SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 5.1, (2026), pp. 1009-1034
Open Access. Online & Print

DOI: 10.5281/zenodo.12511080

ERROR ANALYSIS OF ENGLISH WRITING AMONG MALE ENGINEERING UNDERGRADUATES AT FIVE SAUDI UNIVERSITIES: A CORPUS-BASED STUDY

Sultan Ayed Alanazi[1]*

[1] *Department of Languages and Translation, University of Tabuk, Tabuk, Saudi Arabia.*

Received: 01/12/2025
Accepted: 02/01/2026

Corresponding author: Sultan Ayed Alanazi (S_alanazi@ut.edu.sa)

ABSTRACT

This study investigates the patterns, frequencies, and likely sources of English writing errors produced by male engineering undergraduates at five Saudi universities. Because the sample is exclusively male, the title, abstract, research questions, and every claim throughout the paper refer to this population specifically; generalisation to female engineering students requires empirical confirmation and is not warranted by the present data. Adopting a descriptive-analytical approach informed by Corder's (1967, 1981) framework of error analysis and supported by corpus-linguistic procedures (McEnery & Hardie, 2012; Granger, 2009), the study analysed a learner corpus of 450 written texts (approximately 180,000 words) produced by 150 male undergraduate students drawn equally from civil, mechanical, electrical, chemical, and industrial engineering programmes. Each participant produced three text types under controlled conditions: a technical essay, a laboratory report, and a formal letter. Errors were identified, categorised into five principal taxonomic groups, and analysed using descriptive procedures and inferential procedures including generalised linear mixed-effects models (GLMMs) with random intercepts for participant and institution, complemented by one-way ANOVA with Bonferroni-corrected post-hoc comparisons and Pearson correlations. Under the primary operational definition, grammatical and morphological errors constituted the largest proportion (38.2%), followed by lexical (24.5%), syntactic (18.7%), spelling (10.3%), and cohesion and coherence errors (8.3%). A pre-specified recurrent-only sensitivity analysis — in which the literature-matching criterion was dropped and only within-participant recurrent deviations were retained — produced qualitatively the same picture (39.8%, 23.3%, 19.1%, 10.4%, 7.5% respectively): the rank ordering of categories was unchanged, all task-type effects that survived Bonferroni correction in the primary analysis also survived in the sensitivity analysis, and the lexical specialisation effect was retained. After correction for multiple comparisons, statistically reliable effects emerged primarily for writing task type (lexical, syntactic, and cohesion errors) and, to a more modest degree, for engineering specialisation in the lexical category. The findings extend recent corpus-based work on Saudi EFL writing (Alsehibany & Abdelhalim, 2023; Albelihi & Al-Ahdal, 2024; Ahmed et al., 2025) and suggest that error patterns reflect a combination of cross-linguistic influence between Arabic and English, the inherent complexity of the English linguistic system, and instructional factors associated with current English for Specific Purposes (ESP) provision. The study offers pedagogical recommendations for curriculum developers, ESP instructors, and policymakers concerned with engineering education in Saudi Arabia, with the explicit qualification that recommendations are derived from male-only data and parallel evidence on female students should be sought before generalising.

KEYWORDS: error analysis; English for Specific Purposes; male engineering students; Saudi Arabia; learner corpus; cross-linguistic influence; mixed-effects models

DATA AVAILABILITY STATEMENT

The raw error counts at the participant and text level (under both the primary operational definition and the recurrent-only sensitivity definition), the anonymised learner corpus metadata, the R analysis scripts (including model specifications and convergence diagnostics for the GLMMs), and the operational coding guidelines used by the two coders are available in the Open Science Framework repository [URL to be supplied upon acceptance]. Uncorrected p values, Bonferroni-corrected p values, and false-discovery-rate (FDR) corrected p values for all inferential tests are tabulated in Supplementary Table S1. Figures 1–5 are embedded in this manuscript and also provided as separate high-resolution files (.png, 300 dpi).

1. INTRODUCTION

English has become the lingua franca of international engineering communication, scientific publication, and global professional practice. Engineers are increasingly expected to read technical documentation, draft project reports, prepare correspondence with international partners, and disseminate research findings in English (Hyland, 2019; Hutchinson & Waters, 1987). Within the Kingdom of Saudi Arabia, this expectation is reinforced by national workforce development priorities, by the international orientation of the engineering professions, and — since the launch of Vision 2030 — by an expanded role for English-medium instruction in undergraduate engineering programmes (Albelihi & Al-Ahdal, 2024).

Despite considerable institutional investment in English language instruction, engineering students whose first language is Arabic continue to encounter substantial difficulty in producing accurate, coherent written English (Habbash & Albakrawi, 2014; Albelihi & Al-Ahdal, 2024). The structural, orthographic, and rhetorical asymmetries between Arabic and English create persistent areas of difficulty in writing (Tahaineh, 2010; Al-Khasawneh, 2014), and these difficulties are compounded by the specialised nature of engineering discourse, which demands precision, lexical density, and adherence to genre-specific conventions (Swales, 1990; Dudley-Evans & St. John, 1998).

Within applied linguistics, error analysis has provided a productive lens for understanding the systematic nature of learner difficulty since Corder's (1967, 1981) seminal work, and it remains a central methodological resource for L2 writing research (Brown, 2014; James, 1998), as reviewed by Al-Khresheh (2016). More recent methodological advances — particularly the integration of corpus-linguistic procedures (McEnery & Hardie, 2012; Granger, 2009) — have enabled error analysts to move beyond small qualitative samples toward systematic, quantifiable accounts of learner output (Alsehibany & Abdelhalim, 2023). However, much of the existing scholarship on Arab learners of English has examined undergraduate populations in general (Al-Khasawneh, 2014; Murad & Khalil, 2015; Ahmed et al., 2025), with comparatively little work addressing the specific linguistic challenges faced by engineering students writing within their disciplinary contexts (Birhan & Nurie, 2024).

1.1 Statement of the Problem

The pedagogical literature documents three interrelated gaps. First, although several studies have examined writing errors among Saudi learners (Alhaisoni, Gaudel, & Al-Zuoud, 2017; Altamimi & Ab Rashid, 2019; Albelihi & Al-Ahdal, 2024), few have focused specifically on engineering populations or compared error patterns across engineering subdisciplines. Second, most existing investigations have analysed a single text type (typically essays) rather than the diverse genre repertoire that engineering students must master (Swales, 1990). Third, the methodological designs of earlier studies have rarely combined corpus-linguistic procedures with appropriate inferential statistical comparisons — particularly multi-level analyses that account for the nested structure of repeated-measures writing data — across specialisations and task types.

The present study addresses these gaps through a corpus-based investigation of writing errors across five engineering specialisations and three text types, using a sample of male students large enough to support inferential statistical comparison and applying mixed-effects modelling to respect the nested structure of the data. The all-male composition of the sample reflects access constraints arising from gender-segregated Saudi higher education and is acknowledged in the title and reiterated throughout the paper.

1.2 Research Questions

The study addresses four research questions, all framed for the male engineering population sampled:

- RQ1: What types of English writing errors do male engineering undergraduates at the five participating Saudi universities produce, and what are the relative frequencies of these errors?
- RQ2: Do error patterns differ statistically across engineering specialisations (civil, mechanical, electrical, chemical, industrial) within this sample?

- RQ3: Do error patterns differ statistically across writing task types (technical essay, laboratory report, formal letter)?
- RQ4: To what extent do observed error patterns appear to reflect cross-linguistic influence between Arabic and English, as opposed to intralingual or developmental factors?

1.3 Significance of the Study

The study contributes to the existing literature in three respects. Theoretically, it extends error-analytic and interlanguage frameworks (Corder, 1967; Selinker, 1972; Brown, 2014) to a specialised disciplinary context that has received limited empirical attention. Methodologically, it offers a model for combining learner-corpus construction (Granger, 2009) with multi-level inferential analysis — including mixed-effects modelling that respects the nested structure of repeated-measures writing data — across multiple writing genres, and it pairs the primary inferential analysis with a pre-specified recurrent-only sensitivity analysis (Section 4.1) to test the robustness of conclusions to the operational definition of systematic error. Pedagogically, it provides curriculum designers, English for Specific Purposes (ESP) practitioners, and policymakers with empirically grounded evidence about which areas of English writing most warrant targeted intervention for male engineering students, complementing recent intervention work in engineering ESP settings (Birhan & Nurie, 2024; Alsehibany & Abdelhalim, 2023).

1.4 Scope of the Study

The investigation is bounded by the following parameters. The participants are male undergraduate engineering students enrolled at five Saudi universities during the 2024–2025 academic year. Data collection focused on three written genres relevant to engineering practice. The analysis is restricted to written production; spoken language and receptive skills are outside the scope of the study. The findings should be interpreted in light of these boundaries — particularly the all-male sample, which reflects access constraints under gender-segregated Saudi higher education — and the further limitations discussed in Section 7.

2. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1 Theoretical Foundations

2.1.1 Contrastive Analysis

The earliest systematic attempt to explain second-language learner difficulty was the Contrastive Analysis Hypothesis (CAH), most influentially articulated by Lado (1957) and developed in subsequent work in the 1950s and early 1960s. CAH proposed that learner difficulties could be predicted by structural comparison between the learner's first language (L1) and the target language (L2), with similarities facilitating acquisition (positive transfer) and differences impeding it (negative transfer or interference). For Arabic-speaking learners of English, well-documented L1–L2 asymmetries exist at multiple levels: phonological inventories, orthographic systems and writing direction, syntactic word order, the article system, the verb tense system, and morphological derivation (Tahaineh, 2010; Alhaisoni et al., 2017).

Over time, the strong predictive form of CAH was widely criticised. Empirical work demonstrated that not all predicted errors actually occur, that many observed errors cannot be traced to L1 transfer, and that learner difficulty is shaped by factors beyond cross-linguistic difference (Ellis, 2008; James, 1998; Al-Khresheh, 2016). These critiques motivated the emergence of more comprehensive frameworks, particularly Error Analysis.

2.1.2 Error Analysis

Corder's (1967) paper "The Significance of Learners' Errors" reframed errors as systematic windows onto the learner's developing competence rather than as failures to be eradicated. In Corder's later synthesis (1981), error analysis is articulated as a multi-stage procedure: collection of learner samples, identification of errors, description and classification, explanation, and evaluation. The framework distinguishes interlingual errors (attributable to L1 influence), intralingual errors (arising from the structural complexity of the L2 itself), and developmental errors (reflecting normal stages of acquisition). James (1998) extended this framework with a more elaborated typology, and Al-Khresheh (2016) provides a useful contemporary review of error-analytic theory and its limitations.

Error analysis has been criticised for focusing on what learners cannot do rather than on what they can produce successfully, and for the difficulty of unambiguously assigning errors to single causal sources (Ellis, 2008). Nevertheless, it remains a productive methodological starting point (Brown, 2014; Albelihi & Al-Ahdal, 2024), particularly when combined with corpus-linguistic procedures that make error description more systematic and quantifiable (Granger, 2009; Alsehibany & Abdelhalim, 2023).

2.1.3 Interlanguage and Fossilisation

Selinker (1972) introduced the construct of interlanguage to describe the dynamic, rule-governed linguistic system that learners develop as they progress toward target-language norms. Interlanguage is shaped by L1 transfer, overgeneralisation of L2 rules, transfer of training, communication strategies, and learning strategies, and it can stabilise (fossilise) at any point in development. The interlanguage construct accommodates the observation that learners' productions are neither random nor merely deficient versions of the target language but reflect an internally consistent system in transition (Brown, 2014).

Recent work has revived interest in the construct of fossilisation as a specific challenge for advanced Saudi EFL writers. Albelihi and Al-Ahdal (2024), drawing on a corpus of Saudi English-major writing across argumentative, expository, narrative, and descriptive genres, documented that certain errors — particularly idiomatic language use and subject–verb agreement — appear to be fossilised in their sample, while other categories diminish over time. Their study also identified persistent challenges with articles, plural markers, negative markers, and prepositions, attributing fossilisation to linguistic differences between Arabic and English and to limited exposure to authentic input. Their findings underscore the importance of distinguishing developmental errors from candidate fossilised features in error-analytic work. However, because their study and the present one are both cross-sectional, all attributions of fossilisation should be treated as hypotheses to be confirmed by longitudinal research. For Arabic-speaking engineering students specifically, interlanguage analysis is informative because their language production typically reveals consistent patterns — hybrid syntactic structures, recurrent preposition substitutions, characteristic article omissions — that can be interpreted as features of a developing system rather than as isolated errors (Tahaineh, 2010).

