronoun density)

Table 13. Summary of the GAMM regarding pronoun density

Group	Parametric Coefficients	Estimate	Std. Error	*t*	Pr(>|t|)	Edf	Ref.df	F	*p*
Combined	(Intercept)	0.029	0.001	20.5	<.001				
	Smooth terms								
	s(TIME)					2.792	2.97	9.686	< .001
	s(TIME, PAR)					1.000	1.000	1.695	0.194
Business	(Intercept)	0.032	0.001	21.69	<.001				
	Smooth terms								
	s(TIME)					1.000	1.000	77.21	< .001
	s(TIME, PAR)					2.599	2.899	16.84	<. 001
Engineering	(Intercept)	0.029	0.001	21.78	<.001				
	Smooth terms								
	s(TIME)					1.162	1.251	4.126	0.147
	s(TIME, PAR)					1.682	2.037	0.126	0.863

Note. Edf = effective degrees of freedom, Ref.df = reference degrees of freedom.

5. Discussion

The present study examined cohesion development among Advanced-Mid L2 graduate learners in a five-week intensive EAP program, focusing on local, global, and text-level cohesion. The findings reveal nuanced, non-linear, and context-sensitive developmental trajectories, reflecting both group-level tendencies and individual variability across tasks and cohorts.

5.1 Development of local cohesion

Local cohesion, measured through lexical overlap (AOAS), synonym overlap (SOSN), semantic similarity (LSA cosine similarity), and connective use, displayed complex and cohort-specific trajectories. Lexical overlap did not increase consistently, whereas synonym overlap showed significant gains in the business cohort, suggesting a developmental shift from surface-level repetition to semantically related linking and deeper conceptual integration. This pattern mirrors L1 developmental trends, where writers increasingly rely on semantic cohesion rather than direct lexical repetition (Crossley et al., 2011; Haswell, 2000; McCutchen & Perfetti, 1982).

Semantic similarity increased significantly in the combined and business cohorts, indicating that sentences became conceptually more coherent even without higher lexical overlap. This finding supports He's (2014) distinction between lexical and semantic cohesion and extends L2 evidence showing measurable increases in semantic similarity among advanced EAP learners (Crossley et al., 2016). The observed growth within a brief, five-week program suggests that intensive, task-aligned instruction can foster rapid gains in semantic integration, particularly in analytical and rhetorical tasks.

Connective use also exhibited non-linear trends. Overall decreases in the all-connective index for the combined cohort suggest a shift from additive to more complex subordinating structures, reflecting enhanced syntactic control and more precise discourse signaling. Similar developmental progressions have been observed in L1 and advanced L2 writing, where maturing writers diversify connective functions while reducing overreliance on simple additive devices (Crowhurst, 1987; Haswell, 1990). The selective underuse of certain connectives may reflect cognitive load or stylistic adjustment, consistent with the CDST perspective on adaptive writing development (Abdi Tabari & Wind, 2025).

5.2 Development of global cohesion

At the paragraph level, increases in lexical overlap (AOAP) for extended definition and technical description tasks suggest strategic repetition for maintaining thematic continuity, while decreases in pronoun overlap point to a shift toward explicit referential expressions or nominalizations. This shift aligns with findings from longitudinal L2 research documenting that advanced writers gradually adopt more explicit referential strategies as cohesion mechanisms mature (Abdi Tabari et al., 2024; Crossley et al., 2016).

The development of global cohesion was closely tied to task type and instructional focus. Analytical and rhetorical tasks elicited higher levels of semantic and syntactic integration,

whereas descriptive and definitional tasks produced simpler, more formulaic patterns. These task-dependent variations support the view that cohesion operates as a context-sensitive construct—one that fluctuates in response to rhetorical demands and cognitive load. Consistent with previous studies (Abdi Tabari & Wind, 2025; Crossley et al., 2016), the results confirm that even short-term, intensive instruction can yield measurable growth in global cohesion when tasks require sustained thematic development and cross-paragraph integration.

