

A CORPUS-BASED COMPARISON OF INTERACTIONAL PATTERNS IN EFL STUDENT AND LLM-GENERATED ESSAYS Insights into Authenticity and Audience Awareness

MARCO BAGNI

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Abstract - The emergence of AI chatbots such as ChatGPT has attracted increasing scholarly attention to the affordances and challenges of integrating large language models (LLMs) in EFL learners' writing skills development. Although some studies have investigated AI-generated text detection and characteristics, research on LLMs in EFL education remains limited. This paper seeks to address this gap and advance learners' critical understanding of AI applications. It reports on a small-scale study comparing genuine EFL undergraduate student essays, suspected LLM-generated essays, and essays generated using ChatGPT 4.0 and DeepSeek-V3. Drawing on Hyland's (2005a, 2005b) framework of interactional metadiscourse, the study examined patterns of stance and engagement in the texts. Combining quantitative analysis with qualitative interpretation, the study identified consistent use of engagement as the defining characteristic of genuine student writing, whereas LLMs showed limited capacity to engage readers. Furthermore, although LLMs approximated human stance-taking, they nevertheless favored a detached style, contrasting with the more personal tone of student writing. We believe that these findings can inform both EFL instructors and students by offering insights that may contribute to a deeper understanding of the rhetorical limitations of LLMs in academic writing contexts.

Keywords: EFL; LLM; AI-generated text; academic writing; interactional metadiscourse.

1. Introduction

The emergence of AI chatbots such as the widely popular ChatGPT has drawn attention to the affordances as well as the impact and challenges, in educational practices, of large language models (LLMs) (Cogo *et al.* 2024), a prominent type of Generative AI (GenAI) application. LLMs are trained on large datasets to generate human language, can operate across multiple languages, and are equipped with an intuitive conversational interface that enhances their ease of use and broad accessibility.

Recent research (Barrot 2023; Huang 2023; Su *et al.* 2024; Teng 2025; Yoo-Jean 2024; Warschauer *et al.* 2023; Woo *et al.* 2024) has discussed how ChatGPT as well as other LLMs can support writing skills development for EFL learners. However, the rapid uptake of ChatGPT and its widespread use by students across educational levels for academic tasks (Rong, Chun 2024) has also brought issues of academic integrity to the forefront, particularly due to the risk of learners delegating writing entirely to the tool, thereby undermining their motivation and autonomy (Alexander *et al.* 2023; De Wilde 2024).

LLMs' ability to produce human-like text raises concerns about the authenticity and originality of EFL student writing, with practical implications for assessment. In this regard, existing AI-detection tools have been shown to be limited in distinguishing between human- and AI-generated texts, particularly struggling to identify hybrid writing that combines human and machine-generated elements (Alexander *et al.* 2023). However,

research comparing EFL student essays with ChatGPT-generated texts (Amirjalili *et al.* 2024; Herbold *et al.* 2023) has shown that the latter lack the nuanced authorship and contextual sensitivity characteristic of human-authored texts, while the former more consistently employ epistemic constructions and modals to convey personal stance.

This also aligns with findings from studies (Jiang and Hyland 2025; Mo and Crosthwaite 2025) comparing metadiscourse features in argumentative essays by ChatGPT and English L1 students, which indicate that LLMs lack human-like contextual understanding and genuine audience awareness, constraining their rhetorical adaptability. This limitation has been attributed to insufficient exposure to domain-specific texts in the training data. Indeed, despite their impressive ability to generate human-like text, LLMs rely on statistical patterns and word associations learned from their training data, rather than on genuine understanding of context and audience (Jiang and Hyland 2025).

Metadiscourse (Hyland 2005a; 2005b) is a key component of effective academic writing, as it encompasses the ways authors convey their stance and involve readers as participants, guiding them in interpreting the text. While a few studies (e.g. Ho and Li 2018; Yoon 2021; Zhao 2017) have examined patterns of metadiscourse in EFL argumentative writing, research on LLM-generated text remains in its early stages, with few studies providing textual evidence of the rhetorical differences between AI-generated and EFL student writing.

This paper aims to address this research gap. It reports on a small-scale study that analyzed and compared genuine and suspected LLM-generated essays from first-year Communication Sciences students' English exam at the University of Modena and Reggio Emilia, and essays generated by ChatGPT 4.0 and DeepSeek-V3. The study aimed to explore how metadiscourse resources of authorial stance and reader engagement were used in the texts to shape writer-reader interactions, and identify distinguishing characteristics between LLM-generated text and genuine student writing. It is intended to offer practical insights for EFL teachers who may be confronting instances of academic misconduct involving LLM-generated texts, while also seeking to promote critical AI literacy in the classroom, pointing out the hazards of uncritical and passive reliance on LLMs in academic writing tasks.

2. Interactional metadiscourse

This study focuses on the interactional dimension of metadiscourse, encompassing the linguistic resources used to convey authorial stance and engage the reader for rhetorical purposes (Hyland 2005a; 2005b).

2.1. Stance

Stance is the expression of the writer's voice in the text and is realized through the following features:

- Hedges, reflecting the writer's choice to refrain from expressing full commitment to statements and acknowledge a plurality of perspectives.
- Boosters, demonstrating writers' confidence in their statements, engagement with the subject and connection with their audience.
- Attitude markers, expressing the writer's affective attitude to a statement.
- Self-mentions, including first-person pronouns and possessive adjectives carrying propositional, affective and interpersonal information.

2.2. Engagement

Engagement refers to the ways in which writers involve readers in the discourse, and includes the following features:

- Reader mentions, the pronouns *you*, *your*, addressing readers directly and the inclusive *we*, suggesting shared perspectives.
- Directives, instructing the reader, used to align goals and perspectives.
- Appeals to shared knowledge.
- Personal asides, offering brief comments or reflections.
- Questions inviting reader engagement.

3. Methodology

3.1. Context, participants and data collection

Seventy-four timed essays written by first-year Communication Sciences students at the University of Modena and Reggio Emilia were collected for this study. All students had previously met the B1 English proficiency requirement, although their actual level of proficiency varied substantially. However, learners' proficiency was not considered as a predictor variable, since the primary focus was comparing AI-generated texts with genuine student writing regardless of essay quality. The essays had been composed in pen and paper during the final exam of the mandatory first-year year English language course taught by the author of this paper. The course targeted an exit level situated between the B1 admission level and level B2, corresponding to a B1+ proficiency level, conceived as a "strong threshold" (Council of Europe 2020, p. 164) stage aimed at consolidating the students' independent user skills in preparation for progression to B2.

The course included a "Writing Skills Development" module designed to balance language- and writing-focused instruction by integrating three complementary approaches: learning to write, writing to learn language, and writing to learn content (Manchón 2011). Accordingly, the module aimed to foster the students' English writing skills while using writing as a means to support overall target language development and facilitate the acquisition of disciplinary knowledge. In line with these objectives, the students worked simultaneously on linguistic precision (grammar and vocabulary) and on using English for effective written communication, drawing both on the course content introduced in a parallel module ("English, Englishes and ELF"), and on their personal experience.

