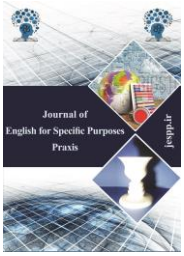
Journal of English for Specific Purposes Praxis ISSN: 3092-7854
https://www.jespp.ir
Vol. 2, No. 3; 2025, July
Article Citation:
Ostad, O., & Heidari, H. (2025). A comparative error analysis of Iranian EFL learners' writing compositions by human evaluators Vs. perplexity AI platform. *Journal of English for Specific Purposes Praxis, 2*(3), 85–104. https://doi.org/10.22034/jespp.2025.546767.1021
JESPP



A Comparative Error Analysis of Iranian EFL Learners' Writing Compositions by Human Evaluators Vs. Perplexity AI Platform

Omid Ostad [1,*], Hadi Heidari [2]

[1,*]Corresponding author, Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran; omidostad007@gmail.com

[2] Department of English language, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran; hheidari67@gmail.com

Original Research Article | Date of Submission: 13 May 2025 | Date of Acceptance: 28 July 2025

Abstract

The advent of the AI as a supplementary tool for error analysis and how it is different from human error analysis seems to be an underexplored and enchanting research area. This study sought to examine the types of errors found in the written compositions of 16 intermediate-level Iranian students— male and female learners included—selected through convenience sampling whose data was gleaned from two language academies in Rasht. All participants were using American English File 3 (2nd Edition) serving as their main course book, with the researchers also being their instructor. The students were tasked with writing a response to a letter, based on a model provided in their textbook. A total of 16 writing samples were collected and analyzed using both qualitative and quantitative methods, guided by Keshavarz's (2013) linguistic error classification framework. The evaluation was conducted by a human rater and the AI tool Perplexity, which was specifically prompted to identify errors according to the same classification system. The results revealed a range of error types with varying frequencies. Morphosyntactic errors tended to be the most common, followed by orthographic, lexicosemantic, and phonological errors, respectively. Moreover, Inter-rater reliability was calculated via Cohen's Kappa, which indicated substantial level of agreement between human and AI raters ($\kappa = 0.78$, $p < 0.001$). Teachers can combine the precision and consistency of AI with the subjective and interpretive depth of human assessment to provide more responsive feedback that supports linguistic accuracy and learner agency. Based upon the findings, we provided a number of empirically grounded suggestions for future research.

Keywords: artificial intelligence; error analysis; human evaluators; L1 interference; morphosyntactic errors; perplexity; writing compositions

1. Introduction

According to Li and Han (2025), writing encompasses both the act of composing carried out by the author of any given text and the final written product. This dual nature is shaped by the context in which writing occurs, including both its intended audience and its genre. In second language (L2) writing, various factors influence performance—such as the writer's proficiency in the second language, literacy in their first language, and cultural or rhetorical conventions—all of which can be enhanced through targeted instruction.

Given English's global status and its widespread use, more L2 learners are engaging in academic activities that demand composition skills. A compelling area of research in this discipline involves examining the types of errors made by non-native English writers. Over the past five decades, scholarly interest has grown around the theoretical understanding of errors—exploring their nature, implications, and communicative roles within texts.

Error Analysis stands as a foundational theory in second language acquisition. It targets identifying and interpreting the errors made by L2 learners through comparing their language output with standard norms of the target language (Parameswari et al., 2024). According to Shakir et al. (2020, p. 108), Error Analysis in language education involves studying the incorrect forms produced by individuals learning a foreign language. According to *Parameswari et al.* (2024, p. 62), EA refers to "the study of linguistic ignorance, the investigation of what people do not know and how they attempt to cope with their ignorance".

This study reports on the frequency of the errors in compositions of Iranian EFL learners at intermediate level of language proficiency. The present contribution is founded on the premise that the knowledge derived from this investigation will provide insights into the nature of the academic compositions of EFL students. The investigation of the role of second language proficiency in second language writing ability continues to be revealing for the better understanding of the nature of second language writing. The first language teaching benefit from the findings of linguistics in many cases includes error analysis.

Error analysis facilitates educators in identifying the underlying causes of learners' mistakes and implementing instructional strategies to address them. From a pedagogical perspective, errors play a crucial role in the learning process, functioning as cognitive tools through which learners actively construct knowledge. (Parameswari et al., 2024). Xu and Zeng (2023) explain the significance of learner errors in three different ways: "The first to the teacher in that they tell him, if he undertakes a systematic analysis, how far towards the goal the learner has progressed, and consequently what remains for him to learn. Second, they provide to the researcher evidence of how language is learned or acquired. Thirdly, they are indispensable to the learner himself, because we can regard the making of errors as a device the learner uses in order to learn."

According to Xu and Zeng (2023), One of the most evident benefits of error analysis lies in its value for educators. Research has demonstrated that learner errors serve as a form of feedback, offering insights into both the extent to which students have internalized previously taught material and the areas of instruction that may require further emphasis. Consequently, language teachers must be well-prepared, attentive, and responsive to the challenges learners encounter throughout the acquisition process.

AI tools can be used to provide feedback to language learners' errors of academic works. As indicated by Park et al. (2019), AI systems can achieve high levels of agreement with human raters and it can correlate strongly with human evaluations, as high as a correlation coefficient of 0.825 in writing assessments. Li et al. (2015) also used automated essay scoring systems, which demonstrated favorable validity evidence that AI can provide consistent and reliable assessments comparable to those of highly trained human examiners.

