Teaching Cohesion Across Genres: Pedagogical Insights from Learners' Academic Writing

Salman Hintaw Abdulhussien – Leila Shoja

DOI: 10.18355/XL.2026.19.04.07

Abstract
Cohesion is a key aspect of writing proficiency. This study examines cohesive device use in 104 paired IELTS Task 1 and Task 2 scripts written by 52 Iranian candidates. Using Halliday and Hasan's (1976) framework, cohesive ties were coded and classified. Results revealed significant genre-mediated differences: Task 1 showed higher cohesion density, reflecting descriptive demands, while Task 2 exhibited greater cohesion accuracy, indicating more selective use. Reference was the most frequent but least accurate device, while lexical cohesion was the most accurate. The findings highlight the need for genre-sensitive writing instruction that emphasizes cohesion efficiency over density.
Key words: cohesion, cohesion density, cohesive device accuracy, academic writing, systemic functional linguistics

Introduction

Cohesion, as a central feature of effective writing, enables readers to construct meaning by tracing cohesive connections across clauses and sentences. In Halliday and Hasan's (1976) seminal model, cohesion results from the use of reference, conjunction, and lexical relations that signal how different parts of a text are connected to one another. These cohesive ties play an essential role in written communication, particularly in academic and expository genres where clarity, coherence, and semantic precision are critical. In second-language (L2) writing, cohesion is regarded as a key indicator of proficiency as well as a challenge for learners, who may struggle to use cohesive devices appropriately or may rely on them excessively (Crewe, 1990; Halliday & Hasan, 1976; Hyland, 2004).

The IELTS writing exam offers a valuable context for studying cohesion as test-takers are required to produce two different writing tasks under timed conditions. Task 1 requires candidates to summarize visual data in a concise, factual manner, placing strong demands on referential clarity, comparative structures, and logical sequencing. On the other hand, Task 2 is an essay that requires argumentation, idea development, and lexical flexibility. Since both tasks are produced by the same writer in the same sitting, they provide a unique opportunity to examine within-writer differences in cohesion across genres and proficiency levels. Despite these advantages, few studies (e.g., Abbaspour & Mathew, 2025; Banerjee et al., 2007) have compared cohesion in IELTS writing tasks using detailed coding and statistical analysis.

Based on previous research, it seems that writers with lower proficiency tend to use cohesion densely, often favoring repetition and explicit linking devices (Crossley & McNamara, 2012; McNamara et al., 2010). However, writers with higher proficiency tend to use cohesion less densely, focusing on lexical variety, conceptual coherence, and implicit rhetorical structuring. Yet, empirical evidence for these patterns, especially in high-stakes testing contexts, remains mixed. This is partly due to the fact that most studies rely on holistic or automated measures of cohesion, which cannot explore certain accuracy, misuse, or underuse patterns that human coding can capture.

The present study is, thus, an attempt to partly bridge these gaps by conducting a cohesion analysis of 104 IELTS Task 1 and Task 2 responses using Halliday and Hasan's (1976) taxonomy. The goal is to examine how cohesive device use differs between the tasks, and what this may reveal about L2 writing development, genre variation, and proficiency. Thus, the study investigates these questions:

1. What are the most frequently used cohesive devices within each IELTS writing task?

2. How does cohesion density differ between Task 1 and Task 2?

3. Is there a significant difference in overall cohesion accuracy between Task 1 and Task 2?

4. Is there a significant difference in cohesion accuracy by category between Task 1 and Task 2?

Review of the Literature

Halliday and Hasan's Framework

Halliday and Hasan's (1976) model remains the canonical framework of cohesion in English. They define cohesion as a set of linguistic resources that create semantic ties between textual elements, thus enabling readers to understand a text as a connected whole. Cohesion is accomplished through five major systems: reference, substitution, ellipsis, conjunction, and lexical cohesion. Substitution and ellipsis occur frequently in spoken discourse, while written discourse tends to rely more on reference, conjunction, and lexical cohesion; the latter group forms the basis for most contemporary cohesion analyses.

Reference includes pronominals, demonstratives, and comparatives that signal how entities in discourse relate to one another. Conjunction works by encoding logical relations such as addition, contrast, cause, and time. Lexical cohesion involves repetition, synonymy, and superordinates that create continuity across clauses. Together, these systems form the backbone of textual cohesion and are used to assess the sophistication of L2 writing.

Cohesion in L2 Writing

Research has revealed that L2 writers are inclined to use explicit cohesive markers more frequently than native speakers do, often producing texts that are cohesive yet too dense or mechanical (McNamara & Kintsch, 1996; Wolfe-Quintero et al., 1998). L2 writers at lower proficiency levels frequently overuse additive and temporal conjunctions, and deploy lexical repetitions to create surface-level coherence. As their proficiency develops, writers typically depend less on explicit cohesive ties, show greater lexical variety, increase accuracy in referential expressions, and draw on more sophisticated connective forms. In fact, with increased proficiency, cohesion is increasingly achieved through rhetorical organization rather than through overt markers alone (Crossley & McNamara, 2012; Norris & Ortega, 2009). Nevertheless, developmental progress is not uniform across cohesion categories; research suggests that reference accuracy often lags behind improvements in lexical cohesion (Liu & Braine, 2005), and conjunctions may remain overused or underused depending on genre and task demands (Field & Oi, 1992; Granger & Tyson, 1996; Hyland, 2004).

