INTERNATIONAL JOURNAL OF
Practical and Pedagogical Issues in English Education
ISSN : 2980-9533 Homepage: www.ijpie.org 3(4), Dec. 2025

# Evaluating Cohesion as a Predictor of Writing Quality: An Analysis of Local, Global, and Text-Level Indices in IELTS Writing Task 2

Ehsan Abbaspour[1,*], Priya Mathew[1]
[1] Centre for Academic Writing, Middle East College, Muscat, Oman
Corresponding Author's Email: abbaspour1986@gmail.com

**KEYWORDS**

Cohesion; IELTS, L2 Writing, Global Cohesion, Local Cohesion, Text Cohesion

**ARTICLE INFO**

**Article type:** Research Article

**Article history:**
Received: 27 July 2025
Revised: 14 September 2025
Accepted: 18 October, 2025
Published online: 18 October 2025

**ABSTRACT**

This study examines the predictive validity of local, global, and text-level cohesion indices for human ratings of Cohesion and Coherence (CC) in IELTS Writing Task 2. A corpus consisting of 105 essays rated by official IELTS examiners was analyzed using the Tool for the Automatic Analysis of Cohesion (TAACO 2.0.4). Forty cohesion indices were extracted and correlated with CC scores using Spearman's Rank Correlation. The results revealed that additive connectives and overall connectives use relate positively to CC scores, while too much redundant local lemma overlap relates negatively. In the global category, a positive correlation was found between the overlap of function lemmas and semantic similarity between paragraphs and higher ratings. At the text level, lexical diversity (MATTR and TTR) correlated most strongly positively, while lexical repetition harmed perceived quality. This suggests that selective and varied use of cohesive devices enhances writing quality, whereas repetition detracts from it. Further implications for automated essay scoring and L2 writing instruction are discussed.

## Introduction

In writing research, one must make a distinction between cohesion and coherence. Nonetheless, these terms are interchangeably used in the literature despite their close relationship (Crossley et al., 2016a; Ebrahimi & Mallaki, 2024; Tasouji Azari & Mohammadi, 2024). Cohesion is best described as explicit links in language by which meaning is imparted to a text (Halliday & Hasan, 1976). While at the other end of the scale, coherence will be a certain global idea of the text nurtured in the reader's mind; such an idea can be influenced by the textual features or even the reader's prior knowledge and skills (McNamara et al., 1996; O'Reilly & McNamara, 2007). As Crossley (2020) puts it, cohesion could be referred to as the property of texts by which linguistic units connect different ideas. In contrast, coherence refers to readers' mental representation of the text. Describing the way a writer defines how elements of writing contribute to the overall quality of text and understanding on the part of the reader constitutes a juxtaposition of these two phenomena. Abdi Tabari and Wind (2023) further elaborate on this contrast by stating that cohesion mainly influences short-memory processes, whereas coherence is needed to reinforce long-term memory recall. This, therefore, implies that writers should first attempt to build up links of cohesion in the surface structure of the text before attempting to propose a solution to the broader-level issues of coherence.

**How to Cite**: 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. DOI: 10.22034/ijpie.2025.535152.1116

Cohesion in writing has been studied to include various types as well as levels of text organization. These are generally divided into local, global, and textual cohesion (Crossley et al., 2016a). Local cohesion is the connection between sentences or ideas in a paragraph. It includes explicit connectives such as “because” or “therefore,” lexical overlap between sentences and referential pronouns (Halliday & Hasan, 1976; Crossley, 2020). Such words help readers make immediate links between ideas while moving through the text. Local cohesion markers are often the earliest writing skills that young learners acquire and second language writers develop (King & Rentel, 1979; McCutchen & Perfetti, 1982).

On the other hand, Global cohesion refers to connections between larger sections of texts, usually at the paragraph level. It can be measured by semantic and lexical overlaps across paragraphs (Abdi Tabari & Johnson, 2023). This type of cohesion is useful for keeping thematic unity across the text and supporting the overall argument or narrative structure. Global cohesion typically develops later than local cohesion and is associated with more advanced writing abilities (Crossley & McNamara, 2011).

Text cohesiveness encompasses elements that span the entirety of a text. Crossley (2020) enumerates various elements, one of which is "givenness," evaluating the equilibrium between novel information and previously presented data within the text. Other examples of cohesion at the textual level include lexical diversity and the use of complex syntactic structures, which connect ideas in an implicit rather than explicit manner (Crossley et al., 2016a; Haswell, 2000).

Each type of cohesion contributes uniquely to the overall quality of a text and the reader's comprehension. While local cohesion aids in immediate understanding, global cohesion and text unity are typically more closely linked to the overall quality of the text, particularly in advanced writing (McNamara et al., 2010; Neuner, 1987). Understanding these different forms of cohesion is crucial for teaching writing, assessing student work, and developing a comprehensive model of coherence and quality in texts.

## Literature Review

The relationship between L2 writing quality and cohesive devices has been an ever-contested matter, as studies have yielded contradictory results. Several researchers have offered explanations suggesting that the use of cohesive devices brings about better scores in instances of second language writing. Jafarpur (1991), for example, suggested a positive correlation between cohesion devices and composition quality in EFL Iranian students. In another study by Yang and Sun (2012), they showed that more advanced learners used more cohesive devices and did so more accurately in argumentative essays.

However, other studies have refuted this claim. Johnson (1992) did not observe any significant difference in the use of cohesive items among strong versus weak essays written by L2 learners. Similarly, while Zhang (2000) did not observe a relationship between essay quality and the relative frequency with which cohesive devices were used by Chinese undergraduates, other studies have yielded contradictory evidence on the questions; hence, some scholars doubt if cohesion and outstanding writing are directly linked among ESL learners (Abdi Tabari & Johnson, 2023).

