Enhancing Register Awareness through a ChatGPT-Mediated Teaching-Learning Cycle in a Vietnamese B1 EFL Co-Teaching Context

Ngo Hung Phu[1*]

[1]VNU-HCM University of Social Sciences and Humanities, Vietnam
[*]Corresponding author's email: phungo.3778@gmail.com
[*] https://orcid.org/0009-0009-8185-8861
https://doi.org/10.54855/ictep.26114

Received: 15/05/2026 Revision: 07/08/2026 Accepted: 08/08/2026 Online: 10/08/2026

ABSTRACT

Register awareness is a primary concern of writing instruction. It requires learners to adapt their language to suit the communicative purpose and degree of formality. In many Vietnamese EFL contexts, however, writing instruction often focuses on grammatical accuracy and text organisation. There is limited support to enhance learners' awareness of register. Within the Teaching–Learning Cycle (TLC), register is expected to be scaffolded through three stages. Yet, utilising ChatGPT to reinforce these stages has remained underexplored, especially in co-teaching classrooms. This exploratory quantitative study adopted a quasi-experimental pre-test-post-test design. Participants included 60 B1-level learners across four classes in an English center with two experimental (EG, n = 30) and comparison group (CG, n = 30) — selected via convenience sampling. Data was collected through Cambridge B1 Preliminary-based article writing tasks, scored using a Register Awareness Rubric (RAR; $\kappa = .84$), and a 15-item four-point Likert-scale questionnaire (Cronbach's $\alpha = .87$). The hypothesis was confirmed: independent-samples t-test revealed $t(58) = 10.08$, $p < .001$, Cohen's $d = 2.59$; ANCOVA controlling for pre-test scores yielded $F(1, 57) = 88.34$, $p < .001$, partial $\eta^2 = .61$, with all ANCOVA assumptions verified (Shapiro-Wilk $p > .05$; Levene's $p = .43$; regression slopes homogeneity $p = .38$). The EG demonstrated significantly greater reductions across all six spoken-like register features. Learner perceptions were strongly positive across all 15 questionnaire items (cluster means: 3.42–3.80 on a four-point scale).

Keywords: Register awareness, Teaching-Learning Cycle, ChatGPT- mediated intervention, co-teaching model, spoken-like features

Introduction

It is undeniable that writing plays an important role in applied linguistics and is one of the most demanding skills in second language acquisition (Hayes, 1996; Kellog, 1996). Beyond grammatical accuracy, proficient writing requires the capacity to make linguistically appropriate choices suited to the communicative context — that is, to demonstrate register

CITATION | Ngo, H. P. (2026). Enhancing Register Awareness through a ChatGPT-Mediated Teaching-Learning Cycle in a Vietnamese B1 EFL Co-Teaching Context. *ICTE Conference Proceedings*, 11, 51-74. ISSN: 2834-0000, ISBN: 979-8-90148-780-8. DOI: https://doi.org/10.54855/ictep.26114

awareness. Register, as theorized within Systemic Functional Linguistics (SFL), refers to the systematic configuration of language varieties associated with particular social situations, encompassing the variables of field, tenor, and mode (Halliday & Hasan, 1985). In EFL writing pedagogy, register awareness has been consistently identified as a key differentiating factor between proficient and less proficient writers (Hyland, 2004).

A well-documented challenge is that EFL learners frequently transfer features of informal, spoken discourse into formal written contexts. Chang and Swales (1999), in their survey of 40 academic style manuals, compiled a taxonomy of ten informal elements most frequently proscribed in formal writing, including first-person pronouns, sentence-initial coordinating conjunctions, contractions, and listing expressions. Building on this foundation, Hyland and Jiang (2017), in a corpus study of 360 research articles across four disciplines and five decades, documented the persistent presence of spoken-like features in academic writing and traced their gradual increase over time. More recently, Tocalo et al. (2023), studying Thai EFL university students across four genre types, found that informal features, particularly first-person pronouns and sentence-initial conjunctions, remained consistently present in formal essay writing despite explicit instruction. Together, these studies establish sentence-initial coordinating conjunctions, first-person pronouns, and contractions as the spoken-like features most resistant to conventional instruction, and these are the features the present study targets in the Vietnamese FEL private-center context.

Teaching and Learning Cycle (TLC), developed within the Sydney School tradition by Martin and Rose (2012), provides a theoretically principled, scaffolded framework for developing register awareness through explicit genre instruction. The emergence of ChatGPT as a generative AI tool capable of producing, comparing, and explaining texts across registers on demand opens new possibilities for enriching each TLC stage (Brown et al., 2020). Large language models such as GPT-3 and its successors are trained to perform tasks with minimal task-specific examples (Brow et al., 2020), an architecture that supports ChatGPT's capacity to produce, compare, and explain texts across registers on demand. These abilities open new possibilities to enrich each TLC stage (Kohnke et al., 2023; Law, 2024). However, as You and You (2026) note, research on AI-mediated genre learning embedded within explicit pedagogical frameworks, especially for integrating into the TLC with B1-level learners in a co-teaching context, has not previously been investigated. The present study addresses this gap within the OLA program at ILA Vietnam, Ho Chi Minh City — a co-teaching context in which VNTs lead the writing hour and are well positioned to deliver explicit register instruction.

Literature review

Systemic Functional Linguistics and Register Theory

Halliday's Systemic Functional Linguistics (SFL) provides the theoretical foundation of this study's conceptualization of register. Within SFL, language is understood as a social semiotic resource whose formal properties are systematically shaped by the context in which it is used (Halliday & Hasan, 1985). Central to SFL is the concept of register, defined as the configuration of linguistic choices that co-vary with three contextual variables: field (the subject matter and nature of the social activity), tenor (the social roles, relationships, and power differentials of participants), and mode (the channel of communication and the role played by language within it) (Halliday & Hasan, 1985). These variables are realized, respectively, through the three metafunctions of the clause: experiential, interpersonal, and textual meaning. Critically, Halliday (1989) demonstrated empirically that written language is characterized by higher

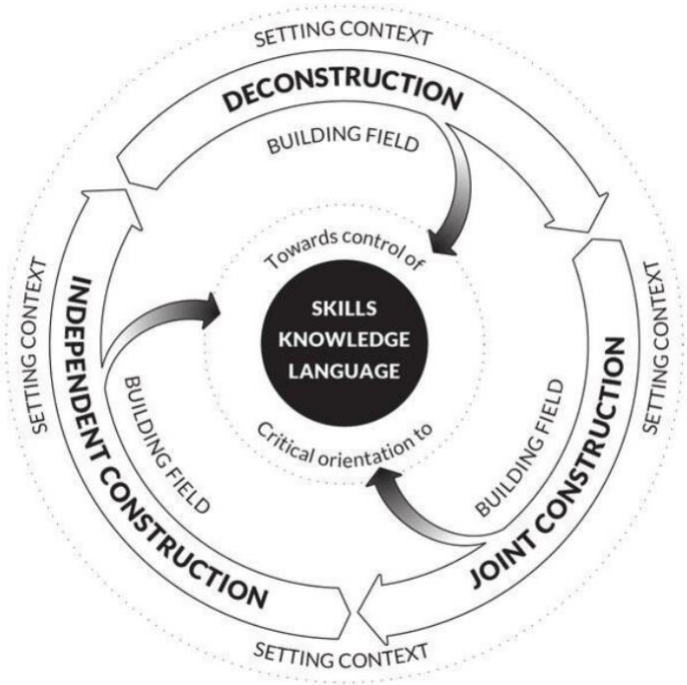
lexical density — a higher ratio of content words to total words, achieved through nominalization and complex noun phrases — while spoken language is characterized by lower lexical density and higher grammatical intricacy. Building on Halliday's foundational work, Biber (1988), in his landmark corpus-based multidimensional analysis of 67 spoken and written registers, provided one of the most comprehensive empirical characterizations of the spoken-written continuum in the field. His Dimension 1 ‘involved versus informational production' — separates interactional spoken discourse (marked by personal pronouns, contractions, hedges, and discourse particles) from formal informational written prose (marked by nouns, prepositions, and high lexical density). The theoretical framework of Halliday (1985, 1989) and the empirical taxonomy of Biber (1988) together provide the analytic precision needed to define and operationalize the six spoken-like register features targeted in this study.