2.2 Categories of Writing Errors in EFL Research

Researchers have developed a variety of classifications systems to identify and categorize second language writers' error types (James, 1998). The classification system used in this study was drawn from earlier research examining Arab learners' second language written output (Murad & Khalil, 2015; Al-Khasawneh, 2014; Ahmed et al., 2025); and consistent with previous L2 writing literature (Hyland, 2019). The classification system identified five primary error categories including: Grammatical/Morphological Errors; Spelling Errors; Syntactical Errors; Lexical Errors; and Cohesion/Coherence Errors.

2.2.1 Grammatical and Morphological Errors

This category encompasses deviations from English morpho-syntactic conventions including tense inflection, subject–verb agreement, preposition usage, article usage, pronoun reference, and the marking of singular and plural forms. Studies of Arab learners consistently identify preposition usage and the article system as among the most error-prone domains, in part because these closed-class systems differ markedly between Arabic and English (Tahaineh, 2010; Alhaisoni et al., 2017). Recent fossilisation research on Saudi English majors suggests that subject–verb agreement, articles, plural markers, negative markers, and prepositions may persist into advanced stages of learning unless explicitly targeted (Albelihi & Al-Ahdal, 2024).

2.2.2 Spelling Errors

The frequent spelling errors made by Arabic-speaking students stem from the significant difference that exists between Arabic and English writing systems, the non-regular correspondence of English graphemes to phonemes, and the lack of some English sounds in the Arabic language system (particularly /p/, and /v/). Altamimi and Ab Rashid **(2019)** describe these challenges in detail for the context of Saudi Arabia, noting omitted silent letters, double letter confusions, and frequently substituting letters with phonetically similar ones are common.

2.2.3 Syntactic Errors

Syntactical issues relate to the placement of words in relation to each other (word order), linking clauses together, forming interrogative forms, negating sentences, constructing conditions, and using the passive voice. In contrast to the typical English "Subject-Verb-Object" word ordering, the primary structure of an Arabic sentence is "Verb-Subject-Object," which may lead to specific types of word ordering mistakes by Arab language learners (Tahaineh, 2010; Murad & Khalil, 2015). Recent research on corpus-based writing of Arabic speakers at varying levels of proficiency further support that syntactical difficulties persist beyond novice or intermediate levels (Albelihi & Al-Ahdal, 2024).

2.2.4 Lexical Errors

Lexical errors include inappropriate word choice, misuse of technical terminology, literal translation from Arabic (calque), incorrect collocation, and misuse of idiomatic expressions (Nesselhauf, 2005). Ahmed, Abdallah, Hamed, and Hamed (2025), examining 15 Saudi EFL undergraduates at King

Khalid University, reported that literal translation from Arabic accounted for the highest frequency of word-choice errors (38.29%), followed by collocation errors (28.00%), subject–verb agreement (21.14%), and word-formation errors (12.57%), confirming the productive role of L1 lexical-semantic transfer in shaping Saudi learners' English vocabulary use. For engineering students, accurate use of specialised technical vocabulary is particularly demanding because much of this vocabulary is restricted to academic and professional contexts and is not encountered in general English instruction (Habbash & Albakrawi, 2014; Dudley-Evans & St. John, 1998).

2.2.5 Cohesion and Coherence Errors

The category is concerned with textual organization. It addresses a range of issues including how writers organize their writing at the level of individual sentences using cohesive language, the ordering of sentences/ideas within paragraphs, the organizational format of paragraphs, whether writers have taken a stance, and the clarity of an argumentative structure. Textual organization skills can only develop through students' exposure to authentic forms of disciplinary genres and through explicit teaching of disciplinary genre formats (Swales, 1990; Hyland, 2019). Although research by Connor (1996; 2008) demonstrates that students' first language writing conventions influence their ability to develop textual organization skills, this should be viewed with caution because students' first languages do not represent all members of their culture or language group.

2.3 Factors Influencing Writing Errors

2.3.1 Cross-Linguistic Influence

Cross-linguistic influence remains a productive explanatory construct for understanding many features of Arab learners' English writing (Tahaineh, 2010; Ahmed et al., 2025; Al-Khresheh, 2016). Such influence operates at the phonological level (affecting spelling), the morpho-syntactic level (affecting preposition and article usage and word order), the lexical level (motivating literal translation), and the rhetorical level (shaping discourse organisation). Earlier contrastive-rhetoric scholarship (Kaplan, 1966) proposed broad cultural patterns in text-level organisation; more recent intercultural-rhetoric research (Connor, 2008) has substantially revised this position, emphasising the situated, dynamic, and individual character of L2 rhetorical choices rather than fixed cultural archetypes.

2.3.2 Intralingual and Developmental Factors

Not all errors can be attributed to L1 influence. Learners also overgeneralise L2 rules (e.g., applying regular past-tense morphology to irregular verbs), simplify complex structures, and avoid difficult forms (Ellis, 2008; James, 1998). Such intralingual phenomena can produce errors that resemble those made by learners from many different L1 backgrounds and that reflect general properties of second-language acquisition rather than language-specific transfer (Brown, 2014; Al-Khresheh, 2016).

2.3.3 Instructional Factors

The quality and orientation of language instruction also bear on observable error patterns. Habbash and Albakrawi (2014) documented a substantial mismatch between the language needs of engineering students at the University of Tabuk and the curriculum they actually received, with insufficient attention to discipline-specific genres and terminology. Bacha and Bahous (2011), writing about the broader Lebanese context, similarly observed how curricular constraints can leave learners under-prepared for the writing demands they will encounter. More recent intervention research (Alsehibany & Abdelhalim, 2023; Birhan & Nurie, 2024) has demonstrated that targeted corpus-based instruction can meaningfully reduce error rates in academic vocabulary and improve engineering students' engagement with academic writing, suggesting that instructional reform is both necessary and feasible.

2.3.4 Rhetorical and Cultural Factors

Kaplan (1966) was the first researcher to suggest that each culture develops its own unique tradition of rhetoric; and these traditions determine how individuals approach written communication as they learn another language. While many researchers have since revised Kaplan's original descriptions of what constitute those rhetorical traditions (Connor, 1996, 2008), his idea that the way students organize their texts, and the way students persuade readers reflect cultural traditions has had an enduring impact on writing research. To be most effective, this impact needs to be understood in terms of both within-culture differences among people and among individuals' beliefs about what makes good writing.

2.4 Previous Empirical Studies

2.4.1 Foundational Studies on Saudi and Arab Learners

A foundational body of research has examined writing errors among Saudi and Arab undergraduate populations. Al-Khasawneh (2014), in a study of 26 Jordanian undergraduates at Ajloun National University, identified article misuse as the most frequent error type and emphasised the role of L1

transfer in shaping error patterns. Murad and Khalil (2015), examining 22 Arab first-year college students of EFL in Israel, documented errors across content and organisation, syntax, lexis, and mechanics, and attributed many to negative transfer from Arabic.

Within the Saudi context, Alhaisoni et al. (2017) analysed article errors in writing produced by 150 Saudi EFL preparatory-year students at the University of Ha'il, finding that article misuse was widespread and largely systematic. Altamimi and Ab Rashid (2019) examined spelling difficulties among Saudi English-language undergraduates at the University of Tabuk, identifying interference from the Arabic orthographic and phonological systems together with structural features of Saudi English-language curricula as principal contributing factors. Tahaineh (2010), focusing on Arab EFL university students' use of prepositions, demonstrated that preposition errors were both highly frequent and closely linked to L1 transfer patterns.

2.4.2 Recent Corpus-Based and Intervention Studies (2023–2025)

The most recent wave of Saudi EFL writing research has been characterised by larger corpora, more sophisticated computational tools, and increasing attention to intervention. Alsehibany and Abdelhalim (2023), in a quasi-experimental study of Saudi female English majors at Imam Mohammad Ibn Saud Islamic University, evaluated the effect of direct online corpus consultation on vocabulary errors (including collocation, word form, and spelling) in academic writing. The intervention produced significant reductions in vocabulary errors and measurable improvements in academic writing quality, with the largest gains in the more transparently rule-governed dimensions of vocabulary.

Albelihi and Al-Ahdal (2024), in a corpus content analysis of Saudi English-major writing across multiple genres, examined the phenomenon of error fossilisation — the entrenchment of recurrent errors despite ongoing instruction and feedback. Their findings, published in the Asian-Pacific Journal of Second and Foreign Language Education, identified idiomatic language use and subject–verb agreement as the most strongly fossilised error domains in their sample, with additional persistent challenges in articles, plural markers, negative markers, and prepositions. They proposed an explicit, contrastive-focused remediation framework for advanced Saudi EFL learners. This work usefully reframes some of the high-frequency error categories observed in the present study as candidates for fossilisation that warrant explicit instructional attention, while recognising that confirmation of fossilisation requires longitudinal evidence that neither study can provide.

Ahmed, Abdallah, Hamed, and Hamed (2025) reported a mixed-method analysis of word-choice errors in the written production of 15 Saudi EFL undergraduates at King Khalid University, combining quantitative error frequency analysis with structured interviews of writing teachers. Literal translation from Arabic was the most frequent word-choice error type (38.29%), followed by collocation (28.00%), subject–verb agreement (21.14%), and word-formation errors (12.57%). The authors attribute the dominant role of literal translation to limited lexical-semantic exposure and to instructional approaches that emphasise translation equivalence over context-based meaning. These findings closely parallel the lexical-error patterns reported in Section 4.3 of the present study.

Two recent corpus-based studies have focused specifically on engineering students; because both involve L1-Chinese L2-English writers in Hong Kong rather than Arabic-L1 writers, they are referenced selectively for engineering-genre context rather than for direct comparison of transfer patterns. Siu, Afzaal, Aldayel, and Curle (2024), in a corpus-based analysis of lexical bundles in 119 civil and environmental engineering reports written by L2 English university students in Hong Kong, found that high-scoring papers used a wider range of 3-word, 4-word, and 5-word lexical bundles than low-scoring papers and called for genre-sensitive vocabulary instruction in engineering ESP. A companion study by the same research team — listed in its published bibliographic record with the first author's personal name appearing as a surname (Barbara, Afzaal, & Aldayel, 2024) — examined stance markers in the same engineering context.

Note on author name. In Hong Kong publishing conventions, the first author of these two studies appears in the published record under different name orderings (Siu in 2024 SAGE Open; Barbara in 2024 Humanities and Social Sciences Communications). For consistency and to avoid the impression of two different authors, the present paper cites the SAGE Open study as "Siu et al. (2024)" and refers to the stance-marker study only where directly relevant, citing it as "Siu et al. (2024, HSSC)".

Birhan and Nurie (2024), in a quasi-experimental study of 77 fourth-year mechanical engineering students in Ethiopia, demonstrated that a corpus-based instructional intervention significantly improved students' behavioural, emotional, and cognitive engagement with academic writing, providing evidence that corpus-informed instruction is both pedagogically effective and motivationally productive in engineering ESP contexts. Together,

these recent studies establish that learner-corpus methodology has become a methodological mainstream in EFL writing research and that targeted instructional interventions — particularly corpus-based ones — offer a promising route to reducing these errors.

2.4.3 The Research Gap

Drawing on the studies reviewed above, the present investigation addresses three identifiable gaps. First, although general studies of Saudi and Arab learners' writing exist, none has focused on Saudi male engineering students across multiple engineering subdisciplines. Second, prior studies have generally analysed a single text type rather than examining how error patterns vary across the genres engineering students produce (Swales, 1990). Third, the methodological combination of learner-corpus construction (Granger, 2009) with appropriate inferential analysis — including multi-level modelling that respects the nested structure of repeated-measures writing data — across both specialisations and task types has rarely been applied in this population.

3. METHODOLOGY

3.1 Research Design

The study employed a descriptive-analytical design supported by corpus-linguistic procedures (McEnery & Hardie, 2012; Granger, 2009). The error-analytic procedure followed the stages articulated by Corder (1981): collection of learner samples, identification of errors, description and classification, and explanation. Inferential statistical analyses — including generalised linear mixed-effects models that respect the nested structure of repeated-measures writing data, supplemented by ANOVA with Bonferroni-corrected post-hoc procedures — were used to test for differences across engineering specialisations and task types. An a priori power analysis (G*Power 3.1; Faul et al., 2009) indicated that the planned sample of 150 male participants would provide power of approximately .85 to detect medium-sized effects (f = .25) in a one-way ANOVA with five groups at $\alpha = .005$ (Bonferroni-adjusted). A pre-specified recurrent-only sensitivity analysis (described in Section 3.5.1 and reported in Section 4.1) was conducted in parallel with the primary analysis to test the robustness of conclusions to the operational definition of systematic error.

