



ERROR PATTERNS IN THE INTERLANGUAGE OF VIETNAMESE EFL LEARNERS: A DUAL-TAXONOMY ANALYSIS OF ESSAY WRITING

Nguyen Minh Ducⁱ,
Tran Thi Huynh Nhu

Faculty of Foreign Languages,
Nguyen Tat Thanh University,
Ho Chi Minh City, Vietnam

Abstract:

Errors constitute an inevitable yet informative feature of second language acquisition, offering direct evidence of a learner's developing interlanguage. Whereas error analysis in Vietnam has predominantly examined undergraduates in degree programmes, learners at intensive foreign language centres remain under-researched, and studies providing an integrated taxonomy of their errors are scarce. The present study, therefore, investigated the categories and prevalence of linguistic errors in argumentative essays by EFL learners in IELTS preparation courses at a foreign language centre in Ho Chi Minh City. Adopting a descriptive mixed-methods design, the study analysed 63 end-of-course essays, in which errors were manually classified using a Linguistic Category Taxonomy synthesised from Politzer and Ramirez (1973), James (1998) and Langan (2001), and re-examined through the Surface Strategy Taxonomy (Dulay, Burt, & Krashen, 1982). Of the 1,685 errors, lexical errors were the most frequent (29.3%), followed by morphology (27.4%), syntax (21.5%), sentence structure and discourse (11.7%) and mechanics (10.1%); misformation (52.58%) dominated the surface strategy dimension. The findings indicate that word choice, number, subject-verb agreement and spelling remain the most persistent difficulties, arguably attributable to L1 negative transfer coupled with intralingual overgeneralisation. It is contended that intensive, exam-oriented instruction does not by itself remedy fundamental linguistic deficits; pedagogical implications for form-focused writing instruction are discussed.

Keywords: error analysis; EFL writing; interlanguage; surface strategy taxonomy; linguistic category taxonomy; Vietnamese learners

ⁱ Correspondence: email ducnm@ntt.edu.vn, reallyduc@gmail.com

1. Introduction

English has become the most influential language of global communication, functioning not only as a medium for exchanging ideas but also as a gateway to education, employment, and international integration. As of 2025, an estimated 1.5 billion people use English as a first or additional language for purposes ranging from business and education to entertainment (Tran-Thanh, 2025). In an increasingly globalized society, English has consolidated its status as a global lingua franca, and in most countries it has become a compulsory component of formal education. Vietnam is not an exception: many universities have adopted English as a medium of instruction, and English remains a required subject that most university students must complete before progressing into their major fields (Chi, 2020).

Among the four language skills, writing occupies a particularly demanding position in the process of learning English. Proficient writing enables learners to express their thoughts clearly and to organize their message cognitively and coherently (Langan, 2001). Moreover, competent writing opens professional opportunities, since staff at multinational companies are frequently required to produce documents in English as the primary language of client communication (Mohammadi & Mustafa, 2020). For these reasons, writing courses feature prominently in the academic schedules of schools and universities. Despite this emphasis, errors remain an inevitable part of the language acquisition process. Rather than being mere signs of failure, errors carry significant value as evidence of a learner's underlying interlanguage system (Corder, 1967; James, 1998).

Although writing is a critical skill, Vietnamese learners frequently struggle to master it, encountering difficulties in ordering ideas coherently and in producing accurate grammatical structures (Lan, 2022). The development of writing is inherently complex, demanding grammatical accuracy, rhetorical technique, and cognitive and evaluative abilities. A further complication is the growing reliance on artificial intelligence writing tools, which, while enhancing surface quality, can leave learners dependent and unable to write accurately without technological assistance (Lan, 2022). Even after attending foreign language centers offering intensive English courses, many learners continue to make errors, struggling with fundamental grammatical structures despite formal instruction. Such errors reduce the clarity of the message, cause misunderstanding, and prevent learners from reaching higher levels of writing proficiency.

The theoretical importance of studying errors lies in what they reveal about the process of language learning. From the perspective of error analysis, a deviation is not a random accident but a window into the rules the learner is currently operating with, and thus a diagnostic resource for both learner and teacher (Corder, 1967; Ellis, 1994). For learners, becoming aware of the typical errors in their writing enables them to recognize their weaknesses and to adopt more effective, personalized learning strategies. For teachers, errors provide distinct feedback that points to what remains to be learned (Corder, 1967), informing adjustments to lesson plans and the design of targeted writing

exercises. It is precisely this dual (diagnostic and pedagogical) value that motivates a systematic analysis of the errors produced in an intensive, high-stakes learning environment.

A considerable body of research has examined grammatical, lexical, and mechanical errors among English learners. However, most previous studies (Chi, 2020; Lan, 2022; Sermsook *et al.*, 2017) have concentrated on university students enrolled in formal academic programs, while learners in intensive, exam-oriented environments such as foreign language centers remain comparatively neglected. The few studies that have addressed language-center contexts (Giang *et al.*, 2023; Mustafa *et al.*, 2017) have tended to focus on organizational or discourse-level errors and have rarely offered a comprehensive taxonomy that captures the full range of linguistic and morphological deviations. This shortage leaves an incomplete picture of the specific error types learners produce in such settings and of how these errors can be systematically categorized.

In response to this gap, the present study investigates the common linguistic errors in the English essay writing of learners at a foreign language center in Ho Chi Minh City and seeks to diagnose the most problematic areas of their written production. The study is guided by two research questions:

- 1) What major categories of writing errors do learners at a foreign language center demonstrate in their essay writing?
- 2) Which error categories are most prevalent in the learners' essay writing?

By addressing these questions, the study contributes to error analysis research in two ways. Pedagogically, it provides teachers with evidence of the actual error categories in learners' writing, informing lesson planning, targeted feedback, and the design of more effective writing exercises. Theoretically, it enriches the literature on error analysis in under-examined intensive learning environments and offers a foundation for future research on error causation, corrective feedback, and cross-group comparison.

