Tabaran Institute of Higher Education
International Journal of Language Testing
Vol. 15, No. 2; 2025, October

ISSN 2476-5880
https://www.ijlt.ir

Examining the Role of Lexical Sophistication, Lexical Diversity, Syntactic Sophistication, Syntactic Complexity, and Cohesion in L2 Speaking Proficiency Assessment

Mahtab Kolahi Ahari[1], Behzad Ghonsooly[2*], Zargham Ghapanchi[3], Hassan Soodmand Afshar[4]

ARTICLE INFO

Article History:
Received: December 2024
Accepted: January 2025

KEYWORDS
Cohesion
L2 speech assessment
Lexis
Syntax

ABSTRACT

The present study developed a model of L2 speaking proficiency investigating how lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion are associated with holistic scores of L2 speaking proficiency employing structural equation modeling (SEM). A corpus of 419 monologues delivered by Iranian EFL learners was compiled and rated to develop the model. Based on the overall scores, the corpus was divided into independent (B1 and B2) and proficient (C1 and C2) users. The results of SEM analysis revealed that the developed L2 speaking proficiency model had an acceptable fit, with partial generalizability across independent and proficient users. Structural regression analysis showed that lexical diversity, lexical sophistication, syntactic sophistication, cohesion, and the indirect effect of syntactic complexity through lexical sophistication explained 34% of the variance in L2 speaking proficiency in descending order of importance. However, their relative importance changed depending on proficiency level. Based on the results, while lexical, syntactic, and cohesive features are sound predictors of L2 speaking proficiency, they function differently across proficiency groups. These findings offer valuable insights for improving speaking proficiency assessment by showing that lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion do not contribute equally to overall L2 speaking proficiency, and their order of importance varies across proficiency levels. Therefore, prioritizing indicators of L2 speaking proficiency in assessment frameworks based on their importance in each proficiency level can add to the validity and reliability of speaking assessments.

1. Introduction

Based on communicative competence models, second language (L2) speaking proficiency can refer to learners' ability to use appropriate linguistic and discoursal features in various communicative

[1] Department of English, Faculty of Letters and Humanities, Ferdowsi University of Mashhad, Mashhad, Iran, Email: mahtabkolahi@gmail.com
[2] Department of English, Faculty of Letters and Humanities, Ferdowsi University of Mashhad, Mashhad, Iran, Email: ghonsooly@um.ac.ir (Corresponding author)
[3] Department of English, Faculty of Letters and Humanities, Ferdowsi University of Mashhad, Mashhad, Iran, Email: ghabanchi@um.ac.ir
[4] Department of English Language, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran, Email: h.soodmand@atu.ac.ir

Cite this paper as: Kolahi Ahari, M., Ghonsooly, B., Ghapanchi, Z., & Soodmand Afshar, H. (2025). Examining the role of lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion in L2 speaking proficiency assessment. *International Journal of Language Testing, 15*(2), 43–70. https://doi.org/10.22034/ijlt.2025.492133.1395

settings (Bachman, 1990; Celce-Murcia et al., 1995; Taş & Khan, 2020). To operationalize the definition of L2 speaking proficiency for assessment purposes and to facilitate the evaluation of these linguistic and discoursal features, different rubrics have been developed (Crossley & McNamara, 2013), most of which emphasize vocabulary, grammar, and cohesion as key components of proficient L2 speech.

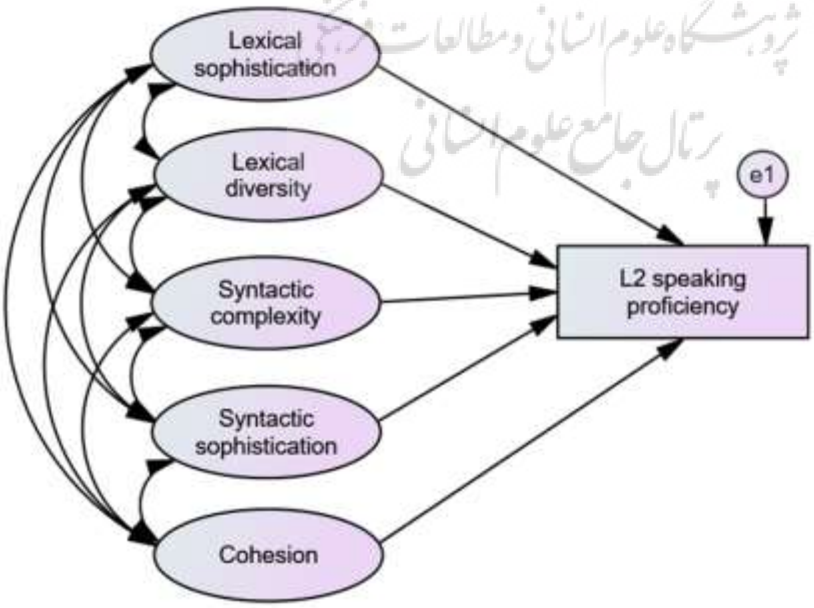
Vocabulary is typically conceptualized in terms of diversity/range and sophistication (Laufer & Nation, 1995; Read, 2000), both significantly predicting L2 speaking proficiency (Bulté & Roothooft, 2020; Eguchi, 2022; Eguchi & Kyle, 2020; Enayat & Derakhshan, 2021). Grammar knowledge is similarly described in terms of the complexity and sophistication of structures used in learner discourse (Biber et al., 2016). Some studies on syntactic complexity and L2 speaking proficiency have illustrated a positive relationship between the two (Hwang & Kim, 2024; Park, 2022; Yazdani, 2018), while others have suggested that syntactic complexity can negatively affect L2 speech (Sadri Mirdamadi & De Jong, 2015). Syntactic sophistication, on the other hand, has been mostly investigated in writing (Kyle & Crossley, 2017; Kyle et al., 2021), revealing itself as a stronger predictor of writing quality than syntactic complexity (Kyle & Crossley, 2017). Research has also shown that cohesion, which binds different parts of texts (Halliday & Hasan, 1976), has associations with speaking proficiency (Crossley & McNamara, 2013).

Despite the importance of lexical, syntactic, and cohesive features, few studies investigated their collective role in L2 speech (Crossley & McNamara, 2013). To our knowledge, no prior study developed a holistic structural model to test the collective impact of these features on L2 speaking proficiency and clarify their role across proficiency levels. This study aims to address the existing gaps by developing an L2 speaking proficiency model (the hypothesized model is shown in Figure 1) examining the relationships between lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, as well as cohesion, and L2 speaking proficiency scores using the SEM approach. It will then explore the generalizability of the developed model across two proficiency groups: independent (B1 and B2) and proficient users (C1 and C2). The present study sought to answer the following two research questions:

1. How do lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion collectively predict L2 speaking proficiency scores?

2. Does the developed model enjoy generalizability across proficiency groups (independent and proficient users)?

Figure 1
The Hypothesized L2 Speaking Proficiency Model



2. Review of Literature

2.1. L2 Speaking Proficiency

The definition of speaking proficiency and its features have always been debated, leading to the development of different scoring rubrics for easier operationalization and assessment of the construct (Crossley & McNamara, 2013). Most rubrics emphasize the centrality of language knowledge in L2 speaking proficiency. According to Bachman and Palmer (1996), grammatical (vocabulary, grammar, and phonology) and textual (cohesion and rhetorical organization) knowledge are two essential components of language knowledge. Research has shown that more grammatical knowledge leads to grammatical accuracy and lexical richness, enhancing the overall judgments of speaking proficiency (De Jong et al., 2012), while advanced textual knowledge leads to more cohesive and logically organized discourse, making speech easier to follow and improving proficiency.