In this study, Perplexity is used as an AI tool to provide many types of feedback to language learners' errors of academic works, analogous to human raters. According to its official platform, Perplexity is a versatile AI engine and is designed to provide accurate, trustworthy, and real-time responses to any question (Perplexity.ai, 2025). However, according to Allahwala (2023), Perplexity's Natural Language Processing (NLP) architecture enables it to analyze text data effectively, and identify grammatical errors, syntactic inconsistencies, and stylistic issues. Allahwala (2023) analyzed how users employed Perplexity to correct grammatical errors on websites and written content as an effective tool for automated language refinement tasks. Hence, Perplexity's deep learning algorithms and its ability to parse sentence structure, detect common ESL errors, and offer corrective suggestions allow it to function as a dedicated error-analysis tool when applied in educational or editorial contexts and makes it a viable resource for language learners and educators seeking automated feedback.

2. Literature Review

2.1. Theoretical Background

The literature on error analysis is informed by different studies which highlight the different theoretical backgrounds they followed at different eras of the development in the field, which are distinguished by how they treat errors and try to resolve it. According to Touchie (2024), In second or foreign language learning, an error refers to the use of a linguistic element—such as a word or grammatical structure—in a way that reflects incomplete or flawed learning, as judged by native or fluent speakers. A distinction is often made between the notion of 'errors', which stem from gaps in knowledge, and 'mistakes', which are typically caused by performance-related factors like inattention, fatigue, or carelessness. Errors can be categorized into various types, including lexical errors which are the incorrect use of vocabulary, phonological errors, i.e. mispronunciation, syntactic errors that are basically grammatical inaccuracies, interpretive errors which can cause misunderstanding a speaker's intended meaning, and finally pragmatic errors that pertain to inappropriate use of speech acts or communicative norms

In the field of second and foreign language acquisition, the study of errors provides insight into the cognitive and linguistic processes learners engage in while acquiring a new language. Error Analysis (EA) is the systematic examination of the mistakes made by second language learners. It is typically conducted to:

a. identifies strategies which learners use in language learning

b. tries to identify the causes of learner errors

c. obtains information on common difficulties in language learning, as an aid to teaching or in the preparation of teaching materials.

Emerging in the 1960s as a subfield of Applied Linguistics, error analysis aimed to reveal that numerous learner errors stemmed not from interference by the first language, but from inherent cognitive strategies common to language acquisition. Positioned as a theoretical counterpoint to Contrastive Analysis, this approach sought to categorize errors based on the presumed cognitive mechanisms underlying them. A foundational dichotomy was established between intralingual errors—those arising within the target language system—and interlingual errors, which result from cross-linguistic influence. Intralingual errors are typically categorized into several types, namely: overgeneralizations, where

learners incorrectly apply target language rules to unsuitable contexts; oversimplifications, involving the use of overly basic linguistic structures; developmental errors, which mirror natural stages in language acquisition; communication-based errors, arising from strategies used to convey meaning; induced errors, resulting from instructional influences; errors of avoidance, where learners refrain from using complex structures they perceive as difficult; and errors of overproduction, involving excessive use of certain forms Despite these classifications, accurately identifying the source of such errors remains challenging. By the late 1970s, the focus of research shifted from error analysis to the broader frameworks of Interlanguage and Second Language Acquisition (SLA).

Up until the late 1960s, behaviorism was the dominant theoretical framework in second language learning. Within this perspective, acquiring a second language was understood as the formation of new verbal habits that are integrated into the learner's existing linguistic repertoire. These newly formed habits of the target language were believed to coexist alongside the established patterns of the first language. Behaviorist scholars contended that the influence of the first language could either obstruct or support the development of these new habits.

Errors were once considered very bad and had to be prevented by various means. As Al-Sobhi (2019) notes, early behaviorist views held that errors should be eliminated through reinforcement of correct responses and discouragement of incorrect ones. Regarding the negative perspective toward errors in L2 learning, Mhamedi and Ibrir (2023) explain that errors were historically viewed as linguistic "sins" to be eradicated, often attributed to the persistence of first language habits in the new language. As a consequence, contrastive analysis—the branch that involves the comparison of native and target languages to predict and explain errors—became a dominant approach in applied linguistics.

As a reaction to this view, error analysis emerged in the 1960s to demonstrate that learner errors may not solely be the outcome of mother tongue interference but also mirrored internal learning strategies (Al-Sobhi, 2019). Error analysis posited that while contrastive analysis could identify linguistic differences and predict potential difficulties, it failed to account for the majority of learner errors. Behaviorist models were unable to explain the full range of errors observed in second language learning.

Error analysis remains useful in identifying and providing and explanation for the systematic errors made by learners. As Mhamedi and Ibrir (2023) emphasize, by describing and classifying the learner's errors in linguistic jargon, a sketch of those features of language causing learning impediments could be painted.

2.2. Empirical Studies

Errors have before been investigated in the context of classroom oral discourse (Al-Sobhi, 2019), informal settings (Yang & Ren, 2025; Parameswari et al., 2024), reading (Touchie, 2024), lectures (Xu & Zeng, 2023), academic genres (Hyland, 2021; Mohebbi & Truscott, 2021), and student writings (Ferris, 2015; Bitchener & Knoch, 2015; Lee, 2020). These investigations have mainly merely targeted their use patterns of frequency and function.

Hyland (2021) analyzed academic writing by postgraduate students using Halliday and Hasan's (1976) cohesion framework. In a similar vein in one study, Ferris (2015) compared essays which were written by native students and those of foreign language learners, focusing on the use of cohesive devices. Yang and Ren (2025) examined revision behaviors of Chinese undergraduates using a corpus-

based error correction method. Xu and Zeng (2023) analyzed learner responses to feedback in lecture-based writing tasks. Lee (2020) studied student writing in classroom contexts, highlighting the role of conjunctions and lexical cohesion.