2. Literature Review

2.1. Errors, Mistakes, and the Shift from Contrastive to Error Analysis

An error is a linguistic form, or combination of forms, that would not be produced by native speakers under comparable conditions (Corder, 1967). Errors, which reflect gaps in the learner's underlying competence, are conventionally distinguished from performance deviations arising from factors external to that competence. Whereas errors were once treated as defects to be eradicated because they signaled the failure of linguistic performance, they are now understood as a valuable learning resource that allows learners to gauge their proficiency in the target language and to design appropriate learning strategies (Richards & Sampson, 1974). Errors likewise furnish teachers with a means of evaluating both learning and teaching processes and of identifying what still requires attention (Richards & Sampson, 1974).

The distinction between error and mistake is central to this framework. An error is a systematic deviation stemming from a lack of knowledge; in other words, it occurs

when the learner has not yet internalized the rule of the target language (Corder, 1967). A mistake, by contrast, is a non-systematic deviation caused by performance limitations such as distraction, fatigue, or memory constraints, and can be self-corrected once the learner's attention is drawn to it. Because errors, and not mistakes, provide direct evidence of the learner's developing system, error analysis focuses on the former (Corder, 1967).

Contrastive Analysis (CA) sought to predict learner errors by comparing structural differences between the native language (L1) and the target language (L2), assuming that errors resulted from L1 interference (Lado, 1957). However, CA could not account for errors arising from the intrinsic complexity of the target language itself (Brown, 2007). Error Analysis (EA) emerged as a response to these limitations. Rather than merely predicting errors, EA analyzes the actual errors learners produce and explores their sources (Corder, 1967). Central to EA is the concept of interlanguage which is the unique linguistic system that second language learners construct, distinct from both L1 and L2 (Selinker, 1972). Where CA treated errors as habits to be avoided, EA reframes them as valuable evidence of a learner's developing interlanguage, providing a more comprehensive diagnostic tool that captures errors stemming from internal learning processes and not only from L1 interference (James, 1998).

This distinction has concrete consequences for classification. A learner who consistently writes "he go" across multiple contexts is committing an error that reflects a gap in competence, whereas a learner who writes "goes" elsewhere but slips into "go" once is making a mistake. Because the present study analyzes a substantial corpus and looks for recurrent patterns rather than isolated slips, it treats systematic, repeated deviations as errors in Corder's (1967) sense. The concept of interlanguage underpins this analytical stance: the learner's output is understood as a rule-governed system in its own right, occupying an intermediate position between the L1 and the target language and evolving as learning proceeds (Selinker, 1972). Viewed this way, the errors documented in this study are not merely a catalogue of failures but a map of the interlanguage that these learners have constructed at an advanced stage of instruction.

2.2. The Procedure of Error Analysis

Error analysis focuses on expressive errors that occur in written or spoken production. Following Corder (1974), the procedure comprises five steps: collecting a sample of learner language, identifying the errors, describing them, explaining them, and evaluating them. Collecting learner language provides the raw data for investigation. Identification requires determining what qualifies as an error by comparing the learner's utterance with its reconstructed well-formed equivalent in the target language (Ellis, 1994), a step that also involves distinguishing errors from mistakes (Corder, 1967). Description classifies errors, typically using a linguistic category taxonomy addressing morphological and syntactic features (Politzer & Ramirez, 1973) or a surface strategy taxonomy describing how the target structure is altered through omission, addition, misformation, and misordering (Dulay *et al.*, 1982). Explanation investigates the source

of the error, such as interlingual or intralingual transfer (Ellis, 1994; Richards, 1971a). Evaluation, finally, considers the effect of errors on the reader in terms of comprehensibility (Ellis, 1994). The present study operationalizes the first four of these steps.

2.3. Taxonomies of Errors

To categorize the deviations found in the learners' writing, this study combines the Surface Strategy Taxonomy of Dulay *et al.* (1982) with a Linguistic Category Taxonomy synthesized from Politzer and Ramirez (1973), James (1998), and Langan (2001). While the linguistic category taxonomy diagnoses the specific linguistic locus of an error, the surface strategy taxonomy illuminates the cognitive mechanism behind it by describing how learners modify the target rule.

2.3.1. Surface Strategy Taxonomy

Dulay *et al.* (1982) identify four error types. Omission errors occur when a required item is absent from a well-formed utterance. These typically affect grammatical morphemes (noun and verb inflections, prepositions, articles, auxiliaries) rather than content morphemes. Addition errors, the converse of omission, occur when an item is redundantly present, and include double marking, regularization, and simple addition. Misformation errors occur when the learner supplies the wrong form of a morpheme or structure (the item is present but incorrect) and encompass regularization, archi-forms, and alternating forms. Misordering errors occur when morphemes or words are correct but incorrectly sequenced.

2.3.2. Linguistic Category Taxonomy

Politzer and Ramirez (1973) divide errors into morphological categories (articles, possessives, number) and syntactic categories (pronouns, prepositions, subject-verb agreement, word order, negation, adjectives, and adverbs). James (1998) supplements these with conjunctions and lexis, in which the latter comprises formal errors (word choice through misselection, misformation, or distortion) and semantic errors (confusion of sense relations and incorrect collocation). Langan (2001) adds sentence-level and mechanical categories relevant to pedagogy, including fragments, run-ons, punctuation, spelling, and capitalization. Verb-related errors are captured through Dulay *et al.*'s (1982) distinction between the auxiliary system and the main verb form. The full integrated taxonomy is summarized in Table 1.

Table 1: Integrated Linguistic Category Taxonomy Used for Coding

Domain	Sub-categories	Focus
Morphology	Articles; Possessive; Number; Main verb form	Word formation and inflection
Syntax	Auxiliary system; Subject-verb agreement; Verb complementation; Word order; Negation; Pronouns; Adjectives; Adverbs	Sentence-level grammatical relations
Sentence structure & discourse	Fragment; Run-ons; Conjunction	Clause combination and cohesion
Lexis	Word choice; Collocation; Prepositions	Selection and combination of words
Mechanics	Spelling; Punctuation; Capitalization	Orthographic conventions

The rationale for combining the two taxonomies is complementary rather than redundant. Used alone, the linguistic category taxonomy tells us where an error occurs: in the article system, in subject-verb agreement, in lexical selection, but is silent about how the learner has departed from the target form. Used alone, the surface strategy taxonomy tells us how the target form has been altered: through omission, addition, misformation, or misordering, but not which part of the grammar is affected. Analyzing the same corpus through both lenses therefore yields a two-dimensional description in which each error is simultaneously located and characterized, enabling stronger inferences about causation: a missing plural suffix, for instance, is at once a number error (linguistic category) and an omission (surface strategy), a combination that points clearly toward interlingual simplification. It is this cross-classification that the present study exploits in interpreting its results.