Findings from analyses by tools like Coh-Metrix (Graesser et al., 2004) and TAACO (Crossley et al., 2016) indicate that cohesion density is not always an indicator of writing quality. Higher-rated essays usually contain fewer cohesive ties overall, show greater variation in cohesive device selection, display increased lexical sophistication, and rely more on implicit cohesion through argumentation rather than explicit connectors (Crossley & McNamara, 2012; McNamara et al., 2010). Therefore, the

common assumption that more linking words necessarily may improve text quality is challenged, and the importance of cohesion accuracy and efficiency is highlighted.

Genres and Cohesion

Cohesion is influenced by genre as different communicative purposes elicit distinct cohesive patterns (Halliday & Hasan, 1989). For example, narrative writing often relies on temporal markers and pronominal reference, whereas expository and argumentative writing tend to employ more causal and logical connectors (Hyland, 2004; Latifi & Gierl, 2021; Taboada, 2004). Academic genres also demand a high degree of lexical cohesion to maintain ideational clarity and to ensure that conceptual relations are clearly marked for readers.

Genre-based variation can become especially salient in standardized, timed writing assessments such as IELTS. Task 1 requires the description and comparison of visual data, and Task 2 is an argument-driven essay; thus, they may depend on different patterns of cohesion (Coffin & Hewings, 2004; Moore & Morton, 2005; Qin & Uccelli, 2016). Therefore, examining cohesion across Task 1 and Task 2 provides valuable insights into how writers adjust their cohesive behavior as they shift from factual reporting to argumentative reasoning. In this regard, Liu et al. (2025) applied an SFL-based framework to examine cohesion across multiple genres. Their findings indicate clear genre differences in cohesion density, with argumentative writing showing the highest density and narrative the lowest. Descriptive writing, structurally parallel to IELTS Task 1, displayed a heavy use of demonstratives and definite articles because of its reliance on referring to visual elements. The study also uncovered frequent overuse of pronominals, repetitions, and demonstratives, along with underuse of conjunctions.

Research on Cohesion in IELTS Writing

IELTS writing research has consistently shown that Task 1 is more formulaic, characterized by predictable cohesive patterns, while Task 2 is less predictable, with greater lexical and rhetorical flexibility (Shaw & Falvey, 2008; Weigle, 2002). Higher band scores are also generally associated with more accurate cohesive use, greater lexical cohesion sophistication, fewer redundant conjunctions, and more implicit conceptual cohesion. Lower-proficiency writers, by contrast, frequently rely on repetition and additive structures, and have inaccuracies in referential expressions (Coffin & Hewings, 2004; Moore & Morton, 2005). Despite these insights, relatively few studies embark on detailed hand-coded cohesion analysis or compare paired texts produced by the same writers under identical testing conditions.

Automated Research on Cohesion

Over the past two decades, natural language processing (NLP) applications have become central to cohesion research. Despite their analytical power, findings from NLP indices have been inconsistent. McNamara et al. (2010), for instance, reported that Coh-Metrix cohesion measures do not significantly predict human ratings of essay quality, whereas later work with TAACO found that certain cohesion indices, especially in integrated writing tasks, correlate with writing quality (Tywoniw & Crossley, 2019).

A persistent critique of NLP-based cohesion analysis, according to Liu et al. (2025), is that frequency-based indices cannot distinguish between appropriate and inappropriate cohesive device use. Their surface-driven, computational approach contrasts with cohesion theory, as highlighted by Halliday and Hasan (1989) and Crewe (1990), which stresses that cohesion must be evaluated in terms of appropriateness as well as its texture contribution. The growing recognition of this limitation has resulted

in a renewed interest in manual, quality-oriented coding approaches grounded in systemic functional linguistics (SFL).

Quality-Based Research: Misuse, Overuse, and Underuse

To address the limitations of frequency-based NLP indices, recent studies have increasingly adopted manual quality-focused coding. They have shown that L2 writers frequently misuse conjunctions (Cheng & Tsang, 2022), and often overuse pronominals and repetitions (Abdul Rahman, 2013; Alyousef, 2020). Underuse of essential cohesive devices, such as articles and demonstratives, has also been documented (Robertson, 2000). Yang and Sun (2012) showed that correct cohesion use correlates with higher writing quality, although error patterns vary substantially across subtypes.

In summary, despite considerable progress in cohesion research, several gaps remain. Many studies use automated frequency-based tools, which cannot distinguish misuse or underuse and overlook appropriateness. Genre-oriented effects are also insufficiently explored, with very few studies comparing two distinct genres written by the same individual.

Methodology

Participants

The participants in this study included 52 Iranian learners who took part in IELTS mock examinations at a private language assessment institute. All participants had obtained an overall IELTS writing band score of 6–6.5, which was deliberately selected to ensure proficiency homogeneity within the sample. By restricting the dataset to this proficiency range and, in turn, minimizing variability attributable to proficiency differences, the study aims to provide a focused and internally consistent analysis of cohesive device use at an upper-intermediate level.

The decision to focus exclusively on Band 6–6.5 candidates was theoretically and methodologically motivated. Band 6 is a crucial developmental threshold in IELTS writing as candidates are expected to demonstrate 'clear overall progression' and 'some effective use of cohesive devices' (IELTS, 2022), yet their writing often still contains inaccuracies. Thus, by choosing this proficiency band, the study aims to develop a detailed profile of cohesion competence at a critical stage, while also creating a baseline against which future studies may compare higher or lower-band performances. It should be pointed out that all participants shared the same first language (Persian) and educational context, which further enhanced the internal consistency of the data.