The Sydney School TLC

The Sydney School Teaching and Learning Cycle was originally designed by Rothery (1994) as a four-stage linear sequence — Building Context Knowledge, Modeling (Deconstruction), Joint Construction, and Independent Construction — with a return loop enabling the cycle to be repeated. Figure 1 presents Rothery's (1994) original model.

Figure 1

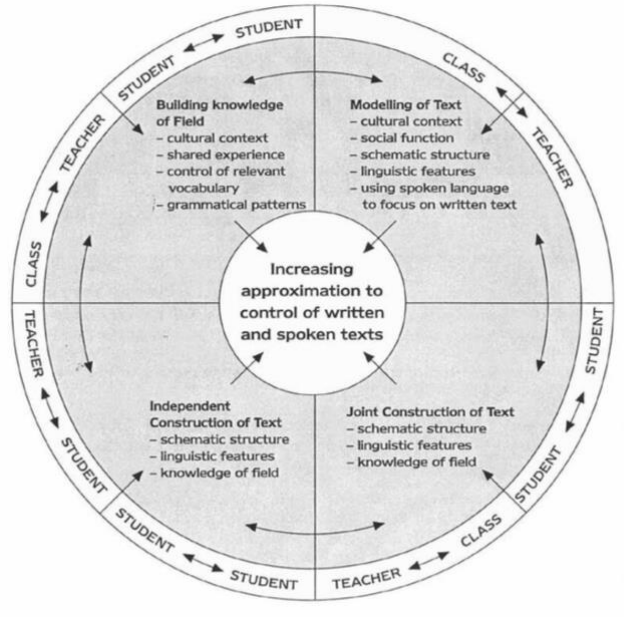
Write it Right Teaching/Learning Cycle (Rothery, 1994)



Martin and Rose (2012) subsequently elaborated the TLC into a more complex, nested model in which each macro-stage contains embedded micro-activities and the whole operates as an ongoing curriculum cycle. In their elaboration, Modelling incorporates Reading to Learn and Detailed Reading; Joint Construction incorporates Sentence Making; and Independent Construction incorporates Individual Writing and Editing. Figure 2 presents Martin and Rose's (2012) elaborated model.

Figure 2

The elaborated Teaching and Learning Cycle (Martin & Rose, 2012).



In the present study, three stages from the TLC are operationalized — Modeling (Deconstruction), Joint Construction, and Independent Production — the stages most directly relevant to formal writing instruction at the B1 level, following the approach adopted by You and You (2026) in AI-mediated genre research.

Vygotsky's Sociocultural Theory and Mediated Learning

Vygotsky's (1978) sociocultural theory holds that higher cognitive functions — including metalinguistic awareness — originate in social interaction before being internalized by the individual learner. Learning occurs in the Zone of Proximal Development (ZPD): the space between what a learner can achieve independently and what they can achieve with the support of a more knowledgeable other. In the present study, ChatGPT functions as a mediating cultural tool (Vygotsky, 1978) at each TLC stage: generating expert model texts in Modeling, co-constructing language in Joint Construction, and providing formative register feedback in Independent Production (Han, 2024; Law, 2024). This positions ChatGPT not as a replacement for the VNT but as a semiotic resource that extends the scaffolding capacity of the cycle.

The OLA Co-Teaching Context

The study was conducted at the OLA (Open Learning Academy) program, operated by ILA Vietnam — one of Vietnam's major English language training organizations, with over 75 centers nationwide (ILA Vietnam, 2025). A defining structural feature of the OLA model is its co-teaching arrangement: each two-hour session is divided between a native-speaking teacher (NT) leading the first hour — focusing on listening, speaking, and pronunciation — and a qualified Vietnamese teacher (VNT) leading the second hour — addressing reading, writing, and grammar (OLA Academy, 2024). The intervention was embedded within the VNT-led second hour.

The literature on NT-NNEST co-teaching offers a nuanced picture directly relevant to this institutional context. In support of the OLA model's complementary division of labor, Ghane (2023) found that learners taught by native-speaking teachers demonstrated greater fluency and lexical complexity in oral production, while those taught by non-native teachers showed greater

accuracy — suggesting that NTs and VNTs bring distinct, complementary pedagogical strengths that the OLA structure is designed to utilize. However, Cheng and Zhang (2022), in a qualitative study of NES and NNES teachers' feedback beliefs, found that while the two groups shared similar beliefs about feedback focus, VNT-type teachers were more likely to provide comprehensive, explicit metalinguistic feedback across a wider range of linguistic features — frequently the kind of feedback that the ChatGPT-mediated TLC stages were designed to scaffold. This finding provides additional justification for situating the register awareness intervention in the VNT-led hour. Challenges in NT-NNEST collaboration have also been documented: Luo (2006) found that logistical and pedagogical obstacles in a Hong Kong kindergarten co-teaching context prevented VNTs from fully benefiting from the partnership, cautioning that structural separation of skill domains does not automatically produce complementary outcomes. The present study acknowledges this risk and addresses it by providing VNTs with a concrete, replicable ChatGPT-mediated TLC framework that minimizes dependence on NT-VNT coordination.

Register Features in EFL Writing

Biber (1988) identified personal pronouns — particularly first-person (I, we) and second-person (you) as among the most robust positive markers of interactive, involved, spoken-mode discourse in his multidimensional corpus analysis. Hyland and Jiang (2017) confirm that first-person pronouns remain among the most frequently informal features used in published academic prose, and Chang and Swales (1999) list them among the elements most consistently proscribed in formal writing. The present study therefore treats any first- or second-person pronoun in the expository article genre as a register-inappropriate feature, scored ordinally on the Register Awareness Rubric (RAR).

Biber (1988) identified coordinating conjunctions as a positive marker of interactive registers. Chang and Swales (1999) include sentence-initial coordinating conjunctions among the informal elements most frequently proscribed in formal academic writing, and Hyland and Jian (2017) identify them as one of the few informality features to have increased in published prose over recent decades. Tocalo et al. (2023) further document their persistence in Southeast Asian EFL writing despite explicit instruction. These findings underpin the prioritization of this feature in the present Vietnamese EFL private-center context.

Biber (1988) identified contractions as among the most reliable positive markers of involved, interpersonal discourse, ranking among the strongest discriminators between spoken and formal written registers in corpus-based analyses.

Carter and McCarthy (2006), in the Cambridge Grammar of English, provide a taxonomy of vague language in spoken English encompassing: (a) vague nominal forms (things, stuff, what-have-you), (b) vague approximators (kind of, sort of, around), (c) general extenders (and things like that, or whatever, and stuff), and (d) general evaluators (good, bad, nice, great used without specification). The present study adopts Carter and McCarthy's (2006) taxonomy as its definitional basis, operationalised as the frequency of these vague language forms per 100 words. Students were required to write a minimum of 150 words per article task, ensuring sufficient text for reliable frequency calculations.