2.4. Sources of Errors

The explanation stage of EA investigates the reasons behind learners' deviations (Corder, 1974). Drawing on the consensus among Ellis (1994), James (1998), and Brown (2007), this study adopts Brown's (2007) framework, which distinguishes four sources. Interlingual transfer, or mother-tongue influence, involves importing elements of the L1 into the L2 (Richards, 1971a) and predominates in the early stages of learning, often through literal translation from L1 (Richards, 1971a). Intralingual transfer originates within the L2 itself and reflects developmental features of rule learning; following Richards (1971a), it subsumes overgeneralization (extending a rule beyond its scope), ignorance of rule restrictions (failing to observe structural limitations), incomplete application of rules, and false concepts hypothesized (misunderstanding target-language distinctions). Context of learning refers to induced errors arising from ineffective teaching, misleading textbook explanations, or incorrect feedback (Brown, 2007; Ellis, 1994). Communication strategies produce errors when learners use techniques such as word coinage, circumlocution, or approximation to compensate for linguistic gaps (Brown, 2007; James, 1998).

2.5. Related Studies and the Research Gap

Error analysis has been applied across a wide range of EFL contexts. In Saudi Arabia, Nuruzzaman *et al.* (2018) found that the majority of errors in students' paragraph writing resulted from intralingual interference and insufficient grammatical knowledge, particularly regarding verb tense and subject-verb agreement. In Southeast Asia, Sermsook *et al.* (2017) examined general grammatical errors among Thai students while Mustafa *et al.* (2017) focused on Indonesian junior high school students, and both concluded that L1 interference was the core cause of syntactic errors. In the Malaysian context, Che Hassan *et al.* (2019) analyzed learners' errors in a reflective writing assessment and showed that interlingual and intralingual sources operate simultaneously, while Manokaran *et al.* (2013), in a corpus study of argumentative essays, documented the systematic misuse of the past-tense auxiliary "be." Synthesizing this scattered evidence, a systematic review by Mohammadi and Mustafa (2020) concluded that ESL/EFL writing errors stem predominantly from a combination of first-language interference and developmental intralingual processes, yet observed that many studies remain narrow in scope and lack an integrated taxonomy of the errors involved.

In the Vietnamese context, Trang (2018) and Chi (2020) investigated the types and frequency of errors in paragraphs and compositions by university students, revealing persistent difficulties with sentence structure, verb tenses, and lexical choice during L1-L2 transfer. However, both studies were confined to single institutions. Man (2020) analyzed errors in research papers at Thu Dau Mot University, finding that structural and grammatical errors persisted even at an advanced academic level, though the sample was restricted to a local journal. Other studies have targeted specific dimensions of writing: Linh and Vien (2020) examined grammatical errors among high school students, while Giang *et al.* (2023) investigated organizational errors in paragraph writing at a language center in Can Tho, identifying weaknesses in coherence and cohesion. Collectively, these studies have illuminated learners' difficulties, yet significant gaps remain in terms of context, methodology, and data scale. In particular, few studies have examined writing errors in the intensive, exam-oriented environment of a foreign language center in Ho Chi Minh City, or have offered an integrated taxonomy spanning morphology, syntax, lexis, discourse, and mechanics alongside a surface-strategy interpretation. The present study addresses this gap.

3. Material and Methods

3.1. Research Design

This study adopted a descriptive mixed-methods design. Examining the linguistic nature of learner errors in authentic written texts constitutes a qualitative approach (Nassaji, 2015); accordingly, errors were identified, described, and explained on the basis of the actual content of learners' essays. At the same time, the frequency and percentage of each error type were calculated using descriptive statistics as a quantitative tool (Gravetter *et al.*, 2021), providing an overview of error distribution and reflecting the prevalence of

each category. Combining qualitative and quantitative elements is encouraged as mixed-methods research (Creswell, 2015) because it affords both interpretive depth and statistical precision. This design therefore captures both the scale and the complexity of learners' writing difficulties, enhancing the objectivity and validity of the findings.

3.2. Participants and Sampling

The data were drawn from English learners enrolled in IELTS preparation courses at a foreign language center in Ho Chi Minh City, aged between 16 and 22. Participants were predominantly high school students, together with a smaller group of university freshmen. All had completed a comprehensive training program of eight courses, and the data were collected at the final stage of this program, immediately preceding participants' actual IELTS examination. This ensured that participants had already been equipped with the lexical resources, grammatical foundations, and writing skills expected for the standardized test, allowing the study to analyze errors that persist despite intensive instruction. Because most participants shared the goal of achieving a strong IELTS band - a certificate that now carries considerable weight for high school graduation exemption and direct university admission in Vietnam (Tran-Thanh, 2025), the essays reflect learners' maximum effort rather than casual practice.

A combined purposive-convenience sampling technique was employed. Participants were selected purposively on the basis of their stage in the program and their possession of the requisite skills (Andrade, 2021), while access was facilitated conveniently through a professional referral from the researcher's supervisor, which secured authorized access to the center's archival data. As shown in Table 2, the final sample comprised 63 learners from four classes, yielding a corpus of 63 essays of at least 250 words each. Although a larger sample would increase generalizability, the study prioritized depth of analysis over breadth, consistent with Corder's (1974) observation that error analysis is a labor-intensive process requiring detailed linguistic interpretation.

