Korean Journal of English Language and Linguistics, Vol 22, October 2022, pp. 1078-1100
DOI: 10.15738/kjell.22..202210.1078

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

The Use of Cohesive Devices in Korean EFL Writing across Different Proficiency Levels

Jungyeon Kim (Yonsei University)

Received: September 02, 2022
Revised: October 10, 2022
Accepted: October 30, 2022

Jungyeon Kim
Postdoctoral Researcher,
Department of English
Language and Literature,
Yonsei University
Tel: 02) 2123-2329
Email: jungyeonkim@yonsei.ac.kr

ABSTRACT

Kim, Jungyeon. 2022. The use of cohesive devices in Korean EFL writing across different proficiency levels. *Korean Journal of English Language and Linguistics* 22, 1078-1100.

This study examines cohesive devices in the essays written by Korean EFL college learners across four common reference levels (CEFR) in the International Corpus Network of Asian Learners of English (ICNALE). The methodology used to analyze all cohesion features in this learner corpus is the assessment of cohesion using the Tool for the Automatic Analysis of Cohesion (TAACO). In order to see whether cohesion would vary across different proficiency levels in Korean EFL writing, this study examined fine-grained indices of four different kinds of components related to cohesive elements, i.e., lexical overlap, connectives, semantic overlap, and givenness. The statistical results suggest that the variable of lexical overlap (i.e., binary adjacent sentence overlap content lemmas) is a stronger predictor of EFL writing performance than the other variables of text cohesion. These findings expand previous corpus-based results regarding the evaluation of EFL writing quality, cohesive features in particular. The current study will bring about the expansion of new research that can investigate the role of cohesion analyses in accounting for foreign language writing proficiency.

KEYWORDS

cohesion, ICNALE, TAACO, Korean EFL writing

1. Introduction

The research of cohesion features has been regarded important in second language study since cohesion is a crucial element for understanding challenging texts that present knowledge demands to the reader (McNamara and Kintsch 1996, McNamara et al. 1996). Cohesion refers to the presence/absence of linguistic cues in the text that allow the reader to make connections between the ideas in the text (Halliday and Natthiessen 2004). Cohesive devices are typically classified into three different categories: local, global and text cohesion (Crossley et al. 2016). While local cohesion cues include overlapping words and concepts between sentences and explicit connectives such as *because*, *thus* and *accordingly*, global cohesion cues include semantic/lexical overlap between paragraphs in a text where words or ideas in one paragraph are repeated in subsequent paragraphs (Foltz 2007, Halliday and Hasan 1976). Additionally, text cohesion cues include givenness in which cohesion is measured across the text based on the number of words that are new or given. Cohesion is critical to the understanding of how language functions and is premised on the notion that the linking of ideas allows for the creation of coherent discourse (Halliday and Hasan 1976), and language features related to cohesion such as the use of connectives and word overlap have been considered productive predictors of L2 writing proficiency (Connor 1990, Ferris 1994, Frase et al. 1997, Grant and Ginther 2000, Reid 1990, Reppen 1994, Silva 1993).

A number of studies that examine the use of cohesive devices for L2 learners have been conducted to investigate the relations with writing quality. Early studies of L2 writers show that the production of local and text cohesive devices is positively related to writing quality. For example, the quality of essays written in English by undergraduate Iranian students was correlated with the number of cohesive types used in the essays (Jafarpur 1991). In addition, Liu and Braine (2005) reported that essay quality scores for undergraduate Chinese L2 writers correlated with the total number of cohesive devices in the text. Yang and Sun (2012) also reported on strong correlations between the total number of cohesive devices and writing quality for argumentative essays written by Chinese writers of English. On the other hand, recent studies using computational tools show negative correlations between writing quality and local cohesion. For example, Crossley and McNamara (2012) found that local and text cohesive devices (e.g., content word overlap between adjacent sentences, semantic similarity between sentences, positive logical connectives, and aspect repetition) were negatively correlated with essay quality for Hong Kong high school students.

Cohesion has been analyzed by automated tools since they afford speed, flexibility and reliability; an automatic approach to assessing text cohesion includes Coh-Metrix (Graesser et al. 2004, McNamara and Graesser 2012) and the Tool for the Automatic Analysis of Cohesion (TAACO, Crossley et al. 2019). Coh-Metrix is a natural language processing tool that measures linguistic features associated with text cohesion. This is a freely available online tool that analyzes a number of linguistic elements related to lexical sophistication, syntactic complexity, and cohesion. Even though Coh-Metrix has had a great influence on our understanding of language and discourse, it has some weaknesses concerning usability and broad measurements of its cohesion indices. Because Coh-Metrix is a web tool, the version available to the public (version 3.0) is not housed on a user's hard drive and batch processing of text is disallowed. Also, the Coh-Metrix cohesion indices mostly focus on local and overall text cohesion rather than global cohesion; the publicly available tool includes a limited number of cohesion indices, i.e., 17 local cohesion indices (latent semantic analysis and connectives), and 12 local cohesion indices (referential cohesion).

One of the previous studies that employed Coh-Metrix includes Crossley and McNamara (2012) where they analyzed a corpus of essays written by graduating Hong Kong high school students for the Hong Kong Advanced Level Examination to predict second language writing proficiency using linguistic features involving cohesive measures. Following Whitten and Frank (2005), they divided the corpus into a training set for selecting linguistic variables and a testing set for calculating the amount of variance that the selected variables explained in an independent corpus. They predicted that higher-rated essays would contain more cohesive devices than lower-rated essays based on the findings of past studies (Connor 1990, Jin 2001, Witte and Faigley 1981). They have found that like first language writers, L2 writers evaluated as being highly proficient do not produce texts that are more cohesive; second language writers judged as highly proficient provide readers with less temporal cohesion and word overlap.

Guo et al. (2013) also used Coh-Metrix to explore whether and how linguistic features related to cohesion help to characterize L2 writing proficiency in the Test of English as a Foreign Language (TOEFL iBT) independent and integrated writing tasks, in which an independent writing task is timed, impromptu writing while an integrated writing task uses reading and/or listening materials as stimuli for composing an essay. They compared differences between integrated and independent essays to better understand the task requirements and expectations for each. The results of their study showed that linguistic features can be used to significantly predict essay scores in the integrated as well as the independent writing. In the case of independent essays, fewer conditional connectives were significant predictors of essay scores; thus, cohesion seemed to negatively affect quality in judgments of independent writing quality, while other cohesion devices like aspect repetition and content word overlap demonstrated a negative correlation with essay quality. This is unlike judgments of integrated essay quality where cohesive devices were positively correlated with essay quality like semantic similarity, noun overlap and tense repetition. Thus, they conclude that cohesion is an important property of human ratings of integrated essays but not independent essays.

On the other hand, Crossley et al. (2016) selected cohesion indices from TAACO as well as from Coh-Metrix since TAACO provides a greater breadth of global cohesion indices, synonym overlap indices, and part of speech tagged cohesion indices. Using these automated cohesion indices that measure both local and global cohesion, they assess cohesion development in descriptive essays written by L2 learners at the beginning, middle and end of a semester long writing course. They reported that L2 writers generally wrote essays that demonstrated greater local, global and text cohesion from the beginning of the semester until the end of the semester. The strongest growth was an increase in noun overlap between paragraphs, indicating an increase in global cohesion across essays. Regarding predicting human judgments of L2 writing organization, two indices that measure function word cohesion were predictive: a positive global predictor (adjacent overlap between paragraphs: function words) and text cohesion predictor (function word TTR).

The present study adopts the formulation of cohesion operationalized in TAACO, which has been known to be advanced predictors of writing quality based on fine-grain analyses of cohesion. The purpose of this research paper is to carry out a cross-sectional study of cohesion in the EFL writing of Korean university level learners. This study employs cohesion measures to examine differences in the linguistic cues written on an essay topic by learners at four CEFR proficiency bands. The current study attempts to answer the following research questions: (i) what is the correlation between fine-grained indices of cohesion and EFL writing proficiency levels? and (ii) if cohesion elements differ across the proficiency levels in Korean EFL learners, which elements distinguish cohesion across levels in the essays written by university level Korean learners? The contribution of this study is that it provides a methodology to predict EFL writing proficiency given a great deal of corpus data along with the state-of-the-art tools for data/statistical analysis. The study of cohesion features is important in that understanding the function of these features and how they relate to writing proficiency significantly impacts our understanding of the importance of linguistic features in essays by explaining the role that text variables play in predicting writing proficiency. These findings can be used to inform writing pedagogy and provide models for computer-assisted language learning.