Various researchers have recently made use of computational tools to analyze cohesion in L2 writing more deeply. In a more delicate manner, Crossley et al. (2016a) argue that the

connection between cohesion and L2 writing quality may vary according to the kind of cohesion examined. Several local cohesive devices were weakly or negatively connected with the quality of L2 compositions, while some global cohesive features had positive effects on the essays' quality.

Abdi Tabari and Johnson (2023) investigated the role of cohesive devices in predicting L2 essay quality across narrative and argumentative genres. Their findings indicated that the relationship of cohesive devices to essay quality is genre-specific and complex. Devices such as temporal connectives and sentence-linking connectives were significant predictors of the quality of narrative essays; using conjunctions too frequently was negatively correlated with this aspect. In the case of argumentative essays, global cohesive devices that contribute to semantic connectivity across paragraphs had a positive relationship with essay quality. This finding thus replicates earlier work by Crossley and McNamara (2010, 2011), which showed a positive relationship between global cohesion and writing quality but a negative or non-significant relationship for local cohesion.

Abdi Tabari, Johnson, and Gao (2024) also targeted genre differences as they studied cohesive devices in L2 narrative and argumentative essays using automated indices. Distinctions were found regarding the use of textual cohesive devices, particularly lexical diversity, in relation to how human raters evaluated essay organization. They observed minimal disparity in the use and evaluation of cohesive devices between genres, contrasting with some previous findings. Their study also found that over time, L2 writers showed a decrease in reliance upon local cohesive devices and an increase in the use of textual devices.

Additionally, recent investigations have focused upon the dynamic aspects of cohesion while developing a text in a second language (L2). Complex Dynamic Systems Theory was implemented by Abdi Tabari and Wind (2023) to discover how cohesion works in time in the case of students' written works in foreign tongues. The results from their study on advanced-level L2 writers indicated that different indices of cohesion improved significantly within the period of one semester, including local, global, and text-based cohesion. However, they also noted that this development is non-linear and dynamic. This non-linear development is also endorsed by Granados and Lorenzo (2021), who state that there exist non-linear relations between proficiency and cohesion use. For instance, local cues may play an important role for beginners as well as intermediate learners, whereas those at advanced levels may develop native-like patterns of cohesion, which tend to include fewer explicit devices indicating local cohesion.

Research also shows that task complexity and type significantly influence the use of cohesive devices in L2 writing. According to Abdi Tabari, Johnson, and Farhanynia (2023), more complex tasks lead to sophisticated cohesive devices. Tywoniw and Crossley (2019) looked at cohesion features by using TOEFL iBT writing tests with integrated and independent L2 tasks and noted significant differences between these two. It is intriguing to note that features distinguishing text types did not always predict higher scores, which points out the complex connection between cohesive device usage and writing quality as perceived. The relationship between cohesion and human scoring in L2 writing is equally complicated. As Abdi Tabari and Johnson (2023) contend, human raters may consider various aspects of cohesion rather than counting its elements. They argue that raters could be more concerned about how ideas flow globally (more related to global cohesion) rather than the presence of specific local cohesive devices.

There are various explanations for the conflicting findings in L2 cohesion research. Some have expressed that the discrepancies may stem from failure to consider all possible facet dimensions in a piece of research (Yang & Sun 2012) or methodological shortcomings. Differences in L2 writers' proficiency also have potential effects on the relationship between cohesion and the quality of their written text (Abdi Tabari & Johnson, 2023), which can lead to difficulties in comparing results of the research studies using participants with different levels of proficiency. These issues demonstrate the need for more diversified and accurate methodologies in the analysis of L2 cohesion; Crossley et al. (2016b) suggested that computer-assisted tools should be employed to capture more accurate and complete measures of cohesiveness at various linguistic levels.

Some of the studies described above dealt with the correlation between cohesion and global writing scores. These scores encompass various factors, such as content, organization, and the accurate and complex employment of grammatical and lexical features. Therefore, it would be worthwhile to compare cohesion indices to a human assessment of texts based solely on this feature.

In conclusion, we can say that there is a rather context-based connection between cohesion and L2 writing quality. Despite their indispensability for producing cohesive writing, though, the effectiveness of cohesive devices is rather contingent on a number of characteristics, such as the proficiency of the writers, type of task, genre, and the very kind of cohesion under examination.

In IELTS writing task 2, the Cohesion and Coherence (CC) criterion is one of the most important linguistic competencies for a candidate. It shows how well they are able to organize information and ideas in a logical order and how well they use different cohesive devices. Mastery in this area is essential for attaining higher scores and is a reliable indicator of overall language proficiency and academic writing skills.

Currently, IELTS essays are scored by human examiners who are trained using IELTS standards. Human examiners are trained extensively and intensively and then retrained from time to time so that the scoring standard remains constant. This procedure is both time-consuming and costly. This procedure is both time-consuming and costly. Moreover, examiners acknowledge that assessing CC is the most challenging criterion when evaluating IELTS essays (Cotton & Wilson, 2011). Thus, developing methods for objectively quantifying and analyzing these structures could significantly enhance the efficiency of essay scoring in this high-stakes test.

Accordingly, the following research questions were proposed:

RQ1: Which indices of local cohesion correlate with human examiners' scores for Cohesion and Coherence (CC) in IELTS Writing Task 2?

R2: Which indices of global cohesion correlate with human examiners' scores for Cohesion and Coherence (CC) in IELTS Writing Task 2?

RQ3: Which indices of text cohesion correlate with human examiners' scores for Cohesion and Coherence (CC) in IELTS Writing Task 2?

INTERNATIONAL JOURNAL OF
Practical and Pedagogical Issues in English Education
ISSN : 2980-9533 Homepage: www.ijpie.org 3(4), Dec. 2025

## Method

### Corpus

The dataset for this research contains 105 IELTS Writing Task 2 essays. The essays were responses to various prompts and rated by official IELTS examiners. These essays were directly sourced from the official IELTS content provided by Cambridge University, the British Council, and IDP Australia, who are the test's three co-owners. To avoid data contamination, researchers corrected spelling errors in the essays, as spelling is assessed under Lexical Resources rather than Cohesion and Coherence in IELTS. As shown in Table 1, the sample essays cover a broad spectrum of band scores for Cohesion and Coherence.