5.3 Development of text-level cohesion

Text-level cohesion, assessed through lexical diversity (MTLD) and referential cohesion (pronoun density), revealed both inter-cohort consistency and individual variation. Increases in pronoun density in the combined and business cohorts indicate improved referential continuity aligned with the Given–New principle (Zhang & Kang, 2023). At the same time, growth in MTLD signals greater lexical sophistication and discourse flexibility.

Task type again influenced developmental patterns: analytical and rhetorical tasks encouraged richer lexical diversity and more cohesive referential tracking, while definitional and descriptive tasks produced constrained lexical patterns. These task-dependent outcomes underscore the adaptive nature of text cohesion development, reinforcing the CDST view that writing proficiency evolves through continuous, context-sensitive adjustments across interconnected linguistic subsystems (Abdi Tabari & Wind, 2025).

Across all three levels (local, global, and text), cohesion development emerged as dynamic, non-linear, and multifaceted. Advanced L2 writers demonstrated measurable progress even within a short instructional period, transitioning from lexical to semantic and syntactic cohesion strategies while varying their approach according to task demands. The findings align with L1 developmental trajectories (Crossley et al., 2011; Haswell, 2000) and extend previous L2 studies (Abdi Tabari et al., 2024; Crossley et al., 2016) by showing that intensive, task-focused EAP instruction can prompt cohesive gains across multiple linguistic dimensions in as little as five weeks.

Viewed from the CDST perspective, these results illustrate that cohesion development in advanced L2 writing is not a linear accumulation of features, but an emergent process shaped by interaction among cognitive, instructional, and contextual variables. Small shifts in cohesion strategies, whether lexical, semantic, or referential, can cascade into broader improvements in text organization and coherence, underscoring the adaptive nature of advanced academic writing development.

5.4 Intensive EAP instruction and cohesion development

A key contribution of this study is that it provides empirical evidence that cohesion development can be meaningfully observable within an intensive EAP context, rather than simply treated as a program descriptor. Unlike semester-long EAP formats, the present five-week program compressed instructional exposure, writing production, and feedback cycles into a short timeframe, creating repeated opportunities for learners to adjust cohesion

strategies across multiple genres. The observed increases in semantic similarity, synonym overlap, and lexical diversity, despite relatively limited change in surface-level repetition, suggest that intensive instruction may facilitate accelerated shifts toward conceptually driven cohesion, particularly when learners engage in frequent task production and receive sustained rhetorical input. Importantly, the non-linear patterns observed across cohorts and tasks indicate that intensive instruction does not uniformly strengthen all cohesion features; instead, it appears to promote selective restructuring, whereby learners reduce reliance on explicit connective marking and increasingly rely on semantic and syntactic mechanisms to maintain continuity. These findings suggest that intensive EAP programs may be especially conducive to rapid reorganization of cohesion resources, although such development remains sensitive to task demands and instructional sequencing. From a CDST perspective, cohesion growth in intensive contexts is, therefore, better understood not as fast linear improvement, but as accelerated, task-mediated adaptation across interacting linguistic subsystems.

6. Conclusion

6.1 Summary of key findings

This study provides a detailed account of cohesion development in advanced L2 graduate learners over a five-week intensive EAP course. Across local, global, and text-level measures, learners exhibited a shift from surface-level cohesion strategies, such as explicit connective marking, toward more conceptually driven cohesion reflected in increased semantic similarity and synonym-based overlap. In particular, semantic similarity increased over time, while overall connective density declined, suggesting that learners relied less on overt discourse markers and more on implicit meaning-based continuity. Gains were also observed in several paragraph-level indices (e.g., synonym overlap and semantic similarity), although these patterns were not fully consistent across cohorts. At the same time, certain aspects of cohesion requiring longer-range discourse integration remained comparatively underdeveloped, indicating that cohesion development in intensive instructional settings may be uneven and non-linear. Taken together, these findings illustrate both advancement and asymmetry in learners' cohesion development and support a CDST-informed view of cohesion as a dynamic, multidimensional construct.

