

Investigation of Verb–Noun Collocation Errors: A Corpus-Based Analysis of the Academic Writing of ESL Learners at FGEI Schools, Pakistan

Zeeshan Hadir¹, Muhammad Kamran Abbas Ismail² and Tayyaba Wakeel³

Abstract

The present study aims to explore syntactic and semantic collocation errors in the academic English writing of secondary-level English as a Second Language (ESL) learners in the Federal Government Educational Institutions (FGEI) school system in Pakistan. The analysis is built on the theoretical bases of the combinatorial taxonomy of lexical collocations by Morton Benson and the error analysis (EA) model by Pit Corder. It uses a corpus-driven methodology to diagnose the deviations in the structure of student writing and the interlingual effects on it. A specially created learner corpus of students' academic essays on the environment for classes 6-10 was constructed, and a detailed diagnostic survey was administered to 100 students and 29 English language teachers from the FGEI network. The quantitative and qualitative analyses indicated that the density of collocational errors was relatively high, including article omission, verb choice, collocation constraints, verb simplification, incomplete collocation constraints, and limited academic vocabulary. The empirical results indicate that negative transfer (mother tongue interference) from Urdu and pedagogical emphasis on the rote learning of grammatical structure and memorization of isolated lexical items have a significant influence on the learner's collocational competence. The results indicate that the current structural teaching mode for ESL academic writing should be replaced with an explicit, corpus-based approach to teaching lexical items, thereby enhancing the idiomaticity and clarity of academic writing among Pakistani ESL students.

Keywords: FGEI Schools, Verb–Noun Collocations, Corpus-Based Error Analysis, Interlanguage Theory, Mother Tongue Interference, Academic Writing

¹⁻²⁻³Department of English, University of Okara – Pakistan

Introduction

Academic English writing is the main barrier to success in higher education, career advancement, and socioeconomic advancement in Pakistan's current sociolinguistic context. Pakistani governmental, judicial, and educational institutions employ English, an official language. Pakistani schoolchildren consider English writing challenging (Akhter et al., 2018). This situation is especially true in the public school and Federal Government Educational Institutions (FGEI) system, which encompasses schools and colleges nationwide and serves a large student body. The major approach has been the traditional grammar-translation approach to English language teaching. This approach emphasizes memorizing isolated lists of vocabulary and abstract grammar rules rather than formulaic language competence. It indicates that, although they may have a good idea of the syntactic forms involved in rules, their writing is often not idiomatic, coherent, or natural. This linguistic shortcoming is readily apparent in their failure to produce the range of lexical collocations that form the building blocks of fluent, natural discourse. The ability to identify and use conventionalized, recurrent word combinations is an important part of second language acquisition (Wray, 2002).

Among the variety of collocational structures, the verb-noun combinations are the syntactically most complex and semantically dense structure in academic writing. Empirical studies in a variety of second language learning settings around the world have consistently demonstrated that even L2 learners who are quite proficient in general grammar have difficulty with these unpredictable constraints. The problem is compounded in Pakistan by the typological and structural differences between English and Urdu, the national language, making it difficult for most students to use both languages as primary medium of their daily lives and in their cognitive functions. ESL learners in Pakistan tend to translate literally or transfer negatively across languages. Negative interlingual transfer is the transfer of the semantic and syntactic properties of Urdu verbs to English nouns in the formation of verb-noun combinations. The written expressions produced by the cognitive approach are often grammatically correct but sound unnatural and non-native to competent readers.

The pedagogical environment of public and semi-public schools, including the FGEI network in Pakistan, further complicates this acquisition. The education system is highly diverse, and the main teaching method in classrooms is the high-stakes summative exam, which emphasizes memorization without prompting or lexical

development in context (Ghani & Hafeez, 2025). Classroom activities typically focus on the individual word and sentence levels, and some teachers have degrees in English literature but not in applied linguistics or second-language writing pedagogy. Students' general lack of systematic knowledge of English's formulaic nature leads to unawareness of natural word co-occurrence in academic registers. Therefore, if students are required to write original academic essays, they have to construct collocations in each word based on their first language's conceptual systems. This dependence produces systematic and non-systematic errors that persist over time and constitute a fundamental obstacle to effective written discourse in the academic and professional fields.

In response to these continuing difficulties in writing, applied linguists have begun to turn to corpus-based approaches to identify and classify learner errors in a more systematic and consistent manner. Corpus linguistics offers data-based, empirical insights to second language writers, rather than the subjective intuitions of the teacher, and offers quantitative evidence. The use of learner corpora can shed light on the point at which L2 errors occur with respect to L2 norms and the underlying cognitive/pedagogical processes that explain the errors (Mahmood et al., 2020). There has been some preliminary research on general grammatical deviations in Pakistan. However, there is a dire need for corpus-based, localized studies on the interaction between LCE and MTI, in particular and pedagogical practices in general, in specific educational settings, such as FGEI (Laufer & Waldman, 2011). The present study aims to fill this gap by conducting an empirical investigation of the verb-noun collocation errors of FGEI ESL learners at the secondary level of English, using corpus diagnostics alongside student and teacher surveys for a comprehensive analysis.

Secondary ESL students at FGEI schools in Pakistan have had official English training for years, but they still struggle to write academically with naturalness and idiomatic precision. These students lack formulaic and collocational comprehension due to the standard grammar-translation approach to English sentence structure. Academic articles often overgeneralize high-frequency verbs, employ their native verb-noun structure, and lack collocational awareness. Students frequently make mistakes by using the wrong verb with the noun in academic papers: overgeneralized high-frequency verbs and collocational ignorance. The language barrier may impede comprehension and communication, leading to academic difficulties. However, their essays lack development and do not employ scholarly or

idiomatic language. The primary concern of educators in Pakistan is not lexical collocation but rather grammatical difficulties that may become rigid. This study is a corpus-based analysis of errors made by secondary school ESL students in their academic writing, specifically focusing on verb-noun collocation errors in FGEI schools, the pedagogical perceptions and practices of their teachers, and the application of the findings to guide instructional reform.

