Interlingual and Intralingual Error Patterns in EFL Students' Argumentative Writing: Saudi Students as a Case Study

Dr. Nayef Shaie Alotaibi

Associate Professor of Applied Linguistics in the Department of English Language and Literature, College of Languages and Translation, Imam Mohammad Ibn Saud Islamic University, Riyadh

(Submitted to the journal on 23/07 /2026 AD, and accepted for publication on 25/08 /2026 AD)

Abstract:

This study analyzes interlingual and intralingual error patterns in the argumentative writing of 43 Level 6 Saudi EFL students (21 male, 22 female) at a public university in Saudi Arabia. The 9,873-word corpus produced 633 grammatical and lexical-syntactic errors, classified by surface category (one of nine grammatical labels) and by source attribution (interlingual, intralingual, or ambiguous) using a deterministic rule applied identically to both coders' span sets. Errors classified as interlingual under the study's deterministic coding rule concentrated in three surface categories: copula and auxiliary omission (60.0% interlingual), article use (53.3%), and sentence structure (36.4%, driven by three-or-more-clause coordination). These three categories accounted for 84.8% of the interlingual classifications in the corpus. Across the corpus as a whole, intralingual errors outnumbered interlingual errors by approximately three to one (68.7% versus 22.9%, with 8.4% ambiguous). All four gender comparisons returned non-significant results: distribution across surface categories (V = .108), 3-category source attribution (V = .012), 2-category sensitivity restricted to IL versus IA (V = .006), and per-essay error count (d = .152). Inter-rater reliability reached substantial agreement on surface category (κ = .626) and on the IL versus IA distinction (κ = .628), with moderate agreement on the full 3-category source attribution (κ = .566). The pedagogical implication is to focus instruction on the three surface categories where errors compatible with Arabic transfer concentrate, with copula and article contrasts as the priority targets.

Keywords: EFL error analysis, L1 transfer, intralingual and interlingual errors, argumentative writing.

أنماط الأخطاء بين اللغات وداخل اللغة الواحدة في الكتابة الحجاجية لطلاب اللغة الإنجليزية كلغة أجنبية: الطلاب السعوديون كدراسة حالة

د. نايف بن شاعي العتيبي

أستاذ اللغويات التطبيقية المشارك في قسم اللغة الإنجليزية في كلية اللغات والترجمة وآدابها، بجامعة الامام محمد بن سعود الإسلامية

المستخلص:

تحلل هذه الدراسة أنماط الأخطاء بين اللغات وداخل اللغة الواحدة في الكتابة الحجاجية لـ 43 طالبًا سعوديًا من طلاب اللغة الإنجليزية كلغة أجنبية في المستوى السادس (21 طالباً، 22 طالبة) في إحدى الجامعات الحكومية في المملكة العربية السعودية. وأسفر تحليل النص المكون من 9,873 كلمة عن 633 خطأً نحويًا ومفرداتيًا–نحويًا، تم تصنيفها حسب الفئة السطحية (واحدة من تسع فئات نحوية) وحسب مصدر الخطأ (بين اللغات، داخل اللغة، أو غامض) باستخدام قاعدة حتمية طُبقت بشكل متماثل على مجموعات النصوص التي قام كل من المرمزين بتحليلها. يتمثل الانتقال من اللغة العربية في ثلاث فئات سطحية: حذف الرابط والمساعد (60.0% بين اللغات)، واستخدام الأداة (53.3%)، وبنية الجملة (36.4%، مدفوعة بالتنسيق بين ثلاث جمل أو أكثر). وشكلت هذه الفئات الثلاث 84.8% من التصنيفات بين اللغات في المجموعة النصية. أما على مستوى المجموعة النصية ككل، فقد فاق عدد الأخطاء داخل اللغة عدد الأخطاء بين اللغات بنسبة ثلاثة إلى واحد تقريبًا (68.7% مقابل 22.9%، مع مانسبته 8.4% يصنف كأخطاء غامضة). وأسفرت جميع مقارنات الأربعة بين الجنسين عن نتائج غير ذات دلالة إحصائية: حيث كان التوزيع عبر فئات السطح (V = 0.108)، ونسبة إسناد المصدر إلى الفئات الثلاث (V = 0.012)، والحساسية المقتصرة على فئتين بين الأخطاء بين اللغات (IL) والأخطاء داخل اللغة (IA) (V = 0.006)، وعدد الأخطاء لكل مقال (d = 0.152). وبلغت موثوقية التقييم بين المقيمين اتفاقًا كبيرًا بشأن الفئة السطحية (κ = 0.626) وبشأن التمييز بين IL و IA (κ = 0.628)، مع اتفاق معتدل بشأن الإسناد الكامل للمصدر في الفئات الثلاث (κ = 0.566). وتتمثل الآثار التربوية في تركيز التدريس على الفئات السطحية الثلاث التي يتركز فيها الأخطاء المتوافقة مع الانتقال اللغوي من اللغة العربية، مع إعطاء الأولوية لمقارنات حروف العطف والأدوات.

الكلمات المفتاحية: تحليل أخطاء اللغة الإنجليزية كلغة أجنبية، الانتقال اللغوي من اللغة الأم، الأخطاء داخل اللغة وبين اللغات، الكتابة الحجاجية.

Introduction

In second language acquisition, learner language is systematic and developmental rather than arbitrary. Selinker (2013) characterized interlanguage as the transitional linguistic system that learners construct as they move toward target-language competence. This system has its own developing rule structure, distinct from both the first and the target language. It is shaped by hypothesis-testing, exposure, and the gradual restructuring of internal rules. Within this framework, learner errors are diagnostic evidence of language development rather than random failures of competence.

Corder (1967) made the foundational argument that systematic learner errors are evidence of the developing interlanguage system. He distinguished errors, which reflect gaps in competence, from mistakes, which reflect performance lapses. He also argued that systematically analyzing errors reveals the strategies learners use to acquire a second language. Lennon (1991) refined this framework by clarifying problems of definition and identification in error analysis. These include the distinction between errors caused by cross-linguistic influence and those that arise within the developing target-language system itself.