Results outlined that conjuncts were overused and lexical cohesion was moderately observed by native speakers (Hyland, 2021), that non-native students of English tended to use more conjunctions in a mechanical way compared to native speakers and often placed them at the beginning of sentences (Ferris, 2015), that there were differences between the type of texts in the way cohesive devices were put to use (Lee, 2020), and that essays receiving higher ratings tended to use more varied cohesive ties.

Banihashem et al. (2024) investigated the comparative impact of ChatGPT-generated feedback and peer feedback on students' argumentative essay writing, employing two distinct coding frameworks—one for essay analysis and another for feedback categorization. The study involved 74 graduate students from a Dutch university. Findings revealed a statistically significant divergence between the two feedback sources. While ChatGPT tended to offer more descriptive commentary, emphasizing the essay's structural strengths and weaknesses, peer feedback was more diagnostic, pinpointing specific issues within the text. Despite these differences, no significant correlation was found between essay quality and the type of feedback received. Consequently, the researchers concluded that ChatGPT and peer feedback serve complementary functions, with ChatGPT acting as a supportive tool rather than a substitute for peer evaluation.

Bui and Barrot, (2024) evaluated the scoring consistency between ChatGPT and a human rater across 200 argumentative essays from various proficiency levels with each essay being scored by both ChatGPT and a human rater. The study aimed to examine the relationship between the scores given by ChatGPT and those assigned by an experienced human examiner. Results indicated very weak to moderate correlation in scoring between the two raters and low intraclass correlation coefficients over multiple evaluations. These inconsistencies were ascribed to ChatGPT's scoring algorithm, training data, and being random which indicated a lack of consistency in ChatGPT's scoring across multiple rounds. The study concluded that while ChatGPT displays potential as an AES tool, it still requires improvements to match the reliability and consistency of human raters. The findings suggest important implications for writing assessment and future research in automated scoring tools.

Building upon the gap in the literature, the current study tries to shed light on the consistencies which were seen between the scores provided by human raters and Perplexity to the participants' written assignment.

3. Research Questions

As the brief review of literature above discloses, some findings were contradictory: while Ferris (2015), Hyland (2021), and Lee (2020) found no significant disparity in the frequency of ties between native and ESL students, Yang and Ren (2025) discovered a correlation between tie frequency and writing quality. Overall, all these studies indicate that language learners can commit more errors—especially in pragmatic functions—compared with native speaker norms.

Based on what has been discussed earlier, this study intended to investigate learners' errors among Iranian students studying English as second language. The present contribution focussed on the 5 steps put forth by Corder. In order to examine the relationship between errors when it comes to composition a piece of writing of EFL students, the following questions were formulated:

RQ1. What are the different types of errors which are committed by EFL learners in their writing?

RQ2. What type of error is the most recurring one among the Iranian EFL learners?

RQ3. Are there any consistencies between the scores provided by human raters and Perplexity to the EFL learners' writing errors?

Method

4.1. Design of the Study

The study adopted a quantitative research method through a repeated measures design to examine how human and Perplexity rated the participants' written works. Each writing sample was evaluated by both types of raters.

4.2. Participants

The study involved 16 adult EFL learners recruited from two language institutes in Rasht: SAMA Language Center and The IELTS Centre. The participants, comprising 7 females and 9 males, had an average age of 26, with ages ranging from 22 to 33. Most were pursuing MA or MS degrees in disciplines unrelated to English. Both classes were at the intermediate level, confirmed through placement tests administered by the institutes. They followed the same course book and were taught by the researcher. Classes were held in the evenings twice a week—either on Saturdays and Mondays or Sundays and Tuesdays. Over the spring semester of 2025, participants attended a total of 20 sessions, each lasting two hours.

4.3. Instrumentation

To collect data, a composition task was employed as the research instrument. The writing prompt was extracted from page 17 of the students' course textbook, American English File 3 by Clive Oxenden and Christina Latham-Koenig (2008), with the relevant page included in Appendix B. Prior to the task, students received the day's instructional content along with explanatory guidance. Subsequently, they were instructed to compose a response to the letter featured in the textbook, mirroring the provided sample while incorporating their own identities, within a 20-minute timeframe. This particular task was selected due to its familiarity among students, as such writing exercises are routinely practiced in classroom settings. Accordingly, it was presumed that the influence of rhetorical structure on learners' writing performance would be minimized. A sample of students' written homework is provided in Appendix A.

4.4. Data Collection Procedure

Following the task introduction, students were instructed to compose a response to a letter, imagining it had been personally addressed to them. They completed the task independently within a 20-minute time frame during the spring semester of 2025, under close observation. Upon finishing, students were given additional time to review and self-correct their compositions. This step allowed them to address surface-level mistakes, enabling the researcher to focus solely on identifying genuine errors rather than distinguishing them from correctable slips.

A total of 16 compositions were submitted to the researcher. No feedback was provided during or after the writing process. The collected samples were then meticulously examined to identify and classify the errors. Both a human evaluator and the AI tool Perplexity conducted the analysis. The initial phase involved a thorough reading of each composition to detect errors, which were subsequently categorized based on their linguistic characteristics using a linguistic taxonomy applicable to corpus analysis. This taxonomy organized errors according to the specific linguistic items affected, and further divided them into subcategories to allow for more nuanced interpretation.

Perplexity was prompted as follows to analyses the errors caused by Persian language interference, along with its frequency and percentage:

Analyze this writing sample for errors that may be caused by Persian language interference. Label each error and explain the likely source.