Pedagogically, these findings suggest that cohesion instruction in intensive EAP writing courses may need to move beyond an emphasis on overt discourse markers toward a stronger focus on meaning-based cohesion development. The observed decline in connective use alongside gains in semantic similarity indicates that learners increasingly relied on conceptual continuity rather than explicit transitional signaling. Accordingly, instead of treating cohesion primarily as a matter of inserting connectives, instructors may need to scaffold learners' ability to sustain semantic continuity through controlled paraphrasing, synonym selection, lexical chaining, and thematic progression across sentences. Moreover, the uneven development observed at the paragraph level highlights the importance of explicit instruction in how cohesion is maintained across larger discourse units, particularly when learners shift between

genres common in EAP curricula (e.g., summary, critique, and argumentation). Teachers may therefore incorporate genre-based modeling and guided activities such as reverse outlining, paragraph mapping, and topic-tracing tasks to help students maintain continuity across paragraphs and rhetorical sections. Such pedagogy may support learners in developing cohesion as a discourse-level competence rather than a sentence-level editing strategy, aligning cohesion instruction more closely with the conceptual demands of graduate-level academic writing.

6.2 Limitations

Several limitations should be considered when interpreting the findings of this study. First, the participants were drawn from a convenience sample of Advanced-Mid EAP students in a single institutional context. While this sample allowed for close tracking of developmental change within an intact instructional setting, it limits the generalizability of the findings to learners at other proficiency levels, in different institutional environments, or across broader L2 populations. Second, the study spanned five weeks. Although this timeframe was sufficient to capture short-term developmental trajectories, it does not allow for strong claims about longer-term growth, stabilization, or possible regression patterns. Future longitudinal research extending across a semester or academic year would provide a more comprehensive picture of cohesion development over time.

Third, task type and time were fully confounded in the research design. All participants completed the same sequence of genre-based writing tasks in a fixed curricular order, such that each task corresponded to a unique time point. As a result, task effects cannot be statistically disentangled from longitudinal development. Any observed differences may reflect task demands, cumulative instructional exposure, increasing genre familiarity, or developmental change. Because these influences overlap in the present design, they cannot be independently estimated. Future studies employing counterbalanced or randomized task orders would be better positioned to isolate task-specific effects from temporal change.

Fourth, although the automated cohesion indices used in this study provide systematic and fine-grained linguistic measures, they cannot fully capture the functional appropriateness or rhetorical effectiveness of cohesive devices in context. Integrating computational indices with expert ratings or discourse-analytic approaches would allow for a more nuanced understanding of how cohesion supports meaning-making in academic writing.

Finally, disciplinary background was not modeled as a separate variable. While all participants were enrolled in EAP courses, differences in academic majors may shape genre expectations and cohesion practices. Examining field-specific developmental trajectories in future research could clarify whether cohesion growth patterns vary across disciplinary contexts.

Overall, these limitations underscore the contextual and design constraints of the study and point to important directions for future research aimed at refining and extending the present findings.

6.3 Directions for future research

Future research can triangulate quantitative cohesion measures with qualitative methods, such as think-aloud protocols, discourse analysis, or interviews, to better understand cognitive and instructional mechanisms underlying cohesion development. Longitudinal studies spanning multiple genres and tasks could clarify how cohesion strategies evolve in response to diverse academic demands. Cross-linguistic investigations could examine the role of L1 transfer and typological differences in discourse management, including referential tracking, thematic continuity, and information structuring. Additionally, exploring discipline-specific writing conventions may reveal nuanced patterns of contextualized cohesion development.

Overall, this study provides empirical evidence that cohesion in advanced L2 writing develops in a dynamic, emergent, and task-sensitive manner, even within a condensed instructional period. By examining the trajectories of local, global, and text-level cohesion, it contributes to an in-depth understanding of how advanced learners refine discourse strategies in response to cognitive, task-related, and instructional constraints.