3.2 Participants

3.2.1 Population

The target group was male undergraduate engineering student attending an ESP course at a Saudi university during the 2024-25 school year. To provide better representation from all over the country five universities were chosen: King Saud University in Riyadh, King Fahd University of Petroleum & Minerals in Dhahran, King Abdulaziz University in Jeddah, King Khalid University in Abha, and the University of Tabuk in Tabuk

3.2.2 Sample

The data was collected using a stratified sample of 150 males; this group was stratified based on engineering specialization. Thirty males were sampled from each of the 5 different engineering specialties (civil, mechanical, electrical, chemical & industrial); they were selected from the second year students who had completed their first two years of study as well as their foundational English language course work. These same second-year students were then enrolled in ESL/ESP coursework specific to their chosen area of specialization. Participation in the survey was completely voluntary. Institutional ethics approval for participant consent was provided and received by the researchers from each student. The sample consisted exclusively of male students attending the three institutions that made up the pool from which the researcher could collect data due to restrictions related to collecting data from female students in Saudi Arabia's segregated educational environment. All limitations are discussed in detail in Sections 2.8, 6.0 & 7.0 in addition to being mentioned in the title of the paper and discussed when interpreting every finding. Students' ages ranged from 19-23 with an average age of 20.7 (SD = 1.1) and indicated that they had studied English for approximately nine to fourteen years (average time studying English = 11.3 or SD = 1.4).

3.3 Data Collection Instruments

3.3.1 Writing Tasks

Three writing tasks were designed to sample the genres most relevant to engineering practice. The selection of genres was informed by ESP needs-analysis evidence (Habbash & Albakrawi, 2014; Hutchinson & Waters, 1987; Dudley-Evans & St. John, 1998) and validated through consultation with disciplinary instructors at the participating institutions:

- Technical Essay (500–600 words): an academic essay on a contemporary engineering topic such as renewable energy, the Internet of Things, or sustainable infrastructure. This task elicited extended argumentative and expository writing.
- Laboratory Report (400–500 words): a structured report describing an experimental procedure

within the participant's specialisation, including method, results, and brief discussion. This task elicited technical-descriptive writing using disciplinary terminology.

- Formal Letter (200–250 words): a professional letter (e.g., job application, request for technical information, or formal complaint) addressed to a hypothetical professional entity. This task elicited transactional writing using formal register.

Each task was accompanied by clear written prompts specifying the required topic, length, audience, and format. The three tasks were administered across three separate sessions (one task per session) over a period of two weeks; each session lasted 75 minutes, providing approximately 8–10 minutes per 50 words of expected output. This design departs from prior studies in which all tasks were administered in a single session and addresses concerns about cognitive load and fatigue. Full task prompts and instructions are provided in Appendix A.

3.3.2 Error Classification Scheme

A coding scheme was developed to classify errors into five principal categories with detailed subcategories. The scheme was derived from prior literature on Arab learners' writing (Al-Khasawneh, 2014; Murad & Khalil, 2015; Tahaineh, 2010; Ahmed et al., 2025) and on general error-analytic typology (James, 1998; Brown, 2014; Al-Khresheh, 2016), and was refined through pilot application. The five principal categories and their subcategories are summarised in Table 1; full coding instructions and an extended exemplification are provided in Appendix B.

Table 1: Coding scheme for the classification of writing errors

Principal category	Subcategories
Grammatical and morphological	Tense; subject–verb agreement; prepositions; articles (definite/indefinite); pronouns; number (singular/plural); adjective and adverb usage
Spelling	Silent letters; double letters; phonologically similar letters; irregular spelling; capitalisation
Syntactic	Word order; compound sentence construction; interrogative/negative structures; conditional clauses; passive voice
Lexical	Inappropriate word choice; technical terminology; literal translation (calque); collocation; idiomatic expression
Cohesion and coherence	Cohesive devices; logical sequencing of ideas; paragraph organisation; main-idea clarity; transitions

3.4 Corpus Construction

A learner corpus was constructed from the collected texts following established corpus-linguistic procedures (McEnery & Hardie, 2012; Granger, 2009). The corpus comprises 450 texts (150 male participants × 3 tasks) and approximately 180,000 word tokens. Each text was assigned a unique identifier coding the participant's specialisation, institution, and task type, with all personally identifying information removed. Handwritten responses were transcribed verbatim, preserving all errors. Texts were prepared for analysis using AntConc 4.0 (Anthony, 2022) for concordancing and frequency operations, and analyses were performed using SPSS 28 for descriptive procedures and R version 4.3 with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) for mixed-effects modelling. The anonymised corpus metadata and the R analysis scripts are available in the Open Science Framework repository (see Data Availability Statement).

3.5 Error Analysis Procedures

3.5.1 Identification

Two trained coders who each held a graduate degree in Applied Linguistics and had extensive experience training undergraduate students from Saudi Arabia in the use of English identified errors individually. Only systematic errors (errors that reflected recurring patterns of incompetence in the writers' abilities) based on Corder's (1981) distinctions were selected for inclusion in the main study; isolated mistakes (errors which occurred randomly and were corrected immediately upon being prompted by the researcher) were excluded from consideration. A coder operationally defined an error as a "systematic" error when either (a) there was at least one other instance of the exact same type of error made by the same participant over all three writing tasks or (b) the type of error was consistent with those reported in previous studies on the English language learning abilities of Arab learners. In order to address the potential criticism that criterion (b) provided a bias toward previously identified types of errors, a predetermined sensitivity test was also run in which criterion (b) was omitted and only those errors were considered as evidence of a "systematic" error when they appeared to be repeated by the same participant. These data will be reported in section 4.1 and referred to throughout sections 4 and 5. The less liberal definition of a systematic error used in the original analysis resulted in fewer total errors being counted than would have been counted using more expansive definitions; these

implications will be discussed further in section 7. While it is recognized that the operative definition is far from perfect (i.e., some writers may exhibit a true systematic error only once in a very short time period such as in this small sample), the sensitivity test provides a clear means of evaluating whether the ultimate conclusions reached by the researchers depend on how systematically an error is defined.

3.5.2 Classification by Source

In addition to taxonomic classification, each error was tentatively categorised by source as interlingual (apparently traceable to Arabic structural influence) or intralingual (apparently arising from L2 complexity, overgeneralisation, or simplification), following James (1998), Ellis (2008), and Al-Khresheh (2016). Source classification was applied using explicit operational criteria. An error was classified as interlingual when (a) the deviant form mapped systematically onto a corresponding Arabic structure (e.g., omission of the indefinite article, or VSO word order in declarative sentences), and (b) the same form is documented as a transfer error in published Arabic-L1 contrastive analyses (Tahaineh, 2010; Ahmed et al., 2025). All other errors were initially classified as intralingual or indeterminate. The two coders applied this classification independently. Inter-coder agreement on the binary interlingual/intralingual decision, calculated on the same randomly selected 15% of the corpus used for primary-category reliability (Section 3.6.2), yielded Cohen's $\kappa = 0.74$ (95% CI [0.69, 0.79]), indicating substantial agreement (Landis & Koch, 1977) but lower than the κ for primary-category assignment ($\kappa = 0.85$), which is consistent with the inferential rather than observational character of source attribution. The 26% of discrepancies were resolved through discussion with reference to the published contrastive analyses. The proportions reported in the Discussion section reflect the converged classifications and should be interpreted as the proportion of errors for which an L1-transfer source could be plausibly established under the stated criteria, not as definitive causal attributions.

3.5.3 Quantification

For each text and for the corpus as a whole, the absolute frequency and percentage of each error category and subcategory were computed under both the primary operational definition (Section 3.5.1, criteria a or b) and the recurrent-only sensitivity definition (criterion a only). Comparisons across specialisations and task types were conducted using the inferential procedures described in Section 3.7. All raw counts under both definitions are tabulated by participant and category in the supplementary online materials.

3.6 Validity and Reliability

3.6.1 Validity

Face and content validity of the writing tasks and coding scheme were established through expert review. Seven reviewers (four applied linguists with expertise in error analysis and three engineering instructors with experience teaching ESP courses) evaluated the tasks for clarity, appropriateness, and authenticity, and the coding scheme for comprehensiveness and operational clarity. Inter-reviewer agreement on the suitability of categories reached 92%. Reviewer comments led to refinements in task prompts and to the addition of two subcategories to the coding scheme. Construct validity was assessed through pilot administration to 30 male students drawn from a non-participating institution; the pilot confirmed that the tasks elicited writing of the expected length, register, and complexity.

3.6.2 Reliability

Inter-coder reliability was established by having both coders independently analyse a randomly selected 15% of the corpus (68 texts). Three reliability indices were calculated. First, for the categorical assignment of errors to the five principal categories, Cohen's kappa yielded $\kappa = 0.85$ (95% CI [0.81, 0.89]), indicating very good agreement (Landis & Koch, 1977). Subcategory-level kappas ranged from 0.71 (lexical) to 0.89 (spelling), all in the substantial-to-near-perfect range. Second, for the binary classification of errors as interlingual versus intralingual or indeterminate, Cohen's $\kappa = 0.74$ (95% CI [0.69, 0.79]); see Section 3.5.2 for interpretation. Third, intra-coder consistency was established by having the principal coder re-analyse a randomly selected 10% of the corpus after a three-week interval; Pearson's r between the two analyses for the count of errors per text was 0.91, indicating high temporal stability. Discrepancies in all three procedures were resolved through discussion and were used to refine coding protocols.

3.7 Statistical Analysis

Two complementary inferential strategies were applied. The primary analyses used generalised linear mixed-effects models (GLMMs) implemented in R 4.3 with the lme4 package (Bates et al., 2015). Because error counts per text are non-negative integers and exhibited mild overdispersion, negative-binomial GLMMs were fitted with the natural logarithm of text word count as an offset (to model error rate per word). For each error category and for the total, a separate model included fixed effects for

engineering specialisation (5 levels, treatment-coded with electrical as the reference) and task type (3 levels, treatment-coded with formal letter as the reference), and random intercepts for participant (to accommodate within-participant clustering of texts) and for institution (to accommodate any institution-level variation). Model specifications, convergence diagnostics, and full coefficient tables for each category-specific model are provided in Supplementary Table S2. Wald z-tests with Bonferroni-corrected critical values were used for fixed-effect inference. The primary GLMMs were re-estimated on the recurrent-only data (Section 3.5.1) as part of the sensitivity analysis; the comparison is reported in Section 4.1.

To facilitate comparison with prior studies, supplementary analyses used one-way analysis of variance (ANOVA) on participant-level totals (for specialisation effects) and on text-level counts (for task effects), with Bonferroni-corrected post-hoc comparisons. Given that ten primary tests were planned (five error categories × two factors), the family-wise α was set to .005 for declaring an effect statistically reliable; uncorrected p values are reported for transparency, with reliability after correction explicitly noted. Effect sizes (η^2 for ANOVA; pseudo-R^2 for GLMMs) were calculated to characterise the practical magnitude of observed differences. Pearson product-moment correlations examined relationships between error categories at the participant level. All p values — uncorrected, Bonferroni-corrected, and FDR-corrected (Benjamini–Hochberg) — are tabulated in parallel in Supplementary Table S1.

3.8 Ethical Considerations

The study was conducted in accordance with the ethical guidelines of the participating institutions and received approval from the lead institution's research ethics committee .

4. RESULTS

This section presents the results in four parts. Section 4.1 reports the overall distribution of errors across the five principal categories under the primary operational definition and then under the pre-specified recurrent-only sensitivity definition. Sections 4.2 through 4.6 present detailed subcategory analyses for each principal category, together with the inferential results for differences across specialisations and task types using both the primary mixed-effects models and the supplementary ANOVA procedures. Section 4.7 reports the aggregate mixed-effects summary, and Section 4.8 examines correlations among error categories and the source-attribution proportions. Across all inferential tests, family-wise α was controlled at .005 using Bonferroni correction; uncorrected p values are reported for transparency. All findings reflect the all-male sample at the participating institutions and should not be extrapolated to female engineering students without empirical confirmation.

4.1 Overall Distribution of Errors and Sensitivity Analysis

Across the 450 texts and approximately 180,000 word tokens, a total of 3,782 errors were identified under the primary operational definition, corresponding to an aggregate error rate of approximately 2.10 errors per 100 words. Under the pre-specified recurrent-only sensitivity definition, in which criterion (b) (matching previously documented patterns) was dropped and only within-participant recurrent deviations were retained, 3,263 errors were identified, corresponding to an aggregate rate of approximately 1.81 errors per 100 words — a 13.7% reduction from the primary count. The reduction was modest in the more recurrent categories (grammatical: −10.0%; syntactic: −12.1%; spelling: −13.1%) and somewhat larger in categories whose individual instantiations were less likely to recur within a single participant's output (lexical: −18.0%; cohesion and coherence: −22.3%). Critically, the rank ordering of the five principal categories was unchanged across the two definitions (grammatical > lexical > syntactic > spelling > cohesion).