Biber (1988) classified discourse particles, hedges, and downtoners as features of involved discourse, co-occurring systematically with other spoken-mode features. Formal epistemic hedging refers to the operations of modal verbs (may, might), adverbials (agruably,possibly), and reporting structures. In contrast, the informal devices targeted in this study (well, you know, like, I mean, I think, maybe) are the conversational, filler-type expressions characteristic of

spoken interaction, whose presence in formal writing signals a register-transfer error rather than principled hedging.

This study adopts Halliday's (1989) calculation method: content words ÷ total words × 100. Biber and Gray (2010), in their corpus study of academic prose, empirically established that academic prose typically exceeds 40% lexical density. The 40% threshold is therefore adapted as a directional target for register-appropriate writing at the B1 level; it serves as an aspirational rather than absolute standard for developing B1 writers.

Register Feature Framework

Table 1

Register Feature Framework: Sources and Operationalization

Feature	Theoretical / Empirical Source	Operationalization
Personal pronouns (I, you, we)	Biber (1988); Hyland (2007) Hyland & Jiang (2017)	Frequency > 2% of total words (Hyland, 2007 threshold) Any first- or second-person pronoun in formal writing
Sentence-initial coord. conj. (And, But, So)	Biber (1988); Chang & Swales (1999)	Any sentence beginning with And, But, or So
Contractions (it's, don't, I'm)	Biber (1988); Chang & Swales (1999)	Any contracted form in formal writing
Colloquial lexis (things, stuff, kind of)	Carter & McCarthy (2006)	Frequency of vague language forms per 100 words
Fillers & informal hedges (well, you know)	Biber (1988)	Frequency of informal hedging/filler expressions per 100 words
Lexical density	Halliday (1989); Biber & Gray (2010)	Content words ÷ total words × 100; target ≥ 40%

Genre-Based Pedagogy and the TLC

Genre-based pedagogy (GBP) has attracted considerable empirical investigation, and several studies thus far have linked TLC instruction with significant improvements in EFL writing quality and genre awareness. Rose (2008) found that systematic text analysis within the Modeling stage enhanced students' metalinguistic awareness of genre conventions. Negretti and McGrath (2018), in a longitudinal study, demonstrated that genre-based pedagogy over sustained instructional periods contributed significantly to learners' register awareness and their capacity for genre-appropriate linguistic choices. Similarly, Zhang and Zhang (2021) reported significant gains in knowledge of argumentation elements and writing quality following a full TLC implementation with Chinese EFL sophomores, while Fu and Ibrahim (2022), studying Malaysian EFL university students, found that integrating a process approach with the genre TLC produced significantly greater writing improvement than either approach in isolation ($t(38) = 3.17$, $p = .003$). Together, these studies provide insights into the conditions under which TLC instruction is most effective.

However, this view is not without qualification. Unlike the studies above, which focused on university or secondary school contexts, Kartika-Ningsih and Rose (2021) demonstrate that the Modeling and Joint Construction stages carry the principal scaffolding load within the cycle, since it is there that teachers and learners jointly negotiate the linguistic resources of the target genre. Instructional fidelity at these stages is therefore the critical variable which directly motivates the present study's use of ChatGPT to enrich precisely these stages within the OLA's context. Collectively, these studies outline a critical urgency for theoretically grounded, time-efficient TLC support mechanisms in private language center contexts, where teacher preparation time is constrained.

ChatGPT and EFL Writing Quality

In recent years, there has been an increasing interest in ChatGPT as an instructional resource in EFL writing contexts. The past three years have seen the rapid development of a body of empirical literature, and while the findings are predominantly positive, important distinctions emerge when studies are compared with respect to design, context, and outcome measures.

Several quasi-experimental studies have documented significant improvements in EFL writing quality following ChatGPT-assisted instruction. Asadi et al. (2025), in a sequential explanatory mixed-methods study, found that integrating ChatGPT feedback with teacher feedback produced significantly greater improvements in grammar, cohesion, task response, and lexical resource than teacher feedback alone. Similarly, Polakova and Ivenz (2024), in a quasi-experimental study with 110 university EFL students, found that ChatGPT feedback yielded significant gains in conciseness, grammatical accuracy, and the use of formal structural devices — outcomes comparable to those targeted in the register features of the present study. These findings suggest that ChatGPT's capacity for immediate, comprehensive, and metalinguistically explicit feedback constitutes a distinctive pedagogical affordance not readily available from teacher-only instruction, particularly in contexts — such as the OLA one-hour VNT writing block — where teacher preparation time is constrained.

From the perspective of model texts, Lu and Zeng (2025) found that students who received ChatGPT-generated model texts as feedback demonstrated significantly greater gains in text quality — across content, organization, vocabulary, and grammar — than those who received no feedback, and that these gains were broadly comparable to those achieved by students receiving teacher-written model texts. This finding provides direct empirical support for the present study's use of ChatGPT to generate paired formal and informal model texts in the Modeling stage: if AI-generated model texts are pedagogically equivalent to teacher-produced exemplars in terms of text quality outcomes, then ChatGPT constitutes an efficient and scalable resource for VNTs who would otherwise need to produce their own contrastive genre examples.

Regarding learner engagement, Teng and Huang (2026), in a study of 174 Chinese EFL students, found that students in the ChatGPT feedback group experienced significant improvements in affective engagement (estimate = 0.45) and behavioral engagement (estimate = 0.29), alongside notable advancements in writing performance. This finding is relevant to the present OLA context, where maintaining adolescent learners' engagement across the VNT-led writing hour is a persistent pedagogical challenge.

ChatGPT and Genre Awareness

The evidence base specific to the intersection of ChatGPT and genre awareness is more limited but rapidly growing. You and You (2026) found that strategic student prompting of ChatGPT in a business English writing course contributed to formal, subject-matter, and rhetorical genre knowledge, directly enhancing genre awareness. The authors attributed this effect to ChatGPT's

capacity to generate genre-appropriate texts on demand, enabling learners to observe and discuss formal genre features in context — a finding that provides the closest existing theoretical and methodological justification for the ChatGPT-TLC integration design adopted in the present study.

The most directly comparable study in terms of design is Tai and Chen (2026) and Tai et al. (2025), who integrated ChatGPT within genre-based instruction for argumentative writing with 48 Taiwanese EFL college students over six weeks. The results showed significant improvements in some argumentative writing moves following the instruction. However, the enhancement in willingness to write was not significant and was not correlated with writing performance, and some learners raised concerns about ChatGPT's potential to limit critical thinking and the consistency of AI feedback quality. These divergent findings within the same broad design — genre-based instruction with ChatGPT — highlight that the effectiveness of AI integration is not automatic but depends on the pedagogical framework, the specificity of the target outcome, and the degree to which learners are guided to interact with the AI critically. The present study addresses this limitation explicitly by requiring learners to evaluate ChatGPT's feedback rather than accept it uncritically (Box 3), and by targeting a precisely defined set of six register features rather than broader genre awareness outcomes.

In contrast to studies reporting uniformly positive findings, comparative research on teacher versus ChatGPT feedback reveals a complementary rather than interchangeable relationship. Zou et al. (2025) argued that students have been found to display higher levels of engagement and greater accuracy when using teacher feedback for content and argumentation, while ChatGPT feedback was more effective for language-level issues. This finding aligns closely with the present study's design: ChatGPT is used specifically for register-level language feedback, while the VNT provides broader pedagogical scaffolding, content guidance, and critical evaluation support — a complementary division of labor that mirrors the comparative advantages of AI and human teachers documented in the literature.