**Table 1**

*The Scored Essays Used as the Corpus of the Study*

| Band Score | Number of Sample Essays | Word Count |
|---|---|---|
| 9 | 22 | 6272 |
| 8 | 4 | 1088 |
| 7 | 31 | 9721 |
| 6 | 29 | 8607 |
| 5 | 14 | 3682 |
| 4 | 5 | 1048 |
| Total | 105 | 30814 |

Table 2 indicates the band descriptors for Coherence and Cohesion across different bands in IELTS Writing Task 2.

**Table 2**

*IELTS Writing Task 2 Band Score Descriptors for Coherence and Cohesion*

| CC Score | Descriptors |
|---|---|
| 9 | The message is effortless to follow, with cohesion rarely drawing attention. Any coherence or cohesion lapses are minimal, and paragraphing is expertly handled. |
| 8 | The message is easy to follow, with well-managed cohesion and logical sequencing. Occasional lapses in coherence or cohesion may occur, but paragraphing is sufficient and appropriate. |
| 7 | Information and ideas are clearly organized, with minor lapses. Cohesive devices are used flexibly but may have inaccuracies or over/underuse, and paragraphing generally supports coherence. |
| 6 | Ideas are generally arranged coherently with clear progression. Cohesive devices are sometimes faulty, and paragraphing may not always be logical or clear. |
| 5 | Organization is evident but not fully logical, with some underlying coherence. Cohesion is limited, sentences may not link fluently, and paragraphing may be inadequate or missing. |
| 4 | Ideas are evident but lack coherent arrangement, with no clear progression. Cohesion is basic and often inaccurate, and there may be no paragraphing or clear main topics. |

**Measures of Cohesion**

This study includes a total of 40 indices of cohesion, which are categorized as follows: 20 local indices, 14 global indices, and 6 text cohesion indices (see Table 3). The indices were essentially extracted through a program known as the "Tool for the Automatic Analysis of Cohesion" 2.0.4 (TAACO; Crossley et al. 2016b, 2019), a sophisticated computing tool known for its ability to measure text cohesion. TAACO generates hundreds of indices related to global cohesion, synonym overlap, and part-of-speech tagging. We intentionally selected only those indices that addressed distinct aspects of cohesion in order to prevent any redundancy in measurement.

**Table 3**

*Cohesion Indices Used in the Study*

| | | Index |
|---|---|---|
| Local Cohesion | Connectives | All causal connectives |
| | | All additive connectives |
| | | All logical connectives |
| | | All positive connectives |
| | | All negative connectives |
| | | All connectives |
| | Lexical overlap | Adjacent sentence overlap all lemmas |
| | | Adjacent sentence overlap content lemmas |
| | | Adjacent sentence overlap function lemmas |
| | | Adjacent sentence overlap noun lemmas |
| | | Adjacent sentence overlap verb lemmas |
| | | Adjacent sentence overlap adjective lemmas |
| | | Adjacent sentence overlap adverb lemmas |
| | | Adjacent sentence overlap pronoun lemmas |
| | | Adjacent sentence overlap noun and pronoun lemmas |
| | Semantic similarity | Synonym overlap (sentence, noun) |
| | | Synonym overlap (sentence, verb) |
| | | LSA cosine similarity (adjacent sentences) |
| | | LDA divergence (adjacent sentences) |
| | | word2vec similarity (adjacent sentences) |
| Global cohesion | Lexical overlap | Adjacent paragraph overlap all lemmas |
| | | Adjacent paragraph overlap content lemmas |
| | | Adjacent paragraph overlap function lemmas |
| | | Adjacent paragraph overlap noun lemmas |
| | | Adjacent paragraph overlap verb lemmas |
| | | Adjacent paragraph overlap adjective lemmas |
| | | Adjacent paragraph overlap adverb lemmas |
| | | Adjacent paragraph overlap pronoun lemmas |
| | | Adjacent paragraph overlap noun and pronoun lemmas |
| | Semantic overlap | Synonym overlap (paragraph, noun) |
| | | Synonym overlap (paragraph, verb) |
| | | LSA cosine similarity (adjacent paragraphs) |
| | | LDA divergence (adjacent paragraphs) |
| | | Word2vec similarity (adjacent paragraphs) |

| | | |
|---|---|---|
| Text cohesion | Lexical Diversity (TTR density) | Lemma TTR |
| | | Lemma MATTR |
| | Givenness | Pronoun density |
| | | Pronoun to noun ratio |
| | | Repeated content lemmas |
| | | Repeated content lemmas and pronouns |

***Local cohesion indices***

Local cohesion concerns cohesion happening inside individual sentences, e.g., the use of connectives. The analysis utilizes various TAACO indices to assess sentence cohesion by examining syntactic tags and phrase overlap. It examines theoretical and rhetorical connectives, as well as semantic similarity, synonym overlap, and lexical overlap between sentences. These indices provide information as to how ideas are related at the sentence level, which helps maintain the coherence of the entire text.

***Connectives***

Connectives serve as linking words that help connect clauses, phrases, or sentences. TAACO categorizes connectives into several theoretical groups, following frameworks like Halliday and Hasan (1976) and Louwerse (2001). These groups include:

· Causal Connectives: Words like "therefore" and "as a result" that show cause-effect relationships.

· Contrastive Connectives: Words like "nevertheless" and "on the other hand" that highlight contrasts between ideas.

· Additive Connectives: Words like "in addition" and "furthermore" that introduce additional information.

· Logical Connectives: Words like "hence" and "thus" that convey logical conclusions.