Data Collection

The data consisted of paired IELTS writing Task 1 and Task 2 scripts produced by each participant under simulated test conditions closely mirroring the official IELTS writing examination. The mock tests were administered by an institute that conducts IELTS preparation and assessment programs.

The participants were assigned one Task 1 prompt and one Task 2 prompt. All the test takers completed both tasks within the given sixty minutes, following the standard IELTS format, without access to electronic devices and external resources. Test administration conditions, including time constraints, task sequencing, and instructions, were designed to replicate the actual IELTS writing test as closely as possible to maximize ecological validity.

Following test completion, all scripts were assessed by trained IELTS mock examiners who had participated in official IELTS examiner training courses and had several years of experience in IELTS writing assessment. The examiners evaluated the scripts using IELTS writing band descriptors. Only the scripts that received an overall

Writing band score of 6–6.5 were kept for analysis. Thus, the final dataset consisted of 52 paired Task 1 and Task 2 texts, yielding a total of 104 writing samples.

Coding Framework

Cohesive device analysis was conducted using Halliday and Hasan's (1976) model that conceptualizes cohesion as a semantic relation realized through linguistic ties within a text. Following this framework, the analysis focused on three major cohesive categories: conjunction, reference, and lexical cohesion. Ellipsis and substitution categories of the model were not included in the study due to their relative rarity in written academic discourse. Within each category, cohesive devices were coded into the following subtypes:

- reference: pronominals, demonstratives and definite articles, and comparatives
- conjunction: temporal, additive, causal, and adversative conjunctions
- lexical cohesion: repetition, synonymy, and superordinates

All cohesive ties were identified manually and coded at the sentence level. Each occurrence was further classified according to its quality of use, drawing on Crewe's (1990) distinction between:

- correct use, where the cohesive device contributed appropriately to textual meaning;
- misuse, where the device was present but impeded clarity or coherence; and
- underuse, where an expected or contextually necessary cohesive link was absent.

Overuse was initially coded during the preliminary analysis but was later excluded from the analysis because the raters did not agree on several instances, particularly in the cases of repetition and definite articles.

Coding Procedure

Two coders, both authors of the present study, conducted the cohesion analysis. Before the main coding procedure, both thoroughly reviewed the key theoretical and empirical literature on cohesion, including Halliday and Hasan (1976). Then, they selected one Task 1–Task 2 pair at random from the dataset. This pair was independently coded while the coders frequently consulted the framework and examples from the literature. This pilot coding allowed the coders to refine category boundaries, clarify subtype distinctions, and align their understandings of correct use, misuse, and underuse.

Then, all 104 scripts were co-coded by the two coders. Rather than coding independently and calculating inter-rater reliability post hoc, the coders worked together to resolve disagreements as they arose. During this process, they referred back to the established coding guidelines to ensure consistent application of the framework. This collaborative approach was adopted to increase analytic precision and consistency, particularly because of the nature of the cohesion categories and the importance of qualitative judgment in distinguishing misuse and underuse.

Data Analysis

After coding was completed, raw frequencies of cohesive devices were calculated for each script and aggregated across each task. Because the tasks differed in length, all frequencies were normalized per 1,000 words before statistical analysis. Accuracy was operationalized as the proportion of correctly used cohesive devices relative to total

cohesive device use (excluding underuse). Descriptive statistics were used for overall cohesion density and category-level cohesion, while inferential analyses including Wilcoxon Signed Rank tests, Friedman tests, and effect size measures were employed to examine differences across cohesion categories.

It should be declared that this work is the result of the writers' own intellectual effort. ChatGPT, as an AI tool, was used for grammar checking and improving the text readability.

Results

To present the results, first, within-task analyses are reported, and then between-task comparisons are done.

Within-Task Cohesion

Cohesive Device Use by Category in Task 1

To examine whether IELTS Task 1 scripts differed in the use of the three major cohesive categories, a Friedman test was conducted. The results showed a statistically significant difference among the categories, $\chi^2(2, N = 52) = 52.24$, $p < .001$. Mean rank values showed that reference was the most common (mean rank = 2.77), followed by lexical cohesion (mean rank = 1.84), while conjunctions were used least frequently (mean rank = 1.39). Follow-up Wilcoxon Signed Ranks Tests revealed that the differences between reference and conjunction use ($Z = -5.65$, $p < .001$), reference and lexical cohesion use ($Z = -5.84$, $p < .001$), as well as lexical cohesion and conjunction use ($Z = -2.75$, $p = .006$) were significant. These show a hierarchy of cohesive device use in Task 1, with reference cohesion dominating overall usage.

Accuracy of Cohesive Device Use in Task 1

Another Friedman test was conducted to examine differences in the accuracy rates of reference, conjunction, and lexical cohesion in Task 1. The test showed a significant difference among the categories, $\chi^2(2, N = 52) = 28.19$, $p < .001$. Mean ranks indicated that conjunctions (mean rank = 2.31) and lexical cohesion (mean rank = 2.23) showed higher accuracy than reference cohesion (mean rank = 1.46). Then, post-hoc Wilcoxon signed-rank tests confirmed that conjunction accuracy was significantly higher than reference accuracy ($Z = -4.41$, $p < .001$), and lexical cohesion accuracy was also significantly higher than reference accuracy ($Z = -4.24$, $p < .001$). The accuracy rates for conjunction and lexical cohesion showed no significant difference ($Z = -0.09$, $p = .92$). Thus, although reference was the most common category in Task 1, it was the least accurate.