Within the interlanguage system, learner errors originate from two main sources. Some come from the influence of the learner's first language and are termed interlingual errors. Others reflect the learner's incomplete or developing knowledge of the target language, not traceable to L1 influence, and are termed intralingual errors (Brown & Lee, 2025; Keshavarz, 2015; Richards, 1971). This study investigates how the two types are distributed in the argumentative writing of Level 6 Saudi EFL students and whether the patterns differ between male and female writers.

Significance of the Study

The Saudi EFL context has seen little recent work using a unified two-axis coding framework. The present study supplies quantitative error distributions where most prior Saudi work offers only narrative description. It also targets argumentative writing specifically. This genre places greater cognitive and rhetorical demands on writers than narrative or descriptive tasks, so it is more likely to expose interlanguage features that lower-stakes writing leaves hidden (El-Dakhs et al., 2025). The gender-segregated structure of Saudi higher education offers a further opportunity: male and female writers can be compared at the same program level, taught from the same materials, in otherwise comparable conditions. Coding every error on both a surface and a source axis yields evidence for curriculum and materials design and for targeted instruction.

Statement of the problem

EFL students usually lack the necessary skills and knowledge to write in English (Ariyanti & Fitriana, 2017, p. 9), and such insufficiencies often remain undetected until later. Likewise, some EFL undergraduates with inadequate English proficiency exhibit unsatisfactory writing performance in academic English courses, resulting in failure or low grades and ultimately leading to frustration. Such frustration may lead those students to reconsider their academic specialization or even leave college altogether. It was found that EFL writing errors were mostly attributed to either interlanguage errors (errors arising from L1 interference, where learners transfer grammatical or phonetic traits from their mother tongue to L2, or intralanguage errors (errors taking place within the target language itself, and caused by the EFL learner's incomplete competence of L2's rules. Hameed (2016) states that writing is the most severely affected English skill for students entering university education in Saudi Arabia, with below-average proficiency in English (p. 2).

Literature Review

Error Analysis and Interlanguage Theory

Error analysis emerged in the late 1960s as a reaction to the limitations of contrastive analysis. Whereas contrastive analysis predicted learner difficulty by comparing L1 and L2 structures, error analysis examined actual learner output. Corder (1967) argued that systematic learner errors reflect an underlying developing competence rather than pedagogical failure. Selinker (2013) extended this argument with interlanguage: a distinct, transitional linguistic system between L1 and L2. It is shaped by five processes: language transfer, overgeneralization of target-language rules, transfer of training, strategies of second-language learning, and strategies of second-language communication. Selinker also introduced fossilization, by which certain non-target features stabilize in a learner's interlanguage and resist further development.

Subsequent work refined the framework. Lennon (1991) addressed persistent problems in defining and classi-

fying errors: he separated errors of competence from surface-level slips and argued for explicit operational criteria. Brown and Lee (2025) and Keshavarz (2015) set out taxonomies that classify errors along two axes: surface category (the grammatical or lexical level at which the error occurs) and source attribution (whether the error reflects L1 transfer or developmental processes). Ellis (1994, 1997) places these analytic frameworks within wider second language acquisition research.

Interlingual and Intralingual Errors

Interlingual errors result from the influence of the learner's first language on the developing target-language system. They occur when learners transfer linguistic features (grammatical, lexical, or rhetorical) from their L1 into their L2 production. The transfer can be facilitative, where structural similarity between L1 and L2 supports correct production, or obstructive, where structural difference produces non-target output. Common interlingual error types in EFL writing include misuse of the article system, preposition errors, word-order phenomena, and the literal translation of idiomatic or stance expressions.

Intralingual errors arise from internal processes of L2 development rather than from L1 influence. Richards (1971) identified four principal categories: overgeneralization of target-language rules, simplification, incomplete application of rules, and the formation of false hypotheses about target-language structure. These errors reflect developmental stages within the target-language system itself. They are usually expected to diminish with more exposure and instruction, although Larsen-Freeman (2006) has shown that progress is not linear: learners may show temporary drops in accuracy when attempting more complex structures.

Recent Empirical Work in EFL Writing

Recent learner-corpus and error-analysis studies in EFL contexts have applied formal coding schemes to substantial corpora and reported quantitative distributions of error types. Ehsanzadeh and Dehnad (2024), drawing on a large learner corpus of Iranian medical students, identified 11 high-frequency error types; omission and misuse of articles and determiners accounted for 14% of all errors. That rate is comparable to article-system error rates in other studies of L1-Arabic and L1-Persian EFL writers (Alhaysony, 2012; Crompton, 2011). Larsson et al. (2020) set out a methodological framework for inter-rater reliability in learner-corpus research. They treat reliability checks as a requirement rather than an option and give guidance on subsample size, coder training, and the reporting of agreement coefficients.

In Arabic-context EFL research, El-Dakhs et al. (2025) examined argumentative writing by Saudi EFL undergraduates and contrasted its rhetorical and lexico-grammatical features with native English norms. Their work supports examining the argumentative genre specifically rather than general composition or narrative, since argumentative writing places greater demand on cohesion, hedging, and stance management. Where they analyze rhetorical structure, the present study examines grammatical and lexical-syntactic accuracy.

Saudi EFL Writing and the Arabic Substrate

Five features of Arabic grammar bear directly on interlingual error patterns in Saudi EFL writers. First, Modern Standard Arabic permits zero copula in equational sentences in the present tense, where English requires an overt form of be (Ryding, 2005). Second, Arabic uses a single coordinating particle, wa (و), to link clauses across a wide range of semantic relations, which can produce English run-on sentences when transferred. Third, Arabic noun phrases use the construct state (idāfa) for possessive and genitival relations, which differs from the English of-phrase and possessive-clitic structures. Fourth, the Arabic definite article al- (ال) is used more freely with generic noun phrases than the English definite article. Fifth, Arabic permits topic-comment constructions in which a fronted noun phrase is co-referenced with a resumptive pronoun in the comment clause. Each predicts a specific surface position where a transfer error can appear, and the source-attribution rule described in the Methods keys its interlingual triggers to exactly these positions.

Gap Addressed by the Present Study

Although Saudi EFL writing has been studied in various forms, few studies have combined three things: a unified two-axis coding scheme applied to argumentative writing in a single proficiency cohort, quantitative distributions across both surface categories and source attributions, and a test for gender differences using inferential statistics rather than narrative comparison.

Research Questions

This study addresses the following research questions:

RQ1. What types of interlingual errors occur in the argumentative writing of Level 6 Saudi EFL students, and how are they distributed across surface categories?

RQ2. What types of intralingual errors occur in the argumentative writing of Level 6 Saudi EFL students, and how are they distributed across surface categories?

RQ3. Do the distribution and frequency of interlingual and intralingual errors differ between male and female student groups?

Methods

Setting and Participants

The study was conducted at a public university in the Kingdom of Saudi Arabia. Participants were enrolled in an English-language major program and were in Level 6, the sixth academic semester of the undergraduate program. At this point, students had completed foundational grammar and writing courses and were taking an Academic Writing course focusing on argumentative essay structure. English is the primary medium of instruction within the department. The sample consisted of 43 students (21 male and 22 female). Under the gender-segregated Saudi higher-education system, male and female students study in separate classes. All participants were native speakers of Arabic. Standard ethical procedures were followed: students were informed of the research purpose, permission was obtained before data collection, and identifying information was anonymized to protect confidentiality. The two classes chosen for study were selected randomly and were supposed to represent (in proficiency and gender) the other L6 classes not included in the study.

Corpus

Each participant wrote one argumentative essay in a single one-hour timed writing session. The prompt asked students to take a position on living in big cities and to defend it. The student-produced text totaled 9,873 words across the corpus (male sub-corpus = 4,738 words; female sub-corpus = 5,135 words). Individual essay length ranged from 98 to 368 words, with a mean of 229.6 words. The corpus reported here excludes the prompt material and any reviewer comments that were physically attached to the original drafts.

Coding Scheme

Throughout this paper, interlingual errors refers to errors attributable to the influence of the learner's first language (L1 transfer); intralingual errors refers to errors arising from the learner's developing knowledge of English itself, independent of the L1; and interlanguage refers to the learner's transitional L2 system as a whole (Selinker, 2013). The abbreviations IL and IA are used for the two error types, and these terms are used consistently in the English and Arabic abstracts.

A two-axis coding scheme classified each error by surface category and by source attribution. It draws on the taxonomies in Brown and Lee (2025) and Keshavarz (2015) and on recent learner-corpus work (Ehsanzadeh & Dehnad, 2024).

Axis 1: Surface category. Each error received one of nine labels: article, preposition, copula and auxiliary, subject-verb agreement, verb tense and form, pronoun and possessive, number and plural, word choice (lexical), and sentence structure (run-on, fragment, and word-order phenomena).

Axis 2: Source attribution. Each error was classified as IL (interlingual), IA (intralingual), or AM (ambiguous) by a deterministic rule applied identically to both coders' span sets, following the methodological framework of Jarvis (2000). The rule replaces coder discretion with category-specific structural triggers, each anchored to a Saudi-specific or Arabic-EFL empirical source. An error is coded IL when it matches an Arabic-anchored trigger: zero-copula equational structures (Alshayban, 2012; Ryding, 2005), generic-plural article overuse (Aboras, 2025; Alhaysony, 2012), indefinite-article omission before bare singular count nouns (Crompton, 2011), three-or-more-clause wa-coordination (Aljuraifani & Alanazi, 2025; Dickins, 2017), construct-state (idāfa) noun-of-noun-of-noun chains (Ryding, 2005; Dickins, 2017), and top-

ic-comment resumption (Ryding, 2005). Errors not matching any IL trigger default to IA. Four patterns trigger AM rather than IL: progressive and passive auxiliary omission (Pienemann, 1998, Stage 2 developmental); two-clause and-coordination (universal L2 paratactic preference, with Aljuraifani and Alanazi, 2025, specifying three-or-more chains as the Saudi-specific pattern); past-time markers paired with imperfective verbs (cross-language congruity without intergroup contrast); and NP-singular generic article use, which Aboras (2025) reports as target-like in Saudi Arabic L1 learners. Article and copula triggers fall back to sentence context if the coder's excerpt is clipped at a word boundary that hides the relevant cue

Rationale for the IA default and its consequences. The rule is deliberately asymmetric. Following Jarvis (2000), an error is attributed to the L1 only when positive evidence of cross-linguistic congruity is available, operationalized here as a match to a structural pattern documented for Arabic and for Arabic-L1 learners of English. Errors that match no such pattern cannot be positively linked to Arabic and are therefore assigned to IA by default. This design favors specificity (few false IL attributions) over sensitivity (some genuine transfer errors will be missed). Two consequences follow for interpretation. First, IA functions partly as a residual category: it combines errors with positive developmental evidence and errors that simply lack an Arabic-specific trigger, so the reported IA share should be read as an upper bound on intralingual errors rather than a precise estimate. Second, and correspondingly, the IL share should be read as a conservative lower bound on errors compatible with Arabic influence.

Inclusion and Exclusion Criteria

Mechanical errors (spelling, capitalization, and punctuation) were excluded from the count, as were discourse-level coherence problems such as paragraph organization and weakness of the thesis statement. The analysis was restricted to grammatical and lexical-syntactic errors at the morphosyntactic, lexical, and clause-structure levels. This restriction follows Lennon (1991), who argues for explicit operational criteria for inclusion in error counts.

Unit of Analysis

Each instance of an error was counted separately, including repeated instances of the same error type within a single essay. This reflects the study's focus on production patterns: counting once per essay would have erased the very patterning the research questions ask about. To handle length differences between essays, per-100-word rates are reported alongside raw counts, and per-essay error counts were used for the inferential test of gender differences. This approach is consistent with recent learner-corpus practice (Ehsanzadeh & Dehnad, 2024).

Statistical Procedures

Frequencies were tabulated overall and broken down by gender. Two chi-square tests of independence tested for gender differences: (a) a 9×2 test on the joint distribution of surface category by gender, and (b) a 3×2 test on the joint distribution of source attribution (IL / IA / AM) by gender, with a 2×2 sensitivity check restricted to IL versus IA. Effect sizes for both tests were reported as Cramer's V. A Welch's t-test (which does not assume equal variances) was used to compare per-essay error counts between male and female samples, with effect size reported as Cohen's d.

Assumptions and statistical power. For both chi-square tests, all expected cell frequencies exceeded 5 (minimum expected frequency = 11.8 for the 9×2 table and 25.0 for the 3×2 table). The chi-square tests nevertheless assume independent observations, whereas errors are nested within essays; this assumption is not fully met (see Limitations), so the chi-square p-values should be treated as approximate. Welch's t-test was used because the group standard deviations differed (5.31 vs. 7.03); the per-essay counts include one high outlier (F1, 32 errors). Sensitivity analyses based on the achieved sample sizes ($\alpha = .05$, two-tailed, power = .80) show that the per-essay comparison ($n = 21$ vs. 22) could reliably detect only large differences ($d \geq 0.88$); its power to detect an effect of the observed size ($d = .152$) was below .10. The error-level chi-square tests could nominally detect small effects ($w \geq .15$ for the 9×2 test; $w \geq .12$ for the 3×2 test), but because they treat non-independent errors as independent, this nominal sensitivity overstates the information actually available from 43 essays.

Reliability

The primary coder and a second coder independently coded all 43 essays. Each coder first identified error spans independently; the source-attribution rule was then applied identically to both coders' span sets. Because the two coders did not always mark identical boundaries for the same error, spans were matched before categorization agreement was computed. Two spans were treated as the same error when they came from the same essay and their word-level overlap, measured by the Jaccard index (shared words divided by all words covered by either span), was at least 0.40; each span could be matched only once. Of the 633 primary spans and 662 second-coder spans, 421 were matched, leaving 212 primary-only and 241 second-coder-only spans. Agreement on error identification, calculated as the proportion of specific positive agreement (2 × matched spans ÷ total spans from both coders), was therefore .65. Categorization agreement was then computed on the 421 matched pairs. Surface-category Cohen's κ was .626, within the substantial range (.61–.80) of Landis and Koch's (1977) benchmarks. Source attribution returned a 3-category κ of .566, within the moderate range (.41–.60), with disagreements concentrated in AM decisions, and a 2-category κ restricted to IL versus IA decisions of .628 (substantial; n = 369). These benchmarks are conventions rather than strict thresholds (Larsson et al., 2020). All κ values apply only to errors that both coders identified; the identification agreement of .65 means that about a third of the spans flagged by either coder were not flagged by the other, and this should be borne in mind when interpreting the category distributions.

Results

The 43 essays yielded 9,873 student-produced words, in which 633 grammatical and lexical-syntactic errors were identified and coded according to the two-axis scheme.

Overall Error Frequency

The corpus contained 633 codeable errors, yielding an overall error density of 6.41 per 100 words. The distribution across the nine surface categories is presented in Table 1.

Table 1

Overall error frequency by surface category

Surface category	Count	% of total	Per 100 words
1. Article	107	16.9%	1.08
2. Preposition	58	9.2%	0.59
3. Copula/auxiliary	25	3.9%	0.25
4. Subject-verb agreement	94	14.8%	0.95
5. Verb tense/form	45	7.1%	0.46
6. Pronoun/possessive	44	7.0%	0.45
7. Number/plural	50	7.9%	0.51
8. Word choice (lexical)	70	11.1%	0.71
9. Sentence structure	140	22.1%	1.42
Total	633	100.0%	6.41

Three categories accounted for over half of all errors: sentence structure (22.1%), article (16.9%), and subject-verb agreement (14.8%). Word choice (11.1%) and preposition (9.2%) formed the next tier. Number/plural (7.9%), verb tense/form (7.1%), and pronoun/possessive (7.0%) clustered in the middle range. Copula/auxiliary omission was the least frequent in absolute terms (3.9%, n = 25).

Source Attribution

Of the 633 errors, 145 (22.9%) were classified as interlingual (IL), 435 (68.7%) as intralingual (IA), and 53 (8.4%) as ambiguous (AM). The cross-tabulation of surface category by source attribution is presented in Table 2.

Table 2

Surface category by source attribution

Surface category	IL	IA	AM	Total	% IL
1. Article	57	41	9	107	53.3%
2. Preposition	7	51	0	58	12.1%
3. Copula/auxiliary	15	6	4	25	60.0%
4. Subject-verb agreement	0	94	0	94	0.0%
5. Verb tense/form	0	45	0	45	0.0%
6. Pronoun/possessive	10	34	0	44	22.7%
7. Number/plural	5	45	0	50	10.0%
8. Word choice (lexical)	0	70	0	70	0.0%
9. Sentence structure	51	49	40	140	36.4%
Total	145	435	53	633	22.9%

The distribution is non-uniform. Three categories carried the strongest interlingual signal:

Copula and auxiliary omission (60.0% IL, 15 of 25): omission of the finite be in obligatory contexts (e.g., "there a lot of compney"; "the focus always on education"; "this why I prefer to live").

Article (53.3% IL, 57 of 107): omission before count nouns (e.g., "living in big city"; "is interesting topic") and overuse of the with generic noun phrases (e.g., "the big cities gets expensive"; "living in the big cities").

Sentence structure (36.4% IL, 51 of 140): three-or-more-clause coordination patterns (e.g., "There is no traffic the city is small everyone know each other"; "the expenses are increasing it comes with many benefits"; rents are low life is simple no jam everyone drives freely). A further 40 sentence-structure errors (28.6%) were classified as ambiguous under the rule's two-clause-coordination concession.

Three categories carried weaker but nontrivial interlingual signals: pronoun and possessive (22.7% IL), preposition (12.1% IL), and number and plural (10.0% IL). Three categories produced no IL classifications under the rule: subject-verb agreement, verb tense and form, and word choice. Across the corpus, copula and auxiliary, article, and sentence structure together account for 123 of the 145 interlingual classifications (84.8%).

Gender Comparison

Three test families assessed whether male and female writers differed in their error patterns: distribution across surface categories, source attribution by gender (with a sensitivity restriction to IL versus IA only), and per-essay error count. Summary statistics by gender are presented in Table 3.

Table 3

Summary statistics by gender

Statistic	Male (n = 21)	Female (n = 22)
Total words	4,738	5,135
Total errors	299	334
Errors per 100 words (corpus rate)	6.31	6.50
Per-essay error count: M (SD)	14.24 (5.31)	15.18 (7.03)
Interlingual share	23.1% (69/299)	22.8% (76/334)
Intralingual share	68.2% (204/299)	69.2% (231/334)
Ambiguous share	8.7% (26/299)	8.1% (27/334)

Distribution across surface categories. A chi-square test of independence on the 9 (categories) × 2 (gender) contingency table returned $\chi^2(8, N = 633) = 7.38$, $p = .496$, Cramer's $V = .108$.

Source attribution. A 3 × 2 chi-square on source attribution (IL / IA / AM) by gender returned $\chi^2(2, N = 633) = 0.10$, $p = .952$, Cramer's $V = .012$. The 2-category sensitivity check restricted to IL versus IA returned $\chi^2(1, N = 580) = 0.02$, $p = .885$, $V = .006$. Both effect sizes fall well below the field-specific threshold for a small effect (Plonsky & Oswald, 2014). The interlingual share was 23.1% in male essays and 22.8% in female essays.

Per-essay error count. Male essays averaged 14.24 errors (SD = 5.31); female essays averaged 15.18 (SD = 7.03). A Welch's t-test (which does not assume equal variances) returned $t(38.99) = -0.50$, $p = .621$, Cohen's $d = .152$.

All four tests returned non-significant results: no statistically significant differences were detected between male and female writers on any dimension assessed. Given the limited power of the per-essay comparison, these results do not establish that the two groups perform equivalently.

Discussion

This section distinguishes three levels of claim. Findings are the distributions and test results reported in the Results section. Interpretations are explanations for those findings, drawn from Arabic grammar and prior research; the present data are compatible with them but cannot test them directly. Pedagogical implications, presented in the following section, are recommendations derived from these interpretations. Each subsection below opens with the relevant finding before turning to its interpretation.

Where Arabic Transfer Concentrates

IL classifications in this corpus are structurally localized: copula and auxiliary omission, article use, and sentence structure carry nearly all of them, while subject-verb agreement, verb tense and form, and word choice receive none under the rule. Because the rule identifies patterns compatible with L1 transfer rather than confirming transfer directly for any individual error, these findings are best read as showing where a transfer explanation is most plausible, not as proof that transfer occurred.

Copula and auxiliary omission carry the strongest within-category signal, although the absolute count is modest (n = 25; 15 IL). Modern Standard Arabic permits zero copula in equational present-tense clauses (Ryding, 2005). Learners producing English equivalents such as "the focus always on education" or "there a lot of compney" are plausibly transferring this option into a context where English requires an overt form of be. Because copula omission is also reported developmentally among learners whose L1 has an overt copula, the pattern is compatible with Arabic influence rather than proof of it.

. Alshayban (2012), in a study of 100 male Saudi students, reported a parallel concentration of copula omission. Steiner (2019) frames the same pattern as a Feature Reassembly process (Lardiere, 2008): advanced learners reassemble the relevant tense and predicate features, while lower-proficiency learners default to the L1 zero-copula option. The 60% IL share rests on 15 absolute classifications, with a 95% confidence interval of about 36–81%; the direction of the signal is well-supported, the magnitude less so.

Article errors form the largest interlingual category in absolute terms (57 of 145 IL classifications). Two patterns predominate: omission of a/an before count nouns ("living in big city"; "is interesting topic") and overuse of the with generic noun phrases ("the big cities gets expensive"; "living in the big cities"). Both are compatible with a contrast between the two languages, although article difficulties are also common among learners from other L1s without an indefinite article. Arabic has no indefinite article and uses al- with generic NPs (Ryding, 2005; Fassi Fehri, 2007; Kashgary, 2015); English marks indefiniteness and also distinguishes NP-level from sentence-level generics (Al-Malki et al., 2014). The 53.3% IL share matches prior Saudi-population findings: Alhaysony (2012) reported 57.4% interlingual in article errors at the University of Ha'il, and Crompton (2011) found article errors involving generic reference the most clearly L1-attributable category in advanced L1 Arabic learners. Aboras (2025) refines this picture: the transfer signature is concentrated in NP-plural and sentence-level generic contexts, while NP-singular generic article use approaches target-like performance, and the rule restricts IL classification correspondingly.

Sentence structure shows the third concentration (36.4% IL, n = 51), driven by the rule's three-or-more-clause coordination trigger. Arabic uses wa to link clauses across a wide semantic range (Ryding, 2005; Dickins, 2017). Corpus instances such as "There is no traffic everyone know each other" show the paratactic strategy that Aljuraifani and Alanazi (2025) identify specifically in Saudi Arabic L1 EFL writing. Two-clause coordination defaults to AM rather than IL because universal L2 paratactic preference cannot be separated from Arabic-specific transfer at that length; only chains of three or more clauses count as the Saudi-specific signature.

For contrastive instruction to be cost-effective, it should target the categories where transfer concentrates. This echoes Aljuraifani and Alanazi (2025), who trace persistent L1-influenced errors in advanced Saudi female learners to weak contrastive-linguistics-informed instruction. It also fits the long-standing argument that error analysis is useful because it pinpoints where instruction will most reduce learner error (Brown & Lee, 2025; Corder, 1967).

The Corpus-Wide Source Distribution

Across the nine categories combined, intralingual errors outnumber interlingual errors by about three to one (68.7% versus 22.9%, with 8.4% ambiguous). The corpus-wide ratio mostly reflects the mix of categories: those where transfer concentrates (copula, article, sentence structure) supply most of the IL classifications, while the three categories that produce no IL under the rule (subject-verb agreement, verb tense and form, word choice) account for roughly a third of total errors.

The 22.9% IL share falls between two clusters in the published Arabic L1 → English error analysis literature. Higher-IL studies focus on a narrow set of transfer-prone categories or apply looser source-attribution criteria: Aljuraifani and Alanazi (2025) report 58.09% IL in 32 Saudi female essays, Qasem (2020) reports 54.03% in University of Bisha undergraduates, and article-restricted studies (Alhaysony, 2012; Crompton, 2011) report similar or higher proportions. Lower-IL studies use broader category scope or stricter coding: Abdelmohsen (2022), Al-Hamzi et al. (2023), Alzahrani (2020), and Ahmed Alsoweed et al. (2025) all report IA-dominant patterns in Arabic-L1 cohorts. The closed prediction-list rule applied here is stricter than the attribution criteria in the higher-IL studies, and its default-to-IA design (see Coding Scheme) is expected to lower the IL share relative to them. Differences between the present 22.9% and figures reported elsewhere may therefore reflect coding criteria as much as learner behavior, and direct numerical comparison across studies should be made cautiously.

Proficiency is a further explanatory factor. SLA work since Taylor (1975) documents a developmental path: elementary learners rely more on transfer, while intermediate-to-advanced learners rely more on overgeneralization (Brown & Lee, 2025; Burhansyah, 2019). One possible interpretation, which the present single-level design cannot test, is that Level 6 learners are past the most transfer-heavy stage of acquisition, which fits a pattern where transfer concentrates in a few structures rather than spreading across the corpus. Ehsanzadeh and Dehnad (2024) report article and determiner errors at 14% of the total in an L1-Persian medical context, comparable to the 16.9% for articles in this corpus. The convergence is consistent with the view that some transfer signatures are typical of EFL learners whose L1 lacks articles, not of Arabic-L1 learners specifically.

Gender

None of the four gender comparisons reached significance, and the observed effect sizes were small (the largest was $V = .108$ for the distribution across surface categories; source attribution $V = .012$; Plonsky & Oswald, 2014). A non-significant result, however, does not demonstrate the absence of a difference. With 21 and 22 essays, the per-essay comparison could detect only large differences ($d \geq 0.88$), and the error-level tests overstate the effective sample size because errors are nested within essays. The appropriate conclusion is that this study found no evidence of gender differences, not that such differences are absent.

This pattern is consistent with the broader Saudi EFL literature on gender comparison. Almusharraf and Al-Otaibi (2021), drawing on a 90,753-word Saudi corpus from 197 students (98 male, 99 female), reported that differences between male and female writing were not significantly related to specific error types. Further, Al-Tameemy and Daradkeh (2019), at Prince Sattam bin Abdulaziz University, found no statistically significant gender differences except in grammar errors specifically. Because the source-attribution rule was applied identically to all essays, differences in coder judgment are unlikely to have produced spurious gender effects here, although this does not compensate for the limited statistical power noted above.

Pedagogical Implications

The findings point to two distinct instructional priorities. Arabic transfer is concentrated in three surface categories (copula, article, sentence structure), which are the targets for explicit contrastive teaching. The remaining categories are predominantly developmental and call for sustained practice rather than L1 contrast.

Addressing Intralingual Development

Subject-verb agreement (14.8% of total), word choice (11.1%), and verb tense and form (7.1%) together account for a third of all errors in the corpus and are entirely intralingual under the rule. These categories should receive sustained instructional attention through systematic practice on inflection, morphology, and lexical choice. Feedback should target the specific developmental pattern rather than treat errors as random.

Word-choice errors (11.1% of total, 100% intralingual) suggest a need for collocational competence rather than

vocabulary expansion alone. Instruction should target the typical environments in which words occur (verb-noun, adjective-noun, preposition-noun) rather than treating vocabulary as isolated lexical items.

Preposition errors (9.2% of total, 87.9% intralingual) are best addressed through chunk-based instruction rather than rule-by-rule teaching, given the idiomaticity of English prepositions.

Addressing Arabic Transfer

Copula and auxiliary omission, although low in absolute frequency (3.9% of total), is the most strongly Arabic-attributed category (60.0% interlingual) and the single category most amenable to explicit contrastive instruction. Learners should be made aware that English requires an overt form of be in contexts where Arabic permits a zero copula.

Article use (16.9% of total, 53.3% interlingual) should be addressed through contrastive instruction that explicitly compares the English and Arabic article systems, focused on (a) the obligatory indefinite article before singular count nouns and (b) the restricted use of the definite article with generic reference in English. Article errors are the most numerous interlingual signature in the corpus in absolute terms (57 IL classifications) and yield the largest absolute return on contrastive teaching.

Sentence structure (22.1% of total, 36.4% interlingual) should be addressed through sentence-combining and sentence-decomposition exercises that develop learners' control over English clause-boundary signals (subordinators, full stops, semicolons) as alternatives to chained coordination. The IL classifications here are driven by three-or-more-clause coordination patterns that match the Arabic wa-paratactic strategy; two-clause coordination is too ambiguous to attribute and is not the target of this recommendation.

Gender

The present data provide no evidence to support gender-differentiated instruction; until larger, adequately powered studies are available, recommendations should follow the category-level patterns rather than gender.

Limitations

The sample is restricted to a single proficiency level (Level 6) at a single institution, with a single argumentative prompt, so the results may not generalize to other proficiency levels, other Saudi institutions, or other argumentative tasks. Inter-rater reliability on the 3-category source attribution is moderate ($\kappa = .566$) rather than substantial. The IL versus IA boundary itself reaches substantial agreement ($\kappa = .628$), and this 2-category sensitivity is the appropriate metric for the source-attribution claims; even so, AM-cell disagreements reduce confidence in the precise AM share. Subject-verb agreement defaults to IA under the rule, following the feature-reassembly account (Lardiere, 2008); a more permissive treatment that scored borderline cases as IL on Arabic verb-subject order grounds would produce a higher IL share. The copula and auxiliary category carries a small absolute count ($n = 25$; 15 IL classifications), so the within-category 60% IL share has a 95% confidence interval of about 36–81%. The rule applies only the cross-language congruity component of Jarvis (2000); it has no non-Arabic L1 comparator to complete the triangulation, so IL classifications should be read as compatible with possible transfer rather than as direct evidence of it. One female essay (F1) produced 32 errors, 2.8 SD above the corpus mean; the gender tests were not re-run with this case excluded, because all four were strong nulls and removing it could only strengthen them. Classroom effects cannot be controlled, since male and female students are taught in separate classes under the gender-segregated higher-education system, so any small residual gender differences would be confounded with classroom-level variation. The 633 errors are also not fully independent observations: they are nested within 43 students (and within essays), so the chi-square tests reported here, which treat each error as an independent unit, may understate the true uncertainty around the surface-category and source-attribution distributions; a student- or essay-level analysis (e.g., a mixed-effects or clustered model) would be a more conservative test of the gender comparisons and is recommended for a follow-up analysis. Placement in Level 6 reflects course sequencing rather than a direct proficiency measure, so students grouped here may not have had fully equivalent English proficiency; no independent proficiency instrument was administered, and this should be read as a limitation on the comparability of the sample rather than as a confirmed equivalence.

Conclusion

This study examined the argumentative writing of 43 Level 6 Saudi EFL learners (21 male and 22 female) at a public university in Saudi Arabia to investigate the nature, distribution, and underlying sources of grammatical

and lexical-syntactic errors. The 9,873-word corpus yielded 633 errors, coded under a two-axis scheme using a deterministic source-attribution rule applied identically to both coders' span sets.

Under the study's conservative attribution rule, errors classified as intralingual outnumber interlingual classifications by roughly three to one, and errors compatible with Arabic transfer concentrate in three categories: copula and auxiliary omission, article use, and sentence structure, which together carry 84.8% of the interlingual classifications. The categories where the rule found no Arabic basis (subject-verb agreement, verb tense and form, and word choice) are best treated as develop-mental.

No gender comparison reached significance on any dimension, and the observed effect sizes were small (source attribution V = .012). Given the limited power of the per-essay comparison, this is an absence of evidence for gender differences rather than evidence of their absence; the result is consistent with the broader Saudi EFL literature on gender comparison (Almusharraf & Al-Otaibi, 2021; AlTameemy & Daradkeh, 2019). These patterns support the view that learner errors are systematic evidence of an evolving interlanguage system. They also indicate that instruction at this level should be matched to the surface categories most affected by each source of error, rather than treating L1 transfer as a single, uniform explanation.

Declaration: The author confirms that he has no conflict of interest with anybody.

References

Abdelmohsen, M. M. (2022). Arab EFL learners' writing errors: A contrastive error analysis study. *British Journal of English Linguistics, 10*(4), 1–17.

Aboras, A. (2025). The acquisition of generic reference by L1 Saudi Arabic learners of L2 English. *Journal of the European Second Language Association. DOI:https://doi.org/10.22599/jesla.102*

Ahmed Alsoweed, A., Banafi, A., & Almesaar, O. (2025). An investigation of writing errors made by Saudi English-major students. *International Journal of English Linguistics, 15*(2).

Al-Hamzi, A. M. S., Nababan, M., Santosa, R., Djatmika, Sumarlam, & Yustanto, H. (2023). Frequent linguistic errors in the writing of Yemeni EFL Arabic-speaking learners. *Studies in English Language and Education, 10*(1), 350–368. https://doi.org/10.24815/siele.v10i1.26022

Alhaysony, M. (2012). An analysis of article errors among Saudi female EFL students: A case study. *Asian Social Science, 8*(12), 55–66.

Aljuraifani, Y. A., & Alanazi, Z. M. (2025). L1-induced grammatical errors affecting Saudi female EFL students' academic writing: A cross-linguistic study of Arabic language interference. *World Journal of English Language, 15*(5), 171–188. https://doi.org/10.5430/wjel.v15n5p171

Al-Malki, E. A., Abdul Majid, N., & Mohd Omar, N. A. (2014). Generic reference in English, Arabic and Malay: A cross-linguistic typology and comparison. *English Language Teaching, 7*(11), 15–27. https://doi.org/10.5539/elt.v7n11p15

Almusharraf, N., & Al-Otaibi, H. (2021). Gender-based EFL writing error analysis using human and computer-aided approaches. *Educational Measurement: Issues and Practice, 40*(2), 60–71. https://doi.org/10.1111/emip.12413

Alshayban, A. S. (2012). *Copula omission by EFL Arab learners* [Master's thesis, Colorado State University].

AlTameemy, F. A., & Daradkeh, A. A. (2019). Common paragraph writing errors made by Saudi EFL students: Error analysis. *Theory and Practice in Language Studies, 9*(2), 178–187. https://doi.org/10.17507/tpls.0902.07

Alzahrani, M. A. (2020). EFL Saudi undergraduates' tenses errors in written discourse due to interlingual and intralingual interference. *International Journal of Linguistics, 12*(5), 95–111.

Brown, H. D., & Lee, H. (2025). *Principles of language learning and teaching: A course in second language acquisition* (7th ed.). Routledge.

Burhansyah. (2019). Analysis of error sources in L2 written English by Indonesian undergraduate students. *Studies in English Language and Education, 6*(1), 71–83.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching, 5*(4), 161–170.

Crompton, P. (2011). Article errors in the English writing of advanced L1 Arabic learners: The role of transfer. *Asian EFL Journal, 50*, 4–34.

Dickins, J. (2017). The pervasiveness of coordination in Arabic, with reference to Arabic > English translation. *Languages in Contrast, 17*(2), 229–254.

Ehsanzadeh, S. J., & Dehnad, A. (2024). Analysis of high-frequency errors and linguistic patterns in EFL medical students' English writing: Insights from a learner corpus. *BMC Medical Education, 24*, 1264.https://doi.org/10.1186/s12909-024-06242-z

El-Dakhs, D. A. S., Yahya, N., & Al Thowaini, B. M. (2025). Persuading in Arabic and English: A study of EFL argumentative writing in contrast with native English norms. *Open Cultural Studies, 9*, 20250056. https://doi.org/10.1515/culture-2025-0056

Ellis, R. (1994). *The study of second language acquisition.* Oxford University Press.

Ellis, R. (1997). *SLA research and language teaching.* Oxford University Press.

Fassi Fehri, A. (2007). Bare, generic, mass, and referential Arabic DPs. In S. Karimi, V. Samiian, & W. K. Wilkins (Eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation* (pp. 40–65). John Benjamins. https://doi.org/10.1075/la.101.04feh

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning, 50*(2), 245–309. https://doi.org/10.1111/0023-8333.00118

Kashgary, A. D. (2015). Meanings and functions of the Arabic article *al-* 'the' with reference to the concept of definiteness. *Journal of Semitic Studies, 60*(1), 93–110. https://doi.org/10.1093/jss/fgu033

Keshavarz, M. H. (2015). *Contrastive analysis, error analysis, and interlanguage hypothesis* (Rev. ed.). Rahnama Press.

Landis, J., & Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometric, 33, 159-

74. https://doi.org/10.2307/2529310

Lardiere, D. (2008). Feature assembly in second language acquisition. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106–141). Lawrence Erlbaum Associates.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590–619.

Larsson, T., Paquot, M., & Plonsky, L. (2020). Inter-rater reliability in learner corpus research: Insights from a collaborative study on adverb placement. *International Journal of Learner Corpus Research, 6*(2), 237–251. https://doi.org/10.1075/ijlcr.20001.lar

Lennon, P. (1991). Error: Some problems of definition, identification, and distinction. *Applied Linguistics, 12*(2),180–196.https://doi.org/10.1093/applin/12.2.180

Pienemann, M. (1998). *Language processing and second language development: Processability theory.* John Benjamins.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning,64*(4),878–912.https://doi.org/10.1111/lang.12079

Qasem, F. A. A. (2020). Crosslinguistic influence of the first language: Interlingual errors in the writing of ESL Saudi learners. *Macrolinguistics, 8*(2), 105–120.

Richards, J. C. (1971). A non-contrastive approach to error analysis. *English Language Teaching, 25*(3), 204–219. https://doi.org/10.1093/elt/XXV.3.204

Ryding, K. C. (2005). *A reference grammar of Modern Standard Arabic.* Cambridge University Press.

Selinker, L. (2013). *Rediscovering interlanguage.* Routledge.

Steiner, J. (2019). Acquisition of the English copula by Arabic-speaking ESL learners: Evidence for feature reassembly. *Linguistic Approaches to Bilingualism, 9*(4), 504–537.

Taylor, B. P. (1975). The use of overgeneralization and transfer learning strategies by elementary and intermediate students of ESL. *Language Learning, 25*(1), 73–107.

Appendix A

Summary of Source-Attribution Rules

Code	Trigger pattern	Basis	Illustrative corpus example
IL	Zero copula in a present-tense equational clause	Alshayban (2012); Ryding (2005)	"the focus always on education"
IL	Indefinite article omitted before a bare singular count noun	Crompton (2011)	"living in big city"
IL	Definite article overused with a generic plural noun phrase	Aboras (2025); Alhaysony (2012)	"the big cities gets expensive"
IL	Three or more clauses chained by coordination without appropriate boundaries (wa-type)	Aljuraifani & Alanazi (2025); Dickins (2017)	"big cities have clear air no one get sick they are healthy"
IL	Construct-state (idāfa) noun-of-noun-of-noun chain	Ryding (2005); Dickins (2017)	"Big cities are crowded and expensive and dangerous".
IL	Topic-comment structure with resumptive pronoun	Ryding (2005)	"This is the kind of life I like it"
AM	Progressive or passive auxiliary omission	Pienemann (1998), developmental Stage 2	"I am live in a small city"
AM	Two-clause coordination without appropriate boundary	Universal L2 parataxis; Aljuraifani & Alanazi (2025)	"When moved to the big city I find it dangerous"
AM	Past-time marker with imperfective (present) verb form	Cross-language congruity without intergroup contrast (Jarvis, 2000)	"Three years ago, we move to live in Riyadh"
AM	Article error with a singular generic noun phrase	Aboras (2025): near target-like in Saudi learners	"When we drink a water in the city it is not clean"
IA	Default: any error matching no IL or AM trigger (all subject-verb agreement, verb tense/form, and word-choice errors fell here)	Richards (1971); Jarvis (2000)	"we lives in the city because we like modren live"

Biographical Statement

Dr. Nayef Shaie Alotaibi an Associate Professor of Applied Linguistics at the Department of English Language and Literature, College of Languages and Translation, Imam Mohammad Ibn Saud Islamic University. I received my PhD degree in 2018 from Imam University. My research interests include… apllied lingustics domains

معلومات عن الباحث

د. نايف بن شاعي العتيبي، أستاذ اللغويات التطبيقية المشارك في قسم اللغة الإنجليزية، بكلية اللغات والترجمة، في جامعة الإمام محمد ابن سعود الاسلامية، (السعودية). حاصل على درجة الدكتوراه في اللغويات التطبيقية من جامعة الإمام محمد بن سعود الإسلامية عام 2018. تدور اهتماماته البحثية حول قضايا اللغويات التطبيقية.

Email: nsalotaibi@yahoo.com