Table 2: Composition of the Sample

Class	Number of learners / essays
Class 1	13
Class 2	22
Class 3	14
Class 4	14
Total	63

3.3. Instruments

The principal research instrument was the corpus of 63 learner essays, which served as the primary data source for identifying and analyzing errors. The essays were the final end-of-course assessment, composed directly on computers through Google Forms under strict invigilation in the center's computer laboratory; the formal, controlled conditions reinforce the authenticity of the samples. Classification followed the Surface Strategy Taxonomy (omission, addition, misformation, misordering) of Dulay *et al.* (1982)

and the integrated Linguistic Category Taxonomy synthesized from Politzer and Ramirez (1973), James (1998), and Langan (2001). To record and analyze the data manually, the researcher designed an Error Coding Sheet containing the fields Ordinal Number, Essay Script, Linguistic Type, and Error Category. The coded dataset was then used to compute the frequency and percentage of each error type and to select representative examples for interpretation.

3.4. Data Collection and Analysis

Access to the research site was facilitated by the researcher's instructor, who introduced the researcher to the center's Board of Directors. Following a formal proposal, the center granted permission to use archived student writing samples on the condition that learner identities remain strictly confidential. Because the data had been produced for coursework rather than for the study, this constitutes secondary research. All identifying information including names, class codes, ages, genders, and learning backgrounds was anonymized. During conversion, the integrity of the original texts, including original spelling and punctuation, was carefully preserved.

Analysis followed Corder's (1974) procedure of collection, identification, description, and explanation. Each essay was examined carefully to identify all errors according to the linguistic categories of Politzer and Ramirez (1973), after which the errors were re-analyzed through the surface strategy taxonomy of Dulay *et al.* (1982) to reveal the mechanisms of error production. This approach reflects directed content analysis, in which initial codes are derived from existing theory and prior findings (Hsieh & Shannon, 2005). Descriptive statistics were then used to calculate the frequency and percentage of each error type and to identify the most problematic areas, which were subsequently explained with reference to the frameworks of Ellis (1994) and Richards (1971a).

Two considerations governed the trustworthiness of the coding. First, where a single deviation could potentially fall into more than one category, such as a misspelled verb form that might involve both spelling and main verb form, the most salient locus of deviation was identified, and the error was coded under the category that best represented its grammatical nature. This approach was consistent with the description stage of Corder's (1974) procedure. Second, the linguistic classification was deliberately kept separate from, and prior to, the surface strategy classification: each error was first located within the linguistic category taxonomy and only then re-examined for the strategy (omission, addition, misformation, or misordering) through which the target rule had been altered. This two-pass procedure ensured that the linguistic and cognitive dimensions of each error were analyzed independently, reducing the risk of circular reasoning in which the presumed cause of an error determines its category.

3.5. Ethical Considerations

Ethical procedures were observed throughout. The use of the 63 essays received official permission from the center's management. All writing samples were fully anonymized

before analysis, and the data were collected indirectly to avoid influencing learners' performance. The data were used solely for academic research and were not shared with third parties. Consistent with Creswell (2015), ethical integrity was maintained by respecting participant confidentiality, ensuring transparency of research objectives, and focusing strictly on linguistic analysis rather than individual performance.

4. Results and Discussion

Analysis of the 63 essays yielded a total of 1,685 errors distributed across five linguistic domains. As Table 3 shows, lexis was the most problematic area, accounting for the highest frequency of errors (493; 29.3%). It was closely followed by morphology (461; 27.4%), indicating substantial difficulties in word formation and inflection, and by syntax (362; 21.5%). Sentence structure and discourse (198; 11.7%) and mechanics (171; 10.1%) recorded lower frequencies. These figures indicate that although learners experience difficulty across all linguistic aspects, lexical choice and morphological accuracy constitute the most dominant challenges.

Table 3: Distribution of Errors Across the Five Linguistic Domains

Domain	Frequency	Percentage (%)
Lexis	493	29.3
Morphology	461	27.4
Syntax	362	21.5
Sentence structure & discourse	198	11.7
Mechanics	171	10.1
Total	1,685	100

4.1. Morphological Errors

Morphological errors accounted for 461 instances. As Table 4 indicates, number errors were the most prevalent, comprising 162 instances (35.14%), which suggests that learners frequently struggle with English pluralization rules, likely influenced by the non-inflectional nature of Vietnamese. Main verb form errors ranked second with 124 instances (26.9%), reflecting difficulties in applying correct tense markers and participle forms. Article errors followed closely with 116 instances (25.16%), a consequence of the complexity of the English determiner system for L2 learners. Possessive errors were least frequent, with 59 instances (12.8%).

Table 4: Sub-categories of Morphological Errors

Sub-category	Frequency	Percentage (%)
Number	162	35.14
Main verb form	124	26.90
Articles	116	25.16
Possessive	59	12.80
Total	461	100

Representative examples illustrate these patterns. For number, learners wrote “Citizens must care about lots of thing” instead of “lots of things,” and “Explanation and examples will be given” instead of “Explanations and examples.” For main verb form, learners produced “They should moved out” (for “should move out”) and “choose to live by themselves while concentrate on their career” (for “while concentrating”). Article errors appeared in “In the recent years” (for “In recent years”) and “Living alone in a cities” (for “in cities”), while possessive errors surfaced in “In the mind of them” (for “In their minds”) and “the house’s keys” (for “the house keys”). Errors in number and possessive form are largely attributable to negative transfer. Because Vietnamese is an isolating language that lacks inflectional morphology, learners tend to omit the suffixes required for pluralization and possession, a pattern that can also be classified as ignorance of rule restrictions (Le, 2004; Richards, 1971b). Main verb form errors reflect overgeneralization, as learners extend a known rule (adding -ed) to inappropriate contexts, while article errors frequently arise from false concepts hypothesized, as learners misjudge the obligatory use of the definite article with generic time expressions.

4.2. Syntactic Errors

Syntactic errors totaled 362 instances. As Table 5 shows, subject-verb agreement was the most frequent syntactic error (114; 31.49%), revealing a fundamental difficulty in maintaining grammatical agreement. The auxiliary system ranked second (74; 20.44%), involving the unnecessary insertion of “be” or incorrect modal usage, followed by verb complementation errors (68; 18.78%), which concern the incorrect structure following a verb, such as using a to-infinitive where a bare infinitive is required. Pronoun errors accounted for 58 instances (16.02%), while adjectives (19; 5.25%), word order (17; 4.70%), negation (6; 1.66%), and adverbs (6; 1.66%) occurred less frequently.