The present study addressed an overlooked topic in L2 writing education in Pakistani public schools. This study examines the complicated issue of lexical accuracy and semantic idiomaticity in the verb-noun collocation system, which is crucial to academic speech. The empirical findings from the FGEI learner corpus will help Pakistani experts and practitioners better understand the issues faced by secondary-level language learners. Textbook and curriculum writers will benefit from the findings. To localize content-specific resources and address Urdu-English lexical inconsistencies, school planners can use this native error profile and the imported wide ESL error profiles.

The study will help FGEI teachers train and improve. The Wah Region Teachers' Training Center and other FGEI regional training institutions offer contemporary facilities, language labs, and computer networks. These courses could help English teachers shift from structural, isolated words to lexical, formulaic language. Teachers can use errors to help students create collocations by providing systematic hints. In the short run, it will help pupils improve their academic writing and bridge school-work vocabulary.

❖ Limitations and Delimitations

The study only included Pakistani FGEI ESL 6–10 students. Text and academic verb-noun collocations form academic articles. Environmental concerns. Popular environmental issues included decreasing pollution, planting trees, dumping waste, and conserving water. Language analysis does not contain grammatical collocations (verb-preposition or noun-preposition) or other word combinations.

Limited student-written corpora limit the study. This regional FGEI network sample cannot represent the dialectal and regional variance in ESL writing throughout Pakistan's provinces. Students and teachers self-report in surveys. These questionnaires may misrepresent classroom attitudes and pedagogical concerns and

fail to capture immediate classroom interactions or cognitive processes during writing. Finally, the study only considered Urdu mother-tongue interference. Pakistan's national language is Urdu, and many pupils speak Punjabi, Pashto, Sindhi, Saraiki, etc., which may affect their English learning without their knowledge.

Review of the Literature

❖ Collocational Competence and Formulaic Language in L2 Acquisition

In the last four decades, prefabricated chunks of language have also been considered keys to native-speaker language production and the generation of new sentences in rule-based approaches in applied linguistics and second language acquisition studies (Wray, 2002). This view of language, usually illustrated at the lexical level, assumes that language consists of grammaticalized lexis rather than lexicalized grammar (Lewis, 1993). The most common of these formulaic sequences is collocation, which is the statistically significant co-occurrence of two or more words in a particular structural pattern (Howarth, 1998). Empirical corpus research on native speaker texts (Cowie, 1992) shows that collocations constitute almost half of the written and spoken discourse in the native speaker's corpus. Therefore, the knowledge of these word combinations is crucial to L2 learners; it is considered a major indicator of their high-level language competence, cognitive processing capacity, and academic fluency (Nesselhauf, 2005).

Although collocations are important in language production, L2 learners tend to approximate native language collocational accuracy even at high levels of general English proficiency (GEP) (Laufer & Waldman, 2011). Research indicates that grammatical competence is acquired naturally with sufficient exposure to the language, whereas collocational competence takes much longer to develop (Bahns & Eldaw, 1993). L2 learners' collocational competence is lower than that of native speakers, resulting in significantly fewer collocations than native speakers, either through underproduction or repeated use of a small number of familiar collocations (Akhter et al., 2018). When they try to use collocations, they often make mistakes. The most difficult collocations for them are the combination of verbs and nouns (Laufer & Waldman, 2011). The difficulty is attributed to a number of cognitive and psycholinguistic reasons, such as the semantic transparency of collocations, which often makes it difficult for learners to identify collocations in academic texts (Peters,

2009); and the arbitrary boundaries for the co-occurrence of certain verbs and nouns, which makes it difficult to directly apply the rules (Hill, 2000).

Moreover, several empirical studies in cognitive linguistics show that the retention/recall of these multi-word items are strongly correlated with exposure frequency in natural contexts. Students repeatedly seeing the traditional connections between words build deeper connections in their minds, allowing them to retrieve words quickly and automatically during the actual academic writing task. On the other hand, in the absence of variety of input, these formulaic units will not be able to move from receptive awareness to active, productive competency. Thus, second-language pedagogy should emphasize the development of strong mental representations of word collocations rather than “word-by-word” learning.

❖ **Lexical Collocations and the Benson Combinatory Taxonomy**

These collocational problems are systematically analyzed with the help of the structural taxonomy developed by Morton Benson, Evelyn Benson, and Robert Ilson in the BBI Combinatory Dictionary of English (Ilson et al., 1997). Benson classifies word combinations into two main types: grammatical collocations and lexical collocations (Benson et al., 2010). Grammatical collocations are a content item and a function word (normally a preposition or grammatical construction). However, lexical collocations are open-class content words used in parallel grammatical configurations and are not dominated by function words (Ghani & Hafeez, 2025). Based on Benson’s taxonomy, lexical collocations can be divided into several structural patterns. The verb + noun pattern is among the most structurally complex and semantically diverse in academic discourse (Howarth, 1998).

One of the main problems in identifying the verb + noun lexical collocation pattern is that it is arbitrary rather than idiomatic. Lexical collocations stand somewhere between the two extremes: on the one hand, they differ from free combinations, whose semantic formal properties are regulated only by general semantic rules (buy a book and read an essay), and on the other hand, they differ from idioms, which are completely incommutable, having limited and unpredictable commutability (Laufer & Waldman, 2011). The verbs “make,” “do,” and “commit” are semantically similar. In some cases, a writer can make a mistake or commit a crime, but making a crime or committing a mistake is not allowed. This opacity of meaning means that L2 learners cannot use general syntactic procedures or the meaning of individual

words to produce natural English collocations but need to learn and store these collocations as a whole unit of meaning in their mental lexicon (Nesselhauf, 2005). This issue is further compounded by the fact that synonymous items can have different collocational ranges, and learners also tend to assume that synonymous items are fully commutable (Biskup, 1992).