2. Methodology

The data used in this study are a subset of the International Corpus Network of Asian Learners of English, ICNALE (Ishikawa 2013). The ICNALE is one of the biggest learner corpora publicly available, which contains more than 10,000 topic-controlled essays and speeches produced by college students in 11 different Asian countries, i.e., Korea, Japan, Taiwan, Singapore, the Philippines, Indonesia, Malaysia, Pakistan, Thailand, Hong Kong, and China. The ICNALE contains four different modules (written essays, edited essays, spoken monologue, and spoken dialogue), and the module of *written essays* is chosen for the present analysis. This module controls prompts in which learners were supposed to show their ideas about two topics, i.e., (i) Smoking has to be completely prohibited at all the restaurants in the country; and (ii) It is important for university students to have a part-time job. The dataset examined in this study includes 69,950 words written by Korean college students in four CEFR-related proficiency bands (Table 1) in answer to the second prompt (*part-time job*). The number of texts per band is equivalent to that of students in each band because each participant wrote a single text for the given topic; that is, a total of 300 different essays was produced by 300 different students. An outline of the learner corpus analyzed per proficiency band is shown in Table 2.

Table 1. Mapping of Test Scores on CEFR Proficiency Bands (Ishikawa 2013)

CEFR	TOEIC	TOEFL PBT	TOEFL iBT	IELTS	STEP	VST
A2	-545	-486	-56	3+	3+	-24
B1_1	550+	487+	57+	4+	2+	25+
B1_2	670+	527+	72+	4+	2+	36+
B2+	785+	567+	87+	5+	Pre1+	47+

Table 2. Summary of the Learner Corpus

CEFR	No. of texts	No. of words	Mean	*SD*
A2	75	16875	220.40	20.84
B1_1	61	13699	219.90	21.81
B1_2	88	20632	230.35	31.64
B2+	76	18744	242.14	35.82
Total	300	69950	228.73	29.93

The methodology this study employed was the measurement of cohesive features using the Tool for the Automatic Analysis of Cohesion 2.0.4 (TAACO, Crossley et al. 2019), which is a freely available text analysis tool like Coh-Metrix introduced in the previous section. These two automatic tools differ in that TAACO is housed on the user's hard drive and enables users to work independently of an Internet connection and external servers, allowing for secure processing of sensitive data. Although TAACO is written in Python, it is implemented in a way that requires no knowledge of Python programming; it is user-friendly, intuitive graphical user interface that can be started by simply double-clicking the TAACO icon. This text analysis tool works on most operating systems such as Windows, Mac and Linux and provides hundreds of automatically computed linguistic features related to text cohesion. TAACO incorporates a part-of-speech tagger from the Natural Language Tool Kit (Bird et al. 2009) and synonym sets from the WordNet lexical database (Miller 1995). The tool is different from Coh-Metrix in that it reports on a greater number and variety of local, global, and overall text cohesion markers.

TACCO allows the fine-grained analysis of cohesion with four different features, i.e., (i) lexical overlap, (ii) connectives, (iii) semantic overlap, and (iv) givenness. First, *lexical overlap* includes 24 indices that measure nouns (e.g., number of noun lemma types), verbs (e.g., number of verb lemma types), adjectives (e.g., number of adjective lemma types), adverbs (e.g., number of adverb lemma types). Next, the *connectives* component includes 25 indices that assess the use of theoretical/rhetorical lists of connectives (e.g., conjunctions, disjunctions, coordinating conjuncts, lexical subordinators). The third cohesion-related component, *semantic overlap*, includes eight indices that compute the overlap of synonyms across sentences and paragraphs (e.g., average sentence to sentence overlap of noun synonyms, average sentence to sentence overlap of verb synonyms). Last, the *givenness* component includes four indices that measure ratio of pronouns to nouns, incidence of demonstratives, and definite articles (e.g., pronoun density, repeated content lemmas).

Although TAACO includes a broad range of fine-grained cohesion measures that capture local, global, and overall text cohesion markers, the current study focuses on local and overall text cohesion since the learner corpus consists of one-paragraph essays. In order to see whether cohesion features differ across the proficiency levels in Korean EFL writing, this study examined four different components related to cohesion using TAACO, i.e., (i) lexical overlap; (ii) connectives; (iii) semantic overlap; and (iv) givenness, as shown in Table 3. The components of *lexical overlap*, *connectives*, and *semantic overlap* include 57 indices of local cohesion; the *givenness* component includes four indices of overall text cohesion. Table 4 lists the entire set of cohesive indices included in the four components.

Table 3. Overview of Cohesion Features (adapted from Crossley et al. 2016: 1231)

Feature	Description	Example of high cohesion
Lexical overlap	Overlap between nouns, arguments, stems, content words, and POS tags for sentences	The *sun* was bright. The day was *sunny*.
Connectives	A number of theoretical/rhetorical list of connectives	*Firstly*, she was happy and excited.
Semantic overlap	Overlap of synonyms across sentences	The *animal* was huge. It was a *dog*.
Givenness	Ratio of pronouns to nouns; incidence of demonstratives; definite articles	The girl was satisfied with what *she* had.

Table 4. Cohesive Indices Analyzed in this Study (adapted from Crossley et al. 2016)

Component	Cohesion type	In text name
Lexical overlap	Local cohesion	Adjacent sentence overlap all lemmas
	Local cohesion	Binary adjacent sentence overlap all lemmas
	Local cohesion	Adjacent two-sentence overlap all lemmas
	Local cohesion	Binary adjacent two-sentence overlap all lemmas
	Local cohesion	Adjacent sentence overlap content lemmas
	Local cohesion	Binary adjacent sentence overlap content lemmas
	Local cohesion	Adjacent two-sentence overlap content lemmas
	Local cohesion	Binary adjacent two-sentence overlap content lemmas
	Local cohesion	Adjacent sentence overlap noun lemmas
	Local cohesion	Binary adjacent sentence overlap noun lemmas
	Local cohesion	Adjacent two-sentence overlap noun lemmas
	Local cohesion	Binary adjacent two-sentence overlap noun lemmas
	Local cohesion	Adjacent sentence overlap verb lemmas
	Local cohesion	Binary adjacent sentence overlap verb lemmas
	Local cohesion	Adjacent two-sentence overlap verb lemmas
	Local cohesion	Binary adjacent two-sentence overlap verb lemmas
	Local cohesion	Adjacent sentence overlap adjective lemmas
	Local cohesion	Binary adjacent sentence overlap adjective lemmas
	Local cohesion	Adjacent two-sentence overlap adjective lemmas
	Local cohesion	Binary adjacent two-sentence overlap adjective lemmas
	Local cohesion	Adjacent sentence overlap adverb lemmas

	Local cohesion	Binary adjacent sentence overlap adverb lemmas
	Local cohesion	Adjacent two-sentence overlap adverb lemmas
	Local cohesion	Binary adjacent two-sentence overlap adverb lemmas
Connectives	Local cohesion	Basic connectives
	Local cohesion	Conjunctions
	Local cohesion	Disjunctions
	Local cohesion	Lexical subordinators
	Local cohesion	Coordinating conjuncts
	Local cohesion	Addition
	Local cohesion	Sentence linking
	Local cohesion	Order
	Local cohesion	Reason and purpose
	Local cohesion	All causal connectives
	Local cohesion	Positive causal connectives
	Local cohesion	Opposition
	Local cohesion	Determiners
	Local cohesion	Demonstratives
	Local cohesion	Attended demonstratives
	Local cohesion	Unattended demonstratives
	Local cohesion	All additive connectives
	Local cohesion	All logical connectives
	Local cohesion	Positive logical connectives
	Local cohesion	Negative logical connectives
	Local cohesion	Temporal connectives
	Local cohesion	Positive intentional connectives
	Local cohesion	All positive connectives
	Local cohesion	All negative connectives
	Local cohesion	All connectives
Semantic overlap	Local cohesion	Synonym overlap (sentence, noun)
	Local cohesion	Synonym overlap (sentence, verb)
	Local cohesion	LSA cosine similarity (adjacent sentences)
	Local cohesion	LSA cosine similarity (two adjacent sentences)
	Local cohesion	LDA divergence (adjacent sentences)
	Local cohesion	LDA divergence (two adjacent sentences)
	Local cohesion	Word2vec similarity (adjacent sentences)
	Local cohesion	Word2vec similarity (two adjacent sentences)
Givenness	Text cohesion	Pronoun density
	Text cohesion	Pronoun to noun ratio
	Text cohesion	Repeated content lemmas
	Text cohesion	Repeated content lemmas and pronouns

3. Results and Discussion

A multiple linear regression analysis was carried out for each index to examine differences in cohesion values across different proficiency levels. First, normality was tested threefold, i.e., (i) numerically with skewness and kurtosis (for a normal distribution both values should be close to zero); (ii) graphically with both distribution and Q-Q plots (when data are normally distributed, they have a symmetrical distribution in a distribution plot and all the points are close to the diagonal reference line in a Q-Q plot); and (iii) statistically with Shapiro-Wilk test (normally distributed data show no significant deviation) (Shapiro & Wilk 1965). Any variables that failed to obey a normal distribution were eliminated from further consideration. Next, Pearson correlations were conducted on the remaining variables to find out whether they were significantly correlated with proficiency bands. Any indices that did not satisfy an absolute correlation value of $r \geq 0.100$ with CEFR level and a significance of $p < 0.001$ were dropped (Cohen 1988). Then, the remaining indices were checked for multicollinearity with both VIF and tolerance values to ensure that the resulting model contained unique indices only and multicollinear indices did not exaggerate the results of the regression analysis (Tabachnick and Fidell 2014). Last, the remaining indices were entered into a stepwise multiple linear regression that used the AIC method (Akaike 1974). All statistical analyses involving normality check, correlation test and multiple linear regression were performed using JASP (JASP Team 2022).