· Temporal Connectives: Words like "meanwhile" and "subsequently" that indicate time relationships.

TAACO also calculates the frequency of more specific types of connectives, such as coordinating connectives (e.g., "accordingly"), semi-coordinators (e.g., "likewise"), and complex coordinators (e.g., "nevertheless"). For example, if a text frequently uses words like "and," "or," and "in contrast to," it would suggest a multidimensional relationship between ideas, adding depth to the text's cohesion.

Example:

In a sample text, you might find sentences like *"The company faced significant losses. As a result, they restructured their management."*

Here, the causal connective "As a result" explicitly links the cause (losses) to the effect (restructuring).

***Lexical Overlap between Sentences***

TAACO measures lexical overlap in various forms to understand how much lexical repetition occurs between adjacent sentences:

· Synonym Overlap: This measures how often synonyms or words with similar meanings appear in consecutive sentences. For example, if one sentence uses the word "happy" and the next uses "joyful," TAACO would identify the phenomenon as synonym overlap.

· Verb Argument Overlap: This measure assesses the repetition of verb argument structures, such as the repeated use of similar subject-verb-object patterns.

· Stem Overlap: These measures overlap based on word stems, like "educate" in one sentence and "education" in the next.

· Content Word Overlap: This counts shared content words, such as nouns, verbs, adjectives, and adverbs, between sentences.

TAACO also analyzes lemma overlap, examining how words with similar meanings but different forms relate. For example, the lemmas "observe" and "watch" would be considered similar.

Example:

In two sentences,
*"The scientist observed the stars."*
*"She carefully watched the night sky."*
TAACO would detect synonym overlap between "observed" and "watched."

***Semantic Similarity between Sentences***

TAACO utilizes Latent Semantic Analysis (LSA) to assess the semantic similarity between sentences. The LSA index evaluates the degree to which the meaning of one sentence relates to the meaning of another. It also creates a statistical representation of world knowledge to assess this semantic closeness. TAACO also incorporates the WordNet database to evaluate the overlap of words within synsets, specifically focusing on nouns and verbs. For instance, words such as "sip" and "imbibe" can be found in the same synset as the word "drink."

Example:

Consider these sentences:
*"She poured herself a drink."*
*"She took a sip from her glass."*
Using LSA, TAACO would recognize that these sentences are semantically similar because they share the underlying concept of consuming a beverage.

***Global cohesion indices***

Global cohesion pertains to cohesion at larger textual units e.g., paragraphs. Selected indices include LSA-based metrics for semantic similarity, lexical overlap metrics for word repetition, and synonym overlap metrics for semantic relatedness among words within paragraphs. These measures help assess how well the text maintains coherence across larger sections.

***Lexical Overlap between Paragraphs***

TAACO measures lexical overlap at the paragraph level, similar to sentence-level overlap. This involves calculating the repetition of lemmas (basic forms of words) across paragraphs. The overlap is measured in terms of content words, function words, and part-of-speech categories.

Example:

If one paragraph discusses "*environmental protection" and the next paragraph continues by mentioning "sustainability efforts,"* TAACO would detect lexical overlap through the repeated focus on related environmental topics.

***Synonym Overlap between Paragraphs***

TAACO uses WordNet to measure how semantically related words (synonyms) are repeated across paragraphs. This helps to understand how themes and ideas carry over from one paragraph to the next.

Example:

Suppose a paragraph describes a "*happy*" childhood experience, and the following paragraph refers to the same experience as "*joyfu*l." TAACO would identify synonym overlap between these paragraphs.

***Indices of text cohesion***

Overall text coherence pertains to cohesive elements throughout the entire text. Selected indices measure spatial cohesion, givenness, causal cohesion, lexical diversity, temporal cohesion, and semantic similarity across the entire text. These measures provide a holistic view of how well the text maintains coherence from beginning to end.

***Lexical Diversity (TTR Density)***

Lexical diversity is often measured using the type-token ratio (TTR), which calculates the ratio of unique words (types) to the total number of words (tokens). A higher TTR indicates more word variety, which could suggest lower cohesion, as cohesive texts often repeat key words and phrases. TAACO provides multiple TTR indices, measuring lexical diversity across all words, content words (nouns, verbs, adjectives, and adverbs), and function words (pronouns,

prepositions, and determiners). TTR is also calculated for word sequences, such as bigrams (e.g., "smart dog") and trigrams (e.g., "smart dog plays").

Example:

In an essay, a paragraph that repeatedly uses variations of the word "*research*" would have a lower TTR compared to a paragraph that uses synonyms like "*study*" and "*investigation*," showing higher lexical diversity.

***Givenness***

Givenness relates to the extent to which certain information or a situation is offered as if it were already known to the reader. Usually, it is somehow indicated by means of pronouns and demonstratives. TAACO calculates the noun-to-pronoun ratio, with a higher use of pronouns suggesting greater givenness. TAACO also tracks the use of demonstratives, like "this" or "those," which often refer back to previously mentioned information.

Example:

In a text, you might see the transition from *"The new policy was introduced yesterday."*
to *"This policy aims to reduce emissions."*
Here, the demonstrative "*this*" signals givenness by referring back to the previously mentioned "*policy*".

**Data Analysis**

In this study, cohesion indices were assessed using TAACO (Crossley et al., 2016b). Given the non-normal distribution of the data for CC band scores (Table 4), Spearman's Rank Correlation was employed using SPSS to examine the relationship between human holistic scoring of IELTS Task 2 essays in terms of CC and indices of cohesion.