Underuse of Cohesive Devices in Task 1

To compare underuse across the categories in Task 1, another Friedman test was conducted and revealed a significant difference, $\chi^2(2, N = 52) = 42.28$, $p < .001$. Mean rank values showed that conjunctions were most frequently underused (mean rank = 2.52), followed by reference cohesion (mean rank = 2.16), while lexical cohesion showed the lowest underuse (mean rank = 1.32). Wilcoxon Signed Rank tests revealed that the difference between conjunction and reference underuse did not reach statistical significance ($Z = -1.87$, $p = .06$). However, lexical cohesion was significantly less underused than reference cohesion ($Z = -4.82$, $p < .001$) and conjunctions ($Z = -5.34$, $p < .001$). Overall, underuse was particularly clear for conjunctions in Task 1, whereas lexical cohesion was comparatively well-supplied.

Cohesive Device Use by Category in Task 2

Another Friedman test was run to examine differences in the overall use of the categories in Task 2. The results revealed a significant effect of category, $\chi^2(2, N = 52) = 70.98$, $p < .001$. Reference cohesion was again the most common category (mean rank = 2.94), followed by conjunctions (mean rank = 1.64), with lexical cohesion used least

frequently (mean rank = 1.41). Specifically, post hoc Wilcoxon Signed-Rank tests indicated that reference cohesion was used significantly more frequently than conjunctions ($Z = -6.24$, $p < .001$) and lexical cohesion ($Z = -6.24$, $p < .001$). Conjunctions were also used significantly more commonly than lexical cohesion ($Z = -2.47$, $p = .013$). These findings point to a strong dominance of reference as a cohesive device in Task 2.

Accuracy of Cohesive Device Use in Task 2

To assess accuracy across cohesive categories in Task 2, a Friedman test was conducted and yielded a significant result, $\chi^2(2, N = 52) = 42.01$, $p < .001$. Lexical cohesion was the most accurate category (mean rank = 2.53), followed by conjunctions (mean rank = 2.08), while reference cohesion showed the lowest accuracy (mean rank = 1.39). Then, Wilcoxon Signed-Rank tests revealed that conjunction accuracy was significantly higher than reference accuracy ($Z = -4.50$, $p < .001$), and lexical cohesion accuracy was also significantly higher than reference accuracy ($Z = -5.60$, $p < .001$). Additionally, lexical cohesion accuracy was significantly higher than conjunction accuracy ($Z = -3.16$, $p = .002$). These results indicate a clear accuracy hierarchy in Task 2, with lexical cohesion being the most accurate and reference the least.

Underuse of Cohesive Devices in Task 2

To examine underuse across categories in Task 2, another Friedman test was conducted. It revealed a significant difference among the categories, $\chi^2(2, N = 52) = 30.46$, $p < .001$. Conjunctions were most frequently underused (mean rank = 2.58), followed by reference cohesion (mean rank = 1.86), while lexical cohesion exhibited the lowest underuse (mean rank = 1.57). Post-hoc Wilcoxon signed-rank tests then indicated that conjunction underuse was significantly greater than reference underuse ($Z = -3.89$, $p < .001$). Statistically, the difference between lexical and reference underuse was not significant ($Z = -1.90$, $p = .057$). However, lexical cohesion was significantly less underused than conjunctions ($Z = -5.04$, $p < .001$). These findings suggest that underuse of conjunctions remains a prominent issue in Task 2, whereas lexical cohesion is comparatively well-developed.

Between-Task Cohesion Comparisons

Overall Cohesion Density

Task 1 texts contained a total of 3,261 cohesive devices across 10,193 words, yielding an overall cohesion density of 319.93 devices per 1,000 words. Task 2 texts had 3,994 cohesive devices across 15,876 words, corresponding to an overall cohesion density of 251.58 devices per 1,000 words. Table 1 presents the use patterns of these in terms of different quality measures.

Table 1

Overall Cohesion Density in Tasks 1 and 2

Measure	Task 1 (Mean)	Task 2 (Mean)
Correct use	297.47	239.64
Misuse	25.11	14.24
Underuse	60.28	34.01
Total use	319.93	251.58

According to the table, while Task 2 had a higher raw number of cohesive devices, Task 1 exhibited a considerably higher normalized cohesion density. Based on these, a series of Wilcoxon Signed Rank tests (Table 2) were utilized to compare the normalized

cohesive device frequencies per 1,000 words for Tasks 1 and 2 across four measures: correct use, misuse, underuse, and total cohesive device use.

Table 2

Wilcoxon Signed Rank Tests for Cohesion Measures Between Tasks 1 and 2

Comparison	Negative ranks (n)	Positive ranks (n)	Z	p
Correct use	45	7	−5.42	< .001
Misuse	34	17	−3.34	< .001
Underuse	42	10	−5.26	< .001
Total cohesion	46	6	−5.65	< .001

The results reveal statistically significant differences between the tasks across all cohesion measures. Regarding correct cohesive device use, 45 out of 52 candidates produced fewer correctly used cohesive devices in Task 2 than in Task 1, indicating a consistent within-writer reduction in correct cohesion density when moving from descriptive to argumentative writing. About misuse, the majority of cases (34 negative ranks) indicated that Task 2 contained fewer misused cohesive devices than Task 1. This suggests that although Task 2 texts included fewer cohesive devices overall, candidates were also less prone to incorrect deployment of cohesion in argumentative writing. The difference in underuse was likewise significant; most candidates (42 negative ranks) exhibited lower underuse rates in Task 2, indicating that Task 1 writing involved more frequent omission of contextually required cohesive ties. Finally, regarding total cohesion density, in 46 of the 52 paired comparisons, Task 2 showed lower total cohesion density than Task 1. This confirms that the descriptive task elicited significantly denser cohesion than the argumentative task.