3.1. Indices of Lexical Overlap

Out of a total of 24 *lexical overlap* indices, 21 were removed since they failed to obey a normal distribution, in which a Shapiro-Wilk analysis exhibited a significant departure from normality, i.e., (i) adjacent sentence overlap all lemmas ($W = 0.922$, $p < 0.001$); (ii) binary adjacent sentence overlap all lemmas ($W = 0.842$, $p < 0.001$); (iii) adjacent two-sentence overlap all lemmas ($W = 0.952$, $p < 0.001$); (iv) binary adjacent two-sentence overlap all lemmas ($W = 0.466$, $p < 0.001$); (v) adjacent sentence overlap content lemmas ($W = 0.953$, $p < 0.001$); (vi) binary adjacent two-sentence overlap content lemmas ($W = 0.933$, $p < 0.001$); (vii) adjacent sentence overlap noun lemmas ($W = 0.967$, $p < 0.001$); (viii) binary adjacent sentence overlap noun lemmas ($W = 0.898$, $p < 0.001$); (ix) binary adjacent two-sentence overlap noun lemmas ($W = 0.975$, $p < 0.001$); (x) adjacent sentence overlap verb lemmas ($W = 0.935$, $p < 0.001$); (xi) binary adjacent sentence overlap verb lemmas ($W = 0.871$, $p < 0.001$); (xii) adjacent two-sentence overlap verb lemmas ($W = 0.974$, $p < 0.001$); (xiii) binary adjacent two-sentence overlap verb lemmas ($W = 0.933$, $p < 0.001$); (xiv) adjacent sentence overlap adjective lemmas ($W = 0.900$, $p < 0.001$); (xv) binary adjacent sentence overlap adjective lemmas ($W = 0.796$, $p < 0.001$); (xvi) adjacent two-sentence overlap adjective lemmas ($W = 0.938$, $p < 0.001$); (xvii) binary adjacent two-sentence overlap adjective lemmas ($W = 0.853$, $p < 0.001$); (xviii) adjacent sentence overlap adverb lemmas ($W = 0.784$, $p < 0.001$); (xix) binary adjacent sentence overlap adverb lemmas ($W = 0.662$, $p < 0.001$); (xx) adjacent two-sentence overlap adverb lemmas ($W = 0.857$, $p < 0.001$); and (xxi) binary adjacent two-sentence overlap adverb lemmas ($W = 0.714$, $p < 0.001$). The remaining three indices (i.e., binary adjacent sentence overlap content lemmas, adjacent two-sentence overlap content lemmas, and adjacent two-sentence overlap noun lemmas) satisfied the minimum thresholds of $p < 0.001$ and $r \geq 0.100$ with proficiency levels and were entered into a stepwise linear regression (see Appendix A). The resulting model, which included a single index (i.e., binary adjacent sentence overlap content lemmas), explained 3.9% ($r = 0.198$, $R^2 = 0.039$) of the variance in CEFR levels (see Table 5).

Table 5 Summary of Multiple Regression Model for *Lexical Overlap* Variables

Entry	Predictors included	*R*	R^2	R^2 change	*β*	*SE*	*B*
1	Binary adjacent sentence overlap content lemmas	0.198	0.039	0.036	1.070	0.307	0.198

The relationship between lexical overlap indices and CEFR levels was significant and it showed a small effect size.[1] Three indices associated to lexical overlap fulfilled the inclusion criteria so that they were entered into a multiple linear regression, i.e., binary adjacent sentence overlap content lemmas, adjacent two-sentence overlap

[1] A multiple regression test calculates an effect size using a multiple correlation coefficient, i.e., trivial $r < 0.100$, small $0.100 < r < 0.300$, medium $0.300 < r < 0.500$, large $r > 0.500$ (Goss-Sampson 2020).

content lemmas, and adjacent two-sentence overlap noun lemmas. The resulting model included a single index (binary adjacent sentence overlap content lemmas) and explained approximately 4% of the variance in proficiency levels. These results present support for the importance of the lexical overlap component in indexing EFL writing proficiency, and are consistent with previous findings where Lee (2021) reported the frequency of lemmas is positively correlated to predict writing quality. Although she failed to mention which type of lemmas was more frequent in the EFL writing, the current result has found that content lemmas overlap more frequently than function lemmas.

A linear regression using this variable explained 3.9% of the variance in proficiency levels (see Table 5). This result indicates that the most advanced level had a tendency to include a substantial number of sentences that contained content lemma overlap with the next sentence in their essays. The writing examples from the ICNALE corpus shown in Table 6 clarifies this tendency, showing that the beginner only had a couple of overlaps for two content words (*job* and *part-time*) while the proficient learner produced more cohesive writing with a total of nine overlaps for three words (*experience*, *work*, and *learn*). This result finds echoes in those of Witte and Faigley (1981) and Ferris (1994) who found that advanced ESL students use a greater number of lexical and syntactic devices in their compositions in addition to a wider range of cohesive devices such as synonymy, antonymy, referencing and the definite article.

Table 6 Examples from ICNALE Essays: Binary Adjacent Sentence Overlap Content Lemmas

Level	Example	Learner code
A2	*Cafeteria working is tired and stressful. Because hard job. Every day smile and fast move. Boss is very nitpicking to me. My feeling is terrifically peed off. Part time job about positive thing. I learned a basic position in service and it is my first experience to earn money myself. Hunt Part time job for reference. We should do what we can to eradicate it.*	W_KOR_PTJ0_043_A2_0
B2+	*However, if they gain experience as a part-time worker, they will be able to meet various people and learn many things which can be only attained from real-world experience. I personally have experiences working at a hotel and a café. I really enjoyed working there even though it was hard, and I learned a lot from those experiences.*	W_KOR_PTJ0_285_B2_0

Next, a series of one-way ANOVAs implemented in JASP were conducted to compare the effects of different proficiency bands; for every model, the dependent variable was each index and the fixed factor was the CEFR proficiency level. Descriptive statistics and ANOVA results of all dependent variables are given in Appendix B. Independent one-way ANOVAs showed a significant effect of ten variables, i.e., (i) binary adjacent sentence overlap lemmas: $F(3, 296) = 9.097$, $p < 0.001$; (ii) adjacent two-sentence overlap all lemmas: $F(3, 296) = 3.376$, $p < 0.05$; (iii) binary adjacent two-sentence overlap all lemmas: $F(3, 296) = 3.669$, $p < 0.05$; (iv) binary adjacent sentence overlap content lemmas: $F(3, 296) = 5.725$, $p < 0.001$; (v) binary adjacent two- sentence overlap content lemmas: $F(3, 296) = 9.168$, $p < 0.001$; (vi) adjacent sentence overlap noun lemmas: $F(3, 296) = 3.095$, $p < 0.05$; (vii) adjacent sentence overlap noun lemmas: $F(3, 296) = 5.657$, $p < 0.001$; (viii) adjacent two-sentence overlap noun lemmas: $F(3, 296) = 3.389$, $p < 0.05$; (ix) binary adjacent two-sentence overlap noun lemmas: $F(3, 296) = 7.483$, $p < 0.001$; and (x) binary adjacent two-sentence overlap verb lemmas: $F(3, 296) = 3.596$, $p < 0.05$.

Post hoc testing then was conducted for the index where the minimum correlation thresholds were fulfilled and the ANOVA results were significant to see whether there is a significant difference between the levels (see Tables 7 through 11). The highest proficiency level had a tendency to show greater use of the lexical overlap markers than the lower proficiency levels, as shown in a representative figure given in Figure 1, where post hoc testing using Tukey's correction showed that the highest level resulted in significantly greater use of content word overlap than the remaining lower levels (B2+ vs. A2/B1_1/B1_2: $p < 0.05$) (See Table 8).