4.5. Data Analysis Procedure

This study employed the linguistic error taxonomy developed by Mohammad Hossein Keshavarz (2013). Detailed descriptions of each error type were presented in accompanying tables. Following the identification and classification of errors, their proportions relative to the total number of errors were manually calculated. Frequency and percentage analyses were conducted to determine which error categories were most prevalent among the participants. The results revealed the category with the highest error rate. The frequency and percentages of errors categorized by human and Perplexity evaluators, as well as the consistency of the scores by human evaluator and Perplexity was calculated provided to show the consistency of the scores assigned by two scoring sets.

5. Results

5.1. Analysis of Errors Evaluated by Human Rates vs. Perplexity

As shown in the following tables, Morphosyntactic errors were most frequently observed errors rated by human rater, among which those errors related to the use of articles were the most frequent one (21%). The detailed frequency of detected errors along with their ratrio is broken down into details in table 1. Moreover, Morphosyntactic errors (77%) were observed more frequently than lexico-semantic or orthographic ones.

Table 1
Analysis of Errors Produced by Iranian EFL Learners Rated by Human Rater

Type of error	Frequency of errors	Percentage (%)
Wrong verb tense	16	11.7
Wrong word order	9	6.6
Subject–verb agreement	11	8.0
Wrong verb group	6	4.4
Spelling	15	10.9
capitalization	13	9.5
Preposition	11	8.0
Articles	28	20.4
Omission of verbs	3	2.2
Wrong plural morpheme	12	8.8
Typical Persian construction	6	4.4
Conditional sentences	3	2.2
TOTAL	133	100.0

Regarding the analysis of errors produced by Iranian EFL learners rated by Perplexity as AI rater, morphosyntactic errors remained the most frequently observed, with article-related errors being the most common (20.4%). The detailed frequency and percentage of observed errors are presented in Table 14. Overall, Morphosyntactic errors accounted for 76.6% of all errors, significantly outnumbering lexico-semantic and orthographic ones. In Table 2, the exact types of errors, their sources, in parallel with an instance of each one is shown in separate tables.

Table 2

Analysis of Errors Committed by Iranian EFL Learners Rated by Perplexity as AI Rater

Type of error	Frequency of errors	Percentage (%)
Wrong verb tense	19	12.84
Wrong word order	10	6.8
Subject–verb agreement	12	8.1
Wrong verb group	6	4.3
Spelling	17	11.5
Capitalization	14	9.5
Preposition	13	8.7
Articles	29	19.6
Omission of verbs	3	2.0
Wrong plural morpheme	14	9.5
Typical Persian construction	7	4.7
Conditional sentences	4	2.7
TOTAL	148	100.0

The total number of errors identified by the human rater was 133, while the AI rater detected 148 errors. Despite this numerical difference, the proportional distribution across error types remained largely consistent, with both raters identifying *Articles*, *Wrong verb tense*, and *Spelling* as the most frequent error categories. This suggests a shared recognition of high-impact grammatical issues, contributing to an overall substantial agreement ($\kappa = .76$).

5.1.1. Verb Tense

Errors involving incorrect verb tense or even the choice of the verb arise when learners use a tense that is inappropriate for the context of the sentence. The findings of this study indicate that participants demonstrated a lack of awareness regarding the proper application of verb tenses, leading to frequent inaccuracies in their written compositions. Table 3 presents an example of verb tense error in learners' writing. The AI rater identified 19 instances (12.84%) compared to 16 (11.7%) by the human rater. This discrepancy may stem from the AI's heightened sensitivity to temporal markers and rule-based tense violations, whereas human raters may consider contextual appropriateness or learner intent.

Table 3

Example of Verb Tense Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Verb tense	1. She is 28… all her siblings were married.	1. She is 28… all her siblings are married.

Type of error: Morphosyntactic – Past simple used instead of present simple Source of error: Unique developmental error – Not traceable to either English or Persian.

Note. This error reflects a non-transfer-based misapplication of tense, suggesting an internalized but incorrect rule.

5.1.2. Wrong Word Order

Table 4 shows the results of word order is defined as the syntactic arrangement of words within a sentence, clause, or even in a phrase. Both raters showed complete alignment in identifying Wrong verb group (6 errors), indicating that these error types are consistently recognizable regardless of rater type.

Table 4

Example of Wrong Word Order Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Wrong word order	1. He with his family goes out.	1. He goes out with his family.

Type of error: Morphosyntactic – SCV structure used instead of SVC Source of error: Interlingual – Transfer from Persian syntactic structuring

Note. This error reflects direct L1 influence, where Persian word order allows pre-verbal placement of adverbial phrases, leading to syntactic inversion in English.

5.1.3. Subject Verb Agreement

In English grammar, subject-verb agreement requires that the verb matches the subject in number. This means that a singular subject—referring to one person or thing—must be paired with a singular verb, while a plural subject—referring to more than one—must take a plural verb form. Table 5 illustrates an instance of subject-verb agreement.

Table 5

Example of Subject–Verb Agreement Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Subject–verb agreement	1. She study very hard.	1. She studies very hard.

Type of error: Morphosyntactic – Intralingual or developmental Source of error: Overgeneralizing second language verb forms.

Note. This error reflects a common developmental pattern in EFL learners, where base verb forms are applied uniformly across subjects, disregarding third-person singular agreement.

5.1.4. Wrong Use of Gerunds and Infinitives

This type of error mainly involves using gerunds and infinitives interchangeably or omitting to or –ing. Table 6 displays an example of such verb group errors.

Table 6

Example of Verb Group Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Wrong use of verb groups	1. I recommend her for rent your room.	1. I recommend her for renting your room.