Grammatical and textual knowledge can be measured in L2 speech through lexis, syntax, and cohesion features. Although these features were previously analyzed manually, they are now automatically computed with great precision by natural language processing (NLP) tools such as the Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Kyle & Crossley, 2015; Kyle et al., 2018), the Tool for the Automatic Analysis of Lexical Diversity (TAALED; Kyle et al., 2020), the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016), and the Tool for the Automatic Analysis of Cohesion (TAACO; Crossley et al., 2016a). Recent studies have also utilized these NLP tools to investigate the role of lexis, syntax, and cohesion in L2 speaking proficiency (Crossley & McNamara, 2013; Eguchi & Kyle, 2020; Enayat & Derakhshan, 2021; Park, 2022). The following sections review key studies in this regard.

2.2. Lexical Sophistication

Lexical sophistication refers to the use of advanced and low-frequency lexical items (Crossley et al., 2016b; Laufer & Nation, 1995). Early measures include word frequency and range, n-gram frequency, range, and association strength, academic language, and psycholinguistic norms (Crossley et al., 2014; Crossley et al., 2015; Kyle & Crossley, 2015). Expanding on these, recent measures incorporate factors such as age of acquisition/exposure (AoA), contextual distinctiveness, word neighborhood information, word recognition norms, and semantic network (Eguchi & Kyle, 2020; Kim et al., 2018; Kyle & Crossley, 2015). Previous research has shown that L2 speakers who use more low-frequency words and n-grams, more academic vocabulary, and more unfamiliar and abstract words are judged to be more proficient (Eguchi, 2022; Kyle & Crossley, 2015). Age of acquisition, contextual distinctiveness, word neighborhood information, word recognition norms, and semantic network can also explain much of the variance in L2 speaking proficiency. Previous research has indicated that as proficiency level increases, learners more often use context-specific, phonologically and orthographically unique, and cognitively demanding words (Eguchi, 2022; Eguchi & Kyle, 2020; Enayat & Derakhshan, 2021).

2.3. Lexical Diversity

Lexical diversity refers to the variety of words, which reflects language users' vocabulary size (Kyle et al., 2020). It is traditionally measured using the type-token ratio (TTR; Lieven, 1978; Bates et al., 1991), which results from dividing the number of unique words (types) by the total number of words (tokens). While TTR and its derivatives (i.e., Root TTR and Log TTR) are highly affected by text length, other indices of lexical diversity are more robust and independent of text length, including the measure of lexical diversity adjusted for text length or the Maas's index (Maas, 1972; Zenker & Kyle, 2021), the mean-segmental type-token ratio (Covington & McFall, 2010; Johnson, 1944), the moving-average type-token ratio (Covington & McFall, 2010), the measure of textual lexical diversity (McCarthy & Jarvis, 2010), and the hypergeometric distribution diversity index (McCarthy & Jarvis, 2007).

Although most rubrics consider lexical diversity a central aspect of speaking proficiency, relatively few studies have explored its relationship with L2 speaking proficiency. These studies indicate that L2 speakers demonstrating a diverse and larger lexicon are judged to be more proficient (Bulté & Roothooft, 2020; Enayat & Derakhshan, 2021; Yu, 2010).

2.4. Syntactic Sophistication

Syntactic sophistication is rooted in usage-based language learning and is most often measured based on the frequency and association strength of verb argument constructions (VAC; Goldberg, 1995; Kyle & Crossley, 2017), with a VAC consisting of a recurring pair of a verb and a specific syntactic pattern. As an example, in the construction give [someone] [something], the verb give has two arguments: the recipient ([someone]) and the object ([something]). Frequency indices calculate the frequency of main verb lemmas (the base form of a verb, such as give), VACs (e.g., give [someone] [something]), and verb-VAC combinations (a specific verb occurring within a VAC, such as the verb give in the give [someone] [something] construction) in a reference corpus. Association strength indices, such as Faith and Delta P scores, measure how probable it is that a main verb lemma and a VAC co-occur within a reference corpus. To the best of our knowledge, syntactic sophistication was mostly investigated in learner writing samples (Kyle & Crossley, 2017; Kyle et al., 2021) and was found to predict writing quality more strongly than syntactic complexity (Kyle & Crossley, 2017). However, the association between syntactic sophistication and L2 speaking proficiency remains unclear.

2.5. Syntactic Complexity

Syntactic complexity refers to the level of subordination and sentence length in a text (Kyle & Crossley, 2017) conceptualized based on clause and noun phrase complexity (Lu, 2012). While clause complexity is determined by the number of subordinate, coordinate, and embedded clauses (Norris & Ortega, 2009), noun phrase complexity is based upon the number of modifiers, embedded clauses, and prepositional phrases attached to a noun. Most studies have measured syntactic complexity based on clause complexity indices, showing that more use of lengthier clauses and a larger number of subordinate clauses indicate more proficient L2 speech (Hwang & Kim, 2024; Park, 2022; Yazdani, 2018). In a very recent cross-proficiency analysis of syntactic complexity, Kim and Lu (2024) found that while this construct is positively associated with speaking proficiency, different measures of syntactic complexity (e.g., finite subordination, mean length of clause, and complex noun phrases) develop differently across proficiency levels. On the other hand, some studies have found that the use of syntactically complex sentences might have a negative impact on fluency due to the greater required cognitive load, especially in lower proficiency levels (Sadri Mirdamadi & De Jong, 2015).

2.6. Cohesion

Cohesion refers to the explicit connection within a text created by sentence connectors, paragraph connectors, and connectives or clausal relationships (Crossley et al., 2016a). Local cohesion, which is created through textual links between sentences (i.e., noun overlaps and connectives), helps communicate similar ideas across sentences and informs the reader/listener of the relations between them (Halliday & Hasan, 1976). Connectives measure local cohesion through various coordinators, subordinators, and conjunctions (Crossley et al., 2016a). Although relatively few studies have investigated cohesion in speaking proficiency (Crossley & McNamara, 2013), they have commonly confirmed that cohesion within learners' discourse and between the prompt and their response can be associated with learners' speaking performance. Although previous studies have shown that lexical, syntactic, and cohesive features are associated with L2 speaking proficiency, and assessment rubrics emphasize their importance to some extent, recent studies have shown that still, assessment frameworks like IELTS and TOEFL fail to authentically represent the reality of L2 speaking proficiency (Souzandehfar, 2024; Kalantar, 2024). Therefore, there is a need to clarify how much of L2 speaking proficiency can truly be explained by these lexical, syntactic, and cohesive features and enhance the precision of assessment frameworks.