Table 7. Post hoc Comparisons of CEFR Level and Binary Adjacent Sentence Overlap All Lemmas

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	-0.016	0.021	-0.770	0.868
A2 vs. B1_2	1.935e-4	0.019	0.010	1.000
A2 vs. B2+	-0.088	0.020	-4.417	<0.001***
B1_1 vs. B1_2	0.016	0.020	0.806	0.852
B1_1 vs. B2+	-0.072	0.021	-3.410	0.004**
B1_2 vs. B2+	-0.088	0.019	-4.601	<0.001***

Table 8. Post hoc Comparisons of CEFR Level and Binary Adjacent Sentence Overlap Content Lemmas

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	-0.011	0.035	-0.301	0.991
A2 vs. B1_2	-0.020	0.032	-0.632	0.922
A2 vs. B2+	-0.021	0.033	-3.667	0.002**
B1_1 vs. B1_2	-0.010	0.034	-0.285	0.992
B1_1 vs. B2+	-0.111	0.035	-3.170	0.009**
B1_2 vs. B2+	-0.101	0.032	-3.177	0.009**

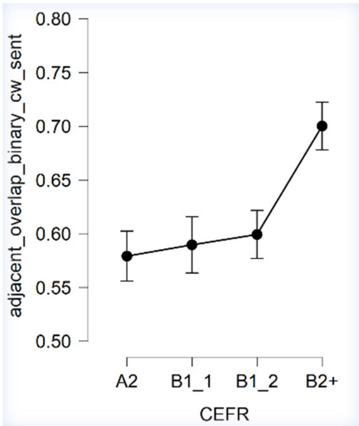

Figure 1. Comparisons of Binary Adjacent Sentence Overlap Content Lemmas between CEFR Levels
(adjacent_overlap_binary_sent is an index name for binary adjacent sentence overlap content lemmas; error bars indicate standard error)

Table 9. Post hoc Comparisons of CEFR Level and Binary Adjacent Two Sentence Overlap Content Lemmas

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	-6.149e-4	0.031	-0.020	1.000
A2 vs. B1_2	0.008	0.028	0.274	0.993
A2 vs. B2+	-0.120	0.029	-4.171	<0.001***
B1_1 vs. B1_2	0.008	0.030	0.279	0.992
B1_1 vs. B2+	-0.120	0.030	-3.929	<0.001***
B1_2 vs. B2+	-0.128	0.028	-4.610	<0.001***

Table 10. Post hoc Comparisons of CEFR Level and Binary Adjacent Sentence Overlap Noun Lemmas

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	0.001	0.039	0.027	1.000
A2 vs. B1_2	-0.018	0.035	-0.505	0.958
A2 vs. B2+	-0.128	0.036	-3.509	0.003**
B1_1 vs. B1_2	-0.019	0.037	-0.504	0.958
B1_1 vs. B2+	-0.129	0.038	-3.349	0.005**
B1_2 vs. B2+	-0.110	0.035	-3.140	0.010*

Table 11. Post hoc Comparisons of CEFR Level and Binary Adjacent Two Sentence Overlap Noun Lemmas

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	0.022	0.040	0.547	0.947
A2 vs. B1_2	0.008	0.036	0.234	0.995
A2 vs. B2+	-0.135	0.038	-3.592	0.002**
B1_1 vs. B1_2	-0.013	0.039	-0.345	0.986
B1_1 vs. B2+	-0.157	0.040	-3.949	<0.001***
B1_2 vs. B2+	-0.144	0.036	-3.968	<0.001***

3.2. Indices of Connectives

The variables of connectives were statistically analyzed regarding normal distribution and Pearson correlation as in the components of lexical overlap. Out of a total of 25 *connectives* indices, 14 were discarded because they showed a significant departure from a normal distribution, i.e., (i) conjunctions (W = 0.978, p < 0.001); (ii) disjunctions (W = 0.727, p < 0.001); (iii) coordinating conjuncts (W = 0.934, p < 0.001); (iv) order (W = 0.885, p < 0.001); (v) reason and purpose (W = 0.956, p < 0.001); (vi) all causal connectives (W = 0.974, p < 0.001); (vii) opposition (W = 0.938, p < 0.001); (viii) demonstratives (W = 0.977, p < 0.001); (ix) attended demonstratives (W = 0.883, p < 0.001); (x) unattended demonstratives (W = 0.961, p < 0.001); (xi) negative logical connectives (W = 0.952, p < 0.001); (xii) temporal connectives (W = 0.934, p < 0.001); (xiii) positive intentional connectives (W = 0.912, p < 0.001); and (xiv) all negative connectives (W = 0.969, p < 0.001). All the remaining indices failed to meet the minimum thresholds with CEFR levels and were eliminated from the analysis, i.e., (i) basic connectives (r = -0.150, p = 0.009); (ii) lexical subordinators (r = 0.047, p = 0.417); (iii) addition (r = -0.081, p = 0.161); (iv) sentence linking (r = -0.070, p = 0.230); (v) positive causal connectives (r = -0.020, p = 0.729); (vi) determiners (r = -0.033, p = 0.571); (vii) all additive connectives (r = -0.002, p = 0.975); (viii) all logical connectives (r = 0.085, p = 0.144); (ix) positive logical connectives (r = -0.066, p = 0.257); (x) all positive connectives (r = -0.058, p = 0.320); and (xi) all connectives (r = 0.007, p = 0.904) (see Appendix C for the entire correlation results).

Next, a series of one-way ANOVAs were conducted to compare the effects of different proficiency levels; for all models, the dependent variable was each index and the fixed factor was the proficiency level. Descriptive statistics and ANOVA results of all dependent variables are given in Appendix D. Independent one-way ANOVAs showed a significant effect of five variables, i.e., (i) basic connectives: $F(3, 296) = 4.303, p < 0.01$; (ii) conjunctions: $F(3, 296) = 4.489, p < 0.01$; (iii) sentence linking: $F(3, 296) = 3.393, p < 0.05$; (iv) order: $F(3, 296) = 8.019, p < 0.001$; and (v) all positive connectives: $F(3, 296) = 3.545, p < 0.05$. Although ANOVA reported a significant difference, since these indices did not meet the correlation inclusion criteria, post hoc testing was not conducted furthermore.

3.3. Indices of Semantic Overlap

All the indices failed to obey normality assumption and were removed from further analysis, i.e., (i) synonym overlap (sentence, noun) ($W = 0.425$, $p < 0.001$); (ii) synonym overlap (sentence, verb) ($W = 0.412$, $p < 0.001$); (iii) LSA cosine similarity (adjacent sentences) ($W = 0.974$, $p < 0.001$); (iv) LSA cosine similarity (two adjacent sentences) ($W = 0.798$, $p < 0.001$); (v) LDA divergence (adjacent sentences) ($W = 0.184$, $p < 0.001$); (vi) LDA divergence (two adjacent sentences) ($W = 0.114$, $p < 0.001$); (vii) word2vec similarity (adjacent sentences) ($W = 0.491$, $p < 0.001$); and (viii) word2vec similarity (two adjacent sentences) ($W = 0.304$, $p < 0.001$) (see Appendix E for the entire results).

To compare the impacts of different proficiency levels, a series of one-way ANOVAs were carried out. For every model, the dependent variable was each of the *semantic overlap* indices and the fixed factor was the writing proficiency band. Appendix F provides the descriptive statistics and ANOVA results of all dependent variables involved in the indices of semantic overlap. Independent one-way ANOVAs showed a significant effect of four variables, i.e., (i) LSA cosine similarity (adjacent sentences): $F(3, 296) = 4.614$, $p < 0.01$; (ii) LSA cosine similarity (two adjacent sentences): $F(3, 296) = 7.873$, $p < 0.001$; (iii) word2vec similarity (adjacent sentences): $F(3, 296) = 7.364$, $p < 0.001$; and (iv) word2vec similarity (two adjacent sentences): $F(3, 296) = 3.340$, $p < 0.05$.

Post hoc tests were carried out for the indices in which the minimum correlation thresholds were satisfied and the ANOVA results were significant in order to see whether there is a significant difference between the levels (see Tables 12 and 13). These two indices (i.e., LSA cosine similarity and word2vec similarity) are considered important elements of discourse cohesion in terms of NLP (Natural Language Processing) techniques, where computational models of semantic memory rely on unsupervised learning methods that measure cohesion between textual fragments (Bestgen and Vincze 2012, Cree and Armstrong 2012). Common models include semantic vector spaces using LSA (latent semantic analysis, Landauer et al. 1998) and word2vec vector space representations (Mikolov et al. 2013). For example, word2vec depends on a neural-network model, in which each word's embedding is computed using the context around it within the training dataset; thus, words with similar contexts are represented as being closer, whereas words co-occurring in dissimilar contexts are represented as being farther apart in different regions of the vector space. As shown in Figures 4 and 5, variables measured by LSA and word2vec have found that the most proficient learners tended to increase cohesion at the sentence level by increasing the semantic similarity between sentences, as compared to the lower level L2 learners.