The "Writing Skills Development" module focused on the essay genre and specifically targeted three sub-genres: compare-and-contrast, argument, and discussion (Smith 2019). These were selected because they offered an accessible framework to support the deployment of different discourse modes (Smith 2003), with particular emphasis on exposition (comparing and contrasting two situations), argumentation (constructing a line of reasoning to support a viewpoint) and counter-argumentation (presenting a balanced discussion of a topic). In line with the course's B1+ exit level, the module did not aim to develop mastery of advanced academic genres, but rather to balance linguistic accuracy with creative expression, consistent with CEFR descriptors indicating that learners at B1+ level "can write short, simple essays on topics of interest" (Council of Europe 2020, p. 77). Consequently, highly specialized conventions of academic formality typically associated with C1-C2 proficiency – such as the use of impersonal language (Smith 2019) – were explicitly set aside. Overall, the module focused on correctness and

originality, with the goal of building strong foundations for future development in both receptive and productive skills.

Two versions of the exam sheets had been created, each including an essay title for a distinct essay type: compare-and-contrast, argument, discussion, as summarized in Table 1. The students had been asked to choose one title and write approximately 200 words.

EXAM SHEET 1	
Compare and contrast	Learning English in the past (before the advent of the Internet) versus learning English nowadays, in the era of the Internet.
Argument	English language teachers should base learning activities on more varieties of English, not only on one Standard variety.
Discussion	The global spread of English is a threat to the world's linguistic and cultural diversity.
EXAM SHEET 2	
Compare and contrast	Learning English as a foreign language in the past – before the advent of the internet – versus learning English today, in the era of the Internet.
Argument	The use of English as a lingua franca makes communication more efficient, in settings that involve speakers of mutually unintelligible languages (such as, for instance, Italian, German, Chinese, Korean, etc.).
Discussion	English learning cannot be based exclusively on Standard English.

Table 1
Essay titles.

During the exam grading process, the teacher identified 16 essays with little or no connection to the course content and exhibiting a level of English proficiency inconsistent with that expected of the students. Additionally, some of these essays, in an otherwise polished academic English style, contained misspellings and conspicuous morpho-syntactic and lexical errors that led the teacher to suspect they may have resulted from inattentive copying of an uncredited GenAI-generated source.

All essays were transcribed verbatim, preserving the errors they contained, using the Microsoft Word Dictate tool. The accuracy of the transcription was then manually checked. The suspect essays were set apart, and two distinct corpora were compiled using Sketch Engine (Kilgariff *et al.* 2014), a web-based corpus-analysis software providing a comprehensive set of tools. One (GSE henceforth) included 58 essays genuinely written by students; the other (SSE henceforth) included 16 suspect essays.

A third corpus (AGE henceforth) was compiled comprising 12 essays generated using the free browser versions of ChatGPT 4.0 and DeepSeek-V3. The inclusion of two LLMs was intended to increase diversity and enhance the reliability of the analysis. At the time of data collection, ChatGPT was widely regarded as the state-of-the-art LLM, while the recently launched DeepSeek-V3 positioned itself as a direct challenger to ChatGPT's dominance. Each LLM was instructed to create one essay for each title included in the exam sheets using zero-shot prompts (Schulhoff *et al.* 2024). As exemplified by (1), the prompts included only the words in the title and the indication of approximately 200 words as desired essay length, without any other contextual information nor any specific examples (see also Annexes).

(1) Write a discussion essay of approximately 200 words: 'The global spread of English is a threat to the world's linguistic and cultural diversity'.

(Prompted to DeepSeek-V3)

This approach was suggested by the in-class exam context, based on the assumption that any student who may have used GenAI tools to cheat would have done so covertly, very likely limiting their input to the essay title alone, without detailed prompting.

Within each corpus three sub-corpora were compiled, each corresponding to a distinct type of essay, as summarized in Table 2.

CORPUS	SUBCORPUS	ESSAY TYPE	# OF DOCS
AGE	A	Compare and contrast	4
AGE	B	Argument	4
AGE	C	Discussion	4
AGE (Total number of essays)			12
SSE	A	Compare and contrast	7
SSE	B	Argument	3
SSE	C	Discussion	6
SSE (Total number of essays)			16
GSE	A	Compare and contrast	20
GSE	B	Argument	20
GSE	C	Discussion	18
GSE (Total number of essays)			58

Table 2
Corpora and subcorpora details.

For illustrative purposes, Tables 8-16 (see Annexes) present sample essays organized by essay type – compare-and-contrast (A), argument (B), discussion (C) – with one representative essay from each corpus (AGE, GSE, and SSE) included for each type.

3.2. Data analysis

This study employed a mixed-methods approach to ensure comprehensiveness and accuracy in metadiscourse items identification. The data analysis thus combined automated item searches for quantitative analysis with qualitative, interpretative examination.

For each corpus and subcorpus, wordlists reporting absolute frequency, relative frequency, document frequency, and relative document frequency were generated using Sketch Engine's Wordlist tool. Based on Hyland's (2005a; 2005b) framework, inventories of interactive and interactional metadiscourse items were compiled and organized into two separate CSV files. ChatGPT 4.0 Plus was then used to quickly retrieve these items from the corpora. However, as the model proved unreliable in determining their discourse function, all results were manually verified using the Sketch Engine Concordance tool to establish their interactive or interactional role. The validated results were compiled into two Excel files – one for interactive and one for interactional metadiscourse – reporting frequency and document frequency measures. These were finally supplemented with additional items identified through close reading of the students' essays. In line with Hyland's (2005a) perspective, metadiscourse was treated as an open-ended repertoire of linguistic resources. Since some of the students' essays displayed academically inappropriate or incorrect uses of certain forms resulting from L1 transfer and L2 proficiency, all students' essays were carefully analyzed to identify any such forms revealing an intention to realize interactional meanings. For instance, some essays contained adjectives expressing the writer's attitudinal stance that would typically be considered inappropriate in an academic register, as illustrated in (2):

- (2) Only a few decades ago learning a new language such as English without the proper tools could be a *nightmarish* experience.

(GSE A essay)

Several students were found to struggle with modal verbs, as exemplified in the following extract, which features a loan translation of an Italian expression corresponding to ‘it must be said’ or ‘it is necessary to say’:

- (3) To conclude, *it is to say* that it is hard to predict what the future will bring.
(GSE A essay)

As it revealed the student’s intention to express deontic modality, this and similar improper constructions were counted among the directives.

4. Results

The item search results were analyzed first across the three main corpora (AGE, SSE, GSE), and subsequently within their respective subcorpora (AGE A–SSE A–GSE A; AGE B–SSE B–GSE B; AGE C–SSE C–GSE C).

Table 3 compares the frequency, relative frequency (per million words), document frequency and document relative frequency (in percentage) of stance and engagement items in the main corpora. The column charts (Fig.1, Fig. 2) below visualize the relative frequency values and relative document frequency values.

	FREQUENCY			RELATIVE FREQUENCY		
	AGE	SSE	GSE	AGE	SSE	GSE
Hedges	29	26	98	10828.98	8885.85	7296.00
Boosters	5	3	78	1867.06	1025.29	5807.03
Attitude	10	9	86	3734.13	3075.87	6402.62
Self mention	3	10	79	1120.24	3417.64	5881.48
STANCE TOT.	47	48	341	17550.41	16404.65	25387.13
Reader mention	0	2	279	0.00	683.53	20771.29
Directives	2	2	55	746.83	683.53	4094.70
Appeals to shared knowledge	1	0	4	373.41	0.00	6.90
Asides	0	1	7	0.00	341.76	12.07
Questions	1	0	15	373.41	0.00	25.86
ENGAGEMENT TOT.	4	5	360	1493.65	1708.82	24910.82

	DOC. FREQUENCY			REL. DOC. FREQUENCY		
	AGE	SSE	GSE	AGE	SSE	GSE
Hedges	12	13	37	100.00%	81.25%	63.79%
Boosters	5	2	36	41.67%	12.50%	62.07%
Attitude	6	6	40	50.00%	37.50%	68.97%
Self mention	3	5	34	25.00%	31.25%	58.62%
STANCE TOT.	12	15	58	100.00%	93.75%	100.00%
Reader mention	0	2	46	0.00%	12.50%	79.31%
Directives	2	2	24	16.67%	12.50%	41.38%
Appeals to shared knowledge	1	0	3	8.33%	0.00%	5.17%
Asides	0	1	7	0.00%	6.25%	12.07%
Questions	1	0	12	8.33%	0.00%	20.69%
ENGAGEMENT TOT.	4	4	51	33.33%	25.00%	87.93%

Table 3
Interactional features in the three main corpora.