Table 5: Sub-categories of Syntactic Errors

Sub-category	Frequency	Percentage (%)
Subject-verb agreement	114	31.49
Auxiliary system	74	20.44
Verb complementation	68	18.78
Pronouns	58	16.02
Adjectives	19	5.25
Word order	17	4.70
Negation	6	1.66
Adverbs	6	1.66
Total	362	100

Learners frequently failed to conjugate the verb according to the number and person of the subject, as in “Living alone are better than living with their family” (for “is better”) and “This trend appear to provide more demerits” (for “appears”). Such agreement errors are largely attributable to interlingual interference: because Vietnamese verbs do not inflect for person or number, learners tend to leave the verb unchanged and to omit

the third-person singular -s (Le, 2004; Richards, 1971b). By contrast, auxiliary and verb complementation errors stem chiefly from intralingual causes. In “If they live alone, people can be do anything” and “I am agree with this statement,” learners mistakenly treat “be” as a universal marker of tense or existence. In “This will make them to build a strong character” and “Living alone makes you to deal with all the issues,” learners overgeneralize the verb-plus-object-plus-to-infinitive pattern to causative verbs that require a bare infinitive (Richards, 1971b).

4.3. Lexical Errors

Lexis was the single most error-prone domain, with 493 instances. As Table 6 shows, word choice was by far the most common lexical error (235; 47.67%), indicating that learners often struggle to select vocabulary appropriate to the context, producing semantic inaccuracy or confusion. Preposition errors were the second most frequent (156; 31.64%), reflecting the persistent challenge of English prepositions, whose usage is often arbitrary and idiomatic. Collocation errors were least frequent (102; 20.69%) and involved unnatural word combinations.

Table 6: Sub-categories of Lexical Errors

Sub-category	Frequency	Percentage (%)
Word choice	235	47.67
Prepositions	156	31.64
Collocation	102	20.69
Total	493	100

Word choice errors included the confusion of homophones and near-synonyms, as in “The affect can make decisions” (for “The effect”). Preposition errors appeared in “Teenagers have to face with social problems” (for “face social problems”) and “your family will worry for you” (for “worry about you”), while collocation errors surfaced in “In another hand” (for “On the other hand”) and “save up money for some basic needs” (for “save money”). Preposition errors are frequently caused by negative transfer: in Vietnamese, the verb *đối mặt* (“face”) is almost always followed by the preposition *với* (“with”), and learners unconsciously attach this L1 structure to English transitive verbs that require a direct object (Le, 2004). Word choice and collocation errors, by contrast, tend to stem from intralingual false concepts hypothesized, as learners use synonyms interchangeably without regard to their collocational restrictions (Richards, 1971b).

4.4. Sentence Structure and Discourse Errors

Errors in sentence structure and discourse totaled 198 instances. As Table 7 shows, run-on errors were the most frequent (102; 51.5%), typically resulting from the failure to join independent clauses correctly, often as comma splices. Conjunction errors ranked second (53; 26.8%), as learners misused, omitted, or added linking words; for instance, beginning a sentence with “But” or using “Because” without a main clause. Fragment errors

accounted for 43 instances (21.7%), in which learners produced incomplete sentences lacking a main verb or subject.

Table 7: Sub-categories of Sentence Structure and Discourse Errors

Sub-category	Frequency	Percentage (%)
Run-ons	102	51.5
Conjunction	53	26.8
Fragment	43	21.7
Total	198	100

Conjunction errors are attributable to interlingual interference. Vietnamese employs paired conjunctions such as “Tuy ... nhưng” and “Bởi vì ... nên” (Nguyen, 2008), and learners transfer these structures directly into English without observing the rule that an English complex sentence takes only one subordinator (Le, 2004). Run-on and fragment errors, by contrast, reflect intralingual ignorance of rule restrictions: learners fail to observe that a comma alone cannot join two independent clauses, and mistake a dependent clause or phrase for a complete sentence on the basis of its semantic content (Richards, 1971b).

4.5. Mechanical Errors

Mechanical errors totaled 171 instances. As Table 8 shows, spelling was the most frequent (103; 60.23%): although learners possess general vocabulary knowledge, they struggle with accurate orthography, omitting letters, transposing characters (e.g., “Fristly” for “Firstly”), or confusing homophones. Punctuation errors accounted for 53 instances (30.99%), such as omitting commas in lists or inserting unnecessary commas. Capitalization errors were least frequent (15; 8.77%), occurring when learners failed to capitalize proper nouns, sentence-initial words, or the pronoun “I.”

Table 8: Sub-categories of Mechanical Errors

Sub-category	Frequency	Percentage (%)
Spelling	103	60.23
Punctuation	53	30.99
Capitalization	15	8.77
Total	171	100

Some punctuation errors reflect interference, since the comma in Vietnamese written discourse frequently links related ideas within a single sentence regardless of grammatical independence, whereas others, particularly spacing errors, stem from limited awareness of English typing conventions (Richards, 1971b). Although capitalization errors are the least severe, they nonetheless diminish the professional quality of academic writing.

To make these categories concrete, Table 9 presents representative pairs of learner errors and their reconstructed corrections drawn from across the five domains. The

examples show how a single essay may combine transfer-based deviations (such as the preposition in “face with social problems”) with developmental deviations (such as the overgeneralized “make them to build”), underscoring that the learners’ interlanguage is shaped simultaneously by cross-linguistic influence and by internal rule construction.

Table 9: Representative Examples of Errors and Their Corrections

Category	Learner error	Corrected form
Number	Citizens must care about lots of thing	Citizens must care about lots of things
Main verb form	They should moved out	They should move out
Articles	In the recent years, many people...	In recent years, many people...
Possessive	In the mind of them, you are...	In their minds, you are...
Subject-verb agreement	Living alone are better than...	Living alone is better than...
Auxiliary system	I am agree with this statement	I agree with this statement
Verb complementation	This will make them to build...	This will make them build...
Word choice	The affect can make decisions	The effect can make decisions
Prepositions	Teenagers have to face with social problems	Teenagers have to face social problems
Collocation	In another hand, living alone...	On the other hand, living alone...
Run-on	Living can cost a lot of money, things such as rent...	Living can cost a lot of money. Things such as rent...
Spelling	Fristly, living alone...	Firstly, living alone...