Moreover, Benson's combinatory taxonomy is arbitrary. It is intriguing to note that this evidence indicates the need to rely on more than structure alone to determine the acceptability of collocations. The combinations may not be semantically logical but are based on language conventions, so the L2 learner cannot use bilingual dictionaries to correctly guess the natural co-occurrence of academic nouns with specific verbs. This uncertainty forces learners to be very sensitive to variations in the register of academic English. Combinations can be recorded, and students' error boundaries can be determined so that teaching materials can be created to work with the lexical restrictions that appear in an unpredictable way in the classroom.

❖ **Interlanguage Theory, Error Analysis, and Mother Tongue Interference**

The errors of language learners are analyzed in light of their interlanguage, based on the two theories first developed by S. Pit Corder (1967) and later by Larry Selinker (1972). Errors in this approach are not simply a breakdown in the language or a direct reflection of the learner's first language, but rather systematic evidence of an active, rule-governed cognitive system, interlanguage, located between the learner's first and second languages. Corder (1981) divides his model of errors into two types: unsystematic errors, which are temporary and occur due to fatigue or lack of concentration, and systematic errors, which reflect the learner's underlying grammatical and lexical competence. Analysis of systematic errors allows the researcher to reconstruct the learner's transitional competence and to identify the strategies underlying the learner's language production.

One of the most important sources of systematic errors in L2 writing is mother tongue interference or negative interlingual transfer (Dulay et al. 1982). Students may fall back on L1 patterns and translate literally from L1 to L2 if they are not familiar with the L2 grammar and vocabulary. The first language-second language (L1-L2) congruence is an important cognitive and psycholinguistic determinant of the rate of collocational acquisition and the extent of mother tongue interference

(MTI) (Yamashita & Jiang, 2010). From the congruence approach, collocations are classified into congruent combinations, which are translated directly with the same semantic boundaries between L1 and L2 (Yamashita & Jiang, 2010), and into incongruent combinations, which exhibit structural or semantic mismatches between L1 and L2 (Yamashita & Jiang, 2010). The identification of incongruent colligations involves a much heavier learning burden for L2 writers, who need to suppress their assimilated L1 conceptual systems and, at the same time, develop non-equivalent lexical associations, which often lead to interlingual Errors (Yamashita & Jiang, 2010).

Due to the negative transfer, Pakistan clearly observes such types of differences between Urdu and English (Ambreen & To, 2021). Urdu's subject-object-verb word order is very flexible, and compound verbs are used (a noun/adjective and light verb such as "karna" [to do] or "dena" [to give]). On the other hand, English is a subject-verb-object language with tight collocational pairs (Mahmood et al., 2020). Pakistani ESL learners translate Urdu compound verbs into English in a way similar to how they are used in Urdu, that is, as a verb + noun. This results in the wrong, incongruent form of the combination. This negative transfer and the school curriculum, with its emphasis on rote memorization, are constant problems for Pakistani ESL learners, making collocational acquisition difficult for them.

Consequently, it can be inferred that interlingual grammatical errors persist despite traditional English grammar instruction among Pakistani learners, suggesting that L1 errors do not adequately resolve through L2 grammar instruction in a regular classroom environment. The ambiguous character of these syntactic and semantic discrepancies compels students to use a literal translation approach, hindering the development of native-fluent academic communication. Translation errors can become entrenched in the learner's interlanguage system over time and are difficult to rectify with conventional error-correction procedures. This raises the question of the impact of pedagogically active approaches to directly addressing first language interference in the secondary classroom.

Methodology

The research design adopted in this study is a corpus-driven mixed-methods approach systematically employed to identify, categorize, and analyze errors in verb–noun collocation in the academic writing of secondary-level ESL learners at

FGEI schools in Pakistan. The empirical data for the present study comes from two main sources: a special learner corpus of student-written academic texts and diagnostic survey questionnaires for students and English teachers of the FGEI network.

The sample of students participating in the study comprised 100 active respondents from grades 6–10, with 53 percent in grades 9–10 and 45 percent in grades 6–8, thus creating a representative profile for the development of secondary education. The teacher sample consisted of 29 valid teacher profiles, a well-experienced and qualified group of teachers. Forty-eight-point three percent of the participating teachers have more than 15 years of experience teaching the English language in the FGEI system, and 69 percent of teachers hold a master's degree, 27 percent an MPhil, and three-point four percent a PhD.

The student participants' L2 production resulted in the creation of a specialized learner corpus, the FGEI Learner Corpus (FLC), consisting of exactly 25,432 tokens of original in-class academic essays written in response to given texts on environmental issues such as pollution, waste management, and ecosystem conservation. Environmental themes were chosen because they naturally lend themselves to academic verb–noun combinations, such as reducing pollution, planting trees, dumping waste and conserving water. Extraction of verb–noun phrases from the corpus was carried out through manual analysis and computational concordance. No multi-word unit could be ignored due to structural variations or syntactical separation (Futagi et al., 2008). For empirical accuracy, cross-referencing was conducted with native-speaker reference corpora, including the BBI Combinatory Dictionary of English, the Corpus of Contemporary American English, and the British National Corpus. Any errors that did not conform to native-speaker collocational patterns were flagged and coded by the research team using a diagnostic coding system that distinguished systematic collocational errors from basic grammatical or mechanical errors (Umair et al., 2023).

Precise validation measures were computed for testing the accuracy and consistency of the collocation extraction and error-coding process (Futagi et al., 2008). We checked the reliability of the extraction process using the precision and recall equations on a pilot sample with two independent coders (Futagi et al., 2008). These metrics are defined mathematically as:

$$\text{Precision} = \frac{|\text{Collocations}_{\text{extracted}} \cap \text{Collocations}_{\text{true}}|}{|\text{Collocations}_{\text{extracted}}|}$$

$$\text{Recall} = \frac{|\text{Collocations}_{\text{extracted}} \cap \text{Collocations}_{\text{true}}|}{|\text{Collocations}_{\text{true}}|}$$

Applying these equations yielded extraction precision and recall rates of 0.98 and 0.97, respectively, verifying that 98% of the extracted word combinations represented true collocations and that only 3% of the target collocations in the sample were missed (Futagi et al., 2008). Furthermore, to ensure that the subsequent error coding and typological categorization were consistent, an inter-coder reliability test was performed on a random 20% sample of the essays using Cohen's Kappa (κ) (Futagi et al., 2008). The statistical analysis yielded a Kappa coefficient of $\kappa = 0.89$ (standard error $SE = 0.017$, $p = 0.000 < 0.05$, indicating an almost perfect level of agreement between coders and confirming the objective reliability of the diagnostic taxonomy.