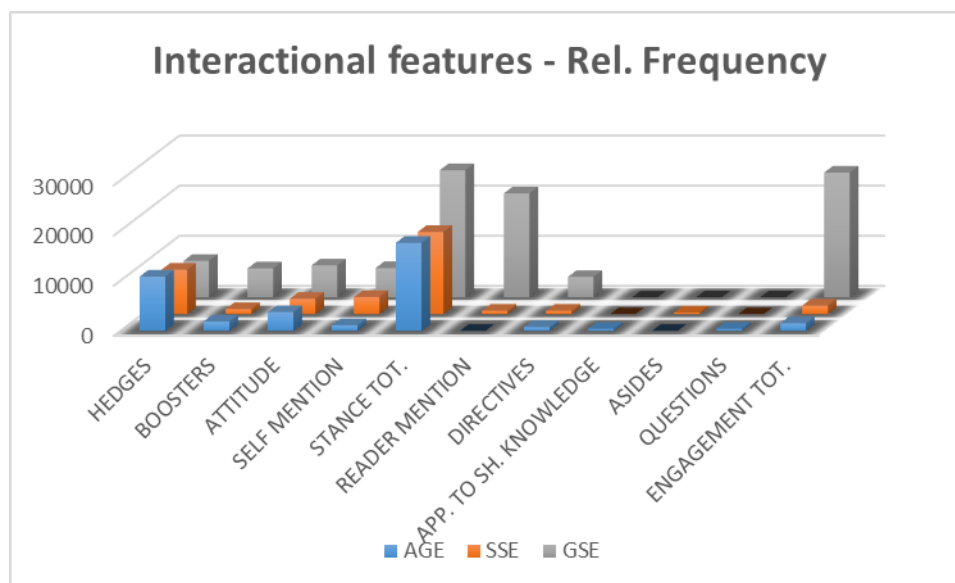


Figure 1
Interactional features relative frequency.

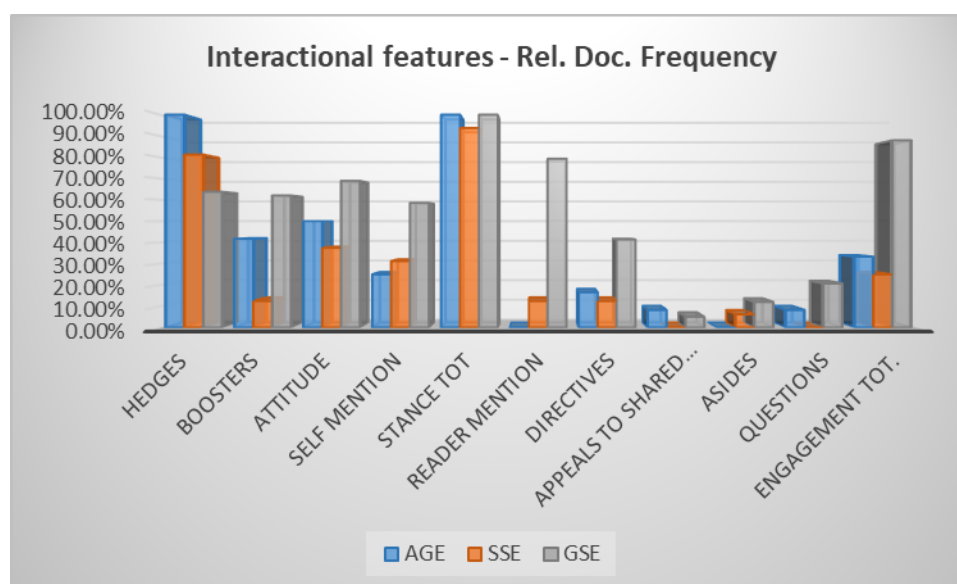


Figure 2
Interactional features relative document frequency.

The comparison revealed a clear disparity in the interactional orientations conveyed across the three corpora, with GSE displaying a relatively more frequent use of stance markers and a markedly higher frequency of engagement features than both AGE and SSE, which revealed instead almost identical patterns.

Stance patterns apparently pointed to the hybrid nature of the SSE essays: hedge frequencies were higher than in GSE but lower than in AGE, while self-mentions occurred more frequently than in AGE yet remained less common than in GSE. Most notably, in contrast to GSE, AGE was characterized by a predominance of hedges. This characteristic suggested that LLMs not only tended to present information in a markedly impersonal manner, with minimal authorial presence, but also avoided strong commitment to propositional content and the expression of affective stance. Investigating the extent to

which LLMs can be fine-tuned to control metadiscourse through prompt engineering (Cain 2024; Schullhoff *et al.* 2025) lies beyond the scope of the present study. Nevertheless, we hypothesize that the prompt strategy adopted may have influenced the observed patterns, as zero-shot prompts may implicitly encourage a detached stance and limit authorial intrusion into the discourse.

Engagement features, and more specifically reader mentions, emerged as the key distinguishing characteristic of genuine student essays. Table 4 shows the reader mentions featured in GSE.

	FREQUENCY	REL. FREQ.	DOC. FREQ.	REL. DOC. FREQ.
we	141	10497.32	40	68.97%
you	72	5360.334	20	34.48%
our	37	2754.616	19	32.76%
us	13	967.838	10	17.24%
your	12	893.3889	10	17.24%
ourselves	2	148.8982	2	3.45%
yourself	1	74.44908	1	1.72%

Table 4
Reader mentions in GSE.

The inclusive pronoun *we* and the pronoun *you*, representing the vast majority of all reader mentions in GSE, typically function to assert both authority and solidarity. While dialogic in nature, in that they acknowledge the reader's viewpoint, suggesting shared perspectives and a common purpose, they also position the writer as a guiding authority, leading the audience through the argument toward a desired conclusion (Jiang, Hyland 2025).

Nevertheless, despite being highly prevalent, it must be noted that reader mentions were not found in all essays, as indicated by the 79.31% relative document frequency value. As observed by Jiang and Hyland (2025), learners may often refrain from engaging readers directly, possibly perceiving this strategy as characteristic of more informal or familiar registers. Indeed, although the students had been explicitly instructed to disregard advanced academic style conventions discouraging the use of personal pronouns, affective adverbs, rhetorical questions, and references to personal experience, some may nonetheless have been inclined to avoid reader mentions to align with perceived academic norms commonly presented in writing textbooks (see Smith 2019).

However, despite this possibility, what stands out is the complete absence of reader mentions in LLM-generated essays. This tendency possibly resulted from the patterns of the models' training data, which include a broad array of registers and genres, each with distinct communicative purposes, structures, and audiences. As it has been noted (Jiang, Hyland 2025), this diversity in training data effectively neutralizes any specific audience, resulting in output that lacks targeted rhetorical adaptation.