**Table 4**

*Test of Normality for CC Band Scores*

| | Shapiro-Wilk | | |
|---|---|---|---|
| | Statistic | df | Sig. |
| CC Band Score | .901 | 105 | .000 |

*Note*. The Sig. value below 0.05 indicates deviation from a normal distribution

## Results

### Research Question 1

The first research question examined the extent of the correlation between holistic scoring in IELTS essays for CC and local cohesion indices. To address this, TAACO was used to calculate the local cohesion indices for the essays. Subsequently, a Spearman's Rank Correlation was conducted using SPSS to analyze the relationship between these indices and the CC bands awarded by the examiners. The results are presented in Table 5:

**Table 5**

*Correlation between Holistic Human Scoring and Local Cohesion Indices*

| Variable | Correlation Coefficient (r) | Sig. (2-tailed) p | N |
|---|---|---|---|
| All causal connectives | -.178 | .069 | 105 |
| All additive connectives | .206* | .035 | 105 |
| All logical connectives | .058 | .557 | 105 |
| All positive connectives | .020 | .841 | 105 |
| All negative connectives | .170 | .084 | 105 |
| All connectives | .231* | .018 | 105 |
| Adjacent sentence overlap all lemmas | -.221* | .023 | 105 |
| Adjacent sentence overlap content lemmas | -.214* | .028 | 105 |
| Adjacent sentence overlap function lemmas | -.093 | .343 | 105 |
| Adjacent sentence overlap noun lemmas | -.193* | .049 | 105 |
| Adjacent sentence overlap verb lemmas | -.056 | .570 | 105 |
| Adjacent sentence overlap adjective lemmas | .066 | .505 | 105 |
| Adjacent sentence overlap adverb lemmas | -.086 | .384 | 105 |
| Adjacent sentence overlap pronoun lemmas | -.119 | .228 | 105 |
| Adjacent sentence overlap noun and pronoun lemmas | -.268** | .006 | 105 |
| Synonym overlap (sentence, noun) | -.062 | .529 | 105 |

| Variable | Correlation Coefficient (r) | Sig. (2-tailed) p | N |
|---|---|---|---|
| Synonym overlap (sentence, verb) | -.171 | .082 | 105 |
| LSA cosine similarity (adjacent sentences) | -.071 | .473 | 105 |
| LDA divergence (adjacent sentences) | -.093 | .343 | 105 |
| word2vec similarity (adjacent sentences) | .054 | .584 | 105 |

*Note.* *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

The data supported several significant correlation indices, showing moderate relations between specific cohesion measures and human ratings. Holistic ratings bore a significant positive correlation with the use of additive connectives (r = .206, p = .035), as well as with the total number of connectives used (r = .231, p = .018), suggesting that the more connectives of this type a text contains, the better its marking for coherence and cohesion may be.

In contrast, certain local cohesion indices showed negative correlations with human scores. Specifically, adjacent sentence overlaps for all lemmas (r = -.221, p = .023) and content lemmas (r = -.214, p = .028) exhibited significant negative correlations. Overlap in noun lemmas also had a marginally significant negative correlation (r = -.193, p = .049). The strongest negative correlation was found with noun and pronoun lemma overlap (r = -.268, p = .006), which indicates that increased lemma overlap may correspond to lower human ratings of cohesion.

Meanwhile, other local cohesion measures that take into account causality, logic, positive and negative connectives, as well as more sophisticated cohesion metrics such as LSA cosine similarity and word2vec similarity, did not show significant correlation with the human scores (p > .05).

**Research Question 2**

The second research question was concerned with the relationship existing between the holistic scores assigned to IELTS essays for Cohesion and Coherence (CC) and the global indices of cohesion. To this end, global cohesion indices were computed in TAACO, making use of cohesion features across larger spans of text, i.e., between paragraphs. The analysis of the relationship between the global indices and the CC scores assigned by the examiners was conducted using Spearman's Rank Correlation in SPSS. The findings are summarized in Table 6:

**Table 6**

*Correlation between Holistic Human Scoring and Global Cohesion Indices*

| **Variable** | **Correlation Coefficient (r)** | **Sig. (2-tailed) p** | **N** |
|---|---|---|---|
| Adjacent paragraph overlap all lemmas | -.042 | .670 | 105 |

| Variable | Correlation Coefficient (r) | Sig. (2-tailed) p | N |
|---|---|---|---|
| Adjacent paragraph overlap content lemmas | -.017 | .864 | 105 |
| Adjacent paragraph overlap function lemmas | .230* | .018 | 105 |
| Adjacent paragraph overlap noun lemmas | -.048 | .624 | 105 |
| Adjacent paragraph overlap verb lemmas | .015 | .883 | 105 |
| Adjacent paragraph overlap adjective lemmas | .011 | .907 | 105 |
| Adjacent paragraph overlap adverb lemmas | .136 | .166 | 105 |
| Adjacent paragraph overlap pronoun lemmas | .079 | .424 | 105 |
| Adjacent paragraph overlap noun and pronoun lemmas | -.119 | .228 | 105 |
| Synonym overlap (paragraph, noun) | .200* | .041 | 105 |
| Synonym overlap (paragraph, verb) | .051 | .602 | 105 |
| LSA cosine similarity (adjacent paragraphs) | .188 | .054 | 105 |
| LDA divergence (adjacent paragraphs) | .060 | .541 | 105 |
| Word2vec similarity (adjacent paragraphs) | .205* | .036 | 105 |

*Note.* *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

The correlations are important, but they are usually not forceful. Human scoring has a positive correlation with the adjacent paragraph overlap of function lemmas (r = .230, p = .018), which means that higher function-lemma overlap between paragraphs may provide an impression of more cohesion and coherence. Statistically significant correlations with the holistic scores were also found for synonym overlaps in noun phrases between paragraphs (r = .200, p = .041) and for word2vec similarity of adjacent paragraphs (r = .205, p = .036). The results suggest that lexical cohesion at the paragraph level, whether via functional elements or semantic similarity, is positively viewed by humans when rating writing quality.