3. Method

3.1. Corpus

This study utilized a learner corpus including 419 speaking samples (monologues) collected from Iranian EFL learners, designed in a format similar to the Part 2 of the IELTS speaking test (IDP IELTS, 2020). The researcher recorded learners' responses after obtaining their consent. Following the recording phase, monologues were evaluated by two raters, including the researcher, assigning each an overall score from 1.0 to 9.0 according to the IELTS speaking band descriptors. The ratings enjoyed a high degree of inter-rater reliability (*r* > .9). The Common European Framework of Reference (CEFR) was then used to categorize speaking samples based on proficiency level. A CEFR level equivalent to the overall score was assigned to each sample. A total of 180 samples scored from 4.0 to 6.5 were categorized as B1 (4.0, 4.5, 5.0) or B2 (5.5, 6.0, 6.5), and 239 scored from 7.0 to 9.0 were classified as C1 (7.0, 7.5, 8.0) or C2 (8.5, 9.0) levels. CEFR considers B1 and B2 learners independent users and C1 and C2 learners proficient users. Following these stages, monologues were transcribed for further analysis. Table 1 shows descriptive statistics related to speaking scores as well as word counts for the samples across proficiency groups.

Table 1
Mean Scores of Speaking Proficiency and Word Counts Across Proficiency Groups

	Independent users			Proficient users		
	B1	B2	Total	C1	C2	Total
Score	4.8 (0.3)	6.0 (0.3)	5.5 (0.7)	7.5 (0.4)	8.6 (0.2)	7.7 (0.5)
Number of words	157.7 (43.9)	193.7 (44.2)	178.7 (47.4)	228.6 (62.6)	248.8 (77.9)	232.0 (65.6)

3.2. Natural Language Processing Tools

The present study utilized four NLP tools for measuring the target variables in speaking outputs: TAALES (Kyle & Crossley, 2015; Kyle et al., 2018), TAALED (Kyle et al., 2020), TAASSC (Kyle, 2016), and TAACO (Crossley et al., 2016a). Each one measures its respective variable based on several indices. Our initial selection of indices was informed by previous research, but then we took an exploratory approach to narrow them down to meet practical considerations.

3.2.1. TAALES Indices. Among TAALES indices, previous studies found word frequency and range, n-gram frequency, range, and association strength, academic language, and psycholinguistic measures to be significant predictors of L2 speaking proficiency (Crossley et al., 2013; Eguchi, 2022; Kyle & Crossley, 2015). However, recent research has indicated that other indices such as contextual distinctiveness, word neighbor information, word recognition norms, AoA, and semantic network can also predict a considerable degree of variance in speaking proficiency scores (Eguchi, 2022; Eguchi & Kyle, 2020; Enayat & Derakhshan, 2021). Since these indices have not been used as measures of lexical sophistication in modeling L2 speaking proficiency, we decided to operationalize the construct accordingly (find validation in Eguchi & Kyle, 2020). That led us to 48 indices for further analysis (see Appendix A).

3.2.2. TAALED Indices. TAALED measures lexical diversity based on traditional and revised varieties of TTR indices. The validity and reliability of TAALED indices were confirmed by Kyle et al. (2021). We excluded indices affected by text length, like simple TTR (McCarthy & Jarvis, 2007), and focused on more robust indices, leading us to seven indices (*n* = 7; see Appendix A).

3.2.3. TAASSC Indices. Since most previous studies operationalized syntactic complexity only based on clause complexity, we measured it based on indices of both clause complexity and noun phrase complexity (*n* = 84; see Appendix A). Regarding syntactic sophistication, as this construct has received little attention from L2 speech research, it was measured based on all indices of syntactic sophistication (*n* = 161; see Appendix A). Their validity and reliability were confirmed by Kyle and Crossley (2017).

3.2.4. TAACO Indices. Cohesion was measured based on connective indices (*n* = 25; see Appendix A) in TAACO (confirmed to be reliable by Crossley et al., 2019) because they help listeners grasp the logical flow and line of reasoning used by the speaker by linking ideas and sentences and thus structuring speech (McCarthy & Carter, 2014). The use of connectives is also emphasized in speaking assessment rubrics and by previous research as indicative of L2 speaking proficiency.

3.3. Statistical Analysis

First, the selected indices were checked for their normal distribution using skewness and kurtosis levels. Skewness and kurtosis values falling between −2 and +2 were used as an indication of normal distribution (George & Mallery, 2010). Next, Pearson's correlation was run between normally distributed indices and overall speaking proficiency scores. Indices correlated with overall speaking scores were then controlled for multicollinearity (defined as $r > .90$). Finally, the remaining non-collinear indices were checked for potential conceptual overlaps.

To generate the hypothesized model for speaking proficiency (Figure 1), we performed confirmatory factor analysis (CFA) and SEM using AMOS software (version 24). CFA was run to validate the latent variables and their indicators. SEM was then used to examine how latent variables predict L2 speaking proficiency scores and assess the overall fit of the model based on four goodness-of-fit measures: CMIN/DF, CFI, GFI, and RMSEA. Indicators of good model fit included CMIN/DF values of less than 3.0; CFI, GFI, and TLI statistics of greater than .95; and RMSEA values of less than .06, while CMIN/DF values of less than 5.0; CFI, GFI, and TLI statistics of ≥ .90; and RMSEA of ≤ .08 indicated an acceptable model fit (Kline, 2011).

A series of measurement invariance tests were then conducted to check whether our proposed model operates similarly in independent (B1 and B2) and proficient (C1 and C2) groups. Comparative chi-square ($\Delta\chi2$) and comparative CFI (ΔCFI) were employed for significance testing of the results. According to Dimitrov (2010), measurement invariance holds in the case of a statistically insignificant $\Delta\chi2$ and a ΔCFI of > -0.01.

4. Results

Based on skewness and kurtosis values, of the 48 lexical sophistication indices considered, 44 were normally distributed. Of these, 35 were significantly correlated ($p < .05$) with the overall speaking proficiency scores, but only 8 indices passed the test of checking multicollinearity. Of the 7 lexical diversity indices considered, all met the assumption of normality and were significantly correlated ($p < .05$) with the overall speaking proficiency scores. Of these, 4 indices passed the test of checking multicollinearity. Of the 84 syntactic complexity indices considered, 19 were normally distributed, 17 of which were significantly correlated ($p < .05$) with the overall speaking proficiency scores and did not show any signs of multicollinearity. Of the 161 syntactic sophistication indices considered, 51 were normally distributed, 13 of which had a statistically significant correlation ($p < .05$) with the overall speaking proficiency scores and were not multicollinear. Of the 25 cohesion indices considered, 7 were normally distributed, 6 of which were significantly correlated ($p < .05$) with the overall speaking proficiency scores without any signs of multicollinearity.