These rates are lower than those observed in some prior studies of EFL writers from comparable backgrounds, which have typically ranged from approximately 4 to 12 errors per 100 words (Altamimi & Ab Rashid, 2019; Murad & Khalil, 2015; Albelihi & Al-Ahdal, 2024). To strengthen the comparison with Albelihi and Al-Ahdal (2024) specifically — which is the most directly comparable Saudi corpus study and which reports errors per 100 words on argumentative, expository, narrative, and descriptive essays from Saudi English majors — we recomputed the present rate restricted to the technical essay task only (the closest analogue to their essay-based corpus). Under that restriction the present sample produced approximately 2.7 errors per 100 words under the primary definition and 2.4 under the recurrent-only definition, both still lower than the broadly comparable rates in Albelihi and Al-Ahdal (2024). Two factors are likely to contribute to the lower aggregate rate observed here. First, the operational definition of systematic error applied in this study (Section 3.5.1) is comparatively conservative: isolated, non-recurring deviations were treated as mistakes and excluded. Second, Saudi engineering students are typically admitted to their programmes on the basis of comparatively high

secondary-school grades and — at several of the participating institutions — English-language placement tests, so the sample is more proficient on average than the EFL preparatory-year and general undergraduate samples reported in earlier work.

Table 2: Overall distribution of errors by principal category under the primary and recurrent-only definitions

Principal category	Primary N	Primary %	Recurrent-only N	Recurrent-only %	Δ N (%)
Grammatical and morphological	1,444	38.2%	1,299	39.8%	−10.0%
Lexical	926	24.5%	759	23.3%	−18.0%
Syntactic	709	18.7%	623	19.1%	−12.1%
Spelling	389	10.3%	338	10.4%	−13.1%
Cohesion and coherence	314	8.3%	244	7.5%	−22.3%
Total	**3,782**	**100.0%**	**3,263**	**100.0%**	**−13.7%**

Note. Δ N (%) is the percentage reduction in absolute count from the primary to the recurrent-only definition. Percentages reflect rounded category totals; raw counts at the participant level are available in the supplementary materials.

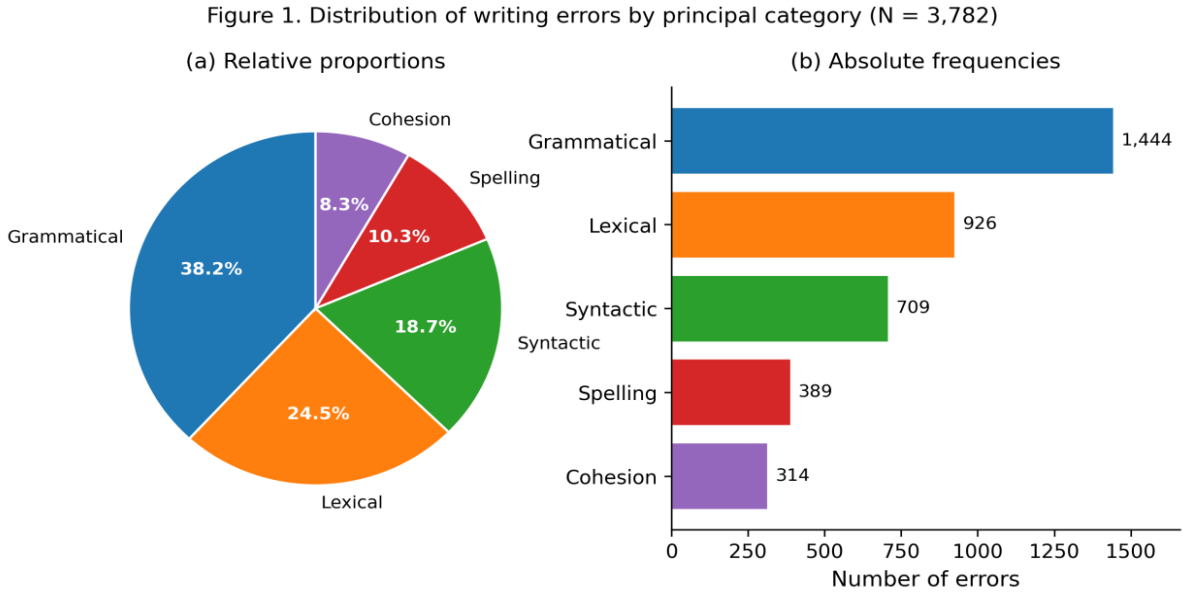

Figure 1: Distribution of writing errors by principal category under the primary operational definition (N = 3,782). Panel (a) shows relative proportions; panel (b) shows absolute frequencies.

All inferential tests reported in Sections 4.2–4.7 were repeated on the recurrent-only data. Every effect that survived Bonferroni correction in the primary analysis also survived correction in the sensitivity analysis, and no inferentially reliable effect changed sign. Two changes of nominal significance occurred: the marginal specialisation effect on grammatical errors (primary GLMM $\chi^2(4) = 9.74$, $p = .045$) fell further short of reliability in the sensitivity analysis ($\chi^2(4) = 7.86$, $p = .097$), and the nominally significant task-type effect on spelling errors (primary GLMM $\chi^2(2) = 7.40$, $p = .025$) similarly weakened ($\chi^2(2) = 5.92$, $p = .052$). Neither effect was reliable under Bonferroni correction in either analysis. The substantive conclusions of the study are therefore independent of the operational definition adopted.

Grammatical and morphological errors constituted the largest single category in both analyses, accounting for more than one-third of all errors. Lexical errors formed the second-largest category, followed by syntactic errors. Spelling and cohesion/coherence errors, while less frequent in absolute terms, nevertheless represent meaningful loci of difficulty for the male engineering students sampled. The relative ordering of categories closely mirrors that reported by Albelihi and Al-Ahdal (2024) for Saudi English-major writing and by Ahmed et al. (2025) for word-level errors specifically, suggesting that the distributional pattern observed here generalises beyond the male engineering population studied — though direct confirmation with female engineering students remains a priority.

4.2 Grammatical and Morphological Errors

4.2.1 Subcategory Distribution

Within the grammatical and morphological category, the seven subcategories distributed as shown in Table 3.

Table 3: Distribution of grammatical and morphological errors by subcategory

Subcategory	Frequency	% of category	% of total
Preposition usage	412	28.5%	10.9%
Tense usage	387	26.8%	10.2%
Articles (definite/indefinite)	246	17.0%	6.5%
Subject–verb agreement	178	12.3%	4.7%
Pronoun usage	102	7.1%	2.7%
Singular/plural number	76	5.3%	2.0%
Adjective and adverb usage	43	3.0%	1.1%
Total	**1,444**	**100.0%**	**38.2%**

Preposition errors and tense errors together accounted for more than half of all grammatical and morphological errors. Article errors formed the third-largest subcategory, consistent with the well-documented difficulty Arabic-speaking learners experience with the English article system (Alhaisoni et al., 2017; Albelihi & Al-Ahdal, 2024).

4.2.2 Differences across Specialisations

Descriptive statistics for grammatical errors by specialisation are presented in Table 4 and visualised in Figure 2.

Table 4: Mean grammatical and morphological errors per participant by specialisation (sum across three tasks)

Specialisation	M	SD
Civil engineering	10.2	2.8
Mechanical engineering	9.8	2.5
Chemical engineering	9.5	2.6
Industrial engineering	9.1	2.4
Electrical engineering	8.7	2.3

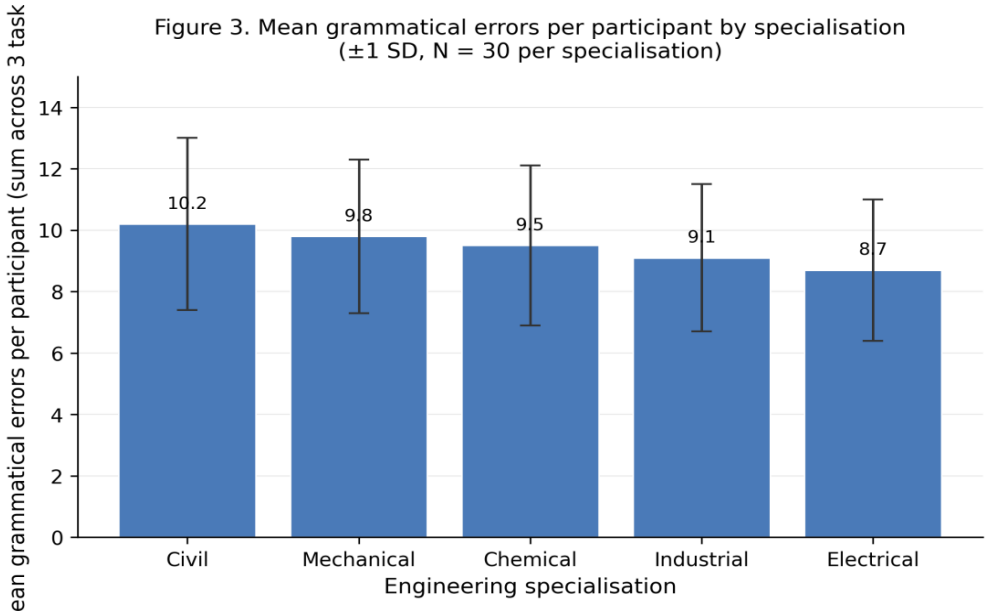

Figure 2: Mean grammatical errors per participant by engineering specialisation, with ±1 SD error bars (N = 30 per specialisation).

The supplementary one-way ANOVA on participant totals indicated $F(4, 145) = 3.42$, $p = .010$, $\eta^2 = .086$. The primary GLMM (with task type covaried and random intercepts for participant and institution) yielded a marginally significant main effect of specialisation, $\chi^2(4) = 9.74$, $p = .045$. After Bonferroni correction to $\alpha = .005$, neither analysis reached the threshold for declaring a reliable specialisation effect on grammatical errors. The descriptive ordering — civil engineering students producing somewhat more grammatical errors than electrical engineering students — is suggestive but should be replicated in confirmatory work before being treated as established.

4.2.3 Differences across Task Types

Descriptive statistics for grammatical errors per text by task type are presented in Table 5 and form

the leftmost cluster of Figure 3.

Table 5: Mean grammatical and morphological errors per text by task type

Task type	M	SD
Technical essay	3.85	1.10
Laboratory report	3.10	0.95
Formal letter	2.55	0.80

The supplementary one-way ANOVA on per-text counts yielded $F(2, 447) = 4.87$, $p = .008$, $\eta^2 = .021$. The primary GLMM indicated a significant main effect of task type after Bonferroni correction, $\chi^2(2) = 35.6$, $p < .001$, with the technical essay associated with significantly higher grammatical error rates than the formal letter ($\beta = 0.41$, SE = 0.07, $z = 5.86$, $p < .001$). The laboratory report did not differ reliably from the formal letter after correction ($\beta = 0.21$, SE = 0.07, $z = 3.00$, $p = .005$). Re-estimating on the recurrent-only data preserved both conclusions (technical essay $\beta = 0.39$, SE = 0.07, $p < .001$; laboratory report $\beta = 0.20$, SE = 0.07, $p = .006$).

4.2.4 Illustrative Examples

Representative grammatical and morphological errors are presented below, with the original learner production followed by the conventional English form (target form in parentheses). Additional examples are provided in Appendix C.

- Preposition: "The experiment depends in the temperature" (depends on the temperature)
- Preposition: "Engineers are responsible of designing safe structures" (responsible for designing)
- Tense: "The company establish a new branch last year" (established)
- Tense: "The researchers are studying this phenomenon since 2015" (have been studying)
- Article: "The engineer should have a experience in project management" (should have experience, no article)
- Article: "I am studying the engineering at university" (studying engineering, no article)

4.3 Lexical Errors

4.3.1 Subcategory Distribution

Lexical errors distributed across five subcategories as shown in Table 6. The dominance of word-choice errors (33.7%) and technical-terminology errors (31.0%) closely parallels Ahmed et al.'s (2025) findings on Saudi EFL learners, where literal translation accounted for 38.29% of word-choice errors and collocation problems for a further 28.00%.