The input is clear and maintains a consistent tone; however, it could benefit from slight rephrasing for improved clarity. Here is a revised version:

"However, no significant correlations were found for the overlap of adjacent paragraphs in terms of content lemmas, noun lemmas, and verb lemmas (p > .05). Additionally, there were no correlations for LSA cosine similarity or LDA divergence (p > .05). These findings suggest that individuals may not heavily weigh certain types of lexical and semantic overlap when assessing global cohesion in essays.

**Research Question 3**

The processes of the third research question investigated the correlation between holistic IELTS essay scores on Cohesion and Coherence (CC) with some text-level cohesion indices. TAACO generated these text cohesion indices by looking at features such as lexical diversity and repetition within the entire essay. A Spearman's Rank Correlation analysis was conducted in SPSS to assess the relationship between these text cohesion indices and the CC scores provided by the examiners. The results are presented in Table 7:

**Table 7**

*Correlation between Holistic Human Scoring and Text Cohesion Indices*

| Variable | Correlation Coefficient (r) | Sig. (2-tailed) p | N |
|---|---|---|---|
| Lemma TTR | .281** | .004 | 105 |
| Lemma MATTR | .447** | .000 | 105 |
| Pronoun density | -.045 | .651 | 105 |
| Pronoun to noun ratio | -.023 | .813 | 105 |
| Repeated content lemmas | -.228* | .019 | 105 |
| Repeated content lemmas and pronouns | -.257** | .008 | 105 |

*Note.* *Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

Several important relationships were discovered, foremost among them lexical diversity and repetition. First, a strong positive correlation between holistic scores and lemma MATTR (moving average type-token ratio) ($r = .447$, $p < .001$) was found. It means that holistically rated essays for cohesion and coherence by humans gain higher lexical diversity levels by this measure. Similarly, the lemma-type token ratio correlated with human scores ($r = .281$, $p = .004$), which further strengthens the argument for lexical variety being a factor for creating a better piece of writing.

Human scores, however, were found to be negatively correlated with content lemma repetition ($r = -.228$, $p = .019$) and with a combination of content lemmas and pronouns ($r = -.257$, $p = .008$). It means that the more a text repeats lexical items, the less cohesion and coherence it is perceived to have.

Other indices, such as pronoun density ($r = -.045$, $p = .651$) and the pronoun-to-noun ratio ($r = -.023$, $p = .813$), did not show significant correlations with human examiners' scores.

## Discussion

### Research Question 1

Study findings reveal noteworthy correlations between some local cohesion indices and holistic scores assigned to IELTS essays for Coherence and Cohesion (CC). The relationships reveal a nuanced connection between various cohesive devices and human judgments of writing quality.

Positive correlation between additive connectives and higher human ratings may be construed to conform to some prior research that advocates the significance of connectives as indicators of writing quality, especially among L2 learners (Abdi Tabari, Johnson, & Gao, 2024; Morallo, 2024). Thus, demonstrating the effective employment of additive connectives may indicate that the candidates at a higher level could very clearly and logically link ideas together, hence being judged on the basis of this ability for the overall coherence of the text. This finding is further supported by the observation that an increase in the total number of connectives is likely associated with higher CC bands. This suggests that a richer variety of connectives may contribute to a more sophisticated and cohesive writing style (Abdi Tabari, Johnson, & Gao, 2024; Crossley et al., 2016a).

However, certain negative correlations for certain local cohesion indices challenge the notion that the use of more cohesive devices inevitably leads to higher-quality writing. The strong negative correlations for adjacent sentence overlap, both for all lemmas and for content lemmas, contradict the findings of some earlier studies that reported positive associations between local cohesion and L2 writing quality (Abdi Tabari, Johnson, & Gao, 2024; Abdi Tabari & Johnson, 2023). This phenomenon may be due to the type of local cohesion studied here. The repetition of lemmas in successive sentences—especially content lemmas—might suggest a lack of lexis and some word sophistication that could possibly lower writing quality ratings. This interpretation is further corroborated by the strongest negative correlation, which was found in noun and pronoun lemma overlap. The overuse of these repetitions might be seen as evidence of a limited lexical repertoire and a tendency to cling to elementary sentence constructions.

The results regarding the lack of significant correlation for various local cohesion measures, including causal and logical connectives (both positive and negative), LSA cosine similarity, and word2vec similarity, warrant further examination. The aforementioned measures may capture certain aspects of cohesion; however, they do not significantly influence human raters' evaluations of coherence and cohesion in IELTS essays compared to other factors (Abdi Tabari & Johnson, 2023; Crossley et al., 2016a). The unique features of the IELTS marking criteria, combined with the limited length of the essays, may limit the effectiveness of these more advanced measures in assessing perceived writing quality.

The observed patterns of correlations suggest that cohesion is not about the mere abundance of cohesive devices: in fact, the very selective use of some kinds of cohesive devices seems to make a lot of difference in shaping raters' perceptions. Although the presence of additive connectives might indicate a higher degree of coherent and organized discourse construction, excessive usages of lemma repetition in two consecutive sentences are more likely to be deemed as a narrowness in vocabulary and a lack of stylistic dynamism. The lack of significant correlations with other measures indicates that the influence of cohesion on human

judgments is shaped by a complex relationship between various factors. These factors encompass the specific context of assessment, genre-related expectations, and the overall maturity of the writing. Certain components, however, fell outside the scope of this study.

**Research Question 2**

The results for the second research question demonstrate intriguing relationships between global cohesion indices and holistic scores for CC in IELTS essays. While several moderate correlations were observed, the findings indicate a degree of subtlety in how different forms of global cohesion are perceived by human raters.

Adjacent-paragraph overlap of function lemmas and the higher holistic score show a positive relationship that warrants special consideration. This finding suggests that repetition may contribute to a sense of coherence and fluency in writing. Specifically, a higher frequency of functional items—such as prepositions, conjunctions, or pronouns—used between paragraphs may enhance the perceived quality of the writing. Such an observation accounts for what previous research underlines, i.e., the importance of global cohesion to overall text coherence (Abdi Tabari, Johnson, & Farahanynia, 2023; Abdi Tabari & Wind, 2023; Crossley et al., 2016a). Functional elements link sentences to one another and paragraphs to each other, guiding readers through the text to ensure cohesion that leads to comprehension. The use of these elements across paragraph boundaries may suggest a more profound grasp of the overall structure of the text, as well as the capability to create cohesive transitions between ideas.