CFA was then run to validate the latent variables and their indicators. Indicators with factor loadings lower than .6 were removed from the model, resulting in a final measurement model with 17 indices (see Appendix B). The final lexical sophistication indices included contextual distinctiveness, word recognition norms, and AoA. Contextual distinctiveness refers to how restricted a word is to its context, with more contextually distinctive words appearing in more distinct contexts (e.g., *sphygmomanometer*; Berger et al., 2017b). Age of acquisition is the estimated age of learning a word by an L1 speaker (Kyle et al., 2018), with some words having a late AoA (e.g., *photosynthesis*) and others an earlier one (e.g., *dog*). Word recognition norms refer to the cognitive effort needed to process a word, with some words being more cognitively demanding than others (e.g., *quintessential* vs. *sun*; Berger et al., 2017a). The final lexical diversity indices included Maas' index, textual lexical diversity, and distribution diversity index, which are more robust measurements of TTR resulting from mathematical transformations. Maas' index is a logarithmic transformation of TTR. Textual lexical diversity is the average number of words needed to reach a TTR of 0.720, which is the standard level (Kyle et al., 2020). Distribution diversity index clarifies the probability that a single word in a text might appear in any other part of that text (Kyle et al., 2020). The final syntactic complexity indices included dependents per direct object (i.e., the average number of adjectives, determiners, or prepositional phrases attached to direct objects), determiners per nominal (i.e., the average number of determiners in nominal phrases), and adjective modifiers per nominal (i.e., the average number of adjectives or adjective modifiers in nominal phrases). For example, in the sentence *She bought the two elegant silk dresses with intricate patterns yesterday*, there are five dependents per direct object (*the*, *two*, *elegant*, *silk*, *with intricate patterns*), two determiners per nominal (*the*, *two*), and two adjectives per nominal (*elegant*, *silk*). Syntactic sophistication indices included Faith and Delta P scores in academic, news, magazine, and fiction reference corpora. A high Faith score means that it is more probable for a specific verb to occur in a specific structure than other constructions based on a reference corpus, making it less syntactically sophisticated (Kyle et al., 2018). For example, *give* has a high Faith score as it frequently occurs in the *give [someone] [something]* construction. A high Delta P score would indicate that a specific verb is particularly predictive of a specific construction based on a reference corpus. A lower Delta P score, on the other hand, would mean that other verbs could just as likely appear in the same structure (Kyle et al., 2018). For instance, the construction *[verb + a break]* has a high Delta P score as the verb *take* predicts it. Cohesion indices included basic connectives, conjunctions, and positive connectives. Connectives link clauses, sentences, and ideas (e.g., *and*, *but*, *so*, *actually*, *after all*) within or across sentences, and conjunctions link words, phrases, and clauses within a sentence (e.g., *although*, *because*, *after*). The structural model was then evaluated to investigate how lexical sophistication, lexical diversity, syntactic complexity, syntactic sophistication, and cohesion predict L2 speaking proficiency. The results showed that the proposed model fits the data acceptably (Table 2). Figure 2 depicts the developed model for L2 speaking proficiency.

Table 2
Goodness of Fit Indices for the Developed Model

χ^2/df	Df	TLI	CFI	RMSEA	GFI
3.4	121	.92	.94	.07	.90

Descriptive statistics, correlations between the selected indices and speaking scores, and the composite reliability (CR) of each latent variable are shown in Table 3. As Table 3 illustrates, all the remaining indices show strong correlations with overall speaking proficiency scores. As for CR, all indices are within the acceptable range of .6 to .7 or > .7 (Hair et al., 2014). Correlation matrices among indicator variables are also shown in Table 4.

Figure 2
The Developed Model for L2 Speaking Proficiency

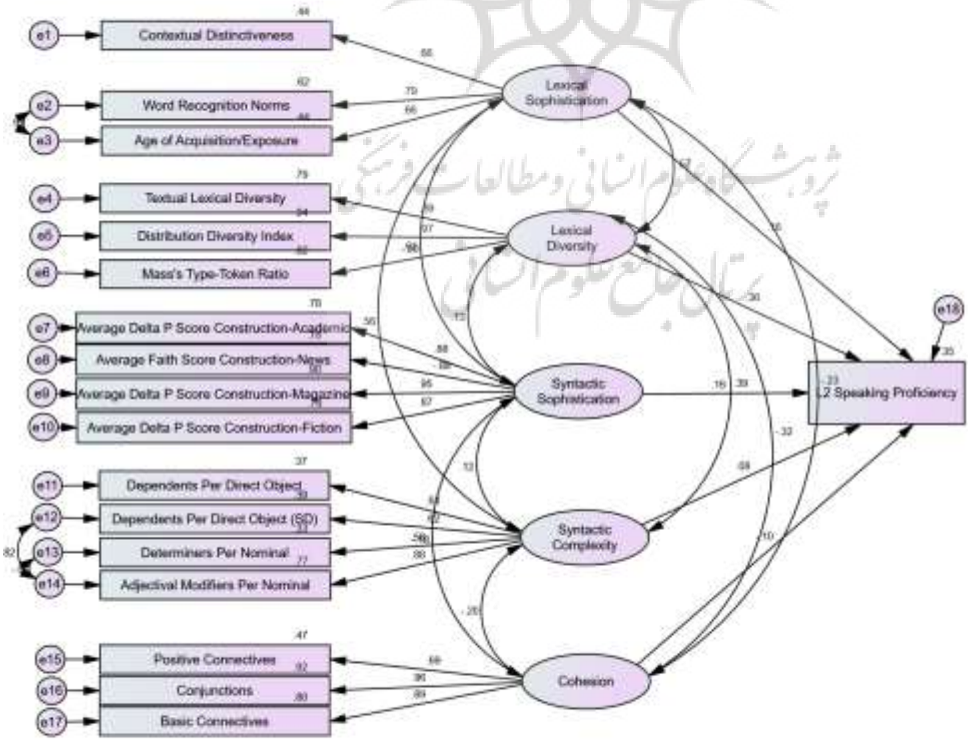


Table 3

Descriptive Statistics for All the Proposed Indices and Speaking Proficiency Scores

Variable		N	Min	Max	Mean	SD	Skewness	Kurtosis	r	CR
	OS	419.00	4.00	9.00	6.81	1.27	-0.36	-0.89		
LS	Contextual distinctiveness	419.00	32.95	44.18	38.49	1.80	0.16	0.53	.27**	.68
	Word recognition norms	419.00	587.38	619.67	601.76	5.43	0.26	0.26	.31**	
	Age of acquisition/exposure	419.00	0.38	4.76	1.66	0.65	0.59	0.88	.34**	
LD	Maas's type-token ratio	419.00	0.04	0.09	0.06	0.01	0.21	-0.03	-.40**	.60
	Distribution diversity index	419.00	0.59	0.85	0.74	0.05	-0.48	0.30	.54**	
	Textual lexical diversity	419.00	14.97	99.07	41.91	13.45	0.86	0.93	.45**	
SC	Dependents per direct object	419.00	0.00	3.00	1.19	0.53	0.00	0.46	.09*	.66
	Dependents per direct object (sd)	419.00	0.00	2.36	0.98	0.38	-0.53	1.38	.22**	
	Determiners per nominal	419.00	0.00	0.50	0.21	0.09	0.19	0.60	.29**	
	Adjectival modifiers per nominal	419.00	0.00	0.46	0.13	0.08	0.82	1.08	.26**	
SS	Average Delta P score construction-academic	419.00	0.00	0.28	0.06	0.05	1.48	1.83	.22**	.88
	Average Faith score construction-news	419.00	0.00	0.29	0.08	0.05	1.18	1.58	.20**	
	Average Delta P score construction-magazine	419.00	0.00	0.24	0.07	0.05	1.23	1.37	.21**	
	Average Delta P score construction-fiction	419.00	0.00	0.24	0.07	0.04	1.15	1.23	.18**	
C	Basic connectives	419.00	0.01	0.15	0.07	0.02	0.34	0.63	-.20**	.76
	Conjunctions	419.00	0.01	0.12	0.06	0.02	0.36	0.37	-.30**	