4.1. Compare-and-contrast essays

Table 5 shows the distribution of stance and engagement features in the compare-and-contrast essays.

	FREQUENCY			RELATIVE FREQUENCY		
	AGE A	SSE A	GSE A	AGE A	SSE A	GSE A
Hedges	10	10	11	11173.18	6835.27	2395.47
Boosters	1	0	11	1117.32	0.00	2395.47
Attitude	0	0	7	0.00	0.00	1524.39

Self mention	0	7	23	0.00	4784.69	5008.71
STANCE TOT.	11	17	52	12290.50	11619.96	11324.04
Reader mention	0	1	132	0.00	683.53	28745.64
Directives	0	0	3	0.00	0.00	653.31
Appeals to shared knowledge	0	0	0	0.00	0.00	0.00
Asides	0	1	2	0.00	683.53	435.54
Questions	0	0	4	0.00	0.00	871.08
ENGAGEMENT TOT.	0	2	141	0.00	1367.05	30705.57
	DOC. FREQUENCY			REL. DOC. FREQUENCY		
	AGE A	SSE A	GSE A	AGE A	SSE A	GSE A
Hedges	4	6	8	100.00%	85.71%	40.00%
Boosters	1	0	7	25.00%	0.00%	35.00%
Attitude	0	0	6	0.00%	0.00%	30.00%
Self mention	0	2	8	0.00%	28.57%	40.00%
STANCE TOT.	4	6	20	100.00%	85.71%	100.00%
Reader mention	0	1	17	0.00%	14.29%	85.00%
Directives	0	0	2	0.00%	0%	10.00%
Appeals to shared knowledge	0	0	2	0.00%	0%	10.00%
Asides	0	1	2	0.00%	14.29%	10.00%
Questions	0	0	0	0.00%	0%	0.00%
ENGAGEMENT TOT.	0	1	17	0.00%	14.29%	85.00%

Table 5
Compare-and-contrast essays.

The nearly identical relative frequency of overall stance features across the three subcorpora, and most crucially their sparser use in comparison to the main corpora may be attributed to the nature of the compare-and-contrast essay, which is predominantly expository in nature and encourages a more factual and objective mode of presentation. The complete absence of attitude markers in AGE A – which otherwise mirrored AGE in terms of stance – with a predominance of hedges and absence of self-mentions, also seemed to confirm this hypothesis. Compared to the main corpus (GSE), GSE A presented noticeably lower frequency values for all features of stance, apart from self-mentions which accounted for almost half of all stance items in the subcorpus. This suggests that some students – as exemplified in (4) – valued projecting their authorial persona even within a compare-and-contrast essay that arguably favored a more objective approach:

(4) In fact now if *I* want to learn English (...) *I* can for example follow an online course (...) *I* can maybe listen some podcast about something that *I* like in English instead of Italian, and *I* can watch TV shows or documentary in English.

(GSE A essay)

The presence of self-mentions in 2 SSE A essays suggested that the students, possibly to evade detection, made a deliberate effort to infuse a personal touch into an otherwise impersonal text:

(5) *I* feel that with today's resources learning English is much easier and more interactive, achieving fluency is faster and more enjoyable than before.

(SSE A essay)

(6) As *I* said online platforms allow learners to connect with native speakers globally, enhancing their speaking and listening skills.

(SSE A essay)

The complete absence of engagement features in AGE A reinforced the observed tendency of LLMs to present propositional content in an impersonal and matter-of-fact manner. In contrast, and even more than in GSE overall, reader mentions were extensively used in GSE A, representing the quasi totality (132 out of 141 in 17 out of 20 essays) of engagement features. This clearly indicated that several students aimed to create a conversational tone, textually positioning themselves and the teacher as participants sharing a common understanding and goals, as exemplified in the following essay extracts:

(7) *We* certainly have more opportunities, but *we* shouldn't forget that the best way to learn a new language is to travel the world and talk with people.

(GSE A essay)

SSE A mirrored AGE A in its lack of engagement features, notwithstanding the presence of 1 reader mention and 1 aside in one SSE A essay, offering strong indications of its author's apparent intent to personalize AI-generated text:

(8) Nowadays *we* could use online platforms, apps and social media too (if they're used in the "right" way).

(SSE A essay)

4.2. Argument essays

Table 6 shows the distribution of stance and engagement features in the argument essays.

	FREQUENCY			RELATIVE FREQUENCY		
	AGE B	SSE B	GSE B	AGE B	SSE B	GSE B
Hedges	5	10	53	5931.20	21231.42	11358.77
Boosters	2	0	41	2372.48	0.00	8786.97
Attitude	5	5	34	5931.20	10615.71	7286.76
Self mention	3	3	44	3558.72	6369.43	9429.92
STANCE TOT.	15	18	172	17793.59	38216.56	36862.41
Reader mention	0	0	51	0.00	0.00	10930.13
Directives	2	2	31	2372.48	4246.28	6643.81
Appeals to shared knowledge	0	0	3	0.00	0.00	642.95
Asides	0	0	4	0.00	0.00	857.27
Questions	0	0	5	0.00	0.00	1071.58
ENGAGEMENT TOT.	2	2	94	2372.48	4246.28	20145.74
	DOC. FREQUENCY			REL. DOC. FREQUENCY		
	AGE B	SSE B	GSE B	AGE B	SSE B	GSE B
Hedges	4	3	18	100%	100%	90%
Boosters	2	0	17	50%	0%	85%
Attitude	3	3	17	75%	100%	85%
Self mention	3	3	15	75%	100%	75%
STANCE TOT.	4	3	20	100%	100%	100%
Reader mention	0	0	12	0%	0%	60%
Directives	2	2	13	50%	66.67%	65%
Appeals to shared knowledge	0	0	2	0%	0%	10%
Asides	0	0	4	0%	0%	20%
Questions	0	0	5	0%	0%	25%
ENGAGEMENT TOT.	2	2	17	50%	66.67%	85%

Table 6
Argument essays.

While the overall stance patterns did not differ much from those observed in the compare-and-contrast essays, in comparison, both LLM-generated and genuine student essays of the argumentative genre presented a few differences. Within AGE B, 3 essays featured each 1 self-mention, corresponding to an expression of agreement in the essays' opening statements, as illustrated in (9), (10), (11):

(9) *I agree that the use of English as a lingua franca makes communication more efficient [...]*
(AGE B essay - ChatGPT)

(10) *I agree that English language teachers should base learning activities on more varieties of English.*
(AGE B essay - ChatGPT)

(11) *I strongly agree that English language teachers should incorporate multiple varieties of English.*
(AGE B essay – DeepSeek)

Also, AGE B presented fewer hedges than AGE A and AGE C (see Section 4.3), suggesting that arguing in support of a thesis prompted the LLMs to reduce their use of such features. Nevertheless, the limited frequency of boosters in the argument essays subcorpus seemed instead to indicate a low degree of assertiveness.

In SSE B, the relatively higher frequency of hedges – in comparison to SSE A – and the complete absence of boosters apparently mirrored the LLMs' tendency to engage with topics in an impersonal manner.