References

Abdi Tabari, M., & Johnson, M. D. (2023). Exploring new insights into the role of cohesive devices in written academic genres. *Assessing Writing, 57*, 100749. https://doi.org/10.1016/j.asw.2023.100749

Abdi Tabari, M., Johnson, M. D., & Farahanynia, M. (2023). Automated analysis of cohesive features in L2 writing: Examining effects of task complexity and task repetition. *Assessing Writing, 58*, 100783. https://doi.org/10.1016/j.asw.2023.100783

Abdi Tabari, M., & Wind, A. M. (2025). Dynamic development of cohesive devices in English as a second language writing. *International Review of Applied Linguistics in Language Teaching, 63*(1), 433–469. https://doi.org/10.1515/iral-2022-0205

Abdi Tabari, M., & Golparvar, S. E. (2024). Cohesion in L2 writing: Assessing the role of pre-task planning and topic familiarity through a task-readiness lens. *The Language Learning Journal,* 1–21. https://doi.org/10.1080/09571736.2024.2437370

Abdi Tabari, M., Johnson, M. D., & Gao, J. (2024). Using automated indices of cohesion to explore the growth of cohesive features in L2 writing. *International Review of Applied Linguistics in Language Teaching, 63*(3), 2169–2200. https://doi.org/10.1515/iral-2023-0185

Barkaoui, K. (2014). Quantitative approaches for analyzing longitudinal data in second language research. *Annual Review of Applied Linguistics, 34,* 65–101. https://doi.org/10.1017/S0267190514000105

Bereiter, C., & Scardamalia, M. (1987). The psychology of written communication. Lawrence Erlbaum.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing, 32*, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The tool for the automatic analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behavioral Research Methods, 51*(1), 14–27. https://doi.org/10.3758/s13428-018-1142-4

Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd annual conference of the cognitive science society (pp. 984–989). Cognitive Science Society.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2010a). The development of semantic relations in second language speakers: A case for Latent Semantic Analysis. *Vigo International Journal of Applied Linguistics, 7*, 55–74.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2010b). The role of lexical cohesive devices in triggering negotiations for meaning. Issues in Applied Linguistics, 18(1), 55–80. https://escholarship.org/uc/item/0n180736

Crossley, S. A., & McNamara, D. S. (2011). Text coherence and judgments of essay quality: Models of quality and coherence. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), Proceedings of the 29th annual conference of the cognitive science society (pp. 1236–1241). Cognitive Science Society.

Crossley, S. A., Roscoe, R. D., McNamara, D. S., & Graesser, A. (2011). Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), Proceedings of the 15th international conference on artificial intelligence in education (pp. 438–440). Springer.

Crowhurst, M. (1987). Cohesion in argument and narration at three grade levels. *Research in the Teaching of English, 21*(2), 185–201. https://doi.org/10.58680/rte198715585

Fitzgerald, J., & Spiegel, D. L. (1986). Textual cohesion and coherence in children's writing. *Research in the Teaching of English, 20*(3), 263–280. https://doi.org/10.58680/rte198615606

Fogal, G. G. (2022). Second language writing from a complex dynamic systems perspective. *Language Teaching, 55*(2), 193–210. https://doi.org/10.1017/s0261444821000410

Foltz, P. W. (2007). Discourse coherence and LSA. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), Handbook of latent semantic analysis (pp. 167–184). Lawrence Erlbaum.

Gernsbacher, M. A. (1990). Language comprehension as structure building. Lawrence Erlbaum.

Godfrey, J. (2024). Prompt and circumstance: Investigating the relationship between college writing and postsecondary policy [Unpublished doctoral dissertation], University of Michigan.

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. Longman.

Haswell, R. H. (1990). Change in undergraduate and post-graduate writing performance (part 2): Problems in interpretation (ERIC No ED323537). ERIC Institute of Education Sciences. Retrieved from http://eric.ed.gov/?id=ED323537i

Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication, 17*(3), 307–352. https://doi.org/10.1177/0741088300017003001

He, Q. (2014). A study of lexical cohesion theory in reading comprehension. International Journal of English Linguistics, 4(6). https://doi.org/10.5539/ijel.v4n6p143

Johnson, M. D., & Abdi Tabari, M. (2024). Linguistic complexity in second language writing: Insight from studies on task planning. *Language Teaching, 58*(1), 1–13. https://doi.org/10.1017/s0261444824000284

King, M. L., & Rentel, V. (1979). Toward a theory of early writing development. *Research in the Teaching of English, 13*(3), 243–253. https://doi.org/10.58680/rte201117862

Kliesch, M., & Pfenninger, S. E. (2021). Cognitive and socioaffective predictors of L2 microdevelopment in late adulthood: A longitudinal intervention study. *The Modern Language Journal, 105*(1), 237–266. https://doi.org/10.1111/modl.12696