Total Cohesion Accuracy Between Tasks

To address the third question of the study about overall cohesion accuracy in Tasks 1 and 2, a Wilcoxon Signed-Rank test was conducted on normalized accuracy scores (Table 3).

Table 3

Wilcoxon Signed Rank Test for Total Cohesion Accuracy

Comparison	Negative ranks (n)	Positive ranks (n)	Ties (n)	Z	*p*
Task 2–task 1 accuracy	21	30	1	−2.43	.015

The analysis revealed a statistically significant difference in total accuracy. As shown in Table 3, a greater number of participants demonstrated higher cohesive accuracy in Task 2 (n = 30) than in Task 1 (n = 21), with only one tie. This indicates that although Task 2 generally involved lower cohesion density, candidates used cohesive devices more accurately in the argumentative task than in the descriptive task.

Cohesive Device Use by Category

Category-level analyses were conducted to examine whether the observed task differences extended across major cohesion types. The normalized frequencies were compared using Wilcoxon Signed Rank tests (Table 4). Results indicated significant task-based differences across all three categories. Reference cohesion was significantly denser in Task 1 than in Task 2. Conjunction use also showed a significant reduction from Task 1 to Task 2. The largest category-level difference was observed for lexical cohesion, where Task 2 exhibited markedly lower values than Task 1. Based on the rank distributions, lexical cohesion exhibited the most sensitive category to task change, followed by conjunction and then reference cohesion.

Table 4

Wilcoxon Signed-Rank Tests for Category-Level Cohesion Density

Category	Negative ranks (Task 2 < Task 1)	Positive ranks (Task 2 > Task 1)	Z	p
Reference	37	15	−3.02	.002
Conjunction	35	17	−3.58	<.001
Lexical cohesion	50	2	−6.12	<.001

Between-Task Categorical Cohesion Accuracy

To address the fourth research question, a Wilcoxon Signed-Rank test (Table 5) was run to compare accuracy proportions across the three categories between Task 1 and Task 2. For reference cohesion, the difference was not statistically significant, indicating no reliable improvement in the accuracy of reference ties. For conjunction cohesion, the Wilcoxon test also revealed no significant change in accuracy, suggesting that in both tasks, conjunctions were used with similar accuracy levels, despite an overall reduction in conjunction density in Task 2. For lexical cohesion, however, accuracy improved significantly from Task 1 to Task 2, indicating a medium effect size. In fact, in Task 2, the writers used repetition, synonymy, and superordinate relations more accurately than in Task 1, reflecting a clearer, more coherent lexical-semantic structure.

Table 5

Wilcoxon Signed-Rank Tests for Category-Level Accuracy

Category	Negative ranks (Task 2 < Task 1)	Positive ranks (Task 2 > Task 1)	Ties	Z	p
Reference	22	29	1	−1.56	.120
Conjunction	17	20	15	−0.75	.455
Lexical cohesion	9	20	23	−3.02	.003

Discussion

The results of this research indicated that in both tasks, reference emerged as the most common cohesive device. In Task 1, it significantly exceeded both conjunctions and

lexical cohesion, and lexical cohesion was also used significantly more than conjunctions. A similar order was observed in Task 2, where reference again dominated, followed by conjunctions and lexical cohesion; of course, the gap between the categories was more pronounced. Similarly, Yang and Sun (2012) as well as Liu et al. (2025) reported that two subtypes of the reference category (demonstratives/definite articles and pronominals) and one subtype of lexical cohesion (repetition) were used most frequently in their corpora. The use of 'the' might explain why reference outnumbered lexical cohesion and conjunction as cohesive devices (Liu & Braine, 2005).

In contrast, accuracy patterns were not parallel in the tasks. In Task 1, reference accuracy was substantially lower than both conjunction and lexical cohesion accuracy, while no significant difference was observed between conjunction and lexical accuracy. This shows that although reference was the most common category in Task 1, it was also the most error-prone; this reveals test takers' persistent difficulty with articles, demonstratives, and pronominal clarity in descriptive reporting. In Task 2, accuracy patterns showed a clearer stratification. Lexical cohesion had the highest accuracy, followed by conjunctions, with reference again ranking the lowest. Unlike Task 1, all pairwise differences in Task 2 were statistically significant, indicating that argumentative writing highlighted accuracy differences between the categories. This suggests that while the candidates had more control over lexical relations and logical connectors in Task 2, referential accuracy was still a persistent challenge even at higher discourse levels. This is in line with previous claims that non-native writers have difficulty using referential devices consistently and effectively, especially in the case of definite articles, which are either overused or are confused with indefinite articles (Liu & Braine, 2005; Master, 1997; Meisuo, 2000).