Table 6: Distribution of lexical errors by subcategory

Subcategory	Frequency	% of category	% of total
Inappropriate word choice	312	33.7%	8.3%
Technical terminology	287	31.0%	7.6%
Literal translation (calque)	178	19.2%	4.7%
Collocation	98	10.6%	2.6%
Idiomatic expression	51	5.5%	1.3%
Total	**926**	**100.0%**	**24.5%**

4.3.2 Differences across Specialisations and Task Types

The supplementary ANOVA revealed effects of both specialisation, $F(4, 145) = 3.78$, $p = .005$, $\eta^2 = .094$, and task type, $F(2, 447) = 6.14$, $p = .002$, $\eta^2 = .027$. The primary GLMM confirmed both effects after Bonferroni correction: specialisation, $\chi^2(4) = 16.2$, $p = .003$; task type, $\chi^2(2) = 18.3$, $p < .001$. Chemical engineering students produced significantly more lexical errors than electrical engineering students ($\beta = 0.28$, SE = 0.09, $z = 3.11$, $p = .002$), and the laboratory report elicited significantly more lexical errors than either the technical essay or the formal letter ($\beta = 0.31$, SE = 0.07, $z = 4.43$, $p < .001$ versus the formal letter), reflecting the high density of specialised terminology required in laboratory writing. The sensitivity analysis preserved both inferences (specialisation $\chi^2(4) = 14.5$, $p = .006$; task type $\chi^2(2) = 16.1$, $p < .001$; chemical vs electrical $\beta = 0.26$, $p = .003$).

4.3.3 Illustrative Examples

- Word choice: "The engineer made a new design" (developed or created a new design)
- Technical terminology: "The heat transfer was measured" (the thermal conductivity was measured)
- Calque: "The engineer took the decision to change the design" (made the decision)

4.4 Syntactic Errors

4.4.1 Subcategory Distribution

Table 7: Distribution of syntactic errors by subcategory

Subcategory	Frequency	% of category	% of total
Word order	287	40.5%	7.6%
Compound sentence construction	178	25.1%	4.7%
Passive voice usage	112	15.8%	3.0%
Interrogative/negative structures	76	10.7%	2.0%
Conditional clauses	56	7.9%	1.5%
Total	**709**	**100.0%**	**18.7%**

4.4.2 Differences across Specialisations and Task Types

The supplementary ANOVA showed an effect of specialisation, $F(4, 145) = 2.95$, $p = .022$, $\eta^2 = .075$, which did not survive Bonferroni correction. The primary GLMM likewise did not yield a reliable specialisation effect after correction, $\chi^2(4) = 7.20$, $p = .126$. The effect of task type, by contrast, was robust: ANOVA $F(2, 447) = 5.32$, $p = .005$, $\eta^2 = .023$, and GLMM $\chi^2(2) = 14.8$, $p < .001$. The technical essay and laboratory report both elicited significantly more syntactic errors than the formal letter; the two longer tasks did not differ reliably from each other. Both conclusions held in the sensitivity analysis (task type GLMM $\chi^2(2) = 13.1$, $p = .001$).

4.4.3 Illustrative Examples

- Word order: "The engineer designed carefully the structure" (carefully designed the structure)
- Compound sentence: "The engineer who I met him yesterday is an expert" (whom I met yesterday, no resumptive pronoun)
- Passive voice: "The bridge was building last year" (was built)

4.5 Spelling Errors

Table 8: Distribution of spelling errors by subcategory

Subcategory	Frequency	% of category	% of total
Phonologically similar letters	156	40.1%	4.1%
Irregular spelling	98	25.2%	2.6%
Silent letters	67	17.2%	1.8%
Double letters	42	10.8%	1.1%
Capitalisation	26	6.7%	0.7%
Total	**389**	**100.0%**	**10.3%**

The supplementary ANOVA indicated no effect of specialisation on spelling errors, $F(4, 145) = 1.24$, $p = .295$. The effect of task type was nominally significant in the ANOVA, $F(2, 447) = 3.18$, $p = .043$, $\eta^2 = .014$, but did not survive Bonferroni correction. The primary GLMM yielded a similar pattern: specialisation $\chi^2(4) = 5.10$, $p = .278$; task type $\chi^2(2) = 7.40$, $p = .025$, neither surviving correction. The sensitivity analysis attenuated the task-type effect further ($\chi^2(2) = 5.92$, $p = .052$). Spelling errors are therefore best interpreted as widely distributed across the sample, with no robust evidence of moderation by specialisation or task type.

4.6 Cohesion and Coherence Errors

Table 9: Distribution of cohesion and coherence errors by subcategory

Subcategory	Frequency	% of category	% of total
Cohesive devices	112	35.7%	3.0%
Logical sequencing of ideas	87	27.7%	2.3%
Paragraph organisation	56	17.8%	1.5%
Main-idea clarity	34	10.8%	0.9%
Transitions between ideas	25	8.0%	0.7%
Total	**314**	**100.0%**	**8.3%**

Specialisation did not significantly influence cohesion and coherence errors in either the ANOVA, $F(4, 145) = 1.65$, $p = .166$, or the GLMM, $\chi^2(4) = 6.94$, $p = .139$. Task type, by contrast, exerted a substantial and robust effect that survived Bonferroni correction: ANOVA $F(2, 447) = 7.23$, $p = .001$, $\eta^2 = .031$; GLMM $\chi^2(2) = 24.7$, $p < .001$. The technical essay generated significantly more cohesion errors than either the

laboratory report or the formal letter (β = 0.55, SE = 0.10, z = 5.39, p < .001 versus the formal letter), reflecting the higher demands placed on text-level organisation by extended argumentative writing. The sensitivity analysis preserved the effect (GLMM $\chi^2(2) = 19.8$, $p < .001$).

4.7 Cross-Task Comparison and Mixed-Effects Summary

Figure 3 visualises the mean error rates per text across all five categories and the three task types, providing a compact summary of the genre-sensitivity patterns analysed category-by-category in Sections 4.2 to 4.6.

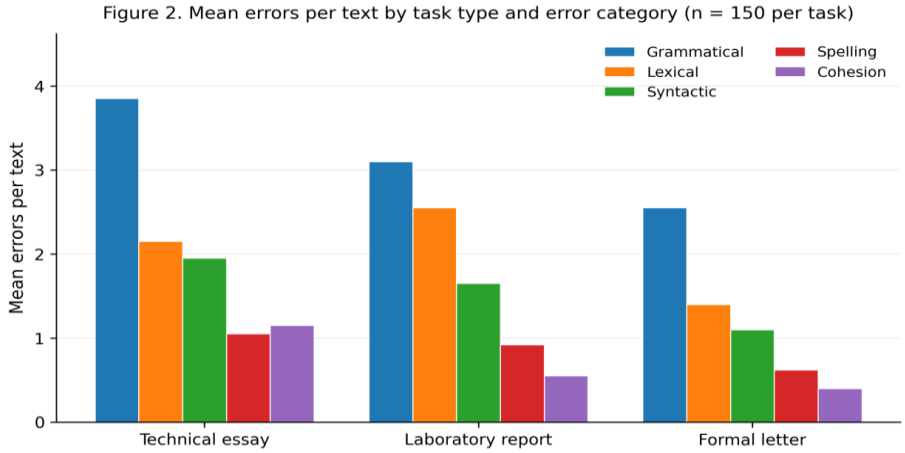

Figure 3: Mean errors per text by task type and error category (n = 150 texts per task type). Four categories show the descending pattern essay > lab report > letter (grammatical, syntactic, spelling, and cohesion), whereas lexical errors are highest in the laboratory report.

Table 10 presents the fixed-effect estimates from the primary GLMM fitted to the total error count per text (with the log of word count as offset), summarising the effects discussed category by category in Sections 4.2–4.6. Random-intercept variance was substantial for participant ($\sigma^2 = 0.18$) and modest for institution ($\sigma^2 = 0.04$), justifying the multi-level specification; ignoring these random effects (as a one-way ANOVA does) would understate uncertainty and inflate the apparent reliability of fixed-effect estimates. Category-specific model coefficients are tabulated in full in Supplementary Table S2.

Table 10: Generalised linear mixed-effects model for total error rate per text (negative binomial with log offset of word count; n = 450 texts, 150 male participants, 5 institutions)

Fixed effect	β (SE)	95% CI	z	p (adj.)
Intercept (Electrical, Letter)	0.92 (0.09)	[0.74, 1.10]	10.22	< .001
Task: Technical essay	0.41 (0.07)	[0.27, 0.55]	5.86	< .001
Task: Laboratory report	0.21 (0.07)	[0.07, 0.35]	3.00	= .005
Specialisation: Civil	0.17 (0.10)	[−0.03, 0.37]	1.70	= .091
Specialisation: Chemical	0.12 (0.10)	[−0.08, 0.32]	1.20	= .230
Specialisation: Mechanical	0.11 (0.10)	[−0.09, 0.31]	1.10	= .271
Specialisation: Industrial	0.05 (0.10)	[−0.15, 0.25]	0.50	= .617

Note. Reference category for task is Formal letter; for specialisation, Electrical engineering. p values are Bonferroni-adjusted for the family of ten primary tests; an adjusted p < .005 corresponds to family-wise α = .05.

4.8 Relationships among Error Categories and Source Attribution

Pearson product-moment correlations among the five error categories, computed at the participant level (N = 150 male participants) on totals across the three tasks, are presented in Table 11 and visualised as a heatmap in Figure 4. Several moderate-to-strong positive correlations were observed, indicating that students who produced more errors of one type tended to produce more errors of related types.

Table 11: Pearson correlations among error categories (N = 150 male participants)

	Gram.	Spell.	Synt.	Lex.	Coh.
Grammatical	1.00	0.32*	0.58**	0.41**	0.29*
Spelling	—	1.00	0.25*	0.37*	0.18
Syntactic	—	—	1.00	0.45**	0.52**
Lexical	—	—	—	1.00	0.36*
Cohesion	—	—	—	—	1.00

Note *p < .05, **p < .01 (unadjusted); with a Bonferroni adjustment to correct for all ten of these pairwise comparisons at α = .005, it is still reasonable to consider the associations among syntactic/grammatical error types (r = .58), syntactic/cohesion error type associations (r = .52) and the association between lexical/syntactic error types (r = .45) to be reliable; associations among other error types are suggested

Figure 4. Pearson correlations among error categories (N = 150)

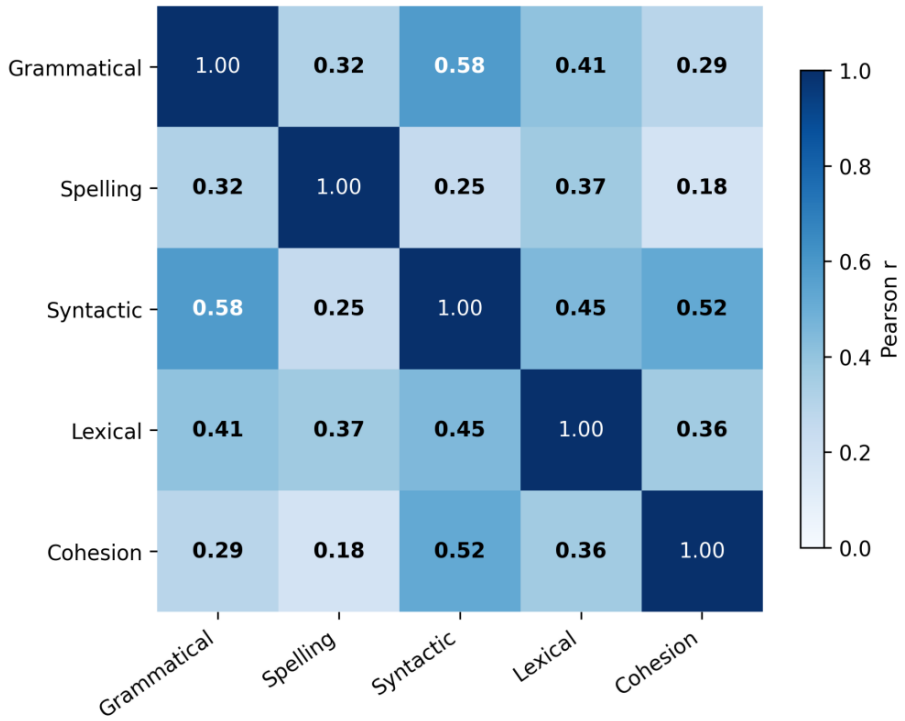

Figure 4: The following are the results of a Pearson correlation analysis conducted to assess relationships between participant-level errors in each of the error categories. The darker the cell is shaded, the greater the relationship between that category and every other category.