Positive Connectives	419.00	0.03	0.16	0.10	0.02	0.14	-0.01	-.14**
Valid N (listwise)	419.00							

Note. LS = lexical sophistication; LD = lexical diversity; SC = syntactic complexity; SS = syntactic sophistication; C = cohesion; CR = composite reliability.
**p* < .05. ***p* < .01.

Table 4
Correlations Among Indicators

Indicators	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Contextual distinctiveness																	
Word recognition norms	-.64**	1															
Age of acquisition/exposure	-.53**	.56**	1														
Maas's type-token ratio	.52**	-.44**	-.36**	1													
Distribution diversity index	-.58**	.43**	.40**	-.87**	1												
Textual lexical diversity	-.57**	.43**	.38**	-.82**	.85**	1											
Dependents per direct object	-.18**	.19**	.14**	-.11*	.13**	.15**	1										
Dependents per direct object (SD)	-.13**	.09*	.15**	-.10*	.16**	.13**	.54**	1									
Determiners per nominal	-.10*	.22**	.19**	-.10*	.13**	.13**	.41**	.27**	1								
Adjectival modifiers per nominal	-.52**	.45**	.40**	-.36**	.40**	.41**	.44**	.27**	.31**	1							
Average Delta P score construction-academic	-0.05	0.02	0.09	-.13**	.14**	.11*	0.09	.11*	.15**	0.04	1						
Average Faith score construction-news	0.01	-0.02	0.04	-.10*	.12*	0.06	.12*	.16**	.16**	0.03	.77**	1					
Average Delta P score construction-magazine	-0.02	-0.01	0.05	-.13**	.13**	.09*	0.08	.18**	.14**	0.01	.84**	.83**	1				
Average Delta P score construction-fiction	0.04	-0.08	-0.01	-0.06	0.06	0.02	0.06	.16**	.14**	-0.01	.77**	.77**	.82**	1			

Basic connectives	.12*	-.19**	-.15**	.17**	-.22**	-.18**	-0.09	-.20**	-.24**	-.11*	-.12*	-.17**	-.18**	-.14**	1		
Conjunctions	.11*	-.14**	-.15**	.24**	-.35**	-.26**	-0.06	-.19**	-.21**	-0.07	-.12**	-.16**	-.17**	-.12*	.85**	1	
Positive Connectives	0.07	-0.08	-.12*	.14**	-.21**	-.14**	-0.05	-.18**	-.22**	-0.07	-0.02	-0.06	-0.05	-0.03	.63**	.65**	1

Note. *$p < .05$. **$p < .01$.

Based on the regression weights, 35% of L2 speaking proficiency variance was explained by lexical diversity (β = .36, $p < .001$), lexical sophistication (β = .16, $p = .03$), syntactic sophistication (β = .16, $p < .001$), and cohesion (β = -.10, $p = .03$). However, syntactic complexity did not directly predict L2 speaking proficiency, possibly due to the stronger correlation it had with lexical sophistication (r = .56), indicating shared variance. Considering the important role of syntactic complexity in L2 speaking proficiency, as confirmed in previous research (Hwang & Kim, 2024; Park, 2022; Yazdani, 2018), and the strong correlation between syntactic complexity and lexical sophistication, a mediation analysis was conducted to explore if syntactic complexity indirectly affects L2 speaking proficiency through lexical sophistication. Based on Baron and Kenny (1986), three mediational assumptions must be met before running a mediation analysis: (a) the independent variable must predict the dependent variable; (b) the mediator must predict the dependent variable; and (c) the independent variable must predict the mediator. In this study, only assumptions (b) and (c) were met. Lexical sophistication predicted L2 speaking proficiency (β = .21, $p < .001$), and syntactic sophistication predicted lexical sophistication (β = .16, $p < .001$). Therefore, we could only test whether syntactic complexity indirectly affects L2 speaking proficiency through lexical sophistication.

The second model (Figure 3), which included a mediation analysis, also had an acceptable fit (Table 5), revealing that 34% of variance in L2 speaking proficiency was described by lexical diversity (β = .40, $p < .001$), lexical sophistication (β = .21, $p < .001$), syntactic sophistication (β = .16, $p < .001$), cohesion (β = -.10, $p = .02$), and the indirect effect of syntactic complexity on L2 speaking proficiency through lexical sophistication (β = .13, $p = .03$). The results of the mediation analysis showed that syntactic complexity indirectly affected L2 speaking proficiency through lexical sophistication. Based on these findings, complex syntactic structures might influence human judgments of L2 speaking proficiency when put into sophisticated lexical items.