4.6. Surface Strategy Taxonomy

Re-analyzing the 1,685 errors through the surface strategy taxonomy reveals the mechanisms underlying the deviations. As Table 10 shows, misinformation was by far the dominant strategy (52.58%), indicating that learners have internalized certain rules but fail to apply them correctly. Word choice and main verb form were the most common linguistic categories within this strategy, reflecting intralingual overgeneralization. Omission ranked second (31.45%), concentrated in number (missing plural -s/-es) and article errors and attributable to interlingual transfer, since Vietnamese does not mark plurality or definiteness morphologically. Addition accounted for 14.66% of errors, most often through double marking or L1 interference, as in the be-insertion error “can be do,” a false concept hypothesized in which “be” is treated as a universal sentence marker. Misordering was rare (1.31%) and consisted almost exclusively of word-order errors arising from direct translation of Vietnamese, especially in complex sentences.

Table 10: Distribution of Errors by Surface Strategy

Surface strategy	Percentage (%)
Misformation	52.58
Omission	31.45
Addition	14.66
Misordering	1.31
Total	100

4.7. Discussion

The findings offer a comprehensive answer to the two research questions. In response to the first, learners at the foreign language center demonstrated errors across five major domains, namely lexis, morphology, syntax, sentence structure and discourse, and mechanics, confirming that writing difficulty is distributed across multiple linguistic levels rather than confined to grammar alone. In response to the second, lexis and morphology emerged as the most prevalent domains, with word choice, number, subject-verb agreement, and spelling constituting the most persistent individual difficulties.

The prominence of lexical errors extends the largely grammar-focused conclusions of earlier Vietnamese studies (Chi, 2020; Trang, 2018). Whereas those studies emphasized sentence structure and verb tense, the present corpus shows that vocabulary selection is the single greatest source of error, suggesting that learners' receptive knowledge of word meaning outpaces their productive command of collocation and register. This pattern is consistent with James's (1998) distinction between formal and semantic lexical errors and indicates that vocabulary instruction in intensive programs may overemphasize breadth of vocabulary at the expense of depth, particularly the collocational and contextual knowledge required for accurate production. Because lexical proficiency is closely tied to the overall quality of ESL compositions (Engber, 1995), this concentration of word-choice errors is especially consequential for the communicative success of the learners' writing. The high frequency of morphological and agreement errors corroborates a consistent finding in the EFL literature: L1 negative transfer strongly shapes the interlanguage of learners whose mother tongue is typologically distant from English. Because Vietnamese is an isolating language without inflectional morphology, learners systematically omit plural suffixes, possessive markers, and the third-person singular -s, echoing the patterns reported by Sermsook *et al.* (2017) and Nuruzzaman *et al.* (2018) for Thai and Saudi learners respectively, and consistent with recent evidence that inflectional morphemes constitute a locus of persistent difficulty for ESL writers (Niez *et al.*, 2025). That these errors persisted among learners nearing the end of an eight-course IELTS program is notable. It suggests that intensive, examination-oriented instruction does not automatically resolve deeply rooted transfer errors, and that such errors may require explicit, form-focused intervention rather than incidental exposure.

The surface strategy analysis deepens this interpretation. The dominance of misformation (52.58%) indicates that the learners' principal difficulty lies not in the absence of grammatical knowledge but in its accurate application, a pattern characteristic of intralingual overgeneralization. Learners know that a rule exists but deploy the wrong form, as when they add -ed after a modal or select a to-infinitive after a causative verb. The substantial share of omission errors (31.45%), by contrast, points to interlingual simplification, whereby learners transfer the morphologically sparse structure of Vietnamese to English and omit markers they perceive as redundant (James, 1998). The coexistence of these two mechanisms confirms that learner errors in this context are the product of both cross-linguistic influence and developmental processes internal to L2 acquisition, in line with Brown's (2007) multi-source framework.

Beyond interlingual and intralingual transfer, the remaining sources in Brown's (2007) framework also help explain particular patterns. Context of learning plausibly contributes to the be-insertion errors and the overgeneralized causative constructions, which may be reinforced by the formulaic drilling characteristic of exam-preparation instruction, where high-frequency patterns are practiced intensively and then over-applied. Communication strategies, meanwhile, are visible in the lexical domain, where learners facing a gap in productive vocabulary resort to approximation, selecting a semantically adjacent but contextually inappropriate word. The persistence of such errors even under the controlled, high-effort conditions of an end-of-course examination suggests that they are not merely artifacts of carelessness but stable features of the learners' current competence, precisely the kind of systematic deviation that error analysis is designed to detect.

The finding that structural, morphological, and lexical errors endure at an advanced stage of preparation resonates with Man's (2020) observation that such errors persist even in the academic writing of experienced Vietnamese authors. Read together, these results imply that the developmental plateau in accuracy is not confined to any single proficiency band but recurs across the Vietnamese EFL trajectory, from language-center learners to published researchers. This has a sobering implication for exam-oriented instruction: raising a learner's band score, or expanding the range of vocabulary and structures they can deploy, does not in itself guarantee a corresponding gain in grammatical and lexical accuracy. Accuracy appears to require its own dedicated, form-focused attention rather than emerging automatically from communicative fluency or test familiarity.

Taken together, these results reinforce the theoretical claim that errors are systematic evidence of an evolving interlanguage rather than random lapses (Corder, 1967; Selinker, 1972). The consistency of the error patterns, concentrated in predictable areas of typological difficulty and rule complexity, demonstrates that learners are operating according to an internally coherent, if incomplete, rule system. For pedagogy, this implies that corrective feedback is most effective when it targets these recurrent, systematic categories rather than treating each error as an isolated event.