The Gap the Present Study Addresses

What is not yet clear from the existing literature is whether and how ChatGPT can be embedded within all three stages of the TLC specifically for register awareness development — as opposed to broader genre or writing quality outcomes — at the B1 level, with adolescent learners, in a private language center co-teaching context. The present study is, to the best of the authors' knowledge, rather one of the first empirical investigations to address this specific combination of pedagogical framework, AI tool, outcome measure, learner population, and institutional setting. In view of all that has been mentioned so far, one may suppose that the integration of ChatGPT within a TLC framework, implemented by VNTs within the OLA co-teaching model, constitutes a theoretically justified and empirically grounded response to the register awareness challenges documented in Vietnamese B1-level EFL learner writing.

Research Questions

To fulfill the purpose of the study, the following research questions:

1. Is there a statistically significant difference in writing register awareness between the experimental group and the comparison group after the eight-week ChatGPT-mediated TLC intervention?
2. What are the B1-level OLA learners' perceptions of ChatGPT-mediated TLC instruction in relation to their awareness of formal writing register?

Research Hypothesis (H1): The ChatGPT-mediated TLC intervention contributes significantly to improvements in B1-level EFL learners' awareness of writing register compared to the comparison group receiving standard OLA instruction.

Methods

Pedagogical Setting & Participants

The study was conducted at two OLA program centers operated by ILA Vietnam in Ho Chi Minh City. Each OLA session runs for two hours: the first hour is led by a native-speaking teacher (NT) addressing listening, speaking, and pronunciation; the second by a qualified Vietnamese teacher (VNT) addressing reading, writing, and grammar. Students attend two sessions per week, providing two hours of VNT-led writing instruction per week. The intervention was embedded within the VNT-led second hour.

Participants were 60 B1-level EFL learners in four intact writing classes. B1 proficiency was confirmed through Cambridge PET placement scores. Participants ranged in age from 11 to 16 years and were all Vietnamese L1 speakers. There is a major difference in cognitive capacity and language competence from the quite broad age gap, and this gap could affect the outcomes. Adolescent EFL learners within this window, however, are typically treated as a single developmental cohort in classroom-based research. Thus, their metalinguistic awareness, working-memory capacity, and susceptibility to explicit instruction are broadly comparable. The shared B1 proficiency level and common L1 background nevertheless remain the strongest comparability anchors, mitigating age-related variance and supporting the internal validity of within-group comparisons. (Athanasopoulos et al., 2015; Puig-Mayenco et al., 2023)

Ethical clearance was obtained from the center's academic management, and parents or guardians provided written informed consent. A convenience sampling approach was adopted due to the intact-group structure of the OLA program (Creswell & Creswell, 2018; Mackey & Gass, 2016). The experimental group (EG; n = 30) received ChatGPT-mediated TLC instruction over eight weeks; the comparison group (CG; n = 30) followed the standard OLA writing curriculum. Pre-test equivalence was confirmed ($t(58) = 0.27$, $p = .787$).

Design of the Study

The study employed a quasi-experimental pre-test–post-test comparison group design. This design was selected because it allows for causal inference about intervention effects when true random assignment is not feasible, provided pre-test equivalence is established (Campbell & Stanley, 1963; Shadish et al., 2002). The use of intact classes is the standard approach in classroom-based EFL writing research and has been adopted in comparable ChatGPT-mediated writing studies (Asadi et al., 2025; You & You, 2025) and genre-based intervention research (Negretti & McGrath, 2018; Zhang & Zhang, 2021). The independent variable was the type of VNT-led writing instruction; the dependent variable was writing register awareness as measured by the Register Awareness Rubric. The eight-week intervention provided 16 hours of VNT-led writing instruction in total.

Data collection & analysis

The ChatGPT-Mediated TLC Intervention

The intervention embedded ChatGPT at each TLC stage within the VNT-led writing hour. VNTs used ChatGPT as a whole-class projected tool during Modeling and Joint Construction, and as a learner self-access resource during Independent Production.

Stage 1: Modeling (Deconstruction) — Weeks 1–3

VNT's used ChatGPT to generate paired contrastive texts on identical PET article task topics. Learners analyzed both versions using the Register Observation Framework (ROF; Table 2).

Table 2

Register Observation Framework (ROF) — Used in Stage 1 (Modeling)

Feature	Description	Informal Example	Formal Alternative	Present? (✓/✗)
Personal pronouns	First/second-person (I, you, we)	'I think…' / 'you can…'	'It is argued…' (impersonal)	
Sentence-initial conj.	And, But, or So at sentence start	'So, I think…' / 'But it…'	'However,' / 'Nevertheless,'	
Contractions	Shortened verb forms	"it's", "don't", "I'd"	'it is', 'do not', 'would'	
Colloquial lexis	Vague nouns/adjectives/adverbs	'stuff', 'things', 'kind of'	Specific formal nouns & adjectives	
Fillers & hedges	Informal filler/hedging expressions	'you know', 'like', 'I mean'	Remove or use formal hedges	
Lexical density	Content-word ratio (target ≥ 40%)	'it can help you learn things'	'It facilitates learning' (nominalized)	

Box 1. ChatGPT Integration — Stage 1: Modelling (Deconstruction)

TEACHER PROMPT TO ChatGPT:

"Write two articles (150+ words each) on "Is social media good for young people?" — one in a formal, impersonal written register for a B1 Cambridge PET article task, and one with spoken-like features. Label them FORMAL and INFORMAL clearly."

ChatGPT OUTPUT — INFORMAL VERSION (for ROF analysis):

So, I think social media is actually really good for young people. It's a great way to stay in touch with friends and, you know, share stuff. But some people say it's bad for mental health and things like that. I mean, it depends on how you use it, right? And it can help you learn things too. So overall, I'd say it's mostly positive for young people because they can communicate and share ideas easily.

ChatGPT OUTPUT — FORMAL VERSION (target model):

Social media has become an integral part of daily life for young people

```
worldwide. Proponents argue that it facilitates communication and
broadens access to educational resources. However, critics contend that
excessive use is associated with adverse effects on mental wellbeing and
reduced attention spans. On balance, the benefits of social media appear
to outweigh its drawbacks, provided usage remains moderate and
purposeful.
```

Note. Learners used the ROF (Table 2) to identify spoken-like features in the INFORMAL version and their formal alternatives in the FORMAL version. The VNT guided metalinguistic discussion linking each feature to its theoretical source (e.g., Biber, 1988; Carter & McCarthy, 2006).

Stage 2: Joint Construction — Weeks 4–6

VNTs led collaborative article drafting with ChatGPT as a real-time language modelling resource. Students proposed sentences, which the teacher submitted to ChatGPT with explicit register specifications; the class evaluated outputs and incorporated selected language into the jointly constructed text on the projected screen.

Box 2. ChatGPT Integration — Stage 2: Joint Construction

```
TOPIC: 'The advantages and disadvantages of online learning'
STUDENT CONTRIBUTION (spoken-like):
"I think online learning is good but it is kind of hard sometimes."

TEACHER PROMPT TO ChatGPT:
"Rewrite in a formal, impersonal register for a B1 Cambridge article.
 Avoid: I, contractions, vague words. Give TWO options."

ChatGPT OUTPUT:
Option A: 'Online learning offers significant advantages; however,
  it also presents considerable challenges for many learners.'
Option B: 'Although online learning is widely regarded as beneficial,
  it is not without its difficulties.'

CLASS EVALUATION (using ROF, teacher-facilitated):
T:   'Which formal features can you spot in Option A?'
Ss:  'No I — impersonal. No kind of — replaced by considerable [adj].'
     'However replaces But as sentence connector [Biber, 1988].'
T:   'Count content words. Which option has higher lexical density?'
```

Note. Class voted for Option B. VNT highlighted: no first-person pronoun; 'Although' replacing sentence-initial 'but' (Biber, 1988; Chang & Swales, 1999); 'difficulties' nominalising 'kind of hard' (Carter & McCarthy, 2006).