Moreover, the positive correlation for synonym overlaps in paragraph-level noun and word2vec similarity of adjacent paragraphs lends evidence to support the conceptualization of lexical cohesion at the paragraph level as being positively correlated with writing quality (Abdi Tabari & Wind, 2023). Synonym overlap rather suggests that a writer may have consciously selected a different word to express the same idea—and thereby bring in one more dimension for the writing and its coherence-making process. The positive correlation for word2vec similarity indicates that the greater the semantic similarity of contiguous paragraphs as measured by this computational index, the higher the ratings given by raters for coherence and cohesion. These results are in accord with previous studies reporting positive relationships between global cohesion and writing quality (Abdi Tabari, Johnson, & Gao, 2024; Crossley et al., 2016a).

The lack of significant correlation for adjacent paragraph overlaps that contain content lemmas (such as nouns and verbs), as well as for LSA cosine similarity and LDA divergence, highlights the complexity of the relationship between global cohesion and human judgments of writing quality (Abdi Tabari & Wind, 2023). The expectation that cohesion is achieved through the repetition of content words across paragraphs contrasts with the findings that show a lack of significant correlations. This suggests, from a certain perspective, that such overlap may not be viewed as an indicator of quality in the context of IELTS essays (Abdi Tabari & Wind, 2023). Additionally, the non-significant results for LSA cosine similarity and LDA divergence imply that human raters do not readily recognize or value these measures during IELTS assessments, as they capture a more abstract level of semantic similarity.

**Research Question 3**

The examination of text-level cohesion indices, particularly lemma MATTR and lemma TTR, reveals a strong positive correlation with holistic scores for CC in IELTS essays. This

finding is in line with previous research, including Abdi Tabari and Johnson (2023), where cohesive devices are studied in written academic genres. The authors recorded a positive correlation between certain cohesive devices and quality ratings and, as such, suggest that a higher lemma diversification—as captured by higher scores on MATTR and TTR—relates to a perception of a more coherent and cohesive text on the part of human raters.

This emphasis on lexical diversity is further supported by Crossley et al. (2016a), who examined the development and use of cohesive devices in L2 writing. They found that indices of cohesion, including measures of lexical diversity, served as significant predictors of human judgments of text organization and overall essay quality. Notably, higher-proficiency L2 writers in their study displayed greater lexical diversity, indicating a preference for using a broader range of vocabulary, which likely contributed to their higher scores.

The observation that repeated content lemmas, the repetition of content lemmas, and pronouns are all negatively correlated with human scores suggests that excessive reliance on repetition undermines perceived cohesion and coherence. Research, however, warns against viewing the results as a blanket rejection of repetition in writing. In certain contexts, repetition can serve a valuable purpose, such as providing emphasis, clarification, or achieving a desired stylistic effect. For instance, research by Abdi Tabari and Johnson (2023) indicates that narrative essays exhibit the highest use of connectives to signal cohesion, while argumentative essays generally rely more on global-level repetition. This underscores the importance of considering genre-specific conventions when assessing the role of repetition in writing.

It is likewise worth considering that the research evidence indicates that skillful writers are known to walk the fine line between repetition and variation, with the repetition being used judiciously without ever threatening the lexical diversity of the text (Crossley et al., 2016a). In this regard, the emphasis in teaching L2 writers must be on the development of an ample vocabulary and its strategic use, coupled with instructing the use of repetition for higher quality in writing.

The non-significant result for pronoun density and the pronoun-to-noun ratio in the present investigation aligns with previous research suggesting that these metrics may imperfectly measure the cohesion and coherence of meaning at the discourse level. While pronouns certainly contribute to cohesive ties, their density and proportion relative to nouns do not adequately capture the complexities of how they are used (Crossley & McNamara, 2012).

## Conclusion

In conclusion, this study considers cohesion measures as predictors of L2 writing quality with regard to the IELTS Writing Task 2. They reveal that some cohesion indices, such as additive connectives and lexical diversity, are positively correlated with high Cohesion and Coherence (CC) scores, while too much lexical repetition tends to lower perceived writing quality. This denotes certain cohesion features that could serve as indicators for automated essay scoring.

Furthermore, the study makes a case for IELTS preparatory courses to teach candidates with a focus on cohesive structures that directly influence Cohesion and Coherence (CC) scores.

Since cohesion features such as additive connectives and lexical diversity correlate positively to CC scores, teaching strategies toward these features will appear to be worthwhile.

For example, if an instructor teaches students how to use additive connectives effectively (such as "in addition" and "furthermore"), these students will be able to make clearer connections between ideas, which will improve the logical flow and clarity of paragraphs. Similarly, exercises that foster lexical diversity will enhance students' ability to vary vocabulary without excessive repetition, which leads to lower cohesion gradings.

Upon incorporation of these cohesion-oriented techniques into IELTS preparation, instructors would be able to assist students with writing according to CC expectations much more efficiently. Within this vein, such instruction could afford students better opportunities to receive a higher IELTS score by mastering how to generate responses that are more cohesive and coherent.

The study similarly tried to bring automated scoring into the assessment of cohesion and coherence for IELTS Writing Task 2 essays as in other major tests, such as TOEFL and PTE. Automated scoring was expected to promise such benefits as objectivity, speed, and cost efficiency.