Table 12. Post hoc Comparisons of CEFR Level and LSA Cosine Similarity (Two Adjacent Sentences)

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	0.015	0.016	0.931	0.788
A2 vs. B1_2	-0.007	0.014	-0.515	0.956
A2 vs. B2+	-0.056	0.015	-3.709	0.001**
B1_1 vs. B1_2	-0.022	0.015	-1.448	0.470
B1_1 vs. B2+	-0.070	0.016	-4.445	<0.001***
B1_2 vs. B2+	-0.048	0.014	-3.339	0.005**

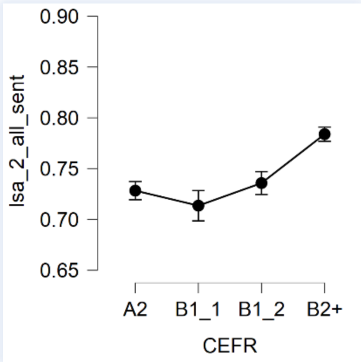

Figure 4. Comparisons of LSA Cosine Similarity (Two Adjacent Sentences) between CEFR Levels
(lsa_2_all_sent is an index name for LSA cosine similarity (two adjacent sentences); error bars indicate standard error)

Table 13. Post hoc Comparisons of CEFR Level and Word2vec Similarity (Adjacent Sentences)

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	0.013	0.011	1.205	0.624
A2 vs. B1_2	-0.011	0.010	-1.062	0.713
A2 vs. B2+	-0.036	0.010	-3.421	0.004**
B1_1 vs. B1_2	-0.024	0.011	-2.248	0.113
B1_1 vs. B2+	-0.049	0.011	-4.447	<0.001***
B1_2 vs. B2+	-0.025	0.010	-2.490	0.064

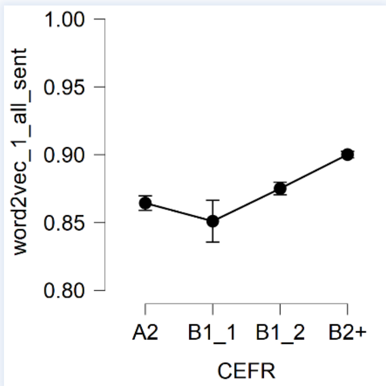

Figure 5. Comparisons of Word2vec Similarity (Adjacent Sentences) between CEFR Levels
(word2vec_1_all_sent is an index name for word2vec similarity (adjacent sentences); error bars indicate standard error)

3.4. Indices of Givenness

Two indices of *givenness* were eliminated because they did not exhibit normal distributions, i.e., pronoun density ($W = 0.953$, $p < 0.001$) and pronoun to noun ratio ($W = 0.930$, $p < 0.001$). One of the remaining indices did not fulfill the minimum thresholds with CEFR levels and were removed from the analysis, i.e., repeated content lemmas. The remaining index (repeated content lemmas and pronouns) was entered into a stepwise multiple linear regression (see Appendix G) and the resulting model that included a single index (repeated content lemmas and pronouns) explained s% ($r = 0.140$, $R^2 = 0.020$) of the variance in proficiency band (see Table 14).

Table 14. Summary of Multiple Regression Model for *Givenness* Variables

Entry	Predictors included	*R*	R^2	R^2 change	*β*	*SE*	*B*
1	Repeated content lemmas and pronouns	0.140	0.020	0.016	2.221	0.910	0.140

Although the relationship between indices of givenness and CEFR proficiency levels was significant, it exhibited a small effect size. A single index associated to givenness satisfied the inclusion criteria and was entered into a stepwise linear regression, i.e., repeated content lemmas and pronouns. The resulting model included this variable and explained 2% of the variance in proficiency levels. These results present support for the important value of the givenness component in indexing EFL writing proficiency. The results also suggest that advanced learners had a tendency to contain greater number of repeated content lemmas and third person pronouns, as illuminated in learner writing samples of repeated content lemmas and pronouns in the ICNALE essays (see Table 15). Examples given in this table show that although no third person pronoun was shown in the lower proficiency level writing, the advanced learner had three of these, while the number of repeated content lemmas was ten for four words in the beginning learner and 20 for eight words in the proficient learner. These findings are consistent with previous studies where Lee (2021) reported that advanced EFL learners showed repeated use of nouns and pronouns in their essays significantly more than beginning learners. This would present further support for the fact that the givenness markers such as content lemmas and pronouns should be considered an important measurement of EFL writing proficiency.

Table 15. Examples from ICNALE Essays: Repeated Content Lemmas and Pronouns

Level	Example	Learner code
B1_1	*After graduating from high school and had hamburgers and pizza at the store part-time job. Most of the women responsible for the counter in the kitchen, but I was responsible for a burger. In the kitchen and everyone except me was the man. Working one month and earned 640,000 won. But the money disappeared in less than one week. I'm making money as a gift to my parents and my grandfather gave to the allowance of 50,000 won.*	W_KOR_PTJ0_049_B1_1
B2+	*Second, having a part-time job deprives the time that college students can develop the personality and knowledge. College students should study a major and English. Spending the time that they study is more important than spending the time that they work part-time. By working part-time, they do not have the time enough to study a major and English.*	W_KOR_PTJ0_227_B2_0

Next, a series of one-way ANOVAs were carried out to compare the outcomes of proficiency levels; for every model, the dependent variable was each of the givenness indices, and the fixed factor was the L2 proficiency level. Appendix H gives descriptive statistics and ANOVA results of all dependent variables that are included in the indices of givenness. Independent one-way ANOVAs showed a significant effect of three variables, i.e., (i) pronoun density: $F(3, 296) = 11.741$, $p < 0.001$; (ii) pronoun to noun ratio: $F(3, 296) = 11.060$, $p < 0.001$; and (iii) repeated content lemmas and pronouns: $F(3, 296) = 3.486$, $p < 0.05$.

Next, post hoc tests were carried out for the indices in which the minimum correlation thresholds were fulfilled and the ANOVA results were significant in order to see whether there is a significant difference between different levels (see Tables 16 and 17). As shown in Table 16, post hoc testing showed that the level B2+ led to significantly greater use of third person pronouns than the level A2 ($p < 0.001$); the most advanced writer also had significantly greater use of third person pronouns than the lower intermediate learner (B1_1: $p < 0.001$) and the higher intermediate learner (B1_2: $p < 0.001$).

The exact same tendency for more proficient writers was also observed in the index of pronoun to noun ratio (A2/B1_1/B1_2 vs. B2+ $p < 0.001$, Table 17). Although there was a significant difference between the use of third person pronouns in the level B2+ and the other levels as shown in Figures 7 and 8, it is rather surprising that they were used less in the B1_1 level than in the A2 level, even though this difference was not statistically significant. This unexpected tendency might be attributed to the range of VST (L2 vocabulary size test) scores. The EFL learners of the ICNALE were required to take a standard VST covering the top 5,000 word levels (Nation and Beglar 2007). Earlier studies have found that it is proper to assess the vocabulary size of non-native speakers with maximum 5,000 words (Meara and Milton 2003, Milton 2010). However, it does not seem that the difference of VST scores is sufficient to distinguish between A2 (VST = -24) and B1_1 (VST = 25+). Another possible reason for this unusual tendency could be that the ICNALE uses the 2010 mapping scheme based on the official mapping guidelines offered by administrators of TOEFL in the year of 2006 though ETS has released a new technical report on the mapping of the TOEFL iBT scores on the CEFR where they relate 42+ to B1 and 72+ to B2 (Papageorgiou et al. 2015).