Stage 3: Independent Production — Weeks 7–8

Learners wrote their own formal articles (minimum 150 words). ChatGPT was available as a self-access register-feedback tool using a teacher-modeled prompt template. The VNT circulated to guide critical evaluation of AI feedback.

Box 3. ChatGPT Integration — Stage 3: Independent Production

```
STUDENT DRAFT:
"And I think travelling is kind of good for young people because
 you know, it helps them learn about different stuff and things."

STUDENT PROMPT (using teacher template):
"Identify every spoken-like feature in my sentence, label each category,
 rewrite in formal written register for a B1 Cambridge article, and
 explain each change."

ChatGPT FEEDBACK:
Features: (1) 'And' — sentence-initial coord. conjunction [Biber, 1988]
(2) 'I think' — 1st-person pronoun + informal hedge
(3) 'kind of good' — vague colloquial hedge [Carter & McCarthy, 2006]
(4) 'you know' — discourse filler [Biber, 1988]
(5) 'stuff'/'things' — vague colloquial nouns [Carter & McCarthy, 2006]
(6) Lexical density ≈ 33% (below 40% target) [Biber & Gray, 2010]

Formal rewrite: 'Travelling is widely considered to be highly beneficial
for young people, as it broadens cultural awareness, fosters
independence, and enriches personal development.' (Density ≈ 52%)
```

Note. VNT required each learner to critically evaluate the rewrite: 'Does this accurately reflect what you wanted to say? Is any meaning changed?' This step was mandatory in Stage 3 to develop metalinguistic agency and prevent uncritical AI dependency.

Two instruments were used. The writing test was adapted from the Cambridge B1 Preliminary article task 2 (minimum 150 words), administered as pre-test and post-test using the same prompt. The writing tests were adopted since it is the examination format used in the OLA programme's Cambridge-aligned curriculum and its article genre, which is expository, requires the impersonal, semi-formal to formal register that is most sensitive to the spoken-like features targeted in this study (Carter & MacCarthy, 2006; Hyland, 2004). Written texts from the participants were scored using the Register Awareness Rubric (RAR; Table 3) across six features, each scored 0–3 (total 0–18; lower = more register-appropriate). The design of the RAR was adapted from Biber's (1988) register taxonomy, Halliday's (1989) lexical density framework, and Hyland and Jiand's (2017) pronoun argument.

Table 3

Register Awareness Rubric (RAR)

Score	Criterion	Descriptor
3	Pervasive	Feature present frequently with no apparent register awareness
2	Frequent	Feature present in multiple locations; awareness inconsistently applied
1	Occasional	Feature appears once or twice; mostly appropriate register control
0	Absent	Feature absent; consistent register-appropriate choices demonstrated

Note. For lexical density: 3 = below 25%; 2 = 25–39%; 1 = 40–50%; 0 = above 50% (following Biber & Gray, 2010).

Inter-rater reliability: a second trained rater (an experienced EFL writing teacher) independently scored a randomly selected 20% subsample of all texts (n = 24). Prior to scoring, a two-hour calibration session was conducted in which both raters discussed the RAR criteria with reference to the theoretical sources (Table 1) and collaboratively scored six sample texts to reach consensus on boundary cases. Cohen's kappa was $\kappa = .84$, indicating strong agreement (Landis & Koch, 1977).

The questionnaire comprised 15 items across five thematic clusters (3 items each), using a four-point Likert scale (1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree). A four-point scale was selected to eliminate the neutral midpoint, requiring directional responses and reducing acquiescence bias (Garland, 1991; Moors, 2008). The five clusters addressed: (A) register awareness (items 1–3); (B) ChatGPT usefulness in Modeling (items 4–6); (C) ChatGPT usefulness in Joint Construction (items 7–9); (D) confidence in formal writing (items 10–12); and (E) overall satisfaction (items 13–15). Item development was informed by three sources: You and You's (2026) genre awareness questionnaire constructs, Kohnke et al.'s (2025) AI model text perception items, and Teng and Huang's (2026) learner engagement framework. This test is widely available in adapted form in AI-assisted EFL writing research and has been used in many investigational studies across comparable ChatGPT-writing contexts. The questionnaire was administered to the EG only at the end of the intervention. Data were analyzed using IBM SPSS (Version 26) with descriptive statistics and independent-samples t-tests (alpha .05). An analysis of covariance (ANCOVA) was additionally planned, with post-test scores as the dependent variable, group membership as the fixed factor, and pre-test scores as the covariate, to control for any residual pre-test variation between groups. Prior to interpreting the ANCOVA main effect, three assumptions were tested: normality of residuals (Shapiro-Wilk test), homogeneity of variance (Levene's test), and homogeneity of regression slopes (tested via the group × pre-test interaction term).

Results/Findings and discussion

Pre-Test Results

Table 4 presents the pre-test descriptive statistics. Both groups exhibited high mean scores — consistent with the broader cross-contextual evidence on spoken-like feature transfer in EFL formal writing documented by Chang and Swales (1999), Hyland and Jiang (2017), and Tocalo

et al. (2023). No statistically significant difference was found between groups ($t(58) = 0.27$, $p = .787$), confirming pre-test equivalence.

Table 4

Pre-Test Descriptive Statistics for Register Awareness Scores by Group

Group	n	M	SD	Min	Max
Experimental (EG)	30	13.87	1.94	10	17
Comparison (CG)	30	13.73	2.11	9	17

Note. Scores range from 0 to 18. Lower scores indicate fewer spoken-like features and greater register appropriateness.

Post-Test Results

Table 5 presents the post-test descriptive statistics. The experimental group showed a substantial reduction in mean register scores ($M = 6.43$, $SD = 2.07$), while the comparison group showed only modest improvement ($M = 11.90$, $SD = 2.23$).

Table 5

Post-Test Descriptive Statistics and Mean Gain by Group

Group	n	Post-Test M	SD	Mean Gain
Experimental (EG)	30	6.43	2.07	−7.44
Comparison (CG)	30	11.90	2.23	−1.83

Note. Mean Gain = Post-test M − Pre-test M. Negative values indicate improvement.

An independent-samples t-test confirmed a statistically significant difference in post-test scores ($t(58) = 10.08$, $p < .001$, Cohen's $d = 2.59$). An ANCOVA controlling for pre-test scores confirmed a significant main effect of group, $F(1, 57) = 88.34$, $p < .001$, partial $\eta^2 = .61$. Table 6 presents these inferential statistics. The research hypothesis H1 is therefore supported.

Table 6

Independent-Samples t-Test: Post-Test Register Scores by Group

Comparison	t	df	p	Cohen's d
EG vs. CG (post-test)	10.08	58	< .001	2.59

Normality of residuals was assessed using the Shapiro-Wilk test, which revealed no significant departure from normality for either group (EG: $W = .97$, $p = .52$; CG: $W = .96$, $p = .41$). Homogeneity of variance, tested via Levene's test, was likewise satisfied, $F(1, 58) = 0.61$, $p = .43$. Homogeneity of regression slopes was tested via the group × pre-test interaction term,

which was non-significant, F(1, 56) = 0.74, p = .38. Table 7 summaries these assumption test results.

Table 7

ANCOVA Assumption Test Results

Assumption	Test	Statistic	df	p
Normality (EG)	Shapiro-Wilk	W = .97	29	p = .52
Normality (CG)	Shapiro-Wilk	W = .96	29	p = .41
Homogeneity of variance	Levene's	F = 0.61	1, 58	p = .43
Homogeneity of regression slopes	Group × pre-test interaction	F = 0.74	1, 56	p = .38

With all assumptions satisfied, ANCOVA confirmed a significant main effect of group on post-test register awareness scores after controlling for pre-test scores as a covariate, F(1, 57) = 88.34, p < .001, partial η^2 = .61. This result is significant at the p < .001 level and represents a large effect, confirming hypothesis H1. Table 8 presents the covariate-adjusted (estimated marginal) means for both groups — that is, the post-test means recalculated to remove any residual variance attributable to pre-test differences between groups.