Parallel diagnostic surveys were designed to collect quantitative and qualitative information on perceptions of the classroom, collocational issues, and teaching methods. Student questionnaires highlighted writing difficulties, direct translation from Urdu to English, and usefulness of the activities in the classroom. The teacher survey measured teachers' knowledge of collocational errors, awareness of mother tongue interference, vocabulary problems and current strategies for providing written corrective feedback. Descriptive statistics were used to analyze survey responses. The frequency distributions and percentages for each Likert-scaled item were calculated to demonstrate a pattern of agreement or disagreement between the two groups. The dual methodology (direct corpus and questionnaire data) allows for comprehensive triangulation of the results, enabling comparison of the students' errors with actual classroom teaching practices and with the classroom perception of these practices.

Theoretical Framework

This study is analytically designed using a dual theoretical approach, namely Pit Corder's Error Analysis model and Morton Benson's combinatory taxonomy of lexical collocations (Benson et al., 1997; Corder, 1981). This study is based on Corder's

model, which offers the main diagnostic perspective and a systematic methodology for the identification, description, explanation and evaluation of linguistic deviations in L2 writing, with a learner-centered approach. This framework makes it possible to consider student writing as a rule-governed cognitive system by separating out its underlying interlanguage competence (which are called systematic errors) from the momentary slips in performance (which are called non-systematic errors). In Corder's framework, collocational errors are considered important evidence of the active hypothesis-testing strategy employed by Pakistani ESL learners in the construction of academic English.

The study classifies and analyzes these collocational errors with reference to Benson's BBI taxonomy, which is a structural taxonomy of collocations. According to Benson, lexical collocations are fixed, identifiable phrases of content words in parallel syntactic structures, but non-idiomatic, with the Verb + Noun pattern as the main focus of this investigation. The study combines Corder's Error Analysis with Benson's taxonomy to develop a diagnostic matrix that helps analyze how the interlanguage systems of FGEI ESL learners are affected by Urdu-English lexical differences, overgeneralization of common verbs and vocabulary simplification.

Data Analysis

The collocational problem in Pakistani ESL classrooms is examined in detail using empirical data from the FGEI student and teacher questionnaires and from the corpus of verb-noun collocations. A complete statistical profile was obtained by counting and analyzing the student survey responses (N = 100; six profiles were completely blanked out) and the teacher survey responses (N = 29; one profile was completely blanked out). Table 1 below summarizes the demographic features of student and teacher participants to contextualize the bounds of the research sample.

Table 1: Demographic Characteristics of Student and Teacher Samples

Demographic Variable	Student Sample (N = 100)	Teacher Sample (N = 29)
FGEI Affiliation (Yes / No)	99% / 1%	93.1% / 6.9%
Class / Grade Level		
Classes 6 to 8	45.0%	Not Applicable

Classes 9 to 10	53.0%	Not Applicable
Unspecified Grade	2.0%	Not Applicable
Academic Qualification		
Master's Degree	Not Applicable	69.0%
MPhil	Not Applicable	27.6%
PhD	Not Applicable	3.4%
Teaching Experience		
Less than 5 years	Not Applicable	10.3%
5 to 10 years	Not Applicable	20.7%
11 to 15 years	Not Applicable	20.7%
More than 15 years	Not Applicable	48.3%

A thematic item analyzed student questionnaire data to explore students' perceptions and self-reported linguistic behaviors. The detailed percentage distribution of students' responses on a 5-point Likert scale is presented in Table 2.

Table 2: Student Questionnaire Response Percentages on Likert Scale

Student Questionnaire Survey Item	Strongly Disagree (%)	Disagree (%)	Neutral (%)	Agree (%)	Strongly Agree (%)
Difficulty choosing correct verb for a noun	6.3	14.6	28.1	41.7	9.4
Translate expressions directly from Urdu to English	11.3	12.4	24.7	45.4	11.3
First language (Urdu)	13.4	13.4	16.5	44.3	19.6

influences English writing					
Often use common verbs like <i>make, do, take, give</i>	6.0	14.6	9.4	49.0	20.8
Write <i>do research</i> or <i>do mistake</i> instead of correct form	6.3	12.6	17.9	41.1	22.1
Translate <i>give exam</i> instead of <i>take an exam</i>	9.4	16.7	7.3	49.0	17.7
Choose wrong noun (e.g., <i>make homework</i> vs. <i>do homework</i>)	13.5	22.9	19.8	29.2	14.6
Classroom activities help improve collocation use	10.4	8.3	50.0	22.9	8.3
Learning collocations helps avoid L1-caused errors	6.3	8.3	15.6	51.0	18.8

The student survey results indicate that a considerable number of the FGEI learners are aware of their problems with collocations and that their L1 has an impact on their academic writing in English. More specifically, 51.1% agreed or strongly agreed with ‘difficulty in selecting appropriate verbs for specific nouns,’ and 56.7% agreed or strongly agreed that ‘during writing tasks they translate Urdu sentences into English.’ Furthermore, 69.8% of students recognize that they overuse common verbs such as *make, do, take, and give*. This overuse is supported by the following percentages of agreement for the literal transfer of light verbs (e.g., “do research” and “make a mistake”) at 63.2% and for the light verb “give” instead of “take an exam” at 66.7%. 69.8% of students feel that explicit collocational teaching helps them avoid L1-induced errors, and 50.0% are undecided about the effectiveness of the activities they are currently having in class, indicating a pedagogical gap.