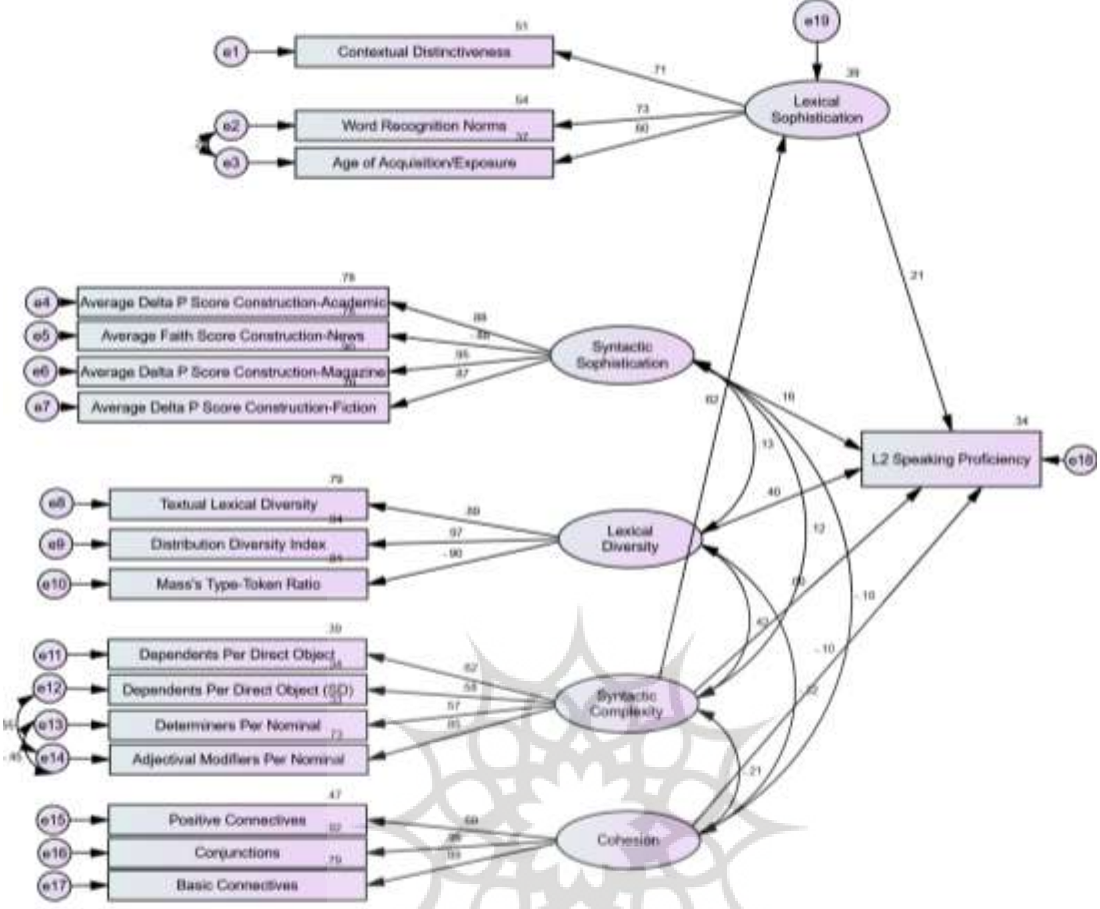
Table 5

Goodness of Fit Indices for the Developed Model with a Mediation Analysis

χ^2/df	Df	TLI	CFI	RMSEA	GFI
3.5	122	.90	.93	.07	.89

In comparison, both models almost fit the data similarly. However, the model with mediation analysis is preferable since it corroborates previous findings on the important role of syntactic complexity in L2 speaking proficiency (Hwang & Kim, 2024; Park, 2022; Yazdani, 2018).

Figure 3
The Developed Model for L2 Speaking Proficiency with a Mediation Analysis



The developed model with mediation analysis (Figure 3) was used for measurement invariance tests to check for the generalizability of the model across independent and proficient users. Table 6 shows the results of the invariance assessment. Configural invariance (Model 0) was supported, indicating the model had a similar overall structure across independent and proficient users. Metric invariance was not fully supported in Model 1. Since factor loadings on contextual distinctiveness (in lexical sophistication), average Delta P score construction-academic (in syntactic sophistication), and textual lexical diversity (in lexical diversity) were substantially different across independent and proficient users, they were freed, and subsequently, partial metric invariance was confirmed (Model 1_{Partial}). Scalar invariance was also not fully supported in Model 2. Since the intercept of textual lexical diversity (in lexical diversity) and dependents per object (syntactic complexity) were substantially different across independent and proficient users, they were freed, and thus partial scalar invariance was confirmed (Model 2_{Partial}). Finally, strict measurement invariance was also not fully supported in Model 3. Since among syntactic complexity indices, the residuals of dependents per direct object and dependents per direct object (SD) were substantially different across independent and proficient users, they were freed, and partial strict invariance was confirmed (Model 3_{Partial}). These results suggest that the proposed model measured observed and latent variables with a partial level of precision across independent and proficient users.

Table 6
Goodness of Fit Indices for Invariance Assessment Across Independent and Proficient Users

Model	χ^2	Df	$\Delta\chi^2$	Δ df	Sig.	CFI	ΔCFI	RMSEA
Model 0	556.43	246	-	-	-	.93	-	.05

Model 1	747.04	258	190.61	12	< .001	.89	-.038	.06
Model $1_{Partial}$	602.51	255	46.07	9	< .001	.92	-.007	.06
Model 2	676.44	267	73.93	12	< .001	.91	-.014	.06
Model $2_{Partial}$	653.55	265	51.04	10	< .001	.91	-.009	.06
Model 3	790.11	283	136.56	16	< .001	.90	-.014	.06
Model $3_{Partial}$	702.92	280	49.37	13	< .001	.90	-.008	.06

Subsequently, a multigroup path analysis was run to analyze differences in the relationships between variables across the two groups. It further revealed that in the case of independent users, 18% of L2 speaking proficiency variance was explained by lexical diversity ($\beta = .29$, $p < .001$), cohesion ($\beta = -.16$, $p = .02$), lexical sophistication ($\beta = .12$, $p = .01$), and syntactic complexity through lexical sophistication ($\beta = .06$, $p = .04$). In this group, syntactic sophistication failed to predict L2 speaking proficiency significantly. In proficient users, 13% of the variance was explained by syntactic sophistication ($\beta = .17$, $p = .007$), lexical sophistication ($\beta = .16$, $p = .01$), lexical diversity ($\beta = .15$, $p = .02$), cohesion ($\beta = -.15$, $p = .02$), and syntactic complexity through lexical sophistication ($\beta = .10$, $p = .03$). These findings indicate that while almost all the variables under investigation contribute to speaking proficiency in both groups, some of them are stronger predictors in one group than the other, further justifying the partial invariance of the model.

5. Discussion

This study aimed to develop a model of L2 speaking proficiency exploring how lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion are associated with holistic scores of L2 speaking proficiency. It also aimed to investigate if the developed model is has generalizability across independent (B1 and B2) and proficient (C1 and C2) users and analyze potential differences in how these variables contribute to L2 speech in each group.

The first research question asked how lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion predict L2 speaking proficiency. The results of SEM analysis showed that 34% of the variance in L2 speaking proficiency was explained by lexical diversity, lexical sophistication, syntactic sophistication, cohesion, and the indirect effect of syntactic complexity through lexical sophistication. While in previous speaking proficiency models, only lexical and cohesive features were significant predictors of L2 speaking proficiency (Crossley & McNamara, 2013), this study showed that syntactic sophistication also plays a crucial role, especially in more advanced levels, and that creating cohesion merely through the use of connectives can negatively impact the assessment of L2 speech. The findings related to each variable are discussed in detail in the following.

Lexical sophistication was the second strongest predictor of L2 speaking proficiency in the whole sample and in proficient users, where it could predict L2 speaking proficiency slightly more strongly than lexical diversity, which also corroborates the findings of previous research (Eguchi, 2022; Eguchi & Kyle, 2020; Kyle & Crossley, 2015). The findings further confirm the multidimensionality of lexical sophistication (Kim et al., 2018; Eguchi & Kyle, 2020) by showing that sophisticated lexical items are not just low-frequency words, as defined traditionally (Laufer & Nation, 1995); rather, they can be defined as lexical items that are contextually distinctive, need more response time, and are acquired at a later age. The results also suggest that for more advanced learners, speaking assessments like IELTS need to place a clear emphasis on lexical sophistication as well, as in this proficiency group, the role of lexical sophistication is as crucial as lexical diversity (if not more). Proficient learners are expected to use appropriate lexical items in various contexts, including academic and professional settings, as in these contexts, the ability to use specialized and subject-specific vocabulary is as important as using a wide range of general words (Schmitt, 2008). McNamara et al. (2010) argue that as proficiency increases, lexical diversity plateaus because learners are expected to use lexical items precisely, appropriately, and accurately, and sophisticated lexical choices help them meet these expectations. Furthermore, the retrieval and use of sophisticated lexical items is a cognitively demanding process that can only be mastered by advanced language users (Crossley et al., 2014). Most rubrics fail to recognize the importance of lexical sophistication in assessing lexical and speaking proficiency, or they refer to the precision of lexical use like the IELTS speaking assessment rubric. The present findings underscore the need for considering lexical sophistication and its indicators in L2 speech assessment. However, as they cannot be assessed through human judgment, automatic assessment tools should be employed, leading to more accurate and more objective assessments of L2 lexical and speaking proficiency across different levels.