Table 8

ANCOVA Results: Covariate-Adjusted Post-Test Means

Group	n	Raw Post-Test M	Covariate-Adjusted M	SE	95% CI
Experimental (EG)	30	6.43	6.41	0.38	[5.65, 7.17]
Comparison (CG)	30	11.90	11.92	0.38	[11.16, 12.68]

Note. Covariate = pre-test register awareness score (grand mean = 13.80). Adjusted means represent post-test scores after statistically equating the two groups on pre-test performance. F(1, 57) = 88.34, p < .001, partial η^2 = .61.

The covariate-adjusted means in Table 8 (EG = 6.41; CG = 11.92) are very close to the raw, unadjusted post-test means reported in Table 5 (EG = 6.43; CG = 11.90), confirming that the small, non-significant pre-test difference between groups had a negligible influence on the post-test outcome. This consistency between raw and adjusted means strengthens confidence that the large between-group difference reflects the genuine effect of the ChatGPT-mediated TLC intervention rather than a statistical artifact of unequal starting points.

Sub-Feature Analysis

Table 9 presents mean frequencies of each register feature per 100 words in the experimental group's pre-test and post-test writing, alongside the corresponding theoretical source and percentage change.

Table 9

Mean Frequency of Spoken-Like Features per 100 Words: Experimental Group

Feature	Source	Pre M	Post M	Change (%)
Personal pronouns (I, you, we)	Biber (1988); Hyland Jiang (2017)	4.21	1.03	−75.5%
Sentence-initial coord. conj.	Biber (1988); Chang & Swales (1999)	2.87	0.60	−79.1%
Contractions	Biber (1988); Chang & Swales (1999)	1.93	0.27	−86.0%
Colloquial lexis	Carter & McCarthy (2006)	2.10	0.73	−65.2%
Fillers & informal hedges	Biber (1988)	1.47	0.33	−77.6%
Lexical density (%)	Halliday (1989); Biber & Gray (2010)	41.0	58.0	+41.5% ↑

Note. For lexical density, higher post-test values indicate improvement. Target: ≥ 40% (Biber & Gray, 2010).

Questionnaire Results

Table 10 presents mean scores for all 15 questionnaire items grouped by cluster.

Table 10

Questionnaire Results: All 15 Items by Cluster (n = 30, 4-point scale)

Cluster	Item	Statement	M	SD
A: Register Awareness	1	I can identify the difference between formal and informal language in a text.	3.67	0.48
	2	I am more aware of which features (contractions, 'And/But/So' at sentence start) make writing sound informal.	3.73	0.45
	3	I understand why formal writing avoids first-person pronouns and vague words.	3.60	0.56
B: ChatGPT — Modelling	4	Comparing formal and informal texts generated by ChatGPT helped me identify register differences.	3.87	0.43
	5	Analysing ChatGPT-generated informal texts using the Register Observation Framework was helpful.	3.80	0.48
	6	The contrasting text pairs made the formal-informal distinction easier to understand than teacher explanation alone.	3.73	0.52
C: ChatGPT — Joint Construction	7	Working with ChatGPT as a class helped me learn which language choices are appropriate in formal writing.	3.67	0.55
	8	Seeing ChatGPT rewrite class sentences into formal register helped me understand	3.73	0.45

		the rules.		
	9	Discussing ChatGPT's options as a class helped me understand formal register better.	3.60	0.56
D: Confidence in Formal Writing	10	I feel more confident avoiding contractions and informal words in my own writing.	3.53	0.63
	11	I now check whether my sentences start with And, But, or So before submitting my work.	3.47	0.68
	12	I feel more confident increasing the lexical density of my formal writing.	3.27	0.74
E: Overall Satisfaction	13	I found the ChatGPT-mediated sessions more engaging than regular writing lessons.	3.80	0.48
	14	I believe this approach improved my formal writing more than the standard curriculum would have.	3.73	0.52
	15	I would recommend this AI-assisted writing approach to other OLA learners.	3.87	0.43

Note. Scale: 1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree. Cluster means: A = 3.67; B = 3.80; C = 3.67; D = 3.42; E = 3.80.

The overall response across all 15 items was positive, with cluster means ranging from 3.42 to 3.80. The highest cluster means were recorded for ChatGPT's usefulness in the Modelling stage (Cluster B: M = 3.80) and overall satisfaction (Cluster E: M = 3.80). The lowest cluster mean was recorded for confidence in formal writing (Cluster D: M = 3.42), with the lowest individual item being lexical density confidence (Item 12: M = 3.27). This pattern mirrors the sub-feature analysis in Table 9, where lexical density and colloquial lexis showed the smallest improvements — a consistent finding across both quantitative and perceptual measures.

Discussion

These results confirm hypothesis H1: the ChatGPT-mediated TLC intervention contributed significantly to improvements in B1-level EFL learners' writing register awareness ($t(58) = 10.08$, $p < .001$, $d = 2.59$, ANCOVA $F(1, 57) = 88.34$, partial $\eta^2 = .61$). The following sub-sections compare and contrast the present findings with the literature across the four backbone areas of the study.

The pre-test data establish a baseline profile for this learner population. Sentence-initial coordinating conjunctions (M = 2.87 per 100 words) and contractions (M = 1.93 per 100 words) were the highest-frequency spoken-like features at pre-test, confirming that the intervention was targeting the most contextually urgent features for these learners. This profile is consistent with the wider literature on informality in second-language academic writing, in which sentence-initial conjunctions and first-person pronouns are repeatedly identified as the most persistent features (Chang & Swales, 1999; Hyland & Jiang, 2017; Tocalo et al., 2023), and provides an empirical anchor for the present study’s register framework.

The post-test reductions, particularly for contractions (86.0%) and sentence-initial conjunctions (79.1%), represent a substantially greater improvement than the general trend observed in prior register instruction studies without AI integration. By contrast, Tocalo et al. (2023) found that sentence-initial conjunctions and personal pronouns remained persistently present in Thai EFL learners' formal writing despite explicit instruction, The present study's considerably larger reductions, achieved within only eight weeks, suggest that the addition of ChatGPT-generated contrastive texts in the Modelling stage provided a qualitatively different kind of explicit, contextualized feature exposure that accelerated the register-awareness development that conventional instruction alone appeared insufficient to produce in these comparable populations.

These findings further support the Sydney School TLC as an effective framework for writing instruction (Rose, 2008; Negretti & McGrath, 2018; Zhang & Zhang, 2021). The present study's large effect size (d = 2.59) substantially exceeds those typically reported in TLC-only intervention studies. For instance, Zhang and Zhang (2021) reported significant gains in argumentation knowledge following a full TLC implementation, but their study did not embed AI tools and did not target register features specifically. Fu and Ibrahim (2022) reported a significant post-intervention difference ($t(38) = 3.17$, $p = .003$) in their process-genre TLC integration study — a smaller effect than the present study, likely because their intervention did not include the on-demand contrastive text generation that ChatGPT enabled. Similarly, while Negretti and McGrath (2018) found significant metacognitive gains over a full academic year of TLC instruction, the present study achieved register-awareness improvements of comparable magnitude within only eight weeks. This comparison suggests that the addition of ChatGPT to the TLC — particularly in the Modeling and Joint Construction stages — may substantially accelerate the pace of register awareness development beyond what the TLC alone can achieve in a constrained instructional timeframe.