Stance features were instead significantly more frequent in GSE B than in GSE A, a pattern likely shaped by the nature of the argumentative genre. This genre arguably encourages greater personal involvement with the topic, resulting in stronger expressions of commitment to the reliability of the propositional content, as well as a more prominent authorial presence in the text and increased affective engagement. Although engagement features patterns in GSE B subcorpus broadly reflected GSE's general patterns, reader mentions were much less frequent in argument essays. As previously observed, some students may have consciously limited direct reader engagement to align with perceived academic norms and to avoid an overly familiar tone. Additionally, the argument essay topics (see Table 1) may have felt rather loosely connected to the students' everyday experiences, potentially prompting a reduced use of hedges.

The influence of genre constraints on rhetorical choices was instead indicated by the higher frequency of directives in GSE B compared to GSE A, as well as by the presence of this key dialogic feature of argumentative writing (Jiang and Hyland 2025) in two AGE B essays and two SSE B essays:

(12) [...] exposure to different varieties *should not* come at the expense of clarity or consistency in foundational learning.
(AGE B essay - ChatGPT)

(13) [...] teachers *should* integrate diverse varieties to create adaptable, culturally aware communicators.
(AGE B essay – DeepSeek)

(14) Mastery of standard grammar rules *should* remain a priority.
(SSE B essay)

(15) *Imagine* an international conference: without a shared language, every interaction would require translation .
(SSE B essay)

Finally, questions and personal asides, although present in only 5 and 4 argument essays respectively, also appeared to distinguish genuine student writing from LLM-generated essays. As exemplified in (16), the students used asides to engage the reader (their teacher) by integrating a personal commentary into the text on the course topics discussed in class:

(16) [...] the speed and efficiency of the use of ELF (*and the rise of the globalized transnational culture that seems to come with it*) is seen by most as a good trade off in the business community.

(GSE B essay)

Questions were used to arouse the reader's curiosity and steer him towards the student's personal interpretation of the title topic, as can be seen in the following extracts:

(17) Of course it's easier communicating between people who speak the same language *but what happens, for example when people travel to another country?*

(GSE B essay)

(18) Today in all of the schools in the world English is taught (...). *But is it correct how teachers explain English to the students?* This is actually a great question, that can look normal, but is not.

(GSE B essay)

(19) English language teachers should base learning activities on more varieties of English, not only on one standard variety. *Should they really do it?*

(GSE B essay)

4.3. Discussion essays

Table 7 shows the distribution of stance and engagement features in the discussion essays.

	FREQUENCY			RELATIVE FREQUENCY		
	AGE C	SSE C	GSE C	AGE C	SSE C	GSE C
Hedges	14	6	34	14941.30	6048.39	8145.66
Boosters	2	3	26	2134.47	3024.19	6229.04
Attitude	5	4	45	5336.18	4032.26	10781.03
Self mention	0	0	12	0.00	0.00	2874.94
STANCE TOT.	21	13	117	22411.95	13104.84	28030.67
Reader mention	0	1	96	0.00	1008.06	22999.52
Directives	0	0	21	0.00	0.00	5031.15
Appeals to shared knowledge	1	0	1	1067.24	0.00	239.58
Asides	0	0	1	0.00	0.00	239.58
Questions	1	0	6	1067.24	0.00	1437.47
ENGAGEMENT TOT.	2	1	125	2134.47	1008.06	29947.29
	DOC. FREQUENCY			REL. DOC. FREQUENCY		
	AGE C	SSE C	GSE C	AGE C	SSE C	GSE C
Hedges	4	4	11	100%	66.67%	61.11%
Boosters	2	2	12	50%	33.33%	66.67%
Attitude	3	3	17	75%	50%	94.44%
Self mention	0	0	11	0%	0%	61.11%
STANCE TOT.	4	6	18	100%	100%	100%
Reader mention	0	1	17	0%	16.67%	94.44%
Directives	0	0	9	0%	0%	50%
Appeals to shared knowledge	1	0	1	25%	0%	5.56%
Asides	0	0	1	0%	0%	5.56%
Questions	1	0	5	25%	0%	27.78%
ENGAGEMENT TOT.	2	1	17	50%	16.67%	94.44%

Table 7
Discussion essays.

Interactional patterns in discussion essays showed no particularly salient differences relative to those observed in the other essay types. However, AGE C featured one question – included in the essay heading that the LLM (DeepSeek-V4) had generated:

(20) The Global Spread of English: *A Threat to Linguistic and Cultural Diversity?*
(AGE C essay)

Also, AGE C presented relatively more hedges than both AGE A and AGE B. LLMs may have been led to use hedges to frame information as opinion rather than fact by genre, as discussion essays are geared towards examining multiple sides of an issue and typically result in a more balanced perspective than argument essays. Genre constraints may also have shaped the use of attitude markers in AGE C essays, as the discussion genre allows for greater expression of evaluative stance. Apart from the higher frequency of hedges in LLM-generated essays, AGE C stance patterns were mirrored in SSE C, further fueling the teacher's suspicions as to the non-genuine nature of the SSE C essays.

In GSE C, the lower frequency of hedges compared to both GSE B and AGE C may have resulted from the students' limited mastery of discussion-genre conventions, as they appeared to prioritize their own perspectives and overlook the importance of addressing opposing viewpoints. Also, a relatively lower frequency of self-mentions set apart GSE C from GSE A and GSE B, possibly indicating influence of genre (as in AGE C) and topic (as in GSE B), with discussion essays titles addressing topics more detached from the students' everyday experience (see Table 1). The high frequency of reader mentions in 16 out of 17 GSE C essays, a substantial number of directives distributed in 50% of the subcorpora, and the presence of questions in 5 essays, point to the perceived relevance of involving the reader as a conversational partner in genuine student discussion writing, as the following extract exemplifies:

(21) The question is, *can foreign learners base their learning exclusively on standard English? We are going to see why varieties are actually so important, even nowadays. [...]* To speak a language, *we have to know* the grammar but *we also have to take into consideration* the varieties of it [...] *It's enough to think¹* about the differences between American and British English or *to think* about the various slangs used by the youth.

(GSE C essay)

5. Discussion

Overall, the study's findings indicate that while LLMs can approximate EFL student writing in terms of writer-oriented features such as stance, they consistently favor an objective, factual mode of expression. Although the overall relative frequency of stance markers was not a notable distinguishing feature between genuine student writing and LLM-generated output, the findings nonetheless pointed to the more personal and individualized tone of student essays. This was reflected in the relatively higher occurrence of attitude markers, boosters, and self-mentions in GSE compared to AGE and SSE. In contrast, the notably higher frequency of hedges in AGE indicated a more detached approach, with minimal engagement of the writer's persona.

¹ The student produced a word-by-word translation of an Italian expression roughly corresponding to 'just think about'. As the loan translation in example (3), this was also counted among the directives.

Most importantly, the LLMs showed clear limitations in audience modelling, as evidenced by their limited ability to establish a connection with the reader. In contrast, genuine EFL student writing, despite not yet meeting academic writing proficiency standards, demonstrated greater sensitivity to the need for explicitly addressing the audience and guiding them through the discourse. This was evidenced by the markedly higher frequency of engagement features in GSE, and particularly the consistent use of reader mentions in 46 out of 51 essays. In sum, this study suggests that it is this dialogic quality that most clearly distinguishes essays generated by LLMs using zero-shot prompts from essays genuinely written by EFL students.