Underuse analyses further uncover task-related contrasts in cohesive control. In Task 1, conjunction underuse was significantly higher than lexical underuse, while reference underuse occupied an intermediate position and did not differ significantly from conjunction underuse. This pattern suggests that Task 1 writers often did not use logical connectors, and instead relied on referential repetition and lexical continuity to structure descriptions. In Task 2, conjunction underuse was again the most prominent, significantly exceeding both reference and lexical underuse. However, unlike Task 1, Task 2 did not show significant differences between reference and lexical underuse; it rather indicated a more balanced but still incomplete use of these resources in argumentative contexts. The persistent conjunction underuse in both tasks underscores a systemic difficulty in employing overt logical relations, even when argumentative discourse would strongly warrant them. Similarly, Liu et al. (2025) observed an underuse of four conjunctive subtypes in their study in all the genres they investigated.

Overall, the within-task category comparisons reveal a mismatch between frequency and accuracy, particularly for reference. While reference dominates cohesive use in both Task 1 and Task 2, its low accuracy may reflect a critical challenge for Band 6–6.5 candidates. This pattern shows that cohesion development in L2 writing is not uniform across categories.

The second question of the study examined overall cohesion density in Tasks 1 and 2. The findings uncover a clear task-based distinction, with Task 1 exhibiting significantly higher normalized cohesion density than Task 2. From an SFL perspective, this difference may reflect task-driven discourse demands. Based on Halliday and Hasan (1976), cohesion is not an abstract property of texts but a functional resource used to attain certain communicative goals. Task 1, as a descriptive visual data report genre, requires writers to constantly re-connect referents, compare things, and maintain explicit continuity between visual elements and linguistic expressions. As a result, high

frequencies of demonstratives, definite articles, comparatives, and repetitions are functionally motivated. The dense cohesion observed in Task 1, therefore, may reflect genre-mediated explicitness, rather than advanced discourse control.

In contrast, Task 2 is an argumentative genre in which meaning is increasingly constructed through logical progression and stance development rather than surface-level textual ties. The lower cohesion density in Task 2 aligns with prior findings that higher-quality argumentative writing relies less on explicit cohesion and more on implicit rhetorical coherence (Crossley & McNamara, 2012; McNamara et al., 2010). However, contrary to our findings, Liu et al. (2025) found that argumentative writing usually has higher numbers of cohesive ties than other genres (i.e., descriptive, narrative, and interactive). This mismatch can be explained by the fact that their participants were from a wider range of proficiency levels (A2-C1), or the fact that their argumentative writing topics were not identical to those of the IELTS second writing task. Moreover, in their study, the four genres had been the focus of classroom activities. These patterns caution against considering higher cohesion density as a marker of superior writing ability. In IELTS writing, dense cohesion in Task 1 is expected, while selectivity in Task 2 is more consistent with higher-level discourse competence.

The third research question of the study addressed if overall cohesive accuracy differed across tasks. The results showed that Task 2 exhibited significantly higher total accuracy than Task 1, despite its lower cohesion density. This finding is theoretically significant because it underscores the distinction between quantity and quality of cohesion, a distinction central to Halliday and Hasan's conception of texture. Within the SFL standpoint, cohesion contributes to texture only when it is functionally appropriate. Over-explicit or mechanically applied cohesive ties may actually disrupt textual flow. The lower accuracy observed in Task 1 may suggest that although writers use many cohesive devices, they constantly struggle with precision in reference, article choice, and comparative structures, leading to misuse or underuse. This is in line with previous research showing that descriptive tasks can magnify article-related difficulties for L2 learners (Ionin et al., 2004; Robertson, 2000; Trenkic, 2007). From a genre perspective, descriptive tasks place unusually high demands on reference tracking, requiring frequent identification, re-identification, and comparison of things (Halliday & Hasan, 1976; Hyland, 2004). In fact, recent SFL-based work has suggested that descriptive genres elicit particularly dense use of demonstratives and definite articles, often along with high rates of misuse and overuse (Liu et al., 2025).

In Task 2, by contrast, writers had greater control over cohesive device use, even while using fewer devices overall. This suggests that argumentative writing, despite being cognitively demanding, may encourage more deliberate and selective cohesion use, especially with causal and adversative conjunctions. From a developmental perspective, this pattern is consistent with the view that advanced cohesion involves reduction, not expansion, of explicit ties as writers rely more on discourse-level organization (Crossley & McNamara, 2012; Norris & Ortega, 2009; Taboada, 2004). While IELTS rubrics do not state this explicitly, Shaw and Falvey (2008) as well as Coffin and Hewings (2004) note that overuse of overt connectors in argumentation can negatively affect perceived quality. This finding matches IELTS band descriptors for Band 6–6.5, which emphasize some effective use of cohesive devices but note that inaccuracies and inefficiencies remain. The data empirically support this descriptor by showing that candidates can exhibit emerging cohesion control in Task 2 while still struggling with accuracy under the referential pressure of Task 1.

The fourth research question explored whether the three major cohesion categories differed in accuracy across the tasks. The results revealed a redistribution of accuracy across categories, highlighting the functional specialization of cohesion types across genres. Lexical cohesion showed the highest accuracy across both tasks, with Task 2 demonstrating greater lexical variety and more effective use of synonyms and superordinates. This pattern is in line with previous research indicating that lexical cohesion is less grammatically dependent and, therefore, more accurately used by L2 writers than reference or conjunctions (Halliday & Hasan, 1976; Liu et al., 2025; Yang & Sun, 2012). Genre-based studies further suggest that argumentative writing encourages lexical reformulation and abstraction, resulting in richer lexical patterns despite lower overall cohesion density (Coffin & Hewings, 2004; Hyland, 2004). From a developmental perspective, lexical cohesion appears to mature earlier and establish faster than grammatical cohesion. These results stabilize the role of lexical cohesion as a key interface between surface cohesion and conceptual coherence in extended academic argumentation. Taken together, these category-level findings reveal that cohesion development is not uniform, but functionally driven by genre demands.