Lexical diversity was the strongest predictor of L2 speaking proficiency, overall and in independent users, but in proficient users, it ranked third (next to cohesion) in its relative importance. In line with previous research (Bulté & Roothooft, 2020; Enayat & Derakhshan, 2021; Yu, 2010), the findings suggest that in the overall assessment of L2 speaking proficiency, especially among independent users, the diversity of lexical items needs to be prioritized over their sophistication and over syntactic features. Research has shown that vocabulary size is critical at this stage of communicative competence development, as learners are expected to perform in different communicative contexts and fulfill various functional language needs (Nation, 2001; Schmitt, 2008). Furthermore, at this stage, independent learners are also expected be able to use lexis to discuss unfamiliar topics fluently (IELTS, 2020), which many of them find challenging (Kalantar, 2024). Learners might rely on less complicated syntactic structures and cohesive tools, but lexical variety expands communication possibilities (Nation, 2001). In speaking assessment rubrics, lexical diversity is conceptualized as the range of words (IELTS, 2020), and like lexical sophistication, it is mainly assessed based on the subjective judgment of raters. In this study, lexical diversity was operationalized based on robust measures and showed itself to be a crucial contributor to overall L2 speaking proficiency. While current rubrics heavily rely on subjective assessment of lexical diversity, the evaluation can be done more accurately and objectively using automatic assessment tools.

Syntactic sophistication was the strongest predictor of L2 speaking proficiency in proficient users, while in independent users, it did not predict speaking proficiency at all. Independent users mainly focus on communicating effectively, for which they need lexical diversity and grammatical accuracy (Norris & Ortega, 2009; Pallotti, 2015). Furthermore, in line with frequency-based learning theories (Ellis, 2002), learners tend to use constructions they are more often exposed to and more comfortable with, which, in the case of independent users, are less sophisticated ones. Reaching the advanced level, learners get exposed to more specialized constructions and start mastering them through practice, which allows them to produce more syntactically sophisticated structures (Kyle & Crossley, 2015). In the overall analysis, the predictive power of syntactic sophistication was less than lexical and more than cohesive features. This hierarchy could reflect the nature of spoken communication, where word choice and lexical diversity often have a more immediate impact on fluency and comprehensibility than sophisticated syntactic forms (Skehan, 2009). This finding highlights that assessing syntactic sophistication in L2 speaking tests is essential for advanced learners, as it differentiates between proficiency levels, particularly among those nearing or achieving higher proficiency (e.g., C1, C2). As current speaking assessment rubrics heavily rely on human judgments and are not capable of capturing measures of syntactic sophistication (i.e., frequency and strength of association of constructions), the incorporation of automatic assessment tools is essential, which allows the measurement of this key indicator and leads to a more data-driven evaluation of speaking proficiency.

Syntactic complexity only predicted L2 speaking proficiency through lexical sophistication. This finding indicates that complex noun phrases can be linked to L2 speaking proficiency when they contain sophisticated lexical items. Previous studies, which mostly measured syntactic complexity based on clausal complexity, found that the length of clauses, the usage of C-Units, and the number of subordinate clauses were directly and positively linked with L2 speaking proficiency (Hwang & Kim, 2024; Kim & Lu, 2024; Park, 2022; Yazdani, 2018). The indirect effect found in the present study can be due to operationalizing the construct based on noun phrase complexity indices. According to Kyle & Crossley (2015), complex noun phrases require learners to retrieve more advanced vocabulary to express meaning with contextual appropriateness. Therefore, the use of complex noun phrases is entangled with using sophisticated lexis and shows learners' capability of integrating advanced lexical knowledge within syntactically complex structures. The findings also showed that this indirect effect was stronger in proficient users than in independent users. Research by Norris and Ortega (2009) suggests that more proficient learners use complex noun phrases to express more precise meanings. Thus, they need to form complex syntactic structures using more sophisticated lexis, allowing them to produce grammatically intricate and lexically advanced speech that can express their intended precise meaning.

Finally, cohesion was a negative predictor of L2 speech, and its predictive power was less than lexical and syntactic features overall and in proficient users. However, in independent users, it ranked third (next to lexical diversity). This could indicate the overuse of overt cohesive devices such as connectives and conjunctions, especially by independent users, signaling a rather mechanically connected speech in which cohesion is achieved merely through connecting the ideas on the surface rather than more implicitly connecting them. Previous studies, mainly exploring cohesion in writing, also suggest that more proficient language users rely less on explicit connectives and more on other complex cohesive strategies such as lexical cohesion and ellipsis (Crossley, 2020; Crossley & McNamara, 2011; Ling & Hari, 2019). Cohesive devices such as conjunctions and connectives facilitate speech production at lower proficiency levels (Halliday & Hasan, 1976), but too much reliance upon them disrupts the natural speech expected of learners at higher proficiency levels (IELTS, 2020; McCarthy, 1991). This implies that proficient speakers are mainly assessed based on their ability to speak with natural fluency and coherence. This finding also suggests that in L2 speaking assessments, particularly at higher proficiency levels, the focus should perhaps move away from emphasizing the use of connectives towards more holistic markers of coherence. It also suggests that the role of connectives might need re-evaluation in speaking rubrics, especially for independent users, adding more emphasis on judicious use of these discoursal features since, as evidenced in this study, the overuse of connectives indicates a lack of mastery over natural coherence.

Although this study operationalized each construct based on a limited set of indicators, resulting in moderate variance explained, the findings still underscore the crucial role of lexical diversity, lexical sophistication, syntactic sophistication, syntactic complexity, and cohesion in L2 speech. Based on these findings, assessment models need to be more precise and objective to capture the nuanced aspects of linguistic and discoursal features. Most speaking assessment rubrics group different aspects of lexical or syntactic knowledge into one construct to be assessed holistically and subjectively. However, the present findings indicate that there is more to the role of lexis, syntax, and cohesion in L2 speech than the available rubrics, which are mainly dependent on human judgments, can capture. Therefore, assessment models should employ automatic and objective assessment tools to offer a more comprehensive, detailed, and reliable assessment of L2 speech. Apart from this, recent studies have pointed out that in many ways, current assessment methods fail to offer a reliable evaluation of speaking proficiency due to the lack of correspondence between their target criteria and the reality of the speaking skill (Mendoza Ramos & Martinez, 2022; Souzandehfar, 2024). Moving from subjective assessment to automatic, objective assessment can approximate assessment frameworks to the reality of speaking proficiency, lead to less variability and bias that potentially arise from subjective judgments (Kyle & Crossley, 2015), enhance fairness (McCarthy & Jarvis, 2010), provide immediate feedback for learners, and align with modern assessment approaches.