Type of error: Morphosyntactic – Misuse of gerund form
Source of error: Intralingual – Faulty categorization of verb patterns

Note. This error reflects confusion between gerund and infinitive constructions, a common intralingual issue among intermediate EFL learners.

5.1.5. Spelling Errors

Spelling refers to the systematic representation of words through the use of letters that conform to established linguistic conventions. A case in point has been provided in Table 7. AI detected more spelling errors (17 vs. 15), reflecting its precision in surface-level orthographic features. Human raters may overlook minor spelling deviations if they do not impede comprehension.

Table 7

Example of Orthographic Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Orthographic errors	1. She graduated from Rasht univercity.	1. She graduated from Rasht university.

Type of error: Orthographic – Confusion of letters with similar sounds
Source of error: Intralingual – False categorization of phonetic spelling

Note. This error reflects a phonological approximation where the learner substitutes a phonetically similar but incorrect spelling, a common issue in intermediate-level EFL writing.

5.1.6. Capitalization

Capitalization refers to the orthographic convention of initiating a word with an uppercase letter, followed by lowercase letters for the remaining characters. Table 8 outlines an example of capitalization error. AI also detected slightly more capitalization errors (14 vs. 13), reflecting its precision in surface-level orthographic features.

Table 8

Example of capitalization Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
capitalization errors	1. She is studying… english with me.	1. She is studying… English with me.

Type of error: Orthographic – Incorrect use of lowercase for a proper noun
Source of error: Intralingual – Ignorance of formal writing conventions

Note. This error reflects a lack of awareness regarding capitalization rules in English, particularly for language names and sentence-initial words.

5.1.7. Preposition

A preposition is a syntactic unit that expresses relational meaning by linking a noun or pronoun—its object—to other elements within a sentence. Serving as a functional connector, it facilitates the expression of temporal, spatial, or logical associations between its object and the surrounding linguistic context. In this study, the majority of participants exhibited noticeable difficulty in using prepositions accurately, as illustrated in the examples provided below in Table 9. The AI system flagged more errors in preposition use (13 vs. 11), likely due to its reliance on syntactic patterns. Human raters may exercise leniency in these areas, especially when errors are common among L2 learners.

Table 9

Example of Redundant Preposition Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Redundant use of preposition	1. He doesn't help to his wife.	1. He doesn't help his wife.

Type of error: Morphosyntactic – Use of "to" where it is not required
Source of error: Interlingual – Transfer of mother tongue grammatical elements

Note. This error reflects direct influence from Persian, where certain verbs require prepositions that are not used in equivalent English constructions.

5.1.8. Articles

An article is a grammatical marker that accompanies a noun to signal the nature of its referential status. In English, there are two types of articles: the and a/an. The definite article (the) denotes specificity, referring to particular entities, whereas the indefinite articles (a/an) introduce non-specific or general references. These forms are conventionally classified as the definite and indefinite articles, respectively. A comparison of this is illustrated in Table 10.

Table 10

Example of Article Omission Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Omission of the indefinite article	1. She is MA student.	1. She is an MA student.

Type of error: Morphosyntactic – Missing indefinite article "an"
Source of error: Interlingual – Transfer of first language grammatical rules

Note. This error reflects influence from Persian, which does not use articles in the same way as English, leading to frequent omission in learner writing.

5.1.9. Omission of Verbs

Both raters showed complete alignment in identifying omission of verbs (3 errors), indicating that these error types are consistently recognizable regardless of rater type. As Table 11 shows, a very rare type of error was not including the main verb.

Table 11

Example of Verb Omission Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Omission of required verb	1. You have careful with her.	1. You have to be careful with her.

Type of error: Morphosyntactic – Missing auxiliary and infinitive verb
Source of error: Unique developmental error – Not traceable to either English or Persian

Note. This error reflects a non-transfer-based structural gap, suggesting an internalized but incomplete syntactic rule not rooted in L1 or L2 norms.

5.1.10. Wrong Plural Morphemes

Certain nouns either lack conventional plural markers or exhibit irregular plural forms, which may lead learners to mistakenly assume that pluralization uniformly requires the addition of a standard morpheme. Table 12 highlights a sample of how this error occurred. The AI system also detected more errors in plural morphemes (14 vs. 12), likely due to its reliance on syntactic patterns. Human raters may exercise leniency in these areas, especially when errors are common among L2 learners.

Table 12

Example of Plural Morpheme Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Wrong use of plural morphemes	1. If you need any more informations…	1. If you need any more information…

Type of error: Morphosyntactic – Addition of plural morpheme to an uncountable noun
Source of error: Interlingual – Transfer from Persian, where “information” may take a plural marker

Note. This error reflects cross-linguistic influence, where learners apply L1 grammatical rules to L2 contexts, resulting in nonstandard pluralization of mass nouns.

5.1.11. Typical Persian Construction

Such constructions emerge when learners rely on their first language as a cognitive framework, translating content into English on a word-for-word basis during written production. As indicated in Table 13, AI identified more errors for Persian constructions (7 vs. 6). These differences may reflect the Perplexity’s broader exposure to structural norms across corpora and its detection of more elaborate differences.

Table 13

Example of Persian Structural Transfer Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Using Persian structures in English	1. He has two years (old).	1. He is two years old.

Type of error: Morphosyntactic – Use of “has” instead of “is” for age expression
Source of error: Interlingual – Transfer from Persian, where age is expressed with the verb “have”

Note. This error reflects direct syntactic transfer from Persian, where age is conveyed through possession rather than identity, leading to nonstandard constructions in English.

5.1.12. Wrong use of Conditional Sentences

AI also identified more errors in the use of conditional sentences (4 vs. 3). These differences may indicate that human raters may interpret such constructions through a pedagogical lens and do not detect as much as Perplexity. Table 14 shows how learners kept either mixing them, or confusing all the 3 conditional forms.