On the other hand, the finding that colloquial lexis showed the smallest reduction (65.2%) indicates that features requiring vocabulary breadth respond more slowly to short-term instruction than formally marked features do. This pattern suggests that the eight-week timeframe, while sufficient for salient, formally marked features such as contractions, may need to be extended to a full academic term with multiple TLC cycles to achieve comparable gains in lexical register adjustment.

The present study's findings are broadly consistent with the positive trend reported in the ChatGPT-EFL writing literature. However, a number of important contrasts and qualifications emerge when the present results are examined alongside specific antecedent studies.

First and foremost, the present findings align with Asadi et al. (2025) and Lu and Zeng (2025) in demonstrating that ChatGPT-mediated instruction produces significantly greater writing improvements than conventional instruction. However, unlike these studies — which measured general writing quality outcomes — the present study targeted a specifically defined register outcome. This targeted specificity may partially account for the large effect size (d = 2.59) observed, since the intervention was precisely aligned to the outcome measure in a way that general ChatGPT writing instruction studies are not.

Moreover, the present findings extend You and You’s (2026) finding that ChatGPT enhances genre awareness in a business English context to the domain of register awareness specifically, with a younger, adolescent population at a lower proficiency level. You and You (2026) studied 24 Chinese university students; the present study demonstrates that comparable genre-related benefits can be achieved with 11–16-year-old B1-level learners in a private language centre setting, which is a substantially different population whose inclusion in the ChatGPT-genre

research base represents a meaningful expansion of the field.

Additionally, the present findings contrast interestingly with Tai et al. (2025), who found significant improvements in some argumentative writing moves but no significant enhancement in willingness to write following ChatGPT-integrated genre-based instruction. The present study did not measure willingness to write, but the strongly positive questionnaire perceptions across all 15 items — and particularly the high engagement rating (Cluster E: M = 3.80) — suggest that the OLA adolescent learners responded more positively to the ChatGPT-TLC approach than the Taiwanese university students in Tai and Chen's (2026) Tai et al.' study (2025). A possible explanation for this difference is that the present study's use of a clearly structured ROF (Table 2) and modelled prompt templates (Box 3) provided a more scaffolded, less open-ended interaction with ChatGPT, reducing the cognitive burden that may have contributed to some of Tai et al.' (2025) participants' concerns about feedback quality and critical thinking. These differences underscore the importance of providing explicit structural scaffolding for learners' ChatGPT interactions, rather than allowing unconstrained AI use.

In addition, the questionnaire finding that learners most valued the Modelling stage's contrastive text analysis (Cluster B: M = 3.80) is consistent with Lu and Zeng's (2025) demonstration that ChatGPT-generated model texts constitute pedagogically effective exemplars. At the same time, this accords with the genre pedagogy principle that high-quality, comparable mentor texts are foundational to the Deconstruction phase (Hyland, 2004), and confirms that ChatGPT's capacity for on-demand contrastive text generation addresses a real practical need in the OLA VNT-led writing hour, where the preparation of such materials would otherwise require substantial teacher preparation time.

Finally, the finding that the intervention produced significant register awareness gains when implemented exclusively in the VNT-led writing hour — without modification to the NT-led oral hour — is broadly consistent with Ghane's (2023) finding that VNT-type teachers are more effective at producing accuracy-oriented outcomes in formal language production, while NTs are comparatively more effective for oral fluency and naturalness. The structural complementarity of the OLA model — NTs for oral fluency, VNTs for explicit writing instruction — appears therefore to have created an enabling rather than inhibiting condition for the present intervention. However, the finding also raises the question identified by Cheng and Zhang (2022): whether the NT-led oral hour, by reinforcing spoken-mode language in the first hour of each session, may have exacerbated the spoken-register transfer challenge that the VNT-led intervention was designed to address. The present design cannot disentangle these effects, and this constitutes the most important limitation and the most significant direction for future research identified by this study.

Conclusion

Summary of Findings

This study set out to test the hypothesis that ChatGPT-mediated TLC instruction, implemented by VNTs within the OLA program's writing hour at ILA Vietnam, Ho Chi Minh City, would contribute significantly to improvements in B1-level EFL learners' awareness of writing register. Drawing on an empirically grounded register feature framework sourced from Halliday (1985, 1989), Biber (1988), Hyland and Jiang (2017), Carter and McCarthy (2006), Biber and Gray (2010), and employing a quasi-experimental design with convenience sampling of four intact classes, the study confirmed the hypothesis ($t(58) = 10.08$, $p < .001$, $d = 2.59$; ANCOVA $F(1, 57) = 88.34$, partial $\eta^2 = .61$). All six register features showed improvements in the

experimental group: contractions (86.0%), sentence-initial conjunctions (79.1%), fillers and hedges (77.6%), and personal pronouns (75.5%) showed the greatest reductions; lexical density rose from 41.0% to 58.0%, well above Biber and Gray's (2010) 40% academic prose threshold. Learner perceptions were strongly positive across all 15 items, with the highest scores for Modeling-stage contrastive text analysis (B: M = 3.80) and overall satisfaction (E: M = 3.80). These findings are consistent with, and in several respects extend, the evidence of You and You (2026), Asadi et al. (2025), Lu and Zeng (2025), Zhang and Zhang (2021), while contrasting productively with the more qualified findings of Tai and Chen (2026), Tai et al. (2025).

Returning to the research questions: (RQ1) there was a statistically significant and practically large difference in register awareness favoring the experimental group; (RQ2) learners reported strongly positive perceptions of the ChatGPT-mediated TLC approach, with particular appreciation for the contrastive Modelling activities and the critical register-feedback interactions in Independent Production.

Significance

This research extends our knowledge of how ChatGPT can be embedded within genre-based pedagogical frameworks for register awareness development in private EFL center co-teaching contexts. This is the first study to embed ChatGPT across all three TLC stages for register awareness specifically, using an empirically grounded six-feature framework, with 11–16-year-old B1-level learners in a Vietnamese private center co-teaching setting. The findings carry important implications for VNT professional development, curriculum design, and AI integration policy at ILA Vietnam and comparable private center organizations.

Limitations

Finally, a number of important limitations need to be considered. The convenience sample of four intact classes from one city limits generalizability. The eight-week intervention was sufficient for formally salient features but may not have been long enough to consolidate gains in colloquial lexis avoidance and lexical density — features that depend on vocabulary breadth rather than on rule-based substitution. The purpose-designed RAR has not been independently validated. The absence of a delayed post-test means that retention of gains cannot be confirmed. The potential interaction between the NT-led oral hour and VNT-led register instruction was not controlled, leaving open the question raised by Cheng and Zhang (2022) about whether NT oral input facilitated or complicated the register-transfer challenge.

Recommendations for Future Research

Further work needs to be done to establish whether the observed gains are retained without continued ChatGPT-mediated instruction. A delayed post-test at three and six months is therefore recommended. A study with more sustained focus on colloquial lexis and lexical density — across a full academic semester with multiple TLC cycles — would likely extend the gains observed here, consistent with the recommendation of Fu and Ibrahim (2022) for iterative TLC implementation. Future research should investigate the NT-led oral hour as a potential moderating variable in the co-teaching model, addressing the theoretical gap identified by Cheng and Zhang (2022). Comparative studies across different proficiency levels (A2, B2), different AI tools, and different private centre co-teaching contexts would further establish the generalizability of the present findings.