It can be further argued that the apparent insertion of metadiscursive features by some students into LLM-generated text, seemingly aimed at evading detection, indicates that the characteristics identified in such essays were perceived by these students as lacking authenticity. In this respect, SSE displayed patterns that alternately aligned with those of AGE and GSE, highlighting the hybrid nature of the suspect essays. The students' efforts to modify LLM-generated output to avoid detection, together with partial or inattentive copying from such sources, likely resulted in texts combining features of both genuine and LLM-generated writing. This interpretation was further supported by the application of non-systematic, intuition-based content-related and linguistic criteria typically employed by teachers when identifying AI-generated text (De Wilde 2024). As previously observed, SSE essays hardly included any reference to the course content and displayed a level of mastery of academic English that was inconsistent with the authors' expected proficiency level, while also exhibiting conspicuous errors.

These findings broadly align with those from comparisons between AI-generated text and English L1 student argumentative writing (Jiang, Hyland 2025; Mo, Crosthwaite 2025) showing that LLMs demonstrate limited capacity to infuse argumentative texts with a personal touch and particularly reproduce human-like engagement features. However, we also suggested that some of the students' understanding of academic expectations may have constrained the extent to which they perceived a need to explicitly mark the reader's presence in the text, particularly avoiding setting a familiar tone and addressing the reader directly. Although the teacher had explicitly instructed the students to disregard advanced academic style conventions, in line with the course objectives (see Section 3.1), it is plausible that some students ignored instructional guidance and nevertheless relied on prescriptive norms encountered in textbooks.

Additionally, different patterns observed across the three essay types suggested possible genre constraints on the use of stance and engagement strategies, and although essay topic was not included as a predictor variable in our analysis, we also suggested that topic choice may have influenced metadiscourse strategies, with more abstract or impersonal topics prompting sparser use of interactional features.

6. Conclusion

This study has extended analyses of stance and engagement features in student- and AI-generated texts to an EFL context. It confirmed the teacher's first impressions regarding the non-genuine nature of some of the essays submitted by the students. In particular, the shifting alignment of SSE patterns with those resembling AGE and GSE strongly indicated the hybrid character of many SSE essays. Given the demonstrated unreliability of existing AI-detection tools with hybrid texts (Alexander *et al.* 2023), and despite possible future improvements, interactional metadiscourse appears to offer an effective interpreting framework for assessing the provenance of student writing.

Confirming findings of research (Jiang and Hyland 2025; Mo and Crostwaithe 2025) carried out in English L1 context, the analysis of interactional metadiscourse features carried out in this study has revealed that genuine EFL student writing is marked by a consistent use of interactional metadiscourse, and particularly reader engagement resources. Given that writing is inherently interactional, the distinctiveness and inventiveness of a text depend on the writer's ability to construct an authorial persona in relation to both content and audience. From this perspective, our findings suggest that, when prompted using zero-shot instructions limited to the essay type, desired length, and title, LLMs lack the capacity to replicate the originality and creativity evident in EFL student writing, regardless of students' proficiency levels.

While intuition and familiarity with one's students' academic profiles often suffice to detect non-genuine writing, raising EFL teachers' awareness of the characteristic features of LLM-generated text can provide additional support. Most importantly, it can inform classroom practice aimed at fostering students' GenAI literacy (Wang *et al.* 2024; Wang *et al.* 2025). As LLMs struggle to model audience or interpret context in a human-like way, the application of critical thinking is essential for their effective and responsible use in academic settings. There is ample reason to believe that students can benefit from developing a deeper understanding of the inherent limitations of LLMs-generated output. EFL teachers aiming to develop learners' writing skills and, most importantly, their creativity, should prioritize guiding students in organizing discourse with a clear awareness of audience, encouraging them to project their persona and mark the reader's presence in the text.

6.1. Limitations and further research

Given its small-scale design, this study offered findings with limited generalizability. Larger-scale investigations may yield deeper insights and more conclusive evidence. The analysis was based on student essays written under time constraints and limited to approximately 200 words, which may have constrained the use of stance and engagement features. Further studies might examine longer, non-timed writing that underwent revision and editing.

The LLM-generated essays in this study were created using zero-shot prompts including only essay type, title and desired length. Further research could also explore how prompt engineering (Cain 2024; Schulhoff *et al.* 2024) may guide LLMs toward generating more human-like texts across different genres and discourse modes. A key question that emerges from this study concerns the extent to which LLMs can be fine-tuned to emulate human capacities for audience awareness and rhetorical involvement. In addition, this study compared EFL student writing with LLM-generated texts but did not compare multiple LLMs output. Only two LLMs were used, and differences in interactional features across a broader range of LLMs were not explored. Further experimental studies may also test the capabilities of diverse and potentially more specialized models.

Finally, given the rapid evolution of GenAI tools, future research could examine the reliability of current AI-detection systems in comparison with qualitative and corpus-assisted analyses, particularly in terms of whether these approaches identify similar metadiscursive features in human and LLM-generated writing, and how effectively they handle hybrid texts.

Bionote: Marco Bagni holds a PhD at the University of Modena and Reggio Emilia and worked on a research grant at the Department of Communication and Economics of the same university. Since 2019, he has taught undergraduate courses in General English, English for Specific Purposes and English linguistics, and has worked for two decades as a secondary school teacher in Italy. His research interests include the impact of GenAI in EFL pedagogy, Global English Language Teaching, English as a Lingua Franca, and English language variation. Among his recent publications, his monograph PhD dissertation: *Students Views and Attitudes Towards English and ELF in an Italian University* (Generis Publishing, 2024).