Beyond these student-reported findings, the teacher survey data (N = 29) offer a reflective perspective on the pedagogical challenges surrounding these issues from an experienced perspective. Table 3 shows the degree of consensus among FGEI English teachers regarding the source and effects of their students' collocational errors in academic writing.

Table 3: Teacher Questionnaire Response Percentages on Likert Scale

Teacher Questionnaire Survey Item	Strongly Disagree (%)	Disagree (%)	Neutral (%)	Agree (%)	Strongly Agree (%)
ESL learners often make VN errors in academic writing	0.0	3.4	13.8	51.7	31.0
Students have difficulty selecting appropriate verbs	0.0	6.9	10.3	13.8	69.0
Students rely on direct translation from Urdu (L1)	3.4	6.9	13.8	24.1	51.7
Mother tongue interference contributes significantly to errors	0.0	10.3	10.3	6.9	72.4
Limited English vocabulary knowledge leads to errors	0.0	3.4	10.3	17.2	69.0
Insufficient exposure to authentic English texts causes errors	0.0	3.4	20.7	55.2	20.7
Students tend to overuse verbs	0.0	3.4	10.3	65.5	20.7

like <i>make, do, take, give</i>					
Students lack awareness of acceptable English VN collocations	0.0	3.4	13.8	69.0	13.8
Explicit instruction of collocations can reduce errors	0.0	7.1	14.3	57.1	21.4
Classroom activities should include focused VN practice	0.0	3.4	13.8	51.7	31.0
Teachers should provide corrective feedback on VN errors	0.0	0.0	6.9	51.7	41.4
Students need more practice using VN collocations in tasks	0.0	3.4	10.3	44.8	41.4
Regular exposure to academic English improves competence	0.0	3.4	10.3	65.5	20.7
Improving VN knowledge enhances overall writing quality	0.0	0.0	6.9	48.3	44.8

The teacher survey data also indicate that the issues FGEI teachers recognize as problematic in student writing are indeed significant and regarded as critical aspects of effective writing. 82.8% of the teachers agree or strongly agree that students have difficulty in choosing suitable verbs, and 75.8% agree that students do heavy direct translation from Urdu. Further, 79.3% of teachers claim that mother tongue

interference is a major factor in causing collocational errors, while 86.2% believe that a lack of English vocabulary is a major cause. There are obstacles, but there is also consensus on pedagogical interventions: 78.5% agree that explicit collocational training can help students avoid errors in their writing, 93.1% believe that written corrective feedback can be used to correct writing errors for collocational errors, and 93.1% agree that the enhancement of collocational competence would greatly contribute to the overall improvement of students' academic writing.

The physical manifestation of the collocational errors was studied by analyzing the specialized environment learner corpus of student essays from the FGEI. Table 4 presents the 20 most frequent two-word verb–noun combinations extracted from the corpus, highlighting their raw collocational frequency and syntactic structures.

Table 4: Frequency and Syntactic Structure of Top 20 Verb–Noun Combinations

Rank	Verb–Noun Collocation	Frequency (Approx.)	Syntactic Pattern
1	reduce pollution	20+	Verb + Noun
2	plant trees	15+	Verb + Noun
3	use plastic	12+	Verb + Noun
4	throw garbage	10+	Verb + Noun
5	dump waste	8+	Verb + Noun
6	cause diseases	8+	Verb + Noun
7	create awareness	6+	Verb + Noun
8	protect environment	6+	Verb + Noun
9	recycle waste	5+	Verb + Noun
10	use transport	5+	Verb + Noun
11	release gases	5+	Verb + Noun
12	burn fuels	4+	Verb + Noun
13	treat waste	4+	Verb + Noun

14	absorb carbon dioxide	4+	Verb + Noun
15	emit smoke	4+	Verb + Noun
16	damage ecosystem	4+	Verb + Noun
17	destroy habitats	3+	Verb + Noun
18	spread diseases	3+	Verb + Noun
19	save energy	3+	Verb + Noun
20	reduce waste	3+	Verb + Noun

The corpus data show very consistent use of a small number of environmental phrases, including common, straightforward, high-frequency collocations such as “reduce pollution,” “plant trees,” and “use plastic.” Some of these combinations are idiomatic; others are examples of common systematic deviations from other standard academic reference corpora. Table 5 provides a typological classification of these collocational errors, detailing the learner form, the preferred academic form, and the underlying linguistic cause of each error category.

Table 5: Typological Classification and Diagnostic Analysis of Verb–Noun Collocation Errors

Error Category	Incorrect / Learner Form	Preferred Form	Diagnostic / Linguistic Explanation
Article Omission	<i>protect environment</i>	<i>protect the environment</i>	Absolute omission of the definite article “the” due to negative transfer from Urdu, which lacks a definite article system.
	<i>save environment</i>	<i>save the environment</i>	
	<i>clean environment</i>	<i>clean the environment</i>	
Wrong Verb Choice	<i>make environment dirty</i>	<i>pollute the environment</i>	Semantic overgeneralization and literal translation of Urdu light verbs (<i>karna</i> mapped to <i>do/make</i> and <i>dena</i> to <i>give</i>).
	<i>make water dirty</i>	<i>contaminate water</i>	

	<i>make bad impact</i>	<i>have a negative impact</i>	
	<i>do pollution</i>	<i>cause pollution</i>	
Restricted Collocation	<i>throw garbage</i> <i>throw waste</i> <i>throw plastic</i>	<i>dispose of garbage</i> <i>dispose of waste</i> <i>dispose of plastic waste</i>	Relying on “lexical approximation” by substituting a high-frequency, generic verb (<i>throw</i>) for a register-specific academic verb.
Simplified Verb	<i>release smoke</i> <i>use plastic bags</i>	<i>emit smoke</i> <i>avoid plastic bags</i>	Underproduction of specialized academic verbs, resulting in simplified, less formal word combinations.
Incomplete Collocation	<i>cut trees</i> <i>use transport</i> <i>reduce pollution</i>	<i>cut down trees</i> <i>use public transport</i> <i>reduce environmental pollution</i>	Omission of essential adverbial particles, adjectives, or structural elements required for complete collocational meaning.
Limited Academic Vocabulary	<i>cause diseases</i> <i>damage ecosystem</i> <i>save energy</i> <i>control pollution</i>	<i>trigger diseases</i> <i>damage the ecosystem</i> <i>conserve energy</i> <i>mitigate pollution</i>	Failure to transition from general, conversational vocabulary to formal, academic lexical registers.