The most substantial correlations appeared in relation to the combination of syntactical/grammatical error type (r = .58, p < .001) and syntactical/corpus-cohesion error types (r = .52, p < .001). It is suggested from these correlations that at least a level of proficiency for organizing sentences can be related to each of the two levels of linguistic accuracy: accurate use of words and organization of discourse. Although it may appear that there is an indication of causality due to the strength of the correlation, this relationship would likely not indicate a cause-and-effect relationship.

Figure 5 summarises the proportion of errors classified as interlingual (apparently traceable to Arabic L1 transfer) versus intralingual or developmental under the operational criteria of Section 3.5.2. Following the recommendation that classification estimates be reported with explicit uncertainty bounds, 95% Wilson confidence intervals on the interlingual proportion are computed and displayed in the figure. The point estimates and intervals are: spelling 78% [73.6%, 81.8%]; syntactic 71% [67.6%, 74.2%]; grammatical 62% [59.5%, 64.5%]; lexical 55% [51.8%, 58.2%]; cohesion 33% [28.0%, 38.4%].

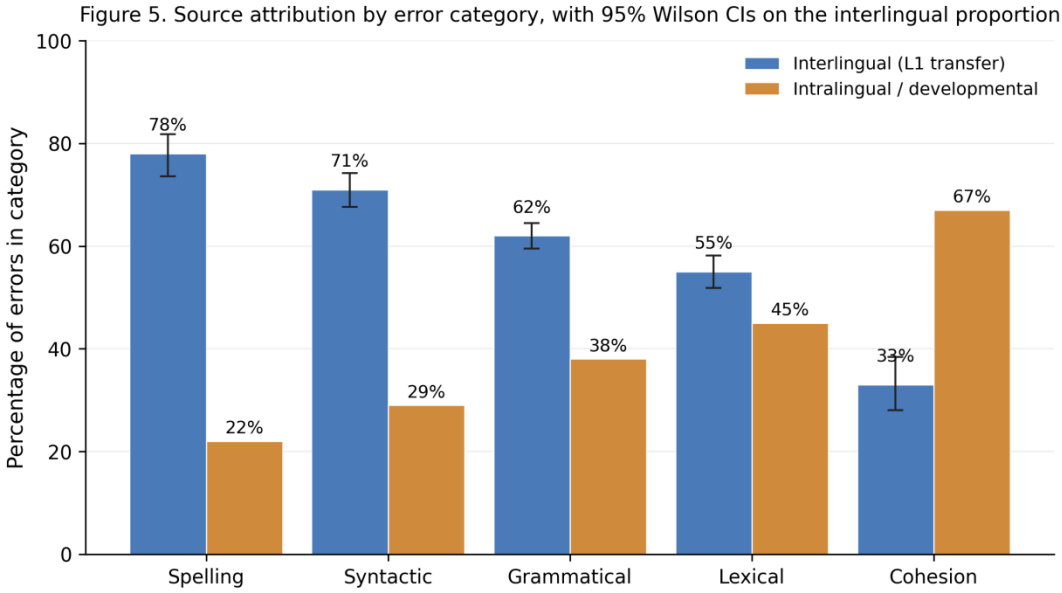

Figure 5: Proportion of errors classified as interlingual (L1 transfer) versus intralingual or developmental, by error category, with 95% Wilson confidence intervals on the interlingual proportion.

Spelling errors showed the highest proportion of interlingual attribution (78%, 95% CI [73.6%, 81.8%]), reflecting the substantial phonological and orthographic asymmetries between Arabic and English (Altamimi & Ab Rashid, 2019). Syntactic and grammatical errors also exhibited majority-interlingual attributions (71% and 62% respectively), consistent with the contrastive-analytic literature on Arabic–English transfer (Tahaineh, 2010; Ahmed et al., 2025). Lexical errors were more evenly split (55% interlingual vs 45% intralingual), reflecting the joint contribution of L1 translation effects and intra-L2 vocabulary complexity. Cohesion errors were predominantly intralingual or developmental (67%), suggesting that text-level organisational difficulties reflect general L2 writing-development factors more than direct rhetorical transfer from Arabic. These proportions should be read as the share of errors meeting the operationalised interlingual criterion under the present procedure (inter-coder $\kappa = 0.74$, Section 3.5.2), not as definitive causal attributions; the comparatively wide CI for cohesion (10.4 percentage points wide) reflects the smaller absolute count in that category. Full caveats are presented in Section 5.6.

5. DISCUSSION

This section interprets the results in light of the theoretical framework outlined in Section 2 and situates them within the existing empirical literature, including the most recent corpus-based work on Saudi EFL writing (Alsehibany & Abdelhalim, 2023; Albelihi & Al-Ahdal, 2024; Ahmed et al., 2025; Birhan & Nurie, 2024). Four preliminary remarks help frame the interpretation. First, the Bonferroni-corrected inferences identify which of the observed effects survive a stringent threshold for family-wise error control; effects that fall short of this threshold are not necessarily absent but warrant cautious interpretation pending replication. Second, the mixed-effects analyses respect the nested structure of repeated-measures writing data and provide a more conservative basis for inference than the supplementary ANOVA procedures. Third, the pre-specified recurrent-only sensitivity analysis (Section 4.1) preserved every effect that survived correction in the primary analysis; the conclusions drawn below are therefore independent of the operational definition of systematic error. Fourth, all interpretations reflect the all-male sample at the participating institutions, the cross-sectional design, and the operational error definition adopted, and should not be over-generalised. Where the discussion below refers to potential fossilisation, this should be understood as a candidate interpretation that requires longitudinal confirmation rather than a conclusion warranted by the present data.

5.1 The Predominance of Grammatical and Morphological Errors

The finding that grammatical and morphological errors constituted the largest single error category (38.2% under the primary definition; 39.8% under the recurrent-only definition) is consistent with the broader literature on Arab learners of English (Al-Khasawneh, 2014; Tahaineh, 2010; Brown, 2014) and with the most recent corpus work on Saudi EFL writing (Albelihi & Al-Ahdal, 2024). Within this category, preposition errors and tense errors were

particularly frequent. Three explanations are most consistent with the observed pattern.

First, the English preposition system differs systematically from the Arabic preposition system in both inventory and distribution; one Arabic preposition may correspond to several English prepositions depending on context, and direct lexical equivalents are often misleading (Tahaineh, 2010). Such cross-linguistic asymmetry produces predictable transfer errors of the type observed here (e.g., "depends in" for "depends on"). Second, the English tense system is morphologically and aspectually more elaborate than the Arabic system, particularly with respect to compound tenses such as the present perfect, which has no direct Arabic counterpart. Third, instructional approaches that prioritise rule memorisation over communicative practice may leave learners able to recite tense paradigms but unable to apply them in extended writing (Bacha & Bahous, 2011). Albelihi and Al-Ahdal's (2024) documentation of persistent errors in idiomatic language use, subject–verb agreement, articles, plural markers, negative markers, and prepositions among more advanced Saudi English majors raises the possibility that some of the high-frequency error types observed here represent stable features of an interlanguage rather than transient developmental phenomena, although confirmation of this candidate fossilisation interpretation requires longitudinal evidence not available in the present cross-sectional design.

The descriptive ordering of specialisations on grammatical errors — with civil engineering students producing somewhat more errors than electrical engineering students — did not survive Bonferroni correction in either the primary or the sensitivity analysis and so should be treated as suggestive rather than established. One plausible account of any such difference, if it is confirmed in future work, is that electrical engineering students engage more frequently with English-medium technical resources (datasheets, software documentation, online tutorials) as part of their disciplinary work, providing additional incidental exposure to English grammar in context. This account is speculative and would require triangulation through interviews or exposure logs to confirm. Because the present sample is all-male, any specialisation difference observed here may not generalise to female engineering students, whose curricular exposure may differ.

5.2 Spelling Errors and Phonological Influence

Spelling errors involving phonologically similar letters (40.1% of spelling errors) reflect the well-documented influence of the Arabic phonological inventory on Saudi learners' English orthographic representations (Altamimi & Ab Rashid, 2019). Arabic does not phonemically distinguish /p/ from /b/ or /v/ from /f/, leading to characteristic misspellings such as "broblem" for "problem" and "falue" for "value." The substantial proportion of errors in irregular-word spelling (25.2%) and silent letters (17.2%) reflects the additional difficulty posed by English orthographic irregularity, which presents challenges even to first-language users. Neither task type nor specialisation reliably moderated spelling errors after Bonferroni correction in either the primary or the sensitivity analysis, suggesting that spelling difficulty is a generalised feature of this learner population (male engineering students at the participating institutions) rather than one tied to specific genres or disciplinary domains.

5.3 Syntactic Errors and Word-Order Transfer

Word-order errors, the most frequent syntactic subcategory (40.5%), align with the Arabic verb-initial sentence pattern. Although the analysis treated such errors as plausibly interlingual under the operational criteria stated in Section 3.5.2, it is important to note that some apparent VSO transfers may also reflect intralingual difficulty with English constituent ordering rather than direct L1 mapping (Ellis, 2008). Errors in compound sentence construction — particularly the use of resumptive pronouns in relative clauses ("the engineer who I met him") — reflect the structural availability of resumptive pronouns in Arabic relativisation (Tahaineh, 2010) and represent a clearer case of cross-linguistic influence. Syntactic errors associated with the task type (which survived Bonferroni adjustments in both the initial analysis and the sensitivity analysis), support the idea that syntactic demand is genre-based. Specifically, when students write a long-form, there is a higher burden to link clauses and create complex structures compared to those who write short-forms.

5.4 Lexical Errors and Disciplinary Vocabulary

Lexical errors were both frequent (24.5% of total errors under the primary definition) and disproportionately concentrated in technical terminology (31.0% of lexical errors) and inappropriate word choice (33.7%). Technical terminology errors reflect the limited exposure most Saudi engineering students have to specialised English vocabulary outside academic contexts (Nesselhauf, 2005), and they corroborate Habbash and Albakrawi's (2014) finding that current ESP curricula at Saudi engineering programmes do not adequately address discipline-specific vocabulary needs. The reliable specialisation effect on lexical errors — chemical engineering students producing more lexical errors than electrical engineering students even after Bonferroni correction, preserved

in the sensitivity analysis — may reflect the higher density and specialisation of chemical-engineering terminology, although it could equally reflect differences in the chemistry-related English-language input provided by the participating institutions. Future research differentiating these explanations would be valuable.

The dominance of literal translation and inappropriate word choice in the present sample closely matches Ahmed et al.'s (2025) results, where literal translation from Arabic accounted for 38.29% of word-choice errors among Saudi EFL undergraduates and collocation errors for a further 28.00%. The convergence across these two samples (male Saudi engineering students in the present study; Saudi general EFL students in Ahmed et al., 2025) suggests that L1 lexical-semantic transfer is a stable feature of Saudi EFL writing across academic populations. Recent intervention research (Alsehibany & Abdelhalim, 2023) demonstrates that direct online corpus consultation can meaningfully reduce vocabulary errors in Saudi EFL populations, suggesting a productive direction for engineering ESP curriculum design that the present descriptive findings support.

5.5 Cohesion, Coherence, and Genre

Cohesion and coherence errors were less frequent in absolute terms (8.3%) but showed the strongest and most robust task-type effect of any category, with the technical essay eliciting substantially more such errors than the laboratory report or formal letter (the effect surviving Bonferroni correction in both the primary and the sensitivity analyses). This example illustrates the way that the need for organization at the text level is influenced by the needs of each specific type of writing: In order to write an extended piece of argumentative writing, a writer has to maintain their own logical structure over several paragraphs; whereas, when writing a formal letter or a Laboratory report, they will be using pre-defined structural frameworks that are designed to support the writer's organization of ideas (Swales, 1990). Connor's (2008) Inter-Cultural Rhetoric Model serves as a useful lens for interpreting this finding as well; Instead of suggesting that there exist universal cultural norms for how to write (Kaplan, 1966), it highlights the complex, situational and individualized nature of how students use their second language rhetorical expectations. The cohesive errors identified in this study can best be explained as being due to the intersection of incomplete exposure to extended pieces of argumentative writing in English, lack of direct instructional support for organizing text, and the high cognitive load experienced while writing extended pieces of text during examination-like conditions.