Author's address: marco.bagni@unimore.it

References

- Alexander K., Savvidou C. and Alexander C. 2023, *Who wrote this essay? Detecting AI-generated writing in second language education in higher education*, in “Teaching English with Technology” 23 [2], pp. 25-43. <https://doi.org/10.56297/BUKA4060/XHLD5365> (13.04.2025).
- Amirjalili F., Neysani M. and Nikbakht A. 2024, *Exploring the boundaries of authorship: a comparative analysis of AI-generated text and human academic writing in English literature*, in “Frontiers in Education” 9. <https://doi.org/10.3389/feduc.2024.1347421> (26.06.2025).
- Barrot J.S. 2023, *Using ChatGPT for Second Language Writing: Pitfalls and Potentials*, in “Assessing Writing” 57, pp. 100745. <https://doi.org/10.1016/j.asw.2023.100745> (21.04.2025).
- Cain W. 2024, *Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education*, in “TechTrends” 68, pp. 47-57. <https://doi.org/10.1007/s11528-023-00896-0> (20.04.2025).
- Cogo A., Patsko L. and Szoke J. 2024, *Generative artificial intelligence and ELT*, in “ELT Journal” 78 [4], pp. 373-377. <https://doi.org/10.1093/elt/ccae051> (27.03.2025).
- Council of Europe, 2020. *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*, Council of Europe Publishing, Strasbourg. www.coe.int/lang-cefr (06.04.2025).
- De Wilde V. 2024, *Can novice teachers detect AI-generated texts in EFL writing?*, in “ELT Journal” 78 [4], pp. 414-422. <https://doi.org/10.1093/elt/ccae031> (13.04.2025).
- Herbold S., Hautli-Janisz A., Heuer U., Kiktev Z. and Trautsch A. 2023, *A large-scale comparison of human-written versus ChatGPT-generated essays*, in “Scientific reports” 13 [1], pp. 18617. <https://doi.org/10.1038/s41598-023-45644-9> (21.04.2025).
- Ho V. and Li C. 2018, *The use of metadiscourse and persuasion: An analysis of first year university students’ timed argumentative essays*, in “Journal of English for Academic Purposes” 33, pp. 53-68. <https://doi.org/10.1016/j.jeap.2018.02.001> (13.04.2025).
- Hyland K. 2005a, *Metadiscourse: Exploring interaction in writing*, Continuum, London.
- Hyland K. 2005b, *Stance and engagement: A model of interaction in academic discourse*, in “Discourse Studies” 7 [2], pp. 173-192. <https://doi.org/10.1177/1461445605050365> (06.04.2025).
- Huang J. 2023, *Engineering ChatGPT prompts for EFL writing classes*, in “International Journal of TESOL Studies” 5 [4], pp.73-79 <https://doi.org/10.58304/ijts.20230405> (06.04.2025)
- Jiang F.K. and Hyland K. 2025, *Does ChatGPT Write Like a Student? Engagement Markers in Argumentative Essays*, in “Written Communication” 42 [3], pp. 463-492. <https://doi.org/10.1177/0741088325132831.1> (21.04.2025).
- Kilgariff A., Baisa V., Bušta J., Jakubíček M., Kovvār V., Michelfeit J., Rychlý P. and Suchomel V. 2014, *The Sketch Engine: Ten Years On*, in “Lexicography” 1 [1], pp. 7-36. <https://doi.org/10.1007/s40607-014-0009-9> (04.06.2025).
- Manchón, R.M. 2011, *Situating the learning-to-write and writing-to-learn dimensions of L2 writing*, in Manchón, R.M. (ed.), *Learning-to-write and writing-to-learn in an additional language*, John Benjamins, Philadelphia, pp. 3-14. <https://doi.org/10.1075/llt.31.03man> (26.06.2025).
- Mo Z. and Crosthwaite P. 2025, *Exploring the affordances of generative AI large language models for stance and engagement in academic writing*, in “Journal of English for Academic Purposes” 75, pp. 101499. <https://doi.org/10.1016/j.jeap.2025.101499> (28.04.2025).
- Rong H. and Chun C. 2024, *Digital Education Council Global AI Student Survey 2024*, <https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024> (20.6.2025).
- Schulhoff S., Ilie M., Balepur N., Kahadze K., Liu A., Si C., Li Y., et al. 2024, *The prompt report: A systematic survey of prompting techniques*, in *arXiv preprint arXiv:2406.06608* 5. <https://doi.org/10.48550/arXiv.2406.06608> (04.06.2025).
- Smith C.S. 2003. *Modes of discourse: The local structure of texts*, Cambridge University Press, Cambridge.
- Smith S. 2019, *Academic writing genres: Essays, reports & other genres*, Evident Press.
- Su Y., Lin Y. and Lai C. 2023, *Collaborating with ChatGPT in argumentative writing classrooms*, in “Assessing Writing”, pp. 100752. <https://doi.org/10.1016/j.asw.2023.100752> (28.04.2025).
- Teng M.F. 2025, *Metacognitive Awareness and EFL Learners’ Perceptions and Experiences in Utilising ChatGPT for Writing Feedback*, in “European Journal of Education” 60 [1], pp. e12811. <https://doi.org/10.1111/ejed.12811> (13.04.2025).
- Wang C., Wang H., Li Y., Dai J., Gu X. and Yu T. 2024, *Factors Influencing University Students’ Behavioral Intention to Use Generative Artificial Intelligence: Integrating the Theory of Planned*

- Behavior and AI Literacy*, in “International Journal of Human–Computer Interaction” 41 [11], pp. 6649-6671. <https://doi.org/10.1080/10447318.2024.2383033> (09.01.2026).
- Wang F., Li N., Cheung A.C.K and Wong G.K.W. 2025, *In GenAI we trust: An investigation of university students' reliance on and resistance to generative AI in language learning*, in “International Journal of Educational Technology in Higher Education” 22 [1], p. 59. <https://doi.org/10.1186/s41239-025-00547-9>. (09.01.2026)
- Warschauer M., Tseng W., Yim S., Webster T., Jacob S., Du Q. and Tate T. 2023, *The Affordances and Contradictions of AI-Generated Text for Writers of English as a Second or Foreign Language*, in “Journal of Second Language Writing” 62. <https://doi.org/10.1016/j.jslw.2023.101071> (21.04.2025).
- Woo D.J., Susanto H., Yeung C.H., Guo K. and Fung A.K.Y. 2024, *Exploring AI-Generated text in student writing: How does AI help?*, in “Language Learning & Technology” 28 [2], pp. 183-208. <https://hdl.handle.net/10125/73577> (28.04.2025).
- Yoo-Jean L. 2024, *Can my writing be polished further? When ChatGPT meets human touch*, in “ELT Journal” 78 [4], pp. 401-413. <https://doi.org/10.1093/elt/ccae039> (13.04.2025).
- Yoon H.J. 2021, *Interactions in EFL argumentative writing: effects of topic, L1 background, and L2 proficiency on interactional metadiscourse*, in “Reading and Writing” 34 [3], pp. 705-725. <https://doi.org/10.1007/s11145-020-10085-7> (21.04.2025).
- Zhao C.G. 2017, *Voice in timed L2 argumentative essay writing*, in “Assessing Writing” 31, pp. 73-83. <https://doi.org/10.1016/j.asw.2016.08.004> (21.04.2025).

Annexes

Tables 8-10 present one representative compare-and-contrast essay from each corpus (AGE, GSE, and SSE).

Essay title: *Learning English in the past (before the advent of the Internet) versus learning English nowadays, in the era of the Internet.*

AGE A essay – ChatGPT	
Prompt:	Write a compare-and-contrast essay of approximately 200 words: “learning English in the past (before the advent of Internet) versus learning English nowadays, in the era of Internet”.
Output:	Learning English: Then and Now Learning English before the advent of the Internet was a slower, more limited process. Students relied heavily on textbooks, printed dictionaries, audio cassettes, and in-person classes. Access to native speakers was rare unless one traveled or had the opportunity to study abroad. Mistakes often went uncorrected for longer periods, and exposure to authentic spoken English was minimal, typically limited to radio or films. In contrast, learning English today is faster, more interactive, and widely accessible. The Internet offers countless free resources: online dictionaries, grammar tools, language learning apps, video tutorials, and interactive platforms like Duolingo or YouTube. Learners can practice speaking with native speakers through video calls or language exchange forums, and AI-driven tools can provide instant feedback on writing and pronunciation. Social media also immerses users in real-time English communication. However, while today’s learners have more tools, they may struggle with distractions, misinformation, and inconsistent practice. Past learners had fewer resources, but perhaps more focused study habits and appreciation for what they had. In conclusion, the Internet has revolutionized English learning by increasing access and variety, but success still depends on the learner’s motivation and how effectively these tools are used.

Table 8
Compare-and-contrast essay generated by ChatGPT.

GSE A essay
Learning English has drastically changed over the years particularly with the advent of the Internet. In the past before the use of the Internet people learned English with textbooks classes or with a face to face interaction. The students had a limited resources. In contrast learning English today is much easy dynamic and accessible because we have Internet and the Internet revolutionated the learning of everything. We have unlimited resources to learn it we can watch videos in every language we prefer, films, TV series, we can search everything we want. We have also social medias like Instagram or TikTok and there we can find many of videos even if we don't search for in others language. In fact I met many of people who learned English through tik toks videos, like my sister, and it's funny but we can reflect the importance of the Internet in our lives. It's not only for “no-important” things we can do in our free times and even if you doing nothing but only use it for dumb videos you are learning indirectly! In conclusion the advent of Internet is the most important invention for learning English in nowadays and everything we want through it.