Results

The results of the data analysis demonstrate that the FGEI environmental essay had a high frequency of verb–noun collocation errors, suggesting the linguistically difficult position of Pakistani ESL learners. Therefore, the results are presented in

accordance with the three research questions posed in this study to systematically explore these findings.

In the learner corpus, deviations were found at a high level across the six most significant categories: article omission, incorrect verb choice, restricted collocational constraints, verb simplification, incomplete collocational patterns, and limited academic vocabulary. According to the analysis, the most frequent type of error in students' written texts is article omission. Students used words such as "protect environment," "save environment," and "clean environment" without the definite article. While students can select appropriate verbs and nouns, they struggle to identify the limits of noun phrases, which are necessary for full collocational accuracy in English.

Secondly, an incorrect verb choice occurs when students use very common, general verbs with academic nouns in an unidiomatic way, such as "make the environment dirty" or "do pollution." This error is closely linked to those of restricted collocations and simplified verbs, such as substituting a more general verb (throw or release) for a specific verb (disposed of or emitted).

To quantify the density of collocational errors relative to general writing, the overall error density (ED) within the corpus was calculated using the following percentage equation:

$$ED = \left(\frac{E_{VN}}{C_{VN}} \right) \times 100$$

Where E_{VN} represents the raw count of erroneous verb–noun collocations detected, and C_{VN} represents the total number of verb–noun collocations produced in the corpus (Laufer & Waldman, 2011). The corpus analysis found that the error density of the FGEI learners was about 45%, and that the verb–noun combinations produced by the FGEI learners showed collocational deviation in at least one of the combinations (Primary Data, 2026). This high error density reveals a continuing problem of the gap between basic grammatical knowledge and natural and idiomatic language use (Akhter et al., 2018).

The analysis of the survey and the corpus reveals that negative interlingual transfer from Urdu is a major source of errors at the collocational level. According to the student survey, 56.7% of students reported translating the expression from Urdu to

English. The student corpus reveals highly systematic interlingual error patterns, which students can self-report (Mahmood et al., 2020). The most common way this negative transfer occurs is through the literal mapping of Urdu compound verbs. In Urdu, compound verbs are created by using a noun or adjective with a light verb, usually *karna* (to do/make) or *dena* (to give). Students get *karna* to correspond directly with “do” and “make” in English and so form unidiomatic collocations such as “do pollution” (from *aloodgi karna*) or “do mistake” (from *ghalti karna*).

Likewise, the Urdu compound verb *imtihaan dena* (literally meaning to give an exam) is translated literally as “give exam,” while in standard English one needs to use “take an exam.” (Anwar, 2012) In contrast, English is used to administer or deliver the exam, and the student takes or writes it. The reverse direction of the mapping is semantically out of normal bounds of English collocations, creating highly unnatural-sounding collocations for native English readers.

Moreover, the absence of an article system in Urdu explains the high number of article omissions in the student corpus, such as “protect environment and save environment.” In the Urdu language, there are no definite and indefinite articles, so learners always use English noun phrases as bare nouns and do not add the definite article “the” to environmental terms.

To evaluate whether this interlingual influence was structurally moderated by development, a Chi-Square (χ^2) test of independence was performed to compare the distribution of congruent versus incongruent collocational errors across grade levels. The active student sample was divided into lower secondary (classes six through eight, $N = 45$) and upper secondary (classes nine and ten, $N = 53$). Across exactly 300 systematic verb–noun errors analyzed, lower secondary students produced 180 errors (40 congruent, 140 incongruent), while upper secondary students produced 120 errors (10 congruent, 110 incongruent). Expected frequencies (E) were calculated under the null hypothesis (H_0) that

$$\text{Expected}_{(\text{Classes 6-8, Congruent})} = \frac{180 \times 50}{300} = 30$$

$$\text{Expected}_{(\text{Classes 6-8, Incongruent})} = \frac{180 \times 250}{300} = 150$$

$$\text{Expected}_{(\text{Classes 9-10, Congruent})} = \frac{120 \times 50}{300} = 20$$

$$\text{Expected}_{(\text{Classes 9-10, Incongruent})} = \frac{120 \times 250}{300} = 100$$

Applying the Chi-Square formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

This calculation yielded a value of $\chi^2 = 10.00$, with degrees of freedom $df = 1$. Since the obtained value exceeds the critical value of 3.841 ($\alpha = 0.05$), the difference is statistically highly significant ($\chi^2 = 10.00, p < 0.005$). This study rejects the null hypothesis, indicating that Urdu and English have distinct structural patterns that affect upper secondary learners' accuracy in reducing congruent collocation errors, which is influenced by vocabulary growth (Primary Data, 2026).

Survey results show a disparity between student needs and teachers' classroom practices. Students (82.8%) and teachers (51.1%) struggle with verb-noun combinations in writing, yet using the right verb for a noun in class is ineffective.

Half the students are unconcerned about the value of class collocational activities. 49% of students claim they overuse "make," "do," "take," and "give" in writing. The majority of teachers (86.2%) attribute collocational errors to 'low vocabulary understanding.'

Although both mother tongue interference (79.3%) and vocabulary limitations (86.2%) are considered major problems, writing instruction in FGEI schools focuses on rote learning and sentence-level mechanics rather than vocabulary development in context (Nesselhauf, 2005). The teachers' responses indicate a high level of agreement with the use of collocational corrective feedback (93.1%) and explicit classroom instruction (78.5%). However, there is little reported evidence of systematic collocational training, and the students' writing approach is largely L1-induced or literal translation.