Table 16. Post hoc Comparisons of CEFR Level and Pronoun Density

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	0.004	0.005	0.897	0.807
A2 vs. B1_2	-0.004	0.004	-0.939	0.784
A2 vs. B2+	-0.022	0.005	-4.704	<0.001***
B1_1 vs. B1_2	-0.009	0.005	-1.814	0.269
B1_1 vs. B2+	-0.026	0.005	-5.353	<0.001***
B1_2 vs. B2+	-0.017	0.004	-3.947	<0.001***

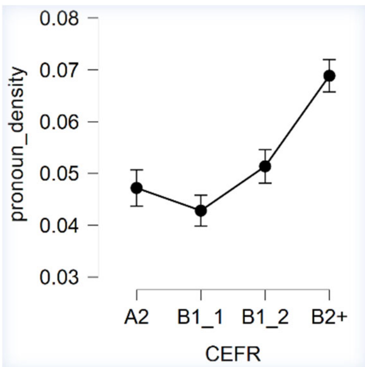

Figure 7. Comparisons of Pronoun Density between CEFR Levels
(pronoun_density is an index name for pronoun density; error bars indicate standard error)

Table 17. Post hoc Comparisons of CEFR Level and Pronoun to Noun Ratio

Comparison	Mean diff.	*SE*	*t*	P_{tukey}
A2 vs. B1_1	0.012	0.022	0.562	0.943
A2 vs. B1_2	-0.014	0.020	-0.697	0.898
A2 vs. B2+	-0.097	0.021	-4.688	<0.001***
B1_1 vs. B1_2	-0.026	0.021	-1.240	0.602
B1_1 vs. B2+	-0.109	0.022	-5.003	<0.001***
B1_2 vs. B2+	-0.083	0.020	-4.173	<0.001***

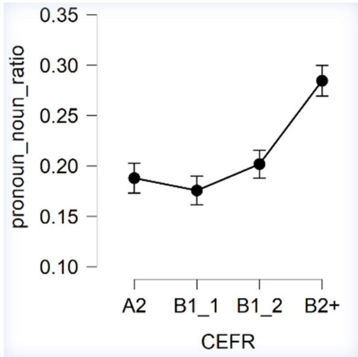

Figure 8. Comparisons of Pronoun to Noun Ratio between CEFR Levels
(pronoun_noun_ratio is an index name for pronoun to noun ratio; error bars indicate standard error)

3.5. Combined Model

The four indices entered into the previous stepwise multiple linear regression models by fulfilling the assumption of normality and the minimum correlation thresholds were taken into account, i.e., (i) binary adjacent sentence overlap content lemmas, (ii) adjacent two sentence overlap content lemmas, (iii) adjacent two sentence overlap noun lemmas, and (iv) repeated content lemmas and pronouns. Because all indices were normally distributed and were not collinear, they were all entered into a stepwise linear regression. The resulting model on the basis of a single index explained 3.9% of the variance (r = 0.198, R^2 = 0.039) in CEFR proficiency levels (see Table 18). The model suggested that an index associated to lexical overlap was a strong predictive of CEFR level and the component of lexical overlap includes a single index, i.e., binary adjacent sentence overlap content lemmas. The results of the combined model indicate that lexical overlap is an important quality observed in proficient L2 writing among different variables of cohesion such as content words including nouns, verbs, adjectives, and adverbs. In other words, more native-like L2 writers tended to compose more cohesive texts with more sentences that contain content lemma overlap with next sentence. Essays composed by advanced writers were specifically more likely to contain more content words that are repeated in the following sentence as compared with those written by beginners.

Table 18. Summary of combined regression model

Entry	Predictor included	R	R^2	R^2 change	β	SE	B
1	binary adjacent sentence overlap content lemmas	0.198	0.039	0.036	1.070	0.307	0.198

Note. The following predictors were considered but not included in the combined model: adjacent two sentence overlap content lemmas, adjacent two sentence overlap noun lemmas, and repeated content lemmas and pronouns

4. Conclusion

The current study investigated the predictive effectiveness of four different elements of cohesive variables including lexical overlap, connectives, semantic overlap and givenness. The component of lexical overlap consists of 24 different indices that calculate nouns, verbs, adjectives, and adverbs. The component of connectives includes 25 variables that evaluate the use of lists of connectives such as conjunctions, disjunctions, lexical subordinators, and so on. The component of semantic overlap comprises eight indices that measure the overlap of synonyms across sentences, and that of givenness includes four variables that consider ratio of pronouns to nouns, incidence of demonstratives, and definite articles. The present statistical results suggested that the fine-grain index of lexical overlap (e.g., binary adjacent sentence overlap content lemmas) was more effective predictors of EFL writing performance than the other indices of cohesion features. The combined investigation also illuminated that the most powerful models will possibly contain variables connected to variables of lexical overlap (binary adjacent sentence overlap content lemmas). The current findings expand previous corpus-based outcomes regarding the assessment of EFL writing quality, cohesive features in particular. Such findings also have an important pedagogical implication; knowing which cohesion variable is a stronger predictor of EFL writing could help better inform teachers about the possible trajectories of their students and potentially allow them to better pinpoint instruction and intervention to target specific areas of cohesion features. Results also have implications for vocabulary teaching and the teaching of writing; in an EFL context, a great deal of importance is placed on the teaching and learning of vocabulary as the belief is that a better vocabulary will contribute to better language proficiency. While this is true, it is not so easy to teach students how to translate this into actual language use in production such as speaking and writing. We would therefore need to show students how their receptive vocabulary can translate into a productive repertoire thereby improving the quality of their writing and speech. This study will hopefully bring about the expansion of new studies that can examine the role of cohesion analyses in accounting for EFL writing proficiency. Follow-up research will conduct a computational text analysis using a digital library such as Project Gutenberg (n.d.) to discuss whether linguistic features of text vary across genres and whether writers have a style that relates to certain dimensions of language.

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716-723.

Bestgen, Y. and N. Vincze. 2012. Checking and boostrapping lexical norms by means of word similarity indexes. *Behavior Research Methods* 44, 998-1006.

Bird, S., K. Klein and E. Loper. 2009. *Natural Language Processing with Python*. Beijing, China: O'Reilly.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Connor, U. 1990. Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English* 24, 67-87.

Cree, G. S. and B. C. Armstrong. 2012. Computational models of semantic memory. In M. Spivey, K. McRae and M. Joanisse, eds., *The Cambridge Handbook of Psycholinguistics*, 259-282. New York, NY: Cambridge University Press.

Crossley, S. A., K. Kyle and M. Dascalu. 2019. The tool for the automatic analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods* 51, 14-27.

Crossley, S. A., K. Kyle and D. S. McNamara. 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(2), 1-16.

Crossley, S. A. and D. S. McNamara. 2012. Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading* 35, 115-135.

Ferris, D. 1994. Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly* 28, 414-420.

Foltz, P. W. 2007. Discourse coherence and LSA. In T. K. Landauer, D. S. McNamara, S. Dennis and W. Kintsch, eds., *Handbook of Latent Semantic Analysis*, 167-184. Mahwah, NJ: Erlbaum.

Frase, L., J. Faletti, A. Ginther and L. Grant. 1997. *Computer Analysis of the TOEFL Test of Written English (TOEFL Research Rep. No. 64)*. Princeton, NJ: Educational Testing Service.

Goss-Sampson, M. A. 2020. Statistical Analysis in JASP 0.14. https://jasp-stats.org/wp-content/uploads/2020/11/Statistical-Analysis-in-JASP-A-Students-Guide-v14-Nov2020.pdf

Grant, L. and A. Ginther. 2000. Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing* 9, 123-145.

Graesser, A. C., D. S. McNamara, M. M. Louwerse and Z. Cai. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers* 36, 193-202.

Guo, L., S. Crossley and D. S. McNamara. 2013. Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing* 18, 218-238.

Halliday, M. A. K. and R. Hasan. 1976. *Cohesion in English*. London, UK: Longman.

Halliday, M. A. K. and C. M. I. Matthiessen. 2004. *An Introduction to Functional Grammar*. London: Routledge.

Ishikawa, S. 2013. The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa, ed., *Learner Corpus Studies in Asia and the World 1*, 91-118. Kobe, Japan: Kobe University.

Jafarpur, A. 1991. Cohesiveness as a basis for evaluating compositions. *Systems* 19, 459-465.

JASP Team. 2022. JASP (Version 0.16.3), Computer software.

Jin, W. 2001. A quantitative study of cohesion in Chinese graduate students' writing: Variations across genres and proficiency levels. (ERIC document reproduction service no. ED 452 726).

Landauer, T. K., D. Laham and P. W. Foltz. 2000. The intelligent essay assessor. *IEEE Intelligent Systems* 15, 27-31.

Lee, Y.-J. 2021. The relationship between text cohesion features and English proficiency for Korean college students. *Korean Journal of English Language and Linguistics* 21, 435-449.

Liu, M. and G. Braine. 2005. Cohesive features in argumentative writing produced by Chinese undergraduates. *System* 33, 623-636.

McNamara, D. S. and W. Kintsch. 1996. Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes* 22, 247-288.

McNamara, D. S., E. Kintsch, N. B. Songer and W. Kintsch. 1996. Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction* 14, 1-43.

McNamara, D. S. and A. Graesser. 2012. Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy and C. Boonthum, eds., *Applied Natural Language Processing and Content Analysis: Identification, Investigation, and Resolution*, 188-205. Hershey, PA: IGI Global.

Meara, P. and J. Milton. 2003. *X_Lex, the Swansea Levels Test*. Newbury, UK: Express Publishing.

Mikolov, T., I. Sutskever, K. Chen, G. Corrado and J. Dean. 2013. *Distributed Representations of Words and Phrases and Their Compositionality*. arXiv: 1310.4546 [cs.CL]

Miller, G. A. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38, 39-41.