Table 9
Compare-and-contrast essay genuinely written by a student.

SSE A essay

Learning English had a significant transformation over the decades, particularly with the advent of the Internet.

ACCESSIBILITY AND RESOURCES

In the past learning English largely depended on traditional methods. Students relied on textbooks, classroom instruction and, the most lucky ones, with audio materials (cassette tapes). Access to these resources was very only a small amount of people could attend language schools or purchase books or tapes. Furthermore learning opportunities were often confined to geographical locations.

Today thanks to the Internet the access to English learning resources has greatly expanded. Nowadays we could use online platforms, apps and social media too (if they're used in the "right" way); so anyone with an Internet connection can learn at his on place.

INTERACTION AND COMMUNICATION

Historically, English learning involved primarily face to face interaction and this created a personal learning environment, but it limited exposure to native speakers and real-world language use. In contrast, the Internet has revolutionized interaction and communication in English learning. As I said online platforms allow learners to connect with native speakers globally, enhancing their speaking and listening skills.

CONCLUSION

The method of learning English have significantly evolved from the past and this transformation has made English learning more inclusive and effective than ever before.

Table 10

Suspected LLM-generated compare-and-contrast essay submitted by a student.

Tables 11-13 present one representative argumentative essay from each corpus (AGE, GSE, and SSE).

Essay title: *English language teachers should base learning activities on more varieties of English, not only on one Standard variety.*

AGE B essay – DeepSeek	
Prompt:	Write an argument essay of approximately 200 words: "English language teachers should base learning activities on more varieties of English, not only on one Standard variety".
Output:	Argument Essay: The Importance of Teaching Multiple Varieties of English I strongly agree that English language teachers should incorporate multiple varieties of English, rather than focusing solely on one standard variety. English is a global language with diverse dialects, accents, and cultural influences, and limiting instruction to a single standard (such as British or American English) does not reflect real-world usage. First, exposing students to different varieties (e.g., Indian English, Australian English, or African Englishes) prepares them for global communication. In an interconnected world, learners will encounter diverse accents and expressions, and rigid adherence to one standard may hinder comprehension. Second, recognizing multiple varieties promotes inclusivity. Many students speak non-standard dialects (e.g., Singaporean English or Caribbean English) as their first language. Validating these varieties fosters confidence and cultural pride, rather than implying that only one form is "correct." Finally, linguistic flexibility enhances learning. Comparing different grammatical structures, vocabulary, and pronunciation improves overall language awareness. While standard English remains important for exams and formal contexts, teachers should integrate diverse varieties to create adaptable, culturally aware communicators. A multilingual world demands a multilingual approach to English teaching.

Table 11

Argument essay generated by DeepSeek.

GSE B essay

I think that English is a good way to communicate with speakers of other languages. In certain works we have to cooperate with people from all around the world and English is the better way to communicate because I think that it is the only language that is learned in every school of the world. Some people can also speak German or French, but not every people you work with can understand this two languages, so English is an easier and faster way to let people know what you're saying. We can also speak in English when we are for instance on holiday. For example you are in a foreign country but you don't know the native language of that country, the easiest way to interact with people that live there is English, especially when you are a tourist.

Another setting where the use of English as a lingua franca makes communication more efficient is in social media. It can be strange that I mention social media but nowadays we are a lot exposed to Internet and it happens that a lot of photos or videos that we watch are in English and the comments below are in English too, just as the songs that we all listen to. It's important to know English in this setting because you have to know what you are reading and listening.

In summary, I think that English is the most important language to know because there are many settings that require this knowledge.

Table 12
Argument essay genuinely written by a student.

SSE B essay	
	English has become the dominant lingua franca worldwide, allowing people from different linguistic background to communicate efficiently. I agree with the statement that the use of English as a lingua franca makes communication more effective in situations where speakers do not share a common native language.
	Firstly, English provides a neutral ground for communication. In international business, academia and diplomacy, a shared language reduces misunderstandings and allows smoother interaction.
	Secondly English is widely taught and accessible. Many educational systems around the world include English as a mandatory subject, ensuring that a large portion of the global population has at least basic proficiency.
	Additionally, the Internet, social media, and entertainment industries reinforce English exposure, making it easier to practice and improve.
	In conclusion English as a lingua franca is an essential tool for global communication, facilitating interactions among people who would otherwise struggle to understand each other.

Table 13
Suspected LLM-generated argument essay submitted by a student.

Tables 14-16 provide one representative discussion essay from each corpus (AGE, GSE, and SSE).

Essay title: *English learning cannot be based exclusively on Standard English.*

AGE C essay – ChatGPT	
Prompt:	Write a discussion essay of approximately 200 words: “English learning cannot be based exclusively on Standard English”.
Output:	Discussion Essay: English Learning Cannot Be Based Exclusively on Standard English While Standard English serves as a foundation for grammar, writing, and formal communication, learning English cannot be based exclusively on it. In today’s globalized world, English is spoken in a wide range of regional and cultural contexts, giving rise to various English varieties such as Indian English, Nigerian English, and Singaporean English. These forms reflect local identities and cultural nuances that Standard English does not fully capture. Furthermore, much of everyday communication, including social media, entertainment, and casual conversation, involves informal English, slang, idioms, and regional expressions. Learners who focus solely on Standard English may struggle to understand or engage in real-world interactions. Exposure to diverse English varieties enables learners to develop listening and comprehension skills across different accents, dialects, and contexts. That said, Standard English remains essential for academic writing, professional communication, and international standards. However, an inclusive approach that balances Standard English with exposure to global Englishes better prepares learners for practical, real-life use of the language. Understanding different forms of English promotes intercultural awareness and enhances

	communicative competence. In conclusion, while Standard English provides an important base, successful English learning must also include the rich diversity of Englishes used worldwide. This approach fosters both linguistic skill and global understanding.
--	---

Table 14
Discussion essay generated by ChatGPT.

GSE C essay
We are going to analyze the real use of the English language. So on the one hand we have the fact of learning standard English, and that is great. But on the other hand, in real life, we have to talk for example with native speakers and they do not talk all the time like the grammar books, so it may happen that we probably would not understand each other. So our English learning cannot be based only on standard English, we have to try to have contact with native speakers, or watch TV in original English language, or listening to the music of English native singers, or – the best decision – make new friends around the world that will help us to increase our English. Overall, yes, it is very important to learn the actual use of English in society, it may help us for example to have more possibilities to find our dreaming job. So, in my opinion we all need to study the standard English, but we have to know how the actual use of the English language around the world.

Table 15
Discussion essay genuinely written by a student.

SSE C essay
English learning cannot be based exclusively on standard English because it fails in capturing the linguistic diversity and cultural richness inherent in global communication. While standard English provides a useful framework for academic, professional and formal context, it represents only one facet of a language that is constantly evolving. In the everyday life communication, speakers employ a variety of registers, dialects and non standard forms that are crucial for an efficient communication in diverse settings. Exposure to regional variations, creoles and pidgins enhances learner's ability to navigate social interactions and appreciate the dynamic nature of English. Moreover, incorporating these forms into language instruction can help bridge the gap between formal education and real-world usage, fostering both linguistic competence and cultural sensitivity. This are more inclusive approach that values all variants of English is essential for learners to fully engage in global discourse and adapt to ever-changing communicative environments in today's world.

Table 16
Suspected LLM-generated discussion essay submitted by a student.