Conclusion

The findings of this study suggest that cohesion in IELTS writing is task-sensitive, and cohesion density does not correspond directly to cohesion quality. In fact, writers adapt their cohesive strategies in response to genre demands, but accuracy gains are uneven across cohesion types. These findings extend previous NLP-based research by showing that frequency-based measures alone obscure critical patterns of misuse and underuse, and they provide empirical support for SFL-based arguments that cohesion must be evaluated in relation to genre and communicative purpose.

This study can advance cohesion research by offering a detailed, quality-oriented account of cohesion in academic writing and can carry implications for academic writing pedagogy. First, the results suggest that teaching cohesion as a matter of *using more linking words* is pedagogically misguided. Instruction should, therefore, shift from encouraging frequency toward fostering precision in referential choice, including controlled practice with definite articles, demonstratives, and comparative structures in descriptive contexts. For Persian-speaking learners, targeted instruction on articles and reference is especially crucial.

Second, the contrast between the tasks indicates that they require different cohesion strategies, and these differences should be made explicit to learners. In Task 1, students need support in managing dense reference chains without redundancy or misuse. In Task 2, instruction should emphasize selective cohesion, helping learners avoid mechanical overuse of conjunctions and focus on logical progression, paragraph development, and lexical reformulation.

Third, the consistently high accuracy of lexical cohesion across tasks highlights the value of lexical development as a core pathway to writing quality. Lexical cohesion exercises can serve as an effective bridge between vocabulary instruction and discourse-level writing skills.

Fourth, the persistent underuse of conjunctions, especially causal and adversative ties, suggests that learners may struggle to use logical relations linguistically. Rather than teaching conjunction lists in isolation, instructors should integrate conjunction use into argument structure training, helping students link claims, reasons, and counterarguments to appropriate logical connectors.

Finally, the study underscores the importance of genre in IELTS preparation. Teachers should avoid applying uniform cohesion criteria across tasks and instead tailor feedback to the communicative demands of each task. Such an approach aligns more

closely with IELTS band descriptors and supports learners in developing context-sensitive cohesive competence.

While the present study offers insights into cohesion device use in IELTS writing, its limitations should be acknowledged. Overuse was excluded from the analysis due to coding disagreements, suggesting the need for further refinement in operationalizing this category. In addition, learners of different proficiency levels were not included in the study. Future research could extend this study to higher and lower proficiency bands, incorporate longitudinal designs, or compare manual SFL-based coding with NLP-derived cohesion indices. Such work would further clarify how cohesion develops across proficiency levels and how assessment practices might better capture cohesion quality rather than density.

Statement on the Use of AI

This work is the result of the writers' own intellectual effort. ChatGPT, as an AI tool, was used for grammar checking and improving the text readability.

Bibliographic references

Abbaspour, E., & Mathew, P. (2025). Evaluating Cohesion as a Predictor of Writing Quality: An Analysis of Local, Global, and Text-level Indices in IELTS Writing Task 2. International Journal of Practical and Pedagogical Issues in English Education, 3(4), 97-117. https://doi.org/10.22034/ijpie.2025.535152.1116

Abdul Rahman, Z. A. A. (2013). The Use of Cohesive Devices in Descriptive Writing by Omani Student-Teachers. Sage Open, 3(4), 1-10. https://doi.org/10.1177/2158244013506715

Alyousef, H. S. (2020). An SF-MDA of the Textual and the Logical Cohesive Devices in a Postgraduate Accounting Course. Sage Open, 10(3). https://doi.org/10.1177/2158244020947129

Banerjee, J., Franceschina, F., & Smith, A. M. (2007). Documenting Features of Written Language Production Typical at Different IELTS Band Score Levels (IELTS Research Report No. 7). IELTS Australia Pty Ltd and British Council.

Cheng, A., & Tsang, A. (2022). Use and Understanding of Connectives: An Embedded Case Study of ESL Learners of Different Proficiency Levels. Language Awareness, 31(2), 155–174. https://doi.org/10.1080/09658416.2021.1871912

Coffin, C., & Hewings, A. (2004). IELTS as Preparation for Tertiary Writing: Distinctive Interpersonal and Textual Strategies. In L. Ravelli & R. Ellis (eds.), Analysing Academic Writing: Contextualized Frameworks (pp. 153-171). Continuum. ISBN 9780826461070

Crewe, W. J. (1990). The Illogic of Logical Connectors. ELT Journal, 44(4), 316–325. https://doi.org/10.1093/elt/44.4.316

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The Development and Use of Cohesive Devices in L2 Writing and Their Relations to Judgments of Essay Quality. Journal of Second Language Writing, 32, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

Crossley, S. A., & McNamara, D. S. (2012). Predicting Second Language Writing Proficiency: The Roles of Cohesion and Linguistic Sophistication. Journal of Research in Reading, 35(2), 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Field, Y., & Oi, Y. L. M. (1992). A Comparison of Internal Conjunctive Cohesion in the English Essay Writing of Cantonese Speakers and Native Speakers of English. RELC Journal, 23(1), 15–28. https://doi.org/10.1177/003368829202300102