Milton, J. 2010. The development of vocabulary breadth across the CEFR levels. In I. Bartning, M. Martin and J. Vedder, eds., *Second Language Acquisition and Testing in Europe*, 211-232. Online: Eurosla.

Nation, P. and D. Beglar. 2007. A vocabulary size test. *The Language Teacher* 31(7), 9-13.

Papageorgiou, S., J. T. Richard, B. Bridgeman and Y. Cho. 2015. *The Association between TOEFL iBT Test Scores and the Common European Framework of Reference (CEFR) Levels* (ETS Research Memorandum RM-

15-06). Princeton, NJ: Educational Testing Service.

Project Gutenberg. (n.d.). Retrieved February 21, 2016, from www.gutenberg.org.

Reid, J. 1990. Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll, ed., *Second Language Writing: Research Insights for the Classroom*, 191-210. Cambridge: Cambridge University Press.

Reppen, R. 1994. *Variation in Elementary Student Language: A Multi-dimensional Pperspective*. Unpublished doctoral dissertation, Northern Arizona University, Flagstaff.

Shapiro, S. S. and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52(3-4), 591-611.

Silva, T. 1993. Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly* 27(4), 657-675.

Tabachnick, B. G. and L. S. Fidell. 2014. *Using Multivariate Statistics*. Harlow, UK: Pearson Education.

Whitten, I. A. and E. Frank. 2005. *Data Mining*. San Francisco, CA: Elsevier.

Witte, S. P. and L. Faigley. 1981. Coherence, cohesion and writing quality. *College Composition and Communication* 22, 189-204.

Yang, W. and Y. Sun. 2012. The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education* 23, 31-48.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Appendix

A. Correlations between CEFR Level and L*exical Overlap* Variables

Variable	Mean (*SD*)	Correlation	*p*
Adjacent sentence overlap all lemmas	0.217 (0.058)	0.121	0.036*
Binary adjacent sentence overlap all lemmas	0.878 (0.127)	0.214	<0.001***
Adjacent two-sentence overlap all lemmas	0.328 (0.070)	0.125	0.031*
Binary adjacent two-sentence overlap all lemmas	0.953 (0.100)	0.113	0.051
Adjacent sentence overlap content lemmas	0.161 (0.069)	0.127	0.028*
Binary adjacent sentence overlap content lemmas	0.618 (0.208)	0.198	<0.001***
Adjacent two-sentence overlap content lemmas	0.247 (0.085)	0.121	0.036*
Binary adjacent two-sentence overlap content lemmas	0.770 (0.184)	0.209	<0.001***
Adjacent sentence overlap noun lemmas	0.204 (0.104)	0.136	0.019*
Binary adjacent sentence overlap noun lemmas	0.475 (0.229)	0.193	<0.001***
Adjacent two-sentence overlap noun lemmas	0.299 (0.125)	0.123	0.033*
Binary adjacent two-sentence overlap noun lemmas	0.615 (0.239)	0.191	<0.001***
Adjacent sentence overlap verb lemmas	0.118 (0.074)	0.091	0.114
Binary adjacent sentence overlap verb lemmas	0.245 (0.190)	0.095	0.100
Adjacent two-sentence overlap verb lemmas	0.197 (0.095)	0.109	0.059
Binary adjacent two-sentence overlap verb lemmas	0.371 (0.220)	0.142	0.014*
Adjacent sentence overlap adjective lemmas	0.101 (0.094)	0.078	0.180
Binary adjacent sentence overlap adjective lemmas	0.158 (0.185)	0.068	0.243
Adjacent two-sentence overlap adjective lemmas	0.162 (0.129)	0.064	0.270
Binary adjacent two-sentence overlap adjective lemmas	0.231 (0.230)	0.047	0.413
Adjacent sentence overlap adverb lemmas	0.068 (0.084)	-0.016	0.789
Binary adjacent sentence overlap adverb lemmas	0.078 (0.120)	-0.015	0.800
Adjacent two-sentence overlap adverb lemmas	0.120 (0.124)	-0.045	0.432
Binary adjacent two-sentence overlap adverb lemmas	0.129 (0.171)	-0.008	0.889

B. Descriptive Statistics and ANOVA Results for *Lexical Overlap* Indices by Proficiency Level

Index	Proficiency level, mean(*SD*)					
	A2	B1_1	B1_2	B2+	F(3, 296)	*p*
Adjacent sentence overlap all lemmas	0.211 (0.050)	0.210 (0.057)	0.214 (0.068)	0.231 (0.052)	2.202	0.088
Binary adjacent sentence overlap lemmas	0.853 (0.130)	0.869 (0.163)	0.853 (0.114)	0.941 (0.078)	9.097	<0.001***
Adjacent two-sentence overlap all lemmas	0.324 (0.069)	0.315 (0.077)	0.321 (0.070)	0.349 (0.062)	3.376	0.019*
Binary adjacent two-sentence overlap all lemmas	0.948 (0.065)	0.940 (0.144)	0.937 (0.120)	0.984 (0.034)	3.669	0.013*
Adjacent sentence overlap content lemmas	0.154 (0.062)	0.150 (0.061)	0.161 (0.075)	0.177 (0.074)	2.188	0.090

Binary adjacent sentence overlap content lemmas	0.579 (0.202)	0.590 (0.204)	0.599 (0.210)	0.700 (0.194)	5.725	<0.001***
Adjacent two-sentence overlap content lemmas	0.242 (0.089)	0.230 (0.084)	0.245 (0.080)	0.268 (0.085)	2.553	0.056
Binary adjacent two- sentence overlap content lemmas	0.741 (0.182)	0.742 (0.186	0.734 (0.195)	0.862 (0.139)	9.168	<0.001***
Adjacent sentence overlap noun lemmas	0.196 (0.102)	0.183 (0.093)	0.200 (0.109)	0.234 (0.105)	3.095	0.027*
Binary adjacent sentence overlap noun lemmas	0.437 (0.222)	0.436 (0.221)	0.455 (0.240)	0.565 (0.207)	5.657	<0.001***
Adjacent two-sentence overlap noun lemmas	0.295 (0.141)	0.268 (0.121)	0.294 (0.123)	0.334 (0.106)	3.389	0.018*
Binary adjacent two-sentence overlap noun lemmas	0.587 (0.244)	0.565 (0.250)	0.579 (0.246)	0.722 (0.179)	7.483	<0.001***
Adjacent sentence overlap verb lemmas	0.113 (0.066)	0.106 (0.066)	0.123 (0.082)	0.128 (0.078)	1.237	0.296
Binary adjacent sentence overlap verb lemmas	0.230 (0.181)	0.231 (0.209)	0.231 (0.195)	0.286 (0.175)	1.595	0.191
Adjacent two-sentence overlap verb lemmas	0.193 (0.097)	0.177 (0.095)	0.196 (0.088)	0.219 (0.098)	2.289	0.079
Binary adjacent two-sentence overlap verb lemmas	0.346 (0.213)	0.349 (0.247)	0.346 (0.207)	0.441 (0.206)	3.596	0.014*
Adjacent sentence overlap adjective lemmas	0.093 (0.097)	0.099 (0.090)	0.097 (0.088)	0.115 (0.101)	0.861	0.462
Binary adjacent sentence overlap adjective lemmas	0.139 (0.191)	0.161 (0.172)	0.153 (0.198)	0.179 (0.175)	0.607	0.611
Adjacent two-sentence overlap adjective lemmas	0.146 (0.126)	0.171 (0.119)	0.159 (0.134)	0.174 (0.134)	0.706	0.549
Binary adjacent two-sentence overlap adjective lemmas	0.209 (0.237)	0.255 (0.243)	0.211 (0.228)	0.255 (0.212)	0.958	0.413
Adjacent sentence overlap adverb lemmas	0.067 (0.087)	0.072 (0.090)	0.071 (0.091)	0.063 (0.068)	0.186	0.906
Binary adjacent sentence overlap adverb lemmas	0.076 (0.123)	0.089 (0.139)	0.075 (0.130)	0.075 (0.087)	0.223	0.880
Adjacent two-sentence overlap adverb lemmas	0.128 (0.141)	0.123 (0.128)	0.117 (0.113)	0.113 (0.117)	0.206	0.892
Binary adjacent two-sentence overlap adverb lemmas	0.121 (0.160)	0.156 (0.240)	0.115 (0.146)	0.130 (0.142)	0.759	0.518