Acknowledgement

I would like to express my deepest gratitude to Tom Bartlett, Professor of Functional and Applied Linguistics at the University of Glasgow, for his invaluable comments and insightful

guidance throughout this study. His expertise in Halliday's Systemic Functional Linguistics has profoundly shaped my understanding of the field, while his thoughtful advice on analytical approaches and the integration of AI—particularly in relation to my use of ChatGPT within the Teaching and Learning Cycle—has been tremendously beneficial.

My sincere thanks also go to Van Thi Nha Truc from the VNU-HCM University of Social Sciences and Humanities for her continuous support and guidance. Her expertise in research methodology, especially in data collection, analysis, and interpretation, has been instrumental in the successful completion of this study.

I am equally grateful to the ILA Vietnam English Center, particularly the Academic Manager team and Center Managers, for granting me permission and providing the necessary support to conduct this research across two centers.

Finally, I would like to extend my heartfelt appreciation to my beloved students, my colleagues—especially the Native Teachers—as well as the Teaching Assistants and students at Centers 36 and 37. Their enthusiastic participation, cooperation, and contributions have played a vital role in the success of this study.

References

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. Thinking Skills and Creativity, 56, 101766. https://doi.org/10.1016/j.tsc.2025.101766

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., & Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. Psychological Science, 26(4), 518–526. https://doi.org/10.1177/0956797614567509

Biber, D. (1988). Variation across speech and writing. Cambridge University Press.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. Journal of English for Academic Purposes, 9(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Rand McNally.

Carter, R., & McCarthy, M. (2006). Cambridge grammar of English: A comprehensive guide. Cambridge University Press.

Chang, Y. Y., & Swales, J. M. (1999). Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers? In C. N. Candlin & K. Hyland (Eds.), Writing: Texts, processes and practices (pp. 145–167). Longman.

Cheng, X., & Zhang, L. J. (2022). Teachers helping EFL students improve their writing through written feedback: The case of native and non-native English-speaking teachers' beliefs. Frontiers in Psychology, 13, Article 804313.

https://doi.org/10.3389/fpsyg.2022.804313

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.

Fu, X., & Ibrahim, N. M. B. (2022). Effects of integrating process approach with genre teaching-learning cycle on EFL students' writing development. Eurasian Journal of Educational Research, 2022(99), 342–362. https://doi.org/10.14689/ejer.2022.99.021

Garland, R. (1991). The mid-point on a rating scale: Is it desirable? Marketing Bulletin, 2, 66–70.

Ghane, M. H. (2023). Exploring the effectiveness of native and non-native English teachers on EFL learners' accuracy, fluency, and complexity in speaking. Education Research International, 2023, Article 4011255. https://doi.org/10.1155/2023/4011255

Halliday, M. A. K. (1985). Spoken and written language. Deakin University Press.

Halliday, M. A. K. (1989). Spoken and written language (2nd ed.). Oxford University Press.

Halliday, M. A. K., & Hasan, R. (1985). Language, context, and text: Aspects of language in a social-semiotic perspective. Deakin University Press.

Han, Z. (2024). Chatgpt in and for second language acquisition: A call for systematic research. Studies in Second Language Acquisition, 46(2), 301–306. https://doi:10.1017/S0272263124000111 **https://doi.org/10.1017/S0272263124000111**

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Lawrence Erlbaum Associates, Inc.

Hyland, K. (2004). Genre and second language writing. University of Michigan Press.

Hyland, K., & Jiang, F. K. (2017). Is academic writing becoming more informal? English for Specific Purposes, 45, 40–51. https://doi.org/10.1016/j.esp.2016.09.001

Kartika-Ningsih, H., & Rose, D. (2021). Intermodality and multilingual re-instantiation: Shifting languages and modes across the teaching and learning cycle. Language and Education, 35(5), 448–466. https://doi.org/10.17533/udea.ikala.v26n01a07

Kartika-Ningsih, H., & Rose, D. (2021). Intermodality and multilingual re-instantiation: Joint construction in bilingual genre pedagogy. *Íkala, Revista de Lenguaje y Cultura, 26*(1), 185–205. http://www.scielo.org.co/scielo.php?pid=S0123-34322021000100185&script=sci_arttext

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy and S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences and applications (pp. 57-71). Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9780203811122

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for EFL/ESL writing: Challenges, strategies, and ethical considerations. RELC Journal, 54(2), 537–550. https://doi.org/10.1177/00336882231162868

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. Computer-Assisted Education Open, 5,

100174. https://doi.org/10.1016/j.caeo.2024.100174

Lu, D. (J.), & Zeng, Y. (2025). Exploring the use of ChatGPT-generated model texts as a feedback instrument: EFL students' text quality and perceptions. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2525341

Luo, W.-H. (2006). Collaboration between native and non-native English-speaking teachers: How does it work? Journal of Asia TEFL, 3(3), 41–58.

Mackey, A., & Gass, S. M. (2016). Second language research: Methodology and design (2nd ed.). Routledge.

Martin, J. R., & Rose, D. (2012). Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney School. Equinox.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. Quality & Quantity, 42(6), 779–794. https://doi.org/10.1007/s11135-007-9125-z

Negretti, R., & McGrath, L. (2018). Scaffolding genre knowledge and metacognition: Insights from an L2 doctoral research writing course. Journal of Second Language Writing, 40, 12–31. https://doi.org/10.1016/j.jslw.2017.12.002

Rose, D. (2008). Writing as linguistic mastery: The development of genre-based literacy pedagogy. In D. Myhill, D. Beard, M. Nystrand, & J. Riley (Eds.), The SAGE handbook of writing development (pp. 151–166). SAGE Publications.

Rose, D., & Martin, J. R. (2012). Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney School. Equinox.

Rothery, J. (1994). Exploring literacy in school English. Metropolitan East Disadvantaged Schools Program.

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Martín-Villena, F. (2023). The LexTALE as a measure of L2 global proficiency: A cautionary tale based on a partial replication of Lemhöfer and Broersma (2012). Linguistic Approaches to Bilingualism, 13(3), 299–314. https://doi.org/10.1075/lab.22048.pui

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. Cogent Education, 11(1). https://doi.org/10.1080/2331186X.2024.2410101

Tai, Y., & Chen, Y., & Lin, W. (2025). Incorporating ChatGPT into genre-based instruction for argumentative writing among EFL college students. International Journal of Applied Linguistics, 36(1), 767–781. https://doi.org/10.1111/ijal.12777

Teng, L. S., & Huang, J. (2026). Assessing ChatGPT feedback for EFL learners' engagement and writing performance. International Journal of Applied Linguistics. Advance online publication. https://doi.org/10.1111/ijal.70155

Tocalo, A., Wattanajinda, N., & Promjun, T. (2023). Formality in the academic writing of Thai EFL English-major students. LEARN Journal: Language Education and Acquisition Research Network, 16(2), 147–175.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

You, X., & You, X. (2026). AI-mediated genre learning in the business English class. English for Specific Purposes, 82, 34-48. https://doi.org/10.1016/j.esp.2025.11.004

Zhang, T. T., & Zhang, L. J. (2021). Taking stock of a genre-based pedagogy: Sustaining the development of EFL students' knowledge of the elements in argumentation and writing improvement. Sustainability, 13(21), Article 11616. https://doi.org/10.3390/su132111616

Zou, S., Guo, K., Wang, J., & Liu, Y. (2026). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: a comparison study. Computer Assisted Language Learning, 39(4), 1062–1091. https://doi.org/10.1080/09588221.2024.2447279

Biodata

Ngo Hung Phu is an MA TESOL student at the University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City (USSH–VNUHCM). He has over six years of experience teaching English as a Foreign Language (EFL) and currently works as a Vietnamese National Teacher at ILA Vietnam. His research interests include genre-based pedagogy, register awareness in EFL writing, and the pedagogical integration of AI in language classrooms.