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity using direct judgments. *Language Assessment Quarterly, 18*(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141–165. https://doi.org/10.1093/applin/18.2.141

Lowie, W. M., & Verspoor, M. H. (2019). Individual differences and the ergodicity problem. *Language Learning, 69*(S1), 184–206. https://doi.org/10.1111/lang.12324

Lu, X., & Hu, R. (2024). Sense-aware connective-based indices of cohesion and their relationship to cohesion ratings of English language learners' written production. *Studies in Second Language Acquisition, 46*(3), 644–662. https://doi.org/10.1017/s0272263124000202

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing, 29*, 3–15. https://doi.org/10.1016/j.jslw.2015.06.004

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. Journal of Memory and Language, 25(4), 431–444. https://doi.org/10.1016/0749-596X(86)90036-7

McCutchen, D., & Perfetti, C. A. (1982). Coherence and connectedness in the development of discourse production. *Text-Interdisciplinary Journal for the Study of Discourse, 2*(1–3), 113–140. https://doi.org/10.1515/text.1.1982.2.1-3.113

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1–43. https://psycnet.apa.org/doi/10.1207/s1532690xci1401_1

Myhill, D. (2008). Towards a linguistic model of sentence development in writing. *Language and Education, 22*(5), 271–288. https://doi.org/10.1080/09500780802152655

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*(2), 121–152. https://psycnet.apa.org/doi/10.1207/s15326950dp4302_2

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rentel, V. M., King, M. L., Pettegrew, B., & Pappas, C. (1983). A longitudinal study of coherence in children's written narratives. The Ohio State University, Research Foundation.

Saggion, H. (2017). Automatic text simplification. Morgan & Claypool Publishers.

Schmid, H.-J. (2020). The dynamics of the linguistic system: Usage, conventionalization, and entrenchment. Oxford University Press. https://doi.org/10.1093/oso/9780198814771.001.0001

Tris, D. C., & Yuan, Q. (2025). ESL academic writing in higher education: The impact of artificial intelligence. In D. C. Tris & Q. Yuan (Eds.), Transformations in digital learning and educational technologies (pp. 293–324). IGI Global Scientific Publishing.

van Rij, J., Wieling, M., Baayen, R. H., van Rijn, H., & van Rij, M. J. (2020). itsadug: Interpreting time series and autocorrelated data using GAMMs (R package; Version 2.4) [Computer software].

Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage-based perspective on L2 writing. *Journal of Second Language Writing, 21*(3), 239–263. https://doi.org/10.1016/j.jslw.2012.03.007

Wei, X. (2023). Exploring the cross-language transfer of L1 rhetorical knowledge in L2 writing: Cognitive and metacognitive perspectives. Springer Nature. https://doi.org/10.1007/978-981-99-7637-9

Wind, A. M. (2021). Nonlinearity and inter- and intra-individual variability in the extent of engagement in self-reflection and its role in second language writing: A multiple-case study. *System, 103*, Article 102672. https://doi.org/10.1016/j.system.2021.102672

Wind, A. M., & Verspoor, M. H. (2026). Variability in second language development. In H. Nesi & P. Milin (Eds.), International Encyclopedia of Language and Linguistics (3rd ed., pp. 0–0). Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.01401-0

Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education, 23*(1), 31–48. https://doi.org/10.1016/j.linged.2011.09.004

Yasuda, S. (2012). The implementation of genre-based tasks in foreign language writing instruction: A longitudinal study of writers' rhetorical awareness, writing quality, and lexicogrammatical choices [Unpublished doctoral dissertation], University of Hawaii at Manoa.

Yde, P., & Spoelders, M. (1985). Text cohesion: An exploratory study with beginning writers. *Applied Psycholinguistics, 6*(4), 407–415. https://doi.org/10.1017/S0142716400006330

Zhang, X., & Kang, S.-G. (2023). Applying the given-before-new principle in L2 English datives development. *Language Acquisition, 32*(1), 52–75. https://doi.org/10.1080/10489223.2023.2216680

Appendix A. Full descriptions of EAP writing tasks across business and engineering disciplines

EAP Writing Tasks for the Business Discipline

Analytical Summaries

For this major assignment, you will read, annotate, and write analytical summaries of two academic articles of your choice. In addition to accurately summarizing each article's main argument and key supporting points, you will analyze how the author develops the argument through rhetorical choices such as organization, evidence, tone, and language. You will be evaluated on the clarity and accuracy of your summaries as well as your ability to apply the rhetorical and linguistic skills introduced in class.