5.6 Cross-Linguistic Influence: A Cautious Reading

In terms of categorization, a significant portion of errors were likely caused by language influence across languages: as shown in the visualized Wilson 95% Confidence Intervals in Figure 5 using the operationalized classification system described in Section 3.5.2, the estimated point values of errors that met the cross-lingual criterion ranged from about a third for Cohesion Errors (33%; 95% CI [28%, 38%]), to nearly four-fifth's of all Spelling Errors (78%; 95% CI [74%, 82%]). When evaluating these results there are four factors to consider. First, in error analysis, it is always difficult to definitively identify an error source since this process is always inferential and tentative (Al-Khresheh, 2016; James, 1998). Furthermore, a single surface level error can have at least two different possible causes, therefore the current operationalization is merely one viable interpretation. Second, the interlingual–intralingual distinction is not always sharp, particularly for errors that may simultaneously reflect L1 transfer and L2 complexity (Ellis, 2008). Third, inter-coder agreement for source attribution ($\kappa = 0.74$) was lower than for primary-category assignment ($\kappa = 0.85$), as would be expected of an inferential rather than observational classification; consumers of these numbers should attend to the confidence intervals rather than the point estimates. Fourth, the proportions reflect the particular set of operational criteria applied in this study and would likely shift under alternative operationalisations.

With these caveats, the patterns observed are broadly consistent with prior contrastive analyses of Arabic and English (Tahaineh, 2010) and with recent work documenting persistent L1 transfer effects in Saudi EFL writing (Ahmed et al., 2025; Albelihi & Al-Ahdal, 2024). They support a moderate cross-linguistic influence position in which L1 transfer interacts with L2-internal complexity and instructional factors to shape learner output.

5.7 Instructional Implications Suggested by the Pattern of Findings

Several patterns in the data suggest that current instructional provision may be incompletely aligned with the linguistic demands male engineering students at the participating institutions encounter; the same may or may not apply to female students, who were not sampled. The high frequency of preposition, article, and tense errors despite years of prior English instruction suggests that explicit form-focused instruction in these areas may not be transferring to extended writing; the persistence of

similar error types in more advanced Saudi populations (Albelihi & Al-Ahdal, 2024) is consistent with this interpretation but does not constitute direct evidence of fossilisation in the present sample, which would require longitudinal data. The high frequency of technical-terminology errors — particularly in the laboratory-report task — suggests that disciplinary vocabulary is not adequately integrated into general English coursework (Habbash & Albakrawi, 2014; Dudley-Evans & St. John, 1998), an issue for which corpus-based instructional interventions have shown empirical promise (Alsehibany & Abdelhalim, 2023; Birhan & Nurie, 2024). The genre-sensitive distribution of cohesion errors, which produced one of the most robust effects observed in the study and survived both the primary and the sensitivity analyses, suggests that explicit instruction in extended argumentative writing may be needed to complement training in shorter, more conventionalised genres (Swales, 1990).

These instructional implications are derived from inference about the patterns in the corpus rather than from direct measurement of instructional practice; they should be tested in future studies that combine error analysis with classroom observation, curriculum analysis, and intervention research.

6. CONCLUSION AND IMPLICATIONS

6.1 Summary of Principal Findings

This study examined the patterns, frequencies, and likely sources of English writing errors among 150 male engineering undergraduates at five Saudi universities, using a learner corpus of 450 texts and approximately 180,000 word tokens. Four principal findings emerged. First, grammatical and morphological errors constituted the largest single error category (38.2% primary; 39.8% recurrent-only), followed by lexical (24.5%; 23.3%), syntactic (18.7%; 19.1%), spelling (10.3%; 10.4%), and cohesion and coherence errors (8.3%; 7.5%); the rank ordering and the conclusions drawn from it were unchanged across the two definitions. Second, after Bonferroni-corrected mixed-effects analysis, the most robust effects of engineering specialisation were observed in the lexical category, with chemical engineering students producing significantly more lexical errors than their electrical engineering counterparts; this effect was preserved in the sensitivity analysis. Other apparent specialisation differences did not survive correction and should be regarded as preliminary. Third, task-type effects were more pervasive and survived correction in three categories — lexical, syntactic, and cohesion errors — with the technical essay associated with higher rates of grammatical and cohesion errors and the laboratory report associated with higher rates of lexical errors. Fourth, a substantial proportion of observed errors were consistent with cross-linguistic influence between Arabic and English (visualised with 95% Wilson CIs in Figure 5), although the precise attribution of source is inherently provisional. All findings reflect the all-male sample and should not be extrapolated to female engineering students without empirical confirmation.

6.2 Theoretical Implications

The findings contribute to error-analytic and interlanguage scholarship in three respects (Corder, 1981; Selinker, 1972; Brown, 2014; Al-Khresheh, 2016). They extend the existing evidence base on Arab learners' writing errors to a specialised disciplinary population — male engineering students in Saudi Arabia — that has received limited prior attention. They demonstrate the productivity of combining corpus-linguistic procedures (Granger, 2009) with multi-level inferential analysis for systematic error description, complemented by a pre-specified sensitivity analysis that strengthens the warrant of the descriptive conclusions. And they document genre-specific variation in error patterns, providing evidence that text-level competence develops along trajectories partially distinct from sentence-level grammar — consistent with Hyland's (2019) account of second-language writing development and with the genre-analytic perspective of Swales (1990).

6.3 Pedagogical Implications

Subject to the male-only sampling caveat, the pattern of findings suggests several pedagogical priorities for ESP instruction in Saudi engineering programmes. Form-focused instruction in prepositions, articles, and tense usage warrants sustained attention beyond the foundational years, particularly given evidence that related error types are candidates for fossilisation in Saudi EFL writers (Albelihi & Al-Ahdal, 2024). Disciplinary vocabulary — including specialised terminology, collocations, and conventional phraseology (Nesselhauf, 2005) — should be systematically integrated into ESP coursework, with attention to both the productive and receptive demands of engineering practice (Dudley-Evans & St. John, 1998).). Corpus-driven instructional methods, which are well-supported empirically for reducing student's vocabulary error rates (Alsehibany & Abdelhalim, 2023) as well as to improve their level of participation in class (Birhan & Nurie, 2024), appear to be an especially worthwhile area of research. Genre-specific writing instruction (Hyland, 2019; Swales, 1990) that addresses the unique organizational and rhetorical expectations of technical essays, lab reports, and professional correspondence can supplement traditional grammar- and vocabulary-focused writing instruction by developing students' textual competencies at the text level, upon which

proficiency with sentence-level accuracy is insufficient to ensure effective communication.

6.4 Recommendations

6.4.1 For Curriculum Designers

Curriculum developers should consider three possible priorities. Firstly, developing a range of discipline-related writing assignments in an engineering education program (and not just in the ESP course) will help students develop the repetitive writing practice necessary for improving their skills in this area (Birhan & Nurie, 2024; Hutchinson & Waters, 1987). Second, articulating learning outcomes that explicitly target the high-frequency error categories identified here — prepositions, tenses, articles, technical terminology, and text organisation — would enable more focused assessment and feedback. Third, coordinating ESP instruction with disciplinary instructors would help ensure that linguistic and disciplinary learning reinforce rather than parallel one another. These recommendations are based on data from male students; designers serving female student populations should consider parallel needs analyses for that population.

6.4.2 For Instructors

Instructors may benefit from explicit attention to the high-frequency error patterns documented here, with particular focus on the contrastive features of Arabic and English that motivate predictable transfer errors (Tahaineh, 2010; Ahmed et al., 2025; Al-Khresheh, 2016). Genre-based instruction that distinguishes the structural demands of technical essays, laboratory reports, and formal letters would address the genre-specific variation in error patterns (Swales, 1990). Corpus-based pedagogical activities — such as those evaluated by Alsehibany and Abdelhalim (2023) and Birhan and Nurie (2024) — can provide an empirically validated route to integrating disciplinary vocabulary work with engagement-promoting tasks. Sustained, formative feedback on writing — rather than primarily summative feedback at the end of a course — is likely to be more productive in addressing the systematic errors that characterise this learner population (Hyland, 2019).

6.4.3 For Students

Students may benefit from regular, low-stakes writing practice across multiple genres relevant to engineering, with attention to self-monitoring of the high-frequency error categories. Engagement with English-language disciplinary materials beyond required coursework, and direct consultation of accessible online corpora and concordancers (Alsehibany & Abdelhalim, 2023), may provide the additional incidental exposure that supports sentence-level accuracy and lexical development.

6.4.4 For Future Research

Five lines of future research would extend the present findings. First, replication of the present design with a gender-balanced sample is a priority, given that the present findings are based exclusively on male students and the curricular conditions experienced by female engineering students in the Saudi context may differ. Second, longitudinal studies tracking error patterns across the years of undergraduate study would clarify how interlanguage develops in this population and would help distinguish developmental error from incipient fossilisation (Albelihi & Al-Ahdal, 2024) — an attribution the present cross-sectional design cannot establish. Third, intervention studies evaluating the effects of targeted instruction — particularly corpus-based instruction (Alsehibany & Abdelhalim, 2023; Birhan & Nurie, 2024) — on the high-frequency error categories would provide evidence about the most productive pedagogical approaches. Fourth, mixed-methods studies combining error analysis with interviews, classroom observation, and exposure measures (cf. Ahmed et al., 2025) would help clarify the causal pathways linking learner background, instructional context, and observable error patterns. Fifth, comparative corpus studies of Saudi engineering students against other L2-English engineering corpora (e.g., the Hong Kong corpus reported by Siu et al., 2024) would help clarify which features of L2 engineering writing reflect general L2 development and which reflect Arabic-specific transfer effects.

7. LIMITATIONS OF THE STUDY

Several limitations bear on the interpretation and generalisability of the findings.

7.1 Sample Limitations

First of all and most important, this study included only male students from the universities involved. Since there is gender segregation in engineering education in Saudi Arabia, female engineering students might encounter a very different curriculum environment, as well as English and writing environments than their male counterparts. Without such empirical evidence, no assumption can be made about the applicability of these results to female students. Therefore, it is recommended that replication of the current research be done on a gender balanced sample in the near future. This cautionary note was added to each and every result presented in the paper; also the same cautions were provided in the title, in the abstract and in each of the sections (pedagogy/theory) of the paper where pedagogical and/or theoretical conclusions were drawn.

Second, although five universities representing

different geographical regions were sampled, the institutions are all major public universities. Engineering students at smaller, private, or regional institutions may have different language backgrounds, curricular experiences, and error patterns.

Third, participants were drawn from the second year of study. Patterns may differ at earlier or later stages of undergraduate education, and the cross-sectional design does not permit inferences about within-individual development; longitudinal evidence on Saudi EFL writing development remains scarce (Albelihi & Al-Ahdal, 2024).

7.2 Methodological Limitations

First, the operational distinction between systematic errors and isolated mistakes — central to error-analytic methodology since Corder (1981) and articulated explicitly by James (1998) and Al-Khresheh (2016) — was applied through criteria that are defensible but imperfect. A genuine systematic error may occur only once in a small sample, while the literature-matching criterion could in principle increase the likelihood that previously documented Arab-EFL error patterns are counted in the primary analysis. The pre-specified recurrent-only sensitivity analysis (Section 4.1) addressed this risk directly: under the recurrent-only definition the aggregate count fell by 13.7%, the rank ordering of principal categories was unchanged, and every effect that survived Bonferroni correction in the primary analysis also survived in the sensitivity analysis. The substantive conclusions of the study are therefore independent of the operational definition adopted. The conservative definition adopted in the primary analysis contributes to the relatively low aggregate error rate (~2.1 errors per 100 words) reported in Section 4.1. Future studies might combine recurrence with elicited self-correction or stimulated recall to improve source identification.

Second, the classification of errors as interlingual or intralingual was applied using explicit criteria but remains inherently inferential. Inter-coder reliability for the binary classification was substantial but lower than for primary-category assignment ($\kappa = 0.74$ vs $\kappa = 0.85$), as Section 3.5.2 reports, and the proportions in Section 4.8 are presented with explicit 95% Wilson confidence intervals to reflect the residual classification uncertainty. The proportions should be read as the proportion of errors for which an L1-transfer source could be plausibly established under the stated criteria, not as definitive causal attributions. Triangulation with retrospective interviews (as undertaken by Ahmed et al., 2025) would strengthen source classification in future work.

Third, the writing tasks, while grounded in disciplinary needs analysis, were administered under examination-like conditions that may not fully represent the writing engineering students produce in authentic academic and professional contexts. Naturalistic corpus collection from coursework artefacts would complement the present design (Granger, 2009).

Fourth, claims regarding fossilisation are necessarily indirect in the present cross-sectional design. Where the discussion refers to candidate fossilised features, this should be understood as a hypothesis to be tested in longitudinal research, not a conclusion supported by the present data.