5. Recommendations

These interpretations translate into concrete instructional priorities. Because a large share of the errors cluster in a small number of high-frequency categories, teachers can achieve disproportionate gains by concentrating remedial attention on number marking, subject-verb agreement, prepositional and collocational usage, and word choice, rather than distributing feedback evenly across all deviations. Contrastive awareness activities that make explicit the typological differences between Vietnamese and English, such as the obligatory marking of plurality and third-person agreement in English but not in Vietnamese, may help learners inhibit the transfer patterns that generate omission errors. For misinformation errors, which reflect misapplied rather than absent rules, guided

noticing and structured output practice that require learners to select the correct form in context are likely to be more effective than rule explanation alone. Finally, the concentration of lexical errors in word choice and collocation argues for teaching vocabulary in phraseological chunks and authentic collocational patterns rather than as isolated items, so that productive knowledge keeps pace with receptive breadth.

Several pedagogical implications follow. Learners would benefit from moving away from word-for-word translation, a key source of preposition and word-order errors, toward greater exposure to authentic English materials that can foster a more natural awareness of collocation and grammatical structure. Developing systematic proofreading habits, with particular attention to mechanical accuracy and agreement at the final stage of the writing process, may also help reduce avoidable errors. For teachers, the findings support greater attention to form-focused instruction targeting the most frequent and systematic error categories, particularly number, agreement, word choice, and prepositions, alongside specific, category-based feedback rather than general commentary. For curriculum designers, the persistence of these errors near the end of an intensive program suggests that vocabulary instruction should place greater emphasis on collocational depth, while grammar practice should specifically address areas of typological difficulty between Vietnamese and English.

6. Conclusion

This study investigated the major categories and prevalence of linguistic errors in the essay writing of EFL learners at a foreign language center in Ho Chi Minh City. Analysis of 63 authentic essays identified 1,685 errors, with lexis and morphology emerging as the most problematic domains. Lexical errors, driven chiefly by word choice, reflect a fundamental difficulty in selecting context-appropriate vocabulary rather than a simple lack of meaning knowledge. Morphological errors reveal how the non-inflectional nature of Vietnamese produces strong negative transfer, as learners omit plural suffixes and verb inflections. Subject-verb agreement remained the most persistent syntactic difficulty, and, within the surface strategy taxonomy, misformation was the dominant mechanism, signaling that learners have internalized rules but fail to apply them accurately owing to both interference and intralingual overgeneralization. Crucially, these errors persisted despite an intensive eight-course training program, indicating that examination pressure and rote learning do not eliminate fundamental linguistic deficits.

The study has limitations that qualify its conclusions. The corpus of 63 essays from a single center constrains generalizability, and the reliance on secondary, product-based data precludes direct insight into the cognitive processes underlying the errors. Future research should therefore expand the scope to multiple centers and larger samples, and should incorporate qualitative instruments such as interviews or stimulated recall to probe error causation more deeply. Comparative designs across proficiency levels, age groups, or institutional types would further clarify how the error patterns identified here vary across the broader landscape of Vietnamese EFL learning. Despite these limitations,

the study contributes an integrated, empirically grounded taxonomy of writing errors in an under-examined intensive-learning context and offers a foundation for more targeted, evidence-based writing instruction.

Acknowledgements

The authors gratefully acknowledge Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam, for supporting this research, and also extend our sincere gratitude to the language center in Ho Chi Minh City for their permission and support with the data collection process.

Creative Commons License Statement

This research work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0>. To view the complete legal code, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.en>. Under the terms of this license, members of the community may copy, distribute, and transmit the article, provided that proper, prominent, and unambiguous attribution is given to the authors, and the material is not used for commercial purposes or modified in any way. Reuse is only allowed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Conflict of Interest Statement

The authors declare no conflicts of interest.

About the Authors

Nguyen Minh Duc, an MA in TESOL from Edith Cowan University, Western Australia, works as a full-time lecturer at the Faculty of Foreign Languages, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam. His research interests include students' motivation, intercultural communication, language assessment and computer-assisted language teaching and learning.

Tran Thi Huynh Nhu is a recent graduate in English Language at Nguyen Tat Thanh University. She currently serves as a customer service staff member at Tam Anh Hospital. Her research interests include Technology & AI in Language Learning, Machine Translation and post-editing.

References

Andrade, C. (2021). The inconvenient truth about convenience and purposive samples. *Indian Journal of Psychological Medicine*, 43(1), 86–88. <https://doi.org/10.1177/0253717620977000>

- Brown, H. D. (2007). *Principles of language learning and teaching*. Longman. Retrieved from https://books.google.ro/books/about/Principles_of_Language_Learning_and_Tea_c.html?id=jUF0QgAACAAJ&redir_esc=y
- Che Hassan, N. S. I., Mohammad Nor, N. H., Rosly, R., & Wan Zakaria, W. N. F. (2019). ESL learners' language errors in a reflective writing assessment. *Issues in Language Studies*, 8(1), 31–43. <https://doi.org/10.33736/ils.1291.2019>
- Chi, P. K. (2020). Error analysis on English compositions and paragraphs of Vietnamese students. *TNU Journal of Science and Technology*, 225(11), 55–63. <https://doi.org/10.34238/tnu-jst.3183>
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching*, 5(4), 161–170. <https://doi.org/10.1515/iral.1967.5.1-4.161>
- Corder, S. P. (1974). *Error analysis*. In J. P. B. Allen & S. P. Corder (Eds.), *The Edinburgh course in applied linguistics* (Vol. 3, pp. 122–131). Oxford University Press. Retrieved from https://books.google.ro/books/about/The_Edinburgh_Course_in_Applied_Linguist.html?id=mmek0QEACAAJ&redir_esc=y
- Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson. Retrieved from https://books.google.com/books/about/Educational_Research.html?id=4PywcQAACAAJ
- Dulay, H., Burt, M. K., & Krashen, S. (1982). *Language two*. Oxford University Press. <https://doi.org/10.2307/327086>
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press. Retrieved from https://books.google.ro/books/about/The_Study_of_Second_Language_Acquisition.html?hl=id&id=sKR7AQAACAAJ&redir_esc=y
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139–155. [https://doi.org/10.1016/1060-3743\(95\)90004-7](https://doi.org/10.1016/1060-3743(95)90004-7)
- Giang, H. T. L., Dan, T. C., & Han, T. P. B. (2023). An investigation into organization errors in EFL learners' paragraph writing: A case in a foreign language center in Can Tho City, Vietnam. *European Journal of Applied Linguistics Studies*, 6(1), 99–125. <https://doi.org/10.46827/ejals.v6i1.420>
- Gravetter, F. J., Wallnau, L. B., Forzano, L.-A. B., & Witnauer, J. E. (2021). *Essentials of statistics for the behavioral sciences*. Cengage. Retrieved from https://books.google.ro/books/about/Essentials_of_Statistics_for_the_Behavio.html?id=hcoYNW4BujYC&redir_esc=y
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. Retrieved from <https://doi.org/10.1177/1049732305276687>