However, the findings revealed that cohesion assessment is more nuanced and subjective than one might have originally thought, also posing some challenges to full automation. The difficulty of judging cohesion elements shows that, for now, people will still be better at it than machines. Nonetheless, insights from this study could guide future developments in automated essay scoring (AES), potentially easing the workload for human raters by allowing them to leverage automatic raters as scoring assistants. In this way, AES could enhance the efficiency and reliability of language proficiency assessments without fully replacing human judgment.

Inevitably, the study has some limitations. The relatively small corpus size restricts the generalizability of the findings. Additionally, the study focused exclusively on cohesion, while the IELTS scoring criteria for CC also encompass coherence, which was not adequately addressed, as it is a mental representation of readers about the text and cannot be measured automatically (at least directly). Another limitation is that the essays were written on varying topics, potentially affecting the writers' use of different cohesion features. Future research should utilize a significantly larger corpus with an equal representation of samples from each band score to more accurately identify the indices associated with each band. Investigating a broader array of genres and task types, particularly with advanced tools or AI-driven analysis, could also provide deeper insights. This study lays the groundwork for such explorations, highlighting cohesion’s multifaceted impact on writing quality and the importance of strategic cohesive device use for L2 writers.

## References

Abdi Tabari, M., & Johnson, M. D. (2023). Exploring new insights into the role of cohesive devices in written academic genres. *Assessing Writing, 57,* 100749. https://doi.org/10.1016/j.asw.2023.100749

Abdi Tabari, M., & Wind, A. (2023). Dynamic development of cohesive devices in English as a second language writing. *International Review of Applied Linguistics in Language Teaching.* https://doi.org/10.1515/iral-2022-0205

Abdi Tabari, M., Johnson, M. D., & Farahanynia, M. (2023). Automated analysis of cohesive features in L2 writing: Examining effects of task complexity and task repetition. *Assessing Writing, 58*, 100783. https://doi.org/10.1016/j.asw.2023.100783

Abdi Tabari, M., Johnson, M., & Gao, J. (2024). Using automated indices of cohesion to explore the growth of cohesive features in L2 writing. *International Review of Applied Linguistics in Language Teaching.* https://doi.org/10.1515/iral-2023-0185

Cotton, F., & Wilson, K. (2011). An investigation of examiner rating of coherence and cohesion in the IELTS academic writing task 2. *IELTS Research Reports, 12*, 1–76.

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research, 11*(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods, 51*(1), 14–27. https://doi.org/10.3758/s13428-018-1142-4

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016a). 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., Kyle, K., & McNamara, D. S. (2016b). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods, 48*(4), 1227–1237. https://doi.org/10.3758/s13428-015-0651-7

Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 984–989). Austin, TX: Cognitive Science Society.

Crossley, S. A., & McNamara, D. S. (2011). Text coherence and judgments of essay quality: Models of quality and coherence. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 29th annual conference of the cognitive science society* (pp. 1236–1241). Austin, TX: Cognitive Science Society.

Ebrahimi, S. F., & Mallaki, A. (2024). The pronoun “this” in civil engineering research articles. *International Journal of Practical and Pedagogical Issues in English Education, 2* (2), 1, 12. https://doi.org/10.22034/ijpie.2024.461451.1021

Granados, A., & Lorenzo, F. (2021). English L2 connectives in academic bilingual discourse: A longitudinal computerized analysis of a learner corpus. *Revista Signos, 54*(106), 626–644. https://doi.org/10.4067/s0718-09342021000200626

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

Haswell, R. H. (2000). Documenting improvement in college writing. *Written Communication, 17*(3), 307–352. https://doi.org/10.1177/0741088300017003001

Jafarpur, A. (1991). Cohesiveness as a basis for evaluating compositions. *System, 19*(4), 459–465. https://doi.org/10.1016/0346-251X(91)90026-L

Johnson, P. (1992). Cohesion and coherence in compositions in Malay and English. *RELC Journal, 23*(2), 1–17. https://doi.org/10.1177/003368829202300201

King, M. L., & Rentel, V. (1979). Toward a theory of early writing development. *Research in the Teaching of English, 13*, 243–253.

Louwerse, M. (2001). An analytic and cognitive parametrization of coherence relations. *Cognitive Linguistics, 12*(3), 291–316. https://doi.org/10.1515/cogl.2002.005

McCutchen, D., & Perfetti, C. A. (1982). Coherence and connectedness in the development of discourse production. *Text-Interdisciplinary Journal for the Study of Discourse, 2*(1–3), 113–140. https://doi.org/10.1515/text.1.1982.2.1-3.113

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*, 57–86. https://doi.org/10.1177/0741088309351547

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

Morallo, A. B. (2024). Cross-cultural and cross-proficiency analysis of selected cohesion indices in student essays. *Journal of English and Applied Linguistics, 3*(1), 7.

Neuner, J. L. (1987). Cohesive ties and chains in good and poor freshman essays. *Research in the Teaching of English, 21*, 92–105.

Tasouji Azari, M. & Mohammadi, K. (2024). Book review: The Routledge Handbook of Discourse Analysis, edited by Michael Handford and James Paul Gee (2023). *International Journal of Practical and Pedagogical Issues in English Education, 2*(4), 91-97. https://doi.org/10.22034/ijpie.2024.474284.1042

Tywoniw, R., & Crossley, S. (2019). The effect of cohesive features in integrated and independent L2 writing quality and text classification. *Language Education and Assessment, 2*(3), 110–134. https://doi.org/10.29140/lea.v2n3.151

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

Zhang, M. (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

**Ehsan Abbaspour** holds a PhD in English Language Teaching from Science and Research Branch, Islamic Azad University, Tehran, Iran. He is a senior lecturer at Centre for Academic Writing, Middle East College, Muscat, Oman. His main research interest includes L2 Writing and Written Corrective Feedback.

**Priya Mathew** holds a PhD in Applied Linguistics from Coventry University, UK. She is an assistant professor at Centre for Academic Writing, Middle East College, Muscat, Oman. Her main research interest includes L2 Writing and Corpus Linguistics.