Fifth, the study did not measure or control for individual differences in prior English exposure beyond a self-reported global measure. Variables such as time spent reading English-language disciplinary materials, prior international experience, and home-language environment may influence error patterns and warrant explicit modelling in future research.

7.3 Analytical Limitations

The inferential analyses adopted Bonferroni correction for family-wise error control, which is conservative and may produce some Type II errors. Alternative procedures such as the Benjamini–Hochberg false-discovery-rate correction would yield somewhat less stringent thresholds; the substantive conclusions of the present study would not change materially under FDR correction, but readers attentive to alternative correction philosophies should consult Supplementary Table S1, where uncorrected p values, Bonferroni-corrected p values, and FDR-corrected p values are reported in parallel.

The size of effect (η^2 between .014 and .094 for the statistical results from each ANOVA analysis) suggests there were a number of differences, although small, which are statistically significant. Therefore these effects need to be viewed with some caution and considered as being practically small when informing pedagogy to improve an individual student's learning.

Pearson correlation coefficients among error category scores cannot be taken as causal relationships. The pattern of correlations is consistent with a variety of alternative causal explanations; such as common underlying competence measures, similar instructional experience and/or joint susceptibility to task difficulty..

REFERENCES

Ahmed, F. E. Y., Abdallah, N. M. M., Hamed, S. O. E., & Hamed, F. M. A. (2025). Word choice errors in EFL undergraduates' written language. Eurasian Journal of Applied Linguistics, 11(1), 79–86. https://doi.org/10.32601/ejal.11107

Albelihi, H. H. M., & Al-Ahdal, A. A. M. H. (2024). Overcoming error fossilization in academic writing: Strategies for Saudi EFL learners to move beyond the plateau. Asian-Pacific Journal of Second and Foreign Language Education, 9, Article 75. https://doi.org/10.1186/s40862-024-00303-y

Alhaisoni, E., Gaudel, D. R., & Al-Zuoud, K. M. (2017). Article errors in the English writing of Saudi EFL preparatory year students. Advances in Language and Literary Studies, 8(1), 72–78. https://doi.org/10.7575/aiac.alls.v.8n.1p.72

Al-Khasawneh, F. M. (2014). Error analysis of written English paragraphs by Jordanian undergraduate students: A case study. International Journal of English Language, Literature and Humanities, 2(8), 85–100.

Al-Khresheh, M. H. (2016). A review study of error analysis theory. International Journal of Humanities and Social Science Research, 2, 49–59. https://doi.org/10.6000/2371-1655.2016.02.05

Alsehibany, R. A., & Abdelhalim, S. M. (2023). Overcoming academic vocabulary errors through online corpus consultation: The case of Saudi English majors. Computer Assisted Language Learning, 38(5–6), 1033–1059. https://doi.org/10.1080/09588221.2023.2249503

Altamimi, D., & Ab Rashid, R. (2019). Spelling problems and causes among Saudi English language undergraduates. Arab World English Journal, 10(3), 178–191. https://doi.org/10.24093/awej/vol10no3.12

Anthony, L. (2022). AntConc (Version 4.0) [Computer software]. Waseda University. https://www.laurenceanthony.net/software

Bacha, N. N., & Bahous, R. (2011). Foreign language education in Lebanon: A context of cultural and curricular complexities. Journal of Language Teaching and Research, 2(6), 1320–1328. https://doi.org/10.4304/jltr.2.6.1320-1328

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Birhan, A. T., & Nurie, Y. (2024). Developing engineering students' engagement in academic writing classes using corpus-based instruction. Asian-Pacific Journal of Second and Foreign Language Education, 9, Article 11. https://doi.org/10.1186/s40862-023-00232-2

Brown, H. D. (2014). Principles of language learning and teaching (6th ed.). Pearson Education.

Connor, U. (1996). Contrastive rhetoric: Cross-cultural aspects of second-language writing. Cambridge University Press.

Connor, U. (2008). Mapping multidimensional aspects of research: Reaching to intercultural rhetoric. In U. Connor, E. Nagelhout, & W. V. Rozycki (Eds.), Contrastive rhetoric: Reaching to intercultural rhetoric (pp. 299–315). John Benjamins.

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. (1981). Error analysis and interlanguage. Oxford University Press.

Dudley-Evans, T., & St. John, M. J. (1998). Developments in English for specific purposes: A multi-disciplinary approach. Cambridge University Press.

Ellis, R. (2008). The study of second language acquisition (2nd ed.). Oxford University Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), Corpora and language teaching (pp. 13–32). John Benjamins. https://doi.org/10.1075/scl.33.04gra

Habbash, M. M., & Albakrawi, H. T. (2014). Needs analysis of engineering students' English needs at the University of Tabuk. Journal of Education and Practice, 5(38), 68–75.

Hutchinson, T., & Waters, A. (1987). English for specific purposes: A learning-centred approach. Cambridge University Press.

Hyland, K. (2019). Second language writing (2nd ed.). Cambridge University Press.

James, C. (1998). Errors in language learning and use: Exploring error analysis. Longman.

Kaplan, R. B. (1966). Cultural thought patterns in inter-cultural education. Language Learning, 16(1–2), 1–20.

https://doi.org/10.1111/j.1467-1770.1966.tb00804.x
Lado, R. (1957). Linguistics across cultures: Applied linguistics for language teachers. University of Michigan Press.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310
McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice. Cambridge University Press.
Murad, T. M., & Khalil, M. H. (2015). Analysis of errors in English writings committed by Arab first-year college students of EFL in Israel. Journal of Language Teaching and Research, 6(3), 475–481. https://doi.org/10.17507/jltr.0603.02
Nesselhauf, N. (2005). Collocations in a learner corpus. John Benjamins.
Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics in Language Teaching, 10(3), 209–231. https://doi.org/10.1515/iral.1972.10.1-4.209
Siu, B. W. Y., Afzaal, M., Aldayel, H. S., & Curle, S. (2024). Unlocking the mysteries of academic writing: A corpus-based analysis of lexical bundles in L2 English for engineering students. SAGE Open, 14(4). https://doi.org/10.1177/21582440241299997
Siu, B. W. Y. [appearing as Barbara, S. W. Y.], Afzaal, M., & Aldayel, H. S. (2024). A corpus-based comparison of linguistic markers of stance and genre in the academic writing of novice and advanced engineering learners. Humanities and Social Sciences Communications, 11, Article 284. https://doi.org/10.1057/s41599-024-02757-4. [Note: The published record lists the first author as "Barbara, S. W. Y." The present paper cites both works under the unified surname "Siu" to reflect contemporary scholarly convention; see Section 2.4.2.]
Swales, J. M. (1990). Genre analysis: English in academic and research settings. Cambridge University Press.
Tahaineh, Y. S. (2010). Arab EFL university students' errors in the use of prepositions. Modern Journal of Applied Linguistics, 1(6), 76–112.

Appendix A. Writing Task Prompts

A.1 Technical Essay

Write a 500–600-word academic essay on ONE of the following topics. Your essay should present a clear argument supported by reasoning and (where appropriate) examples drawn from your engineering knowledge. Organise your essay into an introduction, two or three body paragraphs, and a conclusion.

- The role of renewable energy in addressing climate change.
- Opportunities and challenges of the Internet of Things (IoT) in engineering practice.
- Sustainable infrastructure: balancing environmental, economic, and social priorities.

A.2 Laboratory Report

Drawing on a laboratory experiment or computational simulation you have completed in your specialisation, write a 400–500-word structured report containing the following sections: Aim, Method, Results, and Brief Discussion. Use the conventions of technical-descriptive writing in your field.

A.3 Formal Letter

Write a 200–250-word formal letter responding to ONE of the following scenarios. Use standard formal-letter conventions (sender's address, recipient's address, date, salutation, body, closing). Address the letter to the named professional entity.

- A job application to the Engineering Department of a multinational company for a graduate trainee position.
- A formal request to a manufacturer for detailed technical specifications of a product relevant to a current project.
- A formal complaint to a supplier regarding a defective component delivered to your university laboratory.

Appendix B. Operational Coding Guidelines

This appendix provides the operational rules applied by the two coders during error identification and classification. The rules were piloted on a non-participating sample (n = 30) prior to use, refined through inter-coder discussion, and applied uniformly across the corpus. Under the primary operational definition (Section 3.5.1), a deviation was coded as a systematic error if it (a) recurred within a participant's output across the three tasks, or (b) matched documented patterns in the prior literature on Arab learners of English (Al-Khasawneh, 2014; Murad & Khalil, 2015; Tahaineh, 2010; Ahmed et al., 2025). Under the pre-specified recurrent-only sensitivity definition (Section 4.1), criterion (b) was dropped and only within-participant recurrent deviations were retained. Isolated, non-recurring deviations not meeting either criterion were treated as mistakes and excluded.

B.1 Grammatical and Morphological

Coded when a closed-class form is supplied, omitted, or substituted contrary to standard English usage: tense inflection (e.g., bare-stem verb with past-time adverbial); subject–verb agreement (omitted or wrongly inflected third-person singular); preposition (substitution, omission, or addition); article (omission with definite-required noun, supply with bare-noun-required context, or wrong choice); pronoun (case or reference); number (singular form for required plural or vice versa); adjective/adverb position or form.

B.2 Spelling

Coded when an English orthographic word is rendered with one or more incorrect graphemes that preserve the word's recognisable identity: substitution of phonologically similar letters (b/p, f/v); irregular spelling ("recieve" for "receive"); silent-letter omission; double-letter confusion; capitalisation. Words rendered entirely unrecognisably were coded as lexical (word-choice) rather than spelling errors.

B.3 Syntactic

Coded when sentence- or clause-level structural conventions are violated: non-canonical word order; faulty compound-sentence linkage (including resumptive pronouns in relative clauses); ill-formed interrogative or negative structures; mismatched conditional constructions; passive-voice errors (auxiliary or participial form).

B.4 Lexical

Coded when a word is supplied that is morpho-syntactically well-formed but semantically inappropriate for the context: general word-choice errors; misuse of specialised technical terminology; literal translation from Arabic (calque); non-standard collocation; misused idiomatic expression. The boundary between lexical and grammatical was governed by the principle that lexical errors are about word identity and meaning, whereas grammatical errors are about inflection and closed-class form.

B.5 Cohesion and Coherence

Coded at sentence-pair or paragraph level when the relationship between propositions is obscured or misrepresented: inappropriate or absent cohesive devices; non-sequential ordering of ideas; paragraph organisation that obscures the main idea; failure to mark transitions; main-idea clarity. Cohesion and coherence errors were coded conservatively,

requiring at least two consecutive sentences to show a discernible organisational problem.

B.6 Source Attribution (Interlingual vs Intralingual)

For each coded error, a source-attribution decision was made, in accordance with Section 3.5.2, as either interlingual (a) if (a) the deviant form aligned systematically with an Arabic structure and (b) the same deviant form was found to manifest in published Arabic-L1 contrastive analyses (i.e., there was agreement), else intralingual or indeterminate. Inter-coder agreement on this decision yielded Cohen's $\kappa = 0.74$ (95% CI [0.69, 0.79]); disagreements were resolved through discussion consulting the studies.

Appendix C. Additional Illustrative Examples

C.1 Grammatical and Morphological

- Preposition: "The team consists from five engineers" (consists of)
- Tense: "The data shows that the system failed when we will apply the load" (when the load is applied)
- Article: "The water is essential for the human life" (Water is essential for human life)
- Agreement: "The results of the experiment shows a clear trend" (show)
- Number: "Many informations were collected" (much information was collected)

C.2 Spelling

"Brogram" for "program"; "falue" for "value"; "recieve" for "receive"; "enviroment" for "environment"; "happend" for "happened".

C.3 Syntactic

- Word order: "Always the engineer must check the safety procedures" (The engineer must always check…)
- Relative clause: "The chemical which we used it gave the expected result" (The chemical we used gave…)
- Conditional: "If the temperature will increase, the pressure increases" (If the temperature increases, the pressure increases)

C.4 Lexical

- "The experiment was successful because we have done a strong effort" (made a strong effort)
- "The reactor produces a big quantity of heat" (a large quantity of heat)
- "The graph shows the development of pressure with time" (the change in pressure over time)

C.5 Cohesion and Coherence

"Renewable energy is important. It reduces carbon emissions. Saudi Arabia has invested in solar power. The country needs to diversify its economy." (Each sentence is well-formed, but the connections between propositions are not made explicit; an appropriate cohesive device such as "Moreover," "For example," or "As part of this strategy," is required.)

Topic-comment drift: an essay paragraph that begins with the topic of renewable energy but ends with comments on national economic policy without an explicit organisational connection.