Discussion

The results of this corpus-based exploratory study reveal that secondary-level ESL learners in FGEI schools in Pakistan face significant challenges in producing a native-like, idiomatic combination of verb and noun in school English writing. This is not a random occurrence of slips but a reflection of learners' interlingual development, which has been influenced by negative interlingual transfer from Urdu and the existing teaching-learning traditions (Ghani & Hafeez, 2025). These results are then analyzed by applying Corder's Error Analysis and Benson's combinatory taxonomy to consider the cognitive, linguistic and instructional factors that cause these errors. Noun phrases with Article Omission errors are very common (e.g., 'protect environment', 'save environment'), which shows that there is a very basic typological mismatch between Urdu noun phrases and English noun phrases. Generally, in academic English, singular count nouns require some form of determiner, typically the definite article "the," when referring to a specific ecological structure.

The significant Chi-Square result ($\chi^2 = 10.00, df = 1, p < 0.005$) and the persistent difficulty with incongruent collocations support the contention that collocational acquisition is highly dependent on typological distance and congruence. In the case of upper secondary students (classes 9-10) it is found that the students made progress in the use of congruent collocations; however, they were very prone to make incongruent errors, which indicates that by the time students reach the upper secondary level, the deep mental lexicon is still dominated by Urdu translation models. In the same way, the correspondence between the errors of Wrong Verb Choice (WVC), such as 'do pollution' and 'make environment dirty', and Urdu lexical acquisition is demonstrated. The construction of compound verbs in Urdu is through the combination of a noun with a general auxiliary verb such as 'karna' (to do). 'Karna' is a generic action for which English learners use inappropriately the verbs do and make instead of more precise academic verbs. This overgeneralization is a thinking pattern of lexical simplification/approximation which occurs when a learner employs very common, semantically imprecise verbs to avoid the difficulty of using the limited academic vocabulary he or she has acquired (Biskup, 1992).

There were errors of Restricted Collocation (throw garbage instead of dispose of garbage) and errors of Simplified Verb (release smoke instead of emit smoke), which indicate a pedagogical problem in the Pakistani education system. Rote learning

using textbooks is the main approach in traditional English instruction in public schools, even in the FGEI system, which mainly emphasizes word definitions and neglects formulaic chunks in the teaching and learning process. Failure to acquire vocabulary in isolation means that learners can't learn to be aware of the restrictions on the use of the vocabulary in a particular register. They believe that the terms 'throw' and 'dispose' can be used interchangeably in academic writing because they have general meanings. This lack of register awareness leads to a style of writing that is, in places, grammatically correct but extremely unidiomatic and informal for academic readers. On the cognitive level, non-congruent collocations (those without direct equivalents in Urdu, e.g., take an exam vs. imtihan dena) impose a high cognitive load on second-language writers.

Research on online collocational processing has revealed that, for L2 learners, processing congruent collocations was faster and yielded fewer errors than processing non-congruent collocations, because the latter induce first-language thinking and demand active inhibition of L1 translation paths. This cognitive load is likely to cause high-stakes, timed writing situations (e.g. board exams) to result in students' direct use of the L1, thus creating systematic interlingual errors. The cognitive burden can be reduced if classroom lessons move away from structural drilling and focus on the explicit teaching of collocations as whole units.

Recommendations

These systematic verb–noun collocation errors need to be overcome, and to improve the overall quality of academic writing of the students under the FGEI network, the following pedagogical interventions are suggested:

Explicit Collocational Scaffolding and Materials Design: The authors of the FGEI curriculum and textbook designers should shift from the traditional approach that relies on vocabulary lists to an explicit collocational scaffolding approach. The new vocabulary in environmental and general academic chapters in English textbooks should be presented in whole, natural lexical chunks, explicitly marking the conventional verbs that go with the target academic nouns, e.g., the presentation of "conserve energy" and "mitigate pollution" as chunks, rather than defining "conserve" and "mitigate" separately from the nouns "energy" and "pollution." We need to shift away from word-matching exercises in writing, which

can result in erroneous word pairings in memory, toward contextualized chunk-retrieval writing tasks.

Integrating Urdu-English Contrastive Diagnostics: The systematic introduction of Urdu-English contrastive diagnostics should be incorporated into the vocabulary-teaching process for English language teachers in Pakistani schools. Instead of avoiding mother-tongue interference, teachers should directly teach non-equivalent collocations where the boundaries between Urdu and English do not match (e.g., by directly contrasting *Imtihaan dena* with taking an exam). This will enhance students' metalinguistic awareness and make them aware of the existence of direct L₁ translation paths that they will need to overcome in L₂ writing.

Leveraging the FGEI Teachers' Training Centre: The FGEI Directorate should plan and implement specialized training programs for the teachers' training center in Wah, Region, from the perspectives of lexical approach pedagogy and learner corpus analysis. English teachers should be provided with easy-to-use tools from the corpus and online concordance networks to prepare activities based on data rather than literary analysis, and to shift the focus of classroom activities to functional and academic writing instruction.

Transforming Written Corrective Feedback: The goal is to improve the written corrective feedback provided in FGEI English classes. In marking student essays for such errors as “do pollution,” teachers should provide more focused corrective feedback that indicates the conventional collocational partner (e.g., pointing out the correct form of the verb “pollute” above the error “do pollution”). It is a particular kind of feedback that helps learners detect and repair lexical mismatches, thereby facilitating the development of interlanguage.

Promoting Corpus-Guided Classroom Consulting: FGEI students need to be introduced to simple, straightforward corpus-consulting activities so they can verify word combinations on their own. When writing academic essays, students can verify that they are using valid English verb-noun combinations; e.g., they can determine whether “do mistake” is a valid pattern by consulting an online concordance or collocation dictionary (e.g., the BBI Combinatory Dictionary). This active consultancy approach helps bridge the lexical gap, allows the learner to self-edit, and reduces the use of literal translation strategies.