- James, C. (1998). *Errors in language learning and use: Exploring error analysis*. Longman. <https://doi.org/10.4324/9781315842912>
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press. Retrieved from https://books.google.ro/books/about/LINGUISTICS_ACROSS_CULTURES.html?id=oGKqKok3KtsC&redir_esc=y
- Lan, M. (2022). Factors affecting non-English major students' writing at a university in Vietnam. *Journal of Foreign Language Studies*, (72), 106–123. <https://doi.org/10.56844/tckhnn.72.204>
- Langan, J. (2001). *College writing skills with readings*. McGraw-Hill. Retrieved from https://books.google.ro/books/about/College_Writing_Skills_with_Readings.html?id=mAtZAAAAYAAJ&redir_esc=y
- Le, Q. T. (2004). *Nghiên cứu đối chiếu các ngôn ngữ [Contrastive study of languages]*. Vietnam National University Publishing House.
- Linh, M., & Vien, T. (2020). An investigation into grammatical errors in essays written by Vietnamese EFL high school students. *Journal of Science, Hue University of Education*, 4(56), 7–17. Retrieved from <https://www.semanticscholar.org/paper/AN-INVESTIGATION-INTO-GRAMMATICAL-ERRORS-IN-ESSAYS-Thi-Linh/2c8b8078d099625d0f15b9721626a786d5f9de91>
- Man, N. T. (2020). English language errors in academic writing: An analysis of English research papers. *Thu Dau Mot University Journal of Science*, 2(2), 238–254.
- Manokaran, J., Ramalingam, C., & Adriana, K. (2013). A corpus-based study on the use of past tense auxiliary 'be' in argumentative essays of Malaysian ESL learners. *English Language Teaching*, 6(10), 111–119. Retrieved from <https://doi.org/10.5539/elt.v6n10p111>
- Mohammadi, T., & Mustafa, H. R. (2020). Errors in English writing of ESL/EFL students: A systematic review. *Theory and Practice in Language Studies*, 10(5), 520–526. <https://doi.org/10.17507/tp1s.1005.05>
- Mustafa, F., Kirana, M., & Bahri, S. (2017). Errors in EFL writing by junior high students in Indonesia. *International Journal of Research Studies in Language Learning*, 6(1), 38–52. <https://doi.org/10.5861/ijrsl.2016.1366>
- Nassaji, H. (2015). Qualitative and descriptive research: Data type versus data analysis. *Language Teaching Research*, 19(2), 129–132. <https://doi.org/10.1177/1362168815572747>
- Nguyen, D. D. (2008). *Ngữ pháp logic trong tiếng Việt [Logical grammar in Vietnamese]*. In *Ngữ pháp tiếng Việt: Những vấn đề lý luận* (pp. 147–212). Social Sciences Publishing House.
- Niez, R. A., Cerera, J. P. N., Ardon, J. M., Gillo, A. V., Ralar, T. J. T., & Tancinco, N. P. (2025). Investigating inflectional morpheme errors in ESL learners' writing. *Theory and Practice in Language Studies*, 15(5), 1399–1409. <https://doi.org/10.17507/tp1s.1505.04>

- Nuruzzaman, M., Islam, A., & Shuchi, I. J. (2018). An analysis of errors committed by Saudi non-English major students in English paragraph writing. *Advances in Language and Literary Studies*, 9(1), 31–39. <https://doi.org/10.7575/aiac.all.v.9n.1p.31>
- Politzer, R. L., & Ramirez, A. G. (1973). An error analysis of the spoken English of Mexican-American pupils in a bilingual school and a monolingual school. *Language Learning*, 23(1), 39–61. <https://doi.org/10.1111/j.1467-1770.1973.tb00096.x>
- Richards, J. C. (1971a). *Error analysis and second language strategies*. *Language Sciences*, 17, 12–22. <https://doi.org/10.4324/9781315836003>
- Richards, J. C. (1971b). A non-contrastive approach to error analysis. *ELT Journal*, 25(3), 204–219. Retrieved from https://www.academia.edu/51916352/A_Non_Contrastive_Approach_to_Error_Analysis
- Richards, J. C., & Sampson, G. (1974). *The study of learner English*. In J. C. Richards (Ed.), *Error analysis: Perspectives on second language acquisition* (pp. 3–18). Longman. <https://doi.org/10.4324/9781315836003>
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(1–4), 209–231. Retrieved from https://www.degruyterbrill.com/document/doi/10.1515/iral.1972.10.1-4.209/html?srsId=AfmBOOpTrMgp5UAsuaAun_ro0R8k60nsqQdVSXHlgVwZU665EzBW5OSV
- Sermsook, K., Liamnimitr, J., & Pochakorn, R. (2017). An analysis of errors in written English sentences: A case study of Thai EFL students. *English Language Teaching*, 10(3), 101–110. <https://doi.org/10.5539/elt.v10n3p101>
- Tran-Thanh, V. (2025). Recent trends in English language education in Vietnam: Navigating policy, practice, and societal shifts. *English Today*, 41(2), 121–126. <https://doi.org/10.1017/S0266078424000506>
- Trang, P. T. T. (2018). Analyzing Vietnamese students' errors in writing academic essays. *Van Lang University Journal of Science*, (10). <https://doi.org/10.20431/2347-3134.1403001>