Cover Letter

For this assignment, you will write a professional cover letter for a real internship or job you are interested in applying for. First, find a job posting and carefully review the qualifications and expectations listed. Then, write a cover letter that introduces you as a strong candidate and explains how your skills, experiences, and background match the position. Your cover letter should be tailored to the specific job and company, using appropriate tone, formatting, and professional language.

Rhetorical Analysis of a Popular Press Business Article

The goal of this project is to select, critically read, and analyze a business-related article from a popular press source published within the last five years. You will write a memo presenting a rhetorical analysis of your chosen article. In your analysis, you should discuss the author's use of rhetorical strategies and appeals (e.g., ethos, pathos, and logos). You must also identify the author's intended audience and central message, and evaluate the overall effectiveness of the article's persuasive approach. In other words, does the author successfully convince the intended audience of the message being conveyed? If so, how? If not, what rhetorical choices weaken the argument, and why?

Extended Definition

For this assignment, you will write an extended definition of an important concept, term, or object from your field of study (Business). Unlike a simple dictionary definition, an extended definition explains a term in depth by describing what it is, what it means, how it works, and why it matters in your discipline. Your goal is to help a specific audience clearly understand the concept by using well-organized paragraphs, effective sentence structure, and appropriate domain-specific vocabulary. You should also include relevant examples, descriptions, comparisons, or explanations that illustrate the term and show how it is used in real academic or professional contexts. You may gather information from library sources, credible online sources, and your own experience.

EAP Writing Tasks for the Engineering Discipline

Analytical Summaries

For this major assignment, you will read, annotate, and write analytical summaries of two academic articles of your choice. In addition to accurately summarizing each article's main argument and key supporting points, you will analyze how the author develops the argument through rhetorical choices such as organization, evidence, tone, and language.

You will be evaluated on the clarity and accuracy of your summaries as well as your ability to apply the rhetorical and linguistic skills introduced in class.

Extended Definition

For this assignment, you will write an extended definition of an important concept, term, or object from your field of study (Engineering). Unlike a simple dictionary definition, an extended definition explains a term in depth by describing what it is, what it means, how it works, and why it matters in your discipline. Your goal is to help a specific audience clearly understand the concept by using well-organized paragraphs, effective sentence structure, and appropriate engineering-related vocabulary. You should also include relevant examples, descriptions, comparisons, or explanations that illustrate the term and show how it is used in real academic or professional engineering contexts. You may gather information from library sources, credible online sources, and your own experience.

Rhetorical Analysis of a Popular Press Engineering Article

The goal of this project is to select, critically read, and analyze an engineering-related article from a popular press source published within the last five years. You will write a memo that presents a rhetorical analysis of your chosen article. In your memo, you should examine the author's use of rhetorical strategies and appeals (ethos, pathos, and logos). You must also identify the author's intended audience and central message, and evaluate how effectively the author uses persuasion to achieve the intended purpose. In other words, does the author successfully convince the intended audience of the message being conveyed? If so, how? If not, what rhetorical choices weaken the argument, and why?

Technical Description and Recommendation

Focusing on a controversial technical problem, you will write a detailed description of a device or process for a specific audience and argue for the use of the device or the implementation of the procedure. In other words, you will describe the problem you aim to solve and provide sufficient detail so that readers can understand why the problem is significant, what your proposed solution is, and how it works. The report must include at least one visual aid and cite at least six sources in APA format.

Appendix B. Syntax of the generalized additive mixed models

```
AWL <- bam(AWL ~ s(TIME, k=5)  + s(TIME,PAR,k=5,bs="fs",m=1), data=Our_Data, discrete =
T, nthreads = 2)
plot_smooth(AWL, view="TIME", plot_all="AWL", rug=FALSE,col='blue')
```