C. Correlations between CEFR Level and *Connectives* Variables

Variable	Mean (*SD*)	Correlation	*p*
Basic connectives	0.038 (0.015)	-0.150	0.009**
Conjunctions	0.027 (0.013)	-0.152	0.008**
Disjunctions	0.004 (0.005)	0.120	0.038*
Lexical subordinators	0.023 (0.010)	0.047	0.417
Coordinating conjuncts	0.010 (0.007)	-0.016	0.785
Addition	0.028 (0.012)	-0.081	0.161
Sentence linking	0.035 (0.014)	-0.070	0.230

Order	0.007 (0.006)	0.178	0.002**
Reason and purpose	0.015 (0.009)	0.019	0.740
All causal connectives	0.020 (0.010)	-0.062	0.287
Positive causal connectives	0.026 (0.011)	-0.020	0.729
Opposition	0.009 (0.007)	-0.046	0.424
Determiners	0.078 (0.025)	-0.033	0.571
Demonstratives	0.018 (0.010)	0.013	0.819
Attended demonstratives	0.006 (0.006)	0.028	0.628
Unattended demonstratives	0.012 (0.008)	-0.002	0.971
All additive connectives	0.046 (0.014)	-0.002	0.975
All logical connectives	0.054 (0.017)	0.085	0.144
Positive logical connectives	0.026 (0.011)	-0.066	0.257
Negative logical connectives	0.010 (0.007)	-0.048	0.406
Temporal connectives	0.012 (0.008)	0.111	0.054
Positive intentional connectives	0.010 (0.008)	0.052	0.366
All positive connectives	0.075 (0.020)	-0.058	0.320
All negative connectives	0.014 (0.009)	0.029	0.613
All connectives	0.073 (0.019)	0.007	0.904

D. Descriptive Statistics and ANOVA Results for *Connectives* Indices by Proficiency Level

Index	Proficiency level, mean(*SD*) A2	B1_1	B1_2	B2+	F(3, 296)	*p*
Basic connectives	0.039 (0.016)	0.042 (0.014)	0.038 (0.015)	0.033 (0.013)	4.303	0.005**
Conjunctions	0.028 (0.014)	0.031 (0.012)	0.027 (0.013)	0.023 (0.010)	4.489	0.004**
Disjunctions	0.003 (0.005)	0.004 (0.005)	0.003 (0.005)	0.005 (0.006)	2.169	0.092
Lexical subordinators	0.021 (0.010)	0.024 (0.011)	0.023 (0.010)	0.023 (0.009)	1.017	0.385
Coordinating conjuncts	0.010 (0.008)	0.010 (0.008)	0.010 (0.007)	0.010 (0.007)	0.052	0.984
Addition	0.028 (0.013)	0.030 (0.012)	0.028 (0.013)	0.026 (0.010)	1.128	0.338
Sentence linking	0.034 (0.014)	0.040 (0.015)	0.036 (0.015)	0.032 (0.011)	3.393	0.018*
Order	0.006 (0.007)	0.004 (0.005)	0.007 (0.007)	0.009 (0.005)	8.019	<0.001***
Reason and purpose	0.014 (0.010)	0.017 (0.010)	0.015 (0.009)	0.015 (0.007)	1.127	0.338
All causal connectives	0.020 (0.011)	0.022 (0.011)	0.020 (0.010)	0.019 (0.010)	1.571	0.196
Positive causal connectives	0.025 (0.012)	0.028 (0.012)	0.026 (0.011)	0.025 (0.010)	1.011	0.388
Opposition	0.009 (0.008)	0.010 (0.008)	0.010 (0.007)	0.009 (0.007)	0.790	0.501
Determiners	0.078 (0.024)	0.081 (0.025)	0.076 (0.025)	0.077 (0.026)	0.442	0.723
Demonstratives	0.019 (0.011)	0.018 (0.009)	0.018 (0.009)	0.019 (0.010)	0.213	0.887
Attended demonstratives	0.006 (0.007)	0.006 (0.006)	0.006 (0.005)	0.007 (0.006)	0.258	0.856

Unattended demonstratives	0.013 (0.007)	0.012 (0.008)	0.012 (0.008)	0.013 (0.009)	0.086	0.968
All additive connectives	0.046 (0.015)	0.046 (0.014)	0.046 (0.015)	0.046 (0.012)	0.005	1.000
All logical connectives	0.052 (0.017)	0.052 (0.017)	0.054 (0.017)	0.056 (0.017)	0.824	0.482
Positive logical connectives	0.025 (0.012)	0.027 (0.012)	0.027 (0.011)	0.023 (0.009)	2.281	0.079
Negative logical connectives	0.010 (0.008)	0.010 (0.007)	0.010 (0.007)	0.009 (0.007)	0.439	0.725
Temporal connectives	0.011 (0.008)	0.012 (0.008)	0.013 (0.009)	0.014 (0.008)	1.267	0.286
Positive intentional connectives	0.009 (0.008)	0.011 (0.009)	0.011 (0.008)	0.010 (0.007)	1.484	0.219
All positive connectives	0.073 (0.021)	0.082 (0.019)	0.076 (0.020)	0.071 (0.017)	3.545	0.015*
All negative connectives	0.013 (0.009)	0.014 (0.009)	0.013 (0.008)	0.014 (0.009)	0.443	0.723
All connectives	0.071 (0.019)	0.078 (0.020)	0.073 (0.020)	0.073 (0.016)	1.523	0.209

E. Correlations between CEFR Level and *Semantic Overlap* Variables

Variable	Mean (*SD*)	Correlation	*p*
Synonym overlap (sentence, noun)	1.056 (1.456)	0.042	0.467
Synonym overlap (sentence, verb)	0.790 (1.321)	0.041	0.480
LSA cosine similarity (adjacent sentences)	0.414 (0.123)	0.176	0.002**
LSA cosine similarity (two adjacent sentences)	0.742 (0.095)	0.217	<0.001***
LDA divergence (adjacent sentences)	0.977 (0.059)	0.097	0.093
LDA divergence (two adjacent sentences)	0.983 (0.981)	0.029	0.619
Word2vec similarity (adjacent sentences)	0.874 (0.066)	0.217	<0.001***
Word2vec similarity (two adjacent sentences)	0.901 (0.079)	0.125	0.030*

F. Descriptive Statistics and ANOVA Results for *Semantic Overlap* Indices by Proficiency Level

	Proficiency level, mean(*SD*)					
Index	A2	B1_1	B1_2	B2+	F(3, 296)	*p*
Synonym overlap (sentence, noun)	0.046 (0.961)	0.871 (1.522)	0.839 (1.846)	0.809 (0.458)	0.418	0.740
Synonym overlap (sentence, verb)	0.646 (0.961)	0.871 (1.522)	0.839 (1.846)	0.809 (0.458)	0.418	0.740
LSA cosine similarity (adjacent sentences)	0.396 (0.119)	0.393 (0.122)	0.405 (0.126)	0.458 (0.115)	4.614	0.004**
LSA cosine similarity (two adjacent sentences)	0.728 (0.078)	0.714 (0.116)	0.736 (0.106)	0.784 (0.061)	7.873	<0.001***
LDA divergence (adjacent sentences)	0.974 (0.021)	0.962 (0.126)	0.981 (0.016)	0.986 (0.010)	1.973	0.118
LDA divergence (two adjacent sentences)	0.987 (0.014)	0.973 (0.127)	0.978 (0.106)	0.993 (0.005)	0.877	0.453
Word2vec similarity (adjacent sentences)	0.864 (0.046)	0.851 (0.120)	0.875 (0.043)	0.900 (0.021)	7.364	<0.001***
Word2vec similarity (two adjacent sentences)	0.897 (0.030)	0.883 (0.119)	0.897 (0.100)	0.924 (0.017)	3.340	0.020*

G. Correlations between CEFR Level and *Givenness* Variables

Variable	Mean (*SD*)	Correlation	*p*
Pronoun density	0.053 (0.030)	0.270	<0.001***
Pronoun to noun ratio	0.214 (0.133)	0.261	<0.001***
Repeated content lemmas	0.337 (0.062)	0.024	0.678
Repeated content lemmas and pronouns	0.385 (0.071)	0.140	0.015

H. Descriptive Statistics and ANOVA Results for *Givenness* Indices by Proficiency Level

	Proficiency level, mean (*SD*)					
Index	A2	B1_1	B1_2	B2+	F(3, 296)	*p*
Pronoun density	0.047 (0.030)	0.043 (0.023)	0.051 (0.030)	0.069 (0.027)	11.741	<0.001***
Pronoun to noun ratio	0.188 (0.128)	0.176 (0.111)	0.202 (0.130)	0.285 (0.133)	11.060	<0.001***
Repeated content lemmas	0.337 (0.072)	0.326 (0.061)	0.346 (0.053)	0.335 (0.061)	1.286	0.279
Repeated content lemmas and pronouns	0.379 (0.086)	0.364 (0.066)	0.392 (0.061)	0.399 (0.064)	3.486	0.016*