The second research question explored the generalizability of the developed model across independent and proficient users. Measurement invariance tests only supported partial invariance, meaning that the model performs partially the same in the two groups. More specifically, although the overall structure of the model was the same in the two groups, the relationships between variables and speaking proficiency scores were different. Therefore, the results must be generalized across proficiency levels with great caution, as the importance of some constructs (particularly those related to lexical and syntactic features) differs depending on the user group.

The findings imply that while lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion are significant predictors of L2 speaking proficiency, they influence it differently within proficiency levels. More specifically, while in the case of independent users, lexical features are more significant than syntactic and cohesive ones, in proficient users, along with lexical features, syntactic features play a crucial role in L2 speaking assessment. This suggests that the relative importance of these features changes depending on the proficiency level, which corroborates previous findings (Iwashita et al., 2008), highlighting that the criteria for considering learners proficient enough differ across proficiency levels (Norris & Ortega, 2009). While independent users are expected to use lexis flexibly in various contexts and stay syntactically correct, proficient users are expected to have gained the ability to communicate in more specialized contexts. Therefore, assessment methods need to acknowledge this varying importance by prioritizing features accordingly in each proficiency level instead of assigning equal weight to each rather similarly in proficiency levels. For example, in the IELTS speaking rubric, the range of vocabulary and the complexity of grammatical structures receive equal emphasis for independent users. However, our findings revealed that the former is a stronger predictor of speaking proficiency at this level. The partial invariance also indicates that although the model is generalizable to a degree, adaptations may be needed to capture proficiency across all levels fully.

6. Conclusion

The present study aimed to develop a model of L2 speaking proficiency investigating the relationships between variables of lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion, and human judgments of overall speaking proficiency in independent and proficient users. These variables collectively explained 34% of the variance in speaking proficiency scores. The developed model was partially generalizable across independent and proficient users, with the importance of the investigated variables being different in the two groups. Despite the considerable degree of variance in L2 speech explained by lexical diversity, lexical sophistication, syntactic sophistication, syntactic complexity, and cohesion, a large portion of variance is still unexplained. Future studies could deepen the analysis by exploring the role of additional linguistic features, such as fluency and pronunciation, or even non-linguistic and psycholinguistic features. Furthermore, since this study assessed speaking performance only in monologues, they can analyze L2 speech in various communicative tasks and investigate whether the role of lexical, syntactic, and cohesive features differs accordingly.

The present findings offer several implications for L2 speaking assessment. Speaking assessment frameworks should prioritize linguistic indicators of L2 speaking proficiency, such as lexical sophistication, lexical diversity, syntactic sophistication, syntactic complexity, and cohesion, based on their degree of contribution to L2 speech performance rather than giving them equal weights. Furthermore, the assessment of speaking proficiency should be tailored to each proficiency level by acknowledging that the importance of these linguistic indicators varies across proficiency levels. Finally, incorporating automatic assessment tools in assessment procedures can allow the objective and precise evaluation of nuanced linguistic features, such as lexical and syntactic sophistication, which cannot be fully captured using current subjective assessment rubrics.

Acknowledgments
The authors would like to express their gratitude to everyone who supported and contributed to the completion of this study.

Declaration of Conflicting Interests
The authors declare that there are no conflicting interests regarding the publication of this study.

Funding
The authors received no funding for this study.

Declaration of AI-Generated Content
The authors confirm that no content in this article was generated by artificial intelligence tools. However, Grammarly was used for the final stage of error correction.

References

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford: Oxford University Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bates, E., Bretherton, I., & Snyder, L. S. (1991). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.

Berger, C. M., Crossley, S. A., & Kyle, K. (2017a). Using native-speaker psycholinguistic norms to predict lexical proficiency and development in second language production. *Applied Linguistics*, *40*, 1-22. https://doi.org/10.1093/applin/amx005

Berger, C. M., Crossley, S. A., & Kyle, K. (2017b). Using novel word context measures to predict human ratings of lexical proficiency. *Educational Technology & Society, 20*, 201-212. https://www.jstor.org/stable/90002175

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics, 37*(5), 639-668. https://doi.org/10.1093/applin/amu059

Bulté, B., & Roothooft, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System*, *91*, 102246. https://doi.org/10.1016/j.system.2020.102246

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, *6*(2), 5-35. https://doi.org/10.5070/L462005216

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94-100. https://doi.org/10.1080/09296171003643098

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., Clevinger, A., & Kim, Y. (2014). The role of lexical properties and cohesive devices in text integration and their effect on human ratings of speaking proficiency. *Language Assessment Quarterly*, *11*(3), 250-270. https://doi.org/10.1080/15434303.2014.926905

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. http://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., & Salsbury, T. (2016b). A usage-based investigation of L2 lexical acquisition: The role of input and output. *The Modern Language Journal*, *100*, 702-715. https://doi.org/10.1111/modl.12344

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35*(2), 115-135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., & McNamara, D. (2013). Applications of text analysis tools for spoken response grading. *Language Learning & Technology*, *17*(2), 171-192. https://doi.org/10125/44329

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*(5), 570-590. https://doi.org/10.1093/applin/amt056

Crossley, S. A., Subtirelu, N., & Salsbury, T. (2013). Frequency effects or context effects in second language word learning: What predicts early lexical production? *Studies in Second Language Acquisition*, *35*, 727-755. https://doi.org/10.1017/S0272263113000375

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*(1), 5-34. https://doi.org/10.1017/S0272263111000489

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121-149. https://doi.org/10.1177/0748175610373459

Eguchi, M. (2022). Modeling lexical and phraseological sophistication in oral proficiency interviews: A conceptual replication. *Vocabulary Learning and Instruction*, *11*(2), 1-16. https://doi.org/10.7820/vli.v11.2.Eguchi

Eguchi, M., & Kyle, K. (2020). Continuing to explore the multidimensional nature of lexical sophistication: The case of oral proficiency interviews. *The Modern Language Journal*, *104*(2), 381-400. https://doi.org/10.1111/modl.12637

Enayat, M. J., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, *99*, 102521. https://doi.org/10.1016/j.system.2021.102521

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143-188. https://doi.org/10.1017/S0272263102002024

George, D., & Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference, 17.0 update* (10th ed.). London: Pearson.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks, CA: SAGE Publications, Inc.

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

Hwang, H., & Kim, H. (2024). Korean Syntactic Complexity Analyzer (KOSCA): An NLP application for the analysis of syntactic complexity in second language production. *Language Testing*, *41*(3), 506-529. https://doi.org/10.1177/02655322231222596

IDP IELTS. (2020). *Free speaking practice questions.* https://ielts.idp.com/prepare/article-free-speaking-practice-questions

IELTS. (2020). *Speaking band descriptors*. https://www.ielts.com/results/scores/speaking

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied linguistics*, *29*(1