Table 14

Example of Conditional Sentence Error in Iranian EFL Learner Writing

Error classification	Error identification	Error correction
Using second conditional if-clause with first conditional sentence	1. If you wanted more information, I will give you.	1. If you want more information, I'll give you.

Type of error: Morphosyntactic – Misuse of conditional forms
Source of error: Interlingual – Transfer from Persian, where past tense may express present intention (e.g., I wanted to know instead of I want to know)

Note. This error reflects cross-linguistic influence in conditional constructions, where Persian allows past tense to convey politeness or present relevance, leading to mismatched clause structures in English.

5.2. Reliability Indices

In the next stage, the factors identified in two data sets by human evaluator and Perplexity were inspected for reliability, using Cohen's Kappa. Table 15 shows the reliability estimates for the factors computed by human evaluator and Perplexity through running interrater reliability.

Table 15

Consistency Rate and Inter-Rater Reliability Statistics of the Scores by Human Evaluator and Perplexity

Components	Consistency Rate
Articles	0.84
Wrong Verb Tense	0.84
Spelling	0.88
Subject–Verb Agreement	0.92
Wrong Plural Morpheme	0.86
Capitalisation	0.93
Preposition	0.85
Wrong Word Order	0.90
Wrong Verb Group	1.00
Typical Persian Construction	0.86
Conditional Sentences	0.75
Omission of Verbs	1.00
Inter-rater reliability (Cohen's Kappa)	0.78

The results indicated that learners produced a range of error types, each varying in frequency. Morphosyntactic errors emerged as the most prevalent, followed in descending order by orthographic,

lexicosemantic, and phonological errors. Inter-rater reliability was assessed using Cohen's Kappa. The analysis indicated substantial agreement between human and AI raters, $\kappa = .78$, $p < .001$.

6. Discussion

This investigation aimed to explore how effectively human raters and AI systems identify and categorize writing errors made by Iranian EFL learners. By comparing the evaluations from both sources—showing a strong alignment with a reliability coefficient of 0.94—the study sheds light on the linguistic hurdles learners face in academic writing. The data indicates that morphosyntactic errors dominate, with article misuse (21%), verb tense issues (12.4%), and spelling mistakes (11.6%) being the most prevalent.

The high frequency of article-related errors corresponds with established research in second language acquisition, particularly in cases where the learner's native language lacks an article system. Persian, for instance, lacks a grammatical system that differentiates between definite and indefinite articles in the way that English does, often resulting in omissions or incorrect usage. Both human and AI raters consistently flagged these errors, reinforcing their diagnostic value. Ramazani et al. (2025) similarly noted that AI tools were particularly adept at spotting article misuse among Iranian IELTS candidates, though human raters offered more nuanced insights into how these errors affected rhetorical clarity.

Verb tense and subject–verb agreement errors, accounting for 12.4% and 8% respectively, reflect ongoing developmental challenges in EFL writing. These typically arise from intralingual factors such as overgeneralization or confusion between tense and aspect. AI systems, with their rule-based parsing capabilities, were efficient in detecting such errors, while human raters demonstrated greater sensitivity to context and communicative intent. Bagheri Nevisi and Arab (2023) found that AI-generated feedback did significantly enhance syntactic accuracy in Iranian learners' writing, particularly in areas related to tense and agreement.

Orthographic errors, including spelling (11.6%) and capitalization (10.2%), were also common, suggesting that learners continue to struggle with the transition from Persian script to Latin-based writing. AI systems reliably identified these surface-level mistakes, whereas human raters often overlooked them unless they interfered with meaning. This contrast highlights the complementary strengths of both approaches: AI excels in consistency and speed, while human feedback emphasizes clarity and communicative effectiveness (Almusharraf & Alotaibi, 2023).

Structural errors such as wrong word order (6.5%), incorrect verb group usage (4.3%), and constructions influenced by Persian syntax (4.3%) point to interference from the learners' first language. Persian's flexible word order and verb-final tendencies often lead to misplaced modifiers or inverted sentence structures in English writing. Human raters were more capable of recognizing these patterns as culturally rooted, while AI systems sometimes misclassified them as generic grammatical mistakes. This supports Mirzaeian's (2025) claim that human evaluators are better equipped to interpret errors shaped by cultural and linguistic background.

These types of errors often reveal deeper gaps in learners' understanding of English sentence structure and modality. Notably, both human and AI raters showed strong agreement in identifying even

low-frequency errors, suggesting that with proper training or programming, evaluators can reliably detect a wide range of issues.

The high level of agreement between human and AI evaluations (r = .94) points to the potential of integrated assessment models. AI systems are well-suited for processing large volumes of student writing and providing immediate feedback on recurring error types. Human raters, however, remain essential for interpreting learner intent, assessing pragmatic appropriateness, and evaluating rhetorical coherence. Al-Obaydi and Pikhart (2025) argue that a hybrid feedback model may be the most effective way to promote both linguistic accuracy and learner autonomy in EFL writing.

For Iranian learners, whose writing difficulties are shaped by both linguistic and emotional factors, such a model could enable more targeted instruction. Teachers might use AI tools for initial diagnostics and reserve human feedback for deeper engagement with learner strategies and revision processes.

While the overall reliability between the two evaluation modes was high, differences emerged in how each approached error interpretation. AI systems performed well in identifying surface-level issues—particularly those involving verb tense, article usage, and plural morphemes. These categories accounted for the bulk of flagged errors, with article misuse comprising 20.4%, followed by verb tense (11.7%) and plural morphemes (8.8%). This pattern echoes findings by Almusharraf and Alotaibi (2023), who observed that automated systems tend to outperform human raters in detecting frequent grammatical errors, though they may lack contextual sensitivity.