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of Text on Cohesion and Language. Behavior Research Methods, Instruments, & Computers, 36(2), 193–202. https://doi.org/10.3758/BF03195564

Granger, S., & Tyson, S. (1996). Connector Usage in the English Essay Writing of Native and Non-native EFL Speakers of English. World Englishes, 15(1), 17–27. https://doi.org/10.1111/j.1467-971X.1996.tb00089.x
Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. Longman. ISBN 9780582550315
Halliday, M. A. K., & Hasan, R. (1989). Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective (2nd ed.). Oxford University Press. ISBN 9780194371544
Hyland, K. (2004). Disciplinary Discourses: Social Interactions in Academic Writing. University of Michigan Press. ISBN 9780472030248
IELTS (2022). About the Test, Test Format, Academic Writing**.** Retrieved April 14, 2025, from https://www.ielts.org/for-organisations/ielts-scoring-in-detail
Ionin, T., Ko, H., & Wexler, K. (2004). Article Semantics in L2 Acquisition: The Role of Specificity. Language Acquisition, 12(1), 3–69. https://doi.org/10.1207/s15327817la1201_2
Latifi, S., & Gierl, M. (2021). Automated Scoring of Junior and Senior High Essays Using Coh-Metrix Features: Implications for Large-scale Language Testing. Language Testing, 38(1), 62–85. https://doi.org/10.1177/0265532220929918
Liu, J., Shi, Z., & Li, W. (2025). Understanding Genre Differences in Cohesive Use of EFL Learners: Insights from Systemic Functional Linguistics. System, 132, 103692. https://doi.org/10.1016/j.system.2025.103692
Liu, M., & Braine, G. (2005). Cohesive Features in Argumentative Writing Produced by Chinese Undergraduates. System, 33(4), 623–636. https://doi.org/10.1016/j.system.2005.02.002
Master, P. (1997). The English Article System: Acquisition, Function, and Pedagogy. System, 25(2), 215–232. https://doi.org/10.1016/S0346-251X(97)00010-9
McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic Features of Writing Quality. Written Communication, 27(1), 57–86. https://doi.org/10.1177/0741088309351547
McNamara, D. S., & Kintsch, W. (1996). Learning from Texts: Effects of Prior Knowledge and Text Coherence. Discourse Processes, 22(3), 247–288. https://doi.org/10.1080/01638539609544975
Meisuo, Z. (2000). Cohesive Features in the Expository Writing of Undergraduates in two Chinese Universities. RELC Journal, 31(1), 61–95. https://doi.org/10.1177/003368820003100104
Moore, T., & Morton, J. (2005). Dimensions of Difference: A Comparison of University Writing and IELTS Writing. Journal of English for Academic Purposes, 4(1), 43–66. https://doi.org/10.1016/j.jeap.2004.02.001
Norris, J. M., & Ortega, L. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. Applied Linguistics, 30(4), 555–578. https://doi.org/10.1093/applin/amp044
Qin, W., & Uccelli, P. (2016). Same Language, Different Functions: A Cross-Genre Analysis of Chinese EFL Learners' Writing Performance. Journal of Second Language Writing, 33, 3–17. https://doi.org/10.1016/j.jslw.2016.06.001
Robertson, D. (2000). Variability in the Use of the English Article System by Chinese Learners of English. Second Language Research, 16(2), 135–172. https://doi.org/10.1191/026765800672262975
Shaw, S.D., & Falvey, P. (2008). The IELTS Writing Assessment Revision Project: Towards a Revised Rating Scale. Research Reports, 1.
Taboada, M. (2004). Building Coherence and Cohesion: Task-oriented Dialogue in English and Spanish. John Benjamins. ISBN 9027295050

Trenkic, D. (2007). Variability in Second Language Article Production: Beyond the Representational Deficit vs. Processing Constraints Debate. Second Language Research, 23(3), 289–327. https://doi.org/10.1177/0267658307077643
Tywoniw, R., & Crossley, S. (2019). The Effect of Cohesive Features in Integrated and Independent L2 Writing Quality and Text Classification. Language Education & Assessment, 2(3), 110–134. https://doi.org/10.29140/lea.v2n3.151
Weigle, S. C. (2002). Assessing Writing. Cambridge University Press. ISBN 0521780276
Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity. University of Hawaii Press. ISBN 082482069X
Yang, W., & Sun, Y. (2012). The Use of Cohesive Devices in Argumentative Writing by Chinese EFL Learners at Different Proficiency Levels. Linguistics and Education, 23(1), 31–48. https://doi.org/10.1016/j.linged.2011.09.004

Words: 6989
Characters: 48 686 (27 standard pages)

Salman Hintaw Abdulhussien, PhD.
College of Dentistry
University of Al-Ameed, Karbala
Karbala Governorate, Karbala, Najaf-Karbala Road, 56001
Iraq
hintaws2@gmail.com
ORCID: 0000-0002-4993-9655

Leila Shoja, PhD.
Assistant Professor in TEFL
Department of English Language and Literature
Faculty of Literature and Humanities
Ilam University
Iran, Ilam, Pazhouhesh Blvd, M94H+29P
Iran
l.shoja84@gmail.com
ORCID: 0000-0002-1411-9235