Conclusion

The present study is an application of a corpus approach that enables the empirical study and analysis of errors in the academic English writing of ESL learners at the secondary level in the Federal Government Educational Institutions School System in Pakistan. The study has analyzed the structures, cognitive aspects (LDCs), and pedagogical contexts that affect students' writing through a learner corpus and diagnostic surveys of both students and teachers. The empirical results indicate that errors in collocating verbs and nouns are frequent, with a very high error density, which has a significant negative impact on the idiomatic accuracy, cohesion, and clarity of students' academic texts.

The analysis of errors by structure shows that they are interlanguage patterns, namely: article omission errors, errors in verb choice, limited collocation constraints, verb simplification, incomplete patterns, and limited academic vocabulary. The study has concluded that interference with Urdu in the mother tongue is one of the important cognitive reasons for these errors. The negative interlingual transfer in terms of semantic boundaries of compound verbs and article-free constructions in Urdu is always present in L2 writers that are directly transferred to English noun phrases without any change. One factor that contributes to errors becoming fossilized and persistent is the importation of cognitive structures and the didactic setting, which focuses mainly on memorizing lexemes and working with them in isolation.

Finally, the findings of this study call for a complete overhaul of English language education in English-medium public education networks in Pakistan. To help students move from grammatically rigid to idiomatically natural academic writing, classrooms should shift their focus from a grammatically rigid, formulaic approach to a formulaic, language-chunk-focused lexical approach. We recommend explicit collocational scaffolding, contrastive Urdu-English diagnostics, targeted teacher training at the Wah FGEI training center, and corpus-guided classroom activities to enhance ESL learners' collocational competence in Pakistan for academic and professional purposes.

References

- Akhter, S., Anwar, M. N., & Qureshi, A. H. (2018). An investigation into the grammatical and lexical challenges of Pakistani ESL learners at secondary level. *Journal of Educational Research*, 21(2), 112–125.
- Alruwaili, A. (2020). Verb-noun collocations in the academic writing of EFL learners: A corpus-based study. *International Journal of Linguistics*, 12(4), 1–18.
- Ambreen, S., & To, C. K. S. (2021). First language interference and grammatical errors in the English writing of Pakistani ESL learners. *Journal of Psycholinguistic Research*, 50(3), 589–605.
- Anwar, B. (2012). L1 interference in the acquisition of English syntax by Punjabi-speaking ESL learners (Unpublished doctoral dissertation). University of Faisalabad, Pakistan.
- Bahns, J., & Eldaw, M. (1993). Should we teach collocations? *System*, 21(1), 101–114.
- Benson, M., Benson, E., & Ilson, R. (1997). *The BBI combinatory dictionary of English: A guide to word combinations* (2nd ed.). John Benjamins Publishing.
- Benson, M., Benson, E., & Ilson, R. (2010). *The BBI combinatory dictionary of English: Your guide to collocations and grammar* (3rd ed., rev. by R. Ilson). John Benjamins Publishing.
- Biskup, D. (1992). L1 interference in the production of L2 collocations. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 85–93). Macmillan.
- Carter, R. (2012). *Vocabulary: Applied linguistic perspectives* (2nd ed.). Routledge.
- Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5(4), 161–170.
- Corder, S. P. (1981). *Error analysis and interlanguage*. Oxford University Press.
- Cowie, A. P. (1992). Multiword lexical units and communicative language teaching. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 1–12). Macmillan.
- Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 175–195). Longman.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. Oxford University Press.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.
- Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *IRAL – International Review of Applied Linguistics in Language Teaching*, 33(4), 315–331.
- Ferris, D. R. (2002). *Treatment of error in second language student writing*. Lawrence Erlbaum Associates.
- Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford University Press.
- Futagi, Y., Deane, P., Chodorow, M., & Tetreault, J. (2008). A web-based English-writing assistant for EFL writers: Automated feedback on collocations. *Computer Assisted Language Learning*, 21(4), 349–363.
- Ghani, N. A., & Hafeez, M. R. (2025). A study of problematic areas of English collocations at higher secondary level in Pakistani ESL learners. *Pakistan Languages and Humanities Review*, 9(3), 228–237.
- Granger, S. (1998). *Learner English on computer*. Longman.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

- Hill, J. (2000). Collocation: An essential ingredient in collocation. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47–69). Language Teaching Publications.
- Howarth, P. (1998). The phraseology of academic writing: Towards a learning objectives framework. *Language Teaching*, 31(1), 24–44.
- Jamsaid, R. (2018). Investigating collocation errors of EFL learners at secondary level in Multan (Unpublished master's thesis). National University of Modern Languages, Islamabad, Pakistan.
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647–672.
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Language Teaching Publications.
- Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. Language Teaching Publications.
- Mahmood, A., Anwar, M. S., & Tariq, A. R. (2020). Syntactic interferences of Urdu in the English writing of Pakistani public school students. *Global Social Sciences Review*, 5(2), 145–154.
- Marrium, Z., Kashif, F., & Shahid, S. (2025). Collocational competence in ESL writing: An analysis of common errors among matric students in Lahore. *Pakistan Languages and Humanities Review*, 9(3), 525–542.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223–242.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins Publishing.
- Palmer, H. E. (1933). Second interim report on English collocations. Kaitakusha.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209–231.
- Umair, H. M., Imran, M., & Sarwat, S. (2023). An analysis of the use of collocations in English essay writing at undergraduate level. *Pakistan Languages and Humanities Review*, 7(1), 43–54.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: An empirical study. *TESOL Quarterly*, 44(4), 647–668.

Article Information:

Received	9-Jan-2025
Revised	21-Feb-2026
Accepted	12-Mar-2026
Published	30-Mar-2026

Declarations:

Authors' Contribution:

- All Authors **Conceptualization, and intellectual revisions, Data collection, interpretation, and drafting of manuscript**
- The authors agree to take responsibility for every facet of the work, making sure that any concerns about its integrity or veracity are thoroughly examined and addressed

▪ **Conflict of Interest:** NIL

▪ **Funding Sources:** NIL

Correspondence:

Muhammad Kamran Abbas Ismail

Kamran.ch@uo.edu.pk