Human raters, on the other hand, showed greater discernment in evaluating errors related to pragmatic meaning, rhetorical flow, and cultural transfer. For instance, phrases like "He has two years old" were readily identified by human evaluators as Persian-influenced constructions, whereas AI systems often misclassified them as generic syntax errors. This supports Mirzaeian's (2025) view that human raters are better positioned to recognize culturally embedded patterns in learner writing.

Despite these differences, both evaluator types consistently identified morphosyntactic errors, suggesting that such categories are reliable indicators of learner difficulty across modalities. The strong reliability score reinforces the viability of AI systems for large-scale diagnostic screening. However, their limitations in interpreting error sources and understanding learner intent highlight the continued importance of human judgment in language assessment.

From a pedagogical perspective, the findings support a blended evaluation approach. AI systems offer efficiency and uniformity, making them ideal for initial screening and formative feedback. Human evaluators, meanwhile, provide the depth and contextual awareness necessary for addressing more complex learner needs. As Al-Obaydi and Pikhart (2025) note, while AI-generated feedback is often precise, it lacks the interpersonal engagement and adaptive reasoning that characterize effective human instruction.

For Iranian EFL learners—whose challenges often stem from both linguistic interference and emotional inhibition—this dual approach may be especially beneficial. AI can help learners identify recurring patterns and receive immediate corrections, while human feedback can foster confidence, agency, and strategic awareness during the revision process.

Nonetheless, the study also points to areas where AI evaluation could be improved. Automated systems occasionally misclassified idiomatic expressions and struggled with cohesion-related errors.

Future research should focus on training AI models with Persian-English learner corpora to enhance their sensitivity to interlingual features. Longitudinal studies could also examine how AI-generated feedback influences learner development over time, particularly in terms of error reduction and writing fluency.

7. Conclusion and Implications

The findings indicated that Perplexity as AI rater demonstrated a slightly higher frequency of error detection (e.g., in rule-governed and surface-level categories). However, the overall distribution had significant consistency with the human rater. It was concluded that the differences were due to the to the Perplexity AI's algorithmic precision and lack of contextual flexibility. This can be contrasted with the human rater's subjective pedagogical judgment and interpretive differences with AI. These findings indicate that AI-assisted evaluation can be used as a complementary tool in language assessment, to be balanced with human oversight. At the same time, the study highlights the irreplaceable role of human judgment in interpreting interlingual transfer, pragmatic appropriateness, and rhetorical coherence—areas where AI systems still fall short. Human evaluators demonstrated greater sensitivity to Persian-influenced constructions and were better equipped to contextualize errors within the learner's communicative intent. This distinction suggests that AI can serve as an efficient diagnostic tool. However, it should be pedagogically informed by human feedback.

Although writing is widely regarded as the most challenging of the four English language skills, it is possible for L2 learners to improve their writing accuracy. Much like diagnosing a medical condition before prescribing treatment, effective writing requires first identifying the underlying causes of errors. In line with the focus of this study, producing accurate written work necessitates an understanding of the factors contributing to ineffective writing. The findings clearly indicate that the influence of learners' first language (L1) on their writing varied across genres, particularly in relation to syntactic and semantic features of the L1.

This study has documented the predominant error types observed in the written output of Iranian university-level EFL learners. A considerable proportion of these errors appear to stem from first language (L1) transfer. The explicit influence of Persian on learners' English writing underscores the necessity for language instructors to critically examine the role of L1 interference in both spoken and written language production. As suggested by Ridha (2012), one pedagogically effective approach to addressing such influences involves compiling learner errors and engaging students in reflective analysis and self-correction activities

The findings of this study have contributed to a deeper understanding of how errors in English composition are categorized and diagnosed among Iranian EFL learners. However, like all research, this study has several notable limitations. Firstly, the participants were exclusively drawn from SAMA and IELTS Language Canters, which restricts the generalizability of the results to broader populations of Iranian EFL learners. Secondly, the sample size was relatively small, which may limit the robustness of the conclusions.

Based on the outcomes of this research, several directions for future studies are proposed. It is recommended that further investigations focus on the errors produced by undergraduate students majoring in English. Additionally, exploring both interlingual and intralingual errors among learners of other languages would offer valuable insights. Finally, conducting a comparative analysis of writing

errors among high school and university students may illuminate developmental distinctions in language proficiency across these educational stages.

The comparative analysis of Iranian EFL learners' writing compositions, evaluated by both human and AI raters, reveals a nuanced landscape of learner errors shaped by linguistic, cognitive, and cultural factors. With a reliability coefficient of .94 between evaluators, the study confirms that AI systems can reliably detect a wide range of morphosyntactic and orthographic errors, particularly those involving article misuse, verb tense, and plural morphemes. These findings highlight the potential of AI-assisted evaluation tools in streamlining error identification and supporting large-scale assessment.

Despite the findings, the limited number of the papers rated by the raters as well as the issues related to the human rater (e.g., the issue of inter-rater reliability due to using only one human rater) may introduce bias to the findings and make it hard to generalize the findings. Moreover, the participants of the study were all adults, which renders the results more challenging to be generalizable for adolescents or children if they were included in the current study.

For Iranian EFL learners, whose writing challenges often reflect both structural interference and affective inhibition, a hybrid evaluation model offers the most promise. Teachers can combine the precision and consistency of AI with the subjective and interpretive depth of human assessment to provide more responsive feedback that supports linguistic accuracy and learner agency.

Ultimately, this study advocates for integrative approaches to language assessment to use technological development under the supervision of the human agents. Future research can utilize different learner characteristics (e.g., different proficiency and cultural levels, with more varied written assignments) in different cultures, compared with other AI tools, to study the longitudinal consistency of the AI and human raters under different situations.

Acknowledgment

We all thank the participants for their time, energy, and cooperation.

Authors' Contributions

All authors have conducted the study, collected data, analyzed and interpreted the data, and written up the manuscript.

Funding

The study did not receive any funding.

Competing Interests

The authors declare that there is no conflict of interest.

References

Allahwala, R. (2023). *Use Perplexity.ai and ChatGPT to correct any grammatical errors on your website* [Video]. YouTube. Retrieved from https://www.youtube.com/watch?v=hJJWe0KZ7QA

Almusharraf, N., & Alotaibi, H. (2023). An error-analysis study from an EFL writing context: Human and automated essay scoring approaches. *Technology, Knowledge and Learning, 28*, 1015–1031. https://doi.org/10.1007/s10758-022-09592-z

Al-Obaydi, L. H., & Pikhart, M. (2025). AI partner versus human partner: Comparing AI-based peer assessment with human-generated peer assessment in examining writing skills. *Language Testing in Asia, 15*, Article 38. https://doi.org/10.1186/s40468-025-00375-8

Al-Sobhi, B. M. S. (2019). The nitty-gritty of language learners' errors – Contrastive analysis, error analysis and interlanguage. *International Journal of Education and Literacy Studies, 7*(3), 1–9. https://doi.org/10.7575/aiac.ijels.v.7n.3p.1

Bagheri Nevisi, R., & Arab, N. (2023). Computer-generated vs. direct written corrective feedback and Iranian EFL students' syntactic accuracy and complexity. *Teaching English as a Second Language Quarterly, 42*(2), 111–148.

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, *21*(1), 23.

Bitchener, J., & Knoch, U. (2015). Written corrective feedback in second language acquisition and writing. *Language Teaching, 48*(2), 207–226. https://doi.org/10.1017/S0261444814000069

Bui, N. M., & Barrot, J. S. (2024). ChatGPT as an automated essay scoring tool in the writing classrooms: how it compares with human scoring. *Education and Information Technologies*, *29*(1), 1-18.

Ferris, D. R. (2015). *Treatment of error in second language student writing* (2nd ed.). University of Michigan Press.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Hyland, K. (2021). Second language writing. In *The Routledge Handbook of Second Language Acquisition and Writing* (pp. 23–38). Routledge.

Lee, I. (2020). *Teacher feedback in L2 writing: Contexts and issues*. Cambridge University Press.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (awe) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1-18. https://doi.org/10.1016/j.jslw.2014.10.004

Li, Y., & Han, L. (2025). Corpus-based analysis of second language writing errors and their pedagogical implications. *Applied Mathematics and Nonlinear Sciences, 10*(1). https://doi.org/10.2478/amns-2025-0010

Mhamedi, N., & Ibrir, L. (2023). The place of contrastive analysis and error analysis in second language pedagogy. *Journal of Legal and Social Sciences, 8*(4), 117–140.

Mirzaeian, V. R. (2025). A comparative evaluation of artificial intelligence scoring versus human scoring of EFL students' essays. *Teaching English as a Second Language Quarterly, 44*(1), 97–117.

Mohebbi, H., & Truscott, J. (2021). 25 years on, the written error correction debate continues. *Asian-Pacific Journal of Second and Foreign Language Education, 6*(3). https://sfleducation.springeropen.com/articles/10.1186/s40862-021-00110-9

Park, J., Kim, Y., Lee, K., Yoon, Y., Kang, S., Heesang, K., … & Park, H. (2019). Artificial neural network learns clinical assessment of spasticity in modified Ashworth scale. *Archives of Physical Medicine and Rehabilitation, 100*(10), 1907-1915. https://doi.org/10.1016/j.apmr.2019.03.016

Parameswari, D. A., Manickam, R., Dhas, J. A., Kumar, M. V., & Manikandan, A. (2024). Error analysis in second language writing: An intervention research. *World Journal of English Language, 14*(3), 130–145. https://doi.org/10.5430/wjel.v14n3p130

Perplexity.ai. (2025). *AI-powered answer engine*. Retrieved from https://www.perplexity.ai

Ramazani, A., Bijani, H., & Oroji, M. R. (2025). Comparative analysis of AI vs. human feedback effects on IELTS candidates' writing performance. *Journal of Foreign Language Teaching and Translation Studies, 10*(1), 17–40.

Ridha, N. (2012). The effect of EFL learners' mother tongue on their writings in English: An error analysis study. *Journal of the College of Arts, University of Basrah, 60*, 22–45.

Shakir, M., Rasool, A., & Khan, M. (2020). Error analysis in English as a second language students' writing. *International Journal of Innovation, Creativity and Change*, *14*(8), 151–168. https://ijicc.net/images/Vol_14/Iss_8/151068%20Shakir%202020%20E1%20R.pdf

Touchie, H. Y. (2024). Second language learning errors: Their types, causes, and treatment. *JALT Journal*, *8*(1), 75–84.

Xu, M., & Zeng, S. (2023). Optimal timing of treatment for errors in second language learning: A systematic review of corrective feedback timing. *Frontiers in Psychology*, *14*, Article 1026174. https://doi.org/10.3389/fpsyg.2023.1026174

Yang, Y., & Ren, H. (2025). The efficacy of the corpus-based error correction method on revision in writing classrooms. *PLOS ONE*, *20*(3), e0317574. https://doi.org/10.1371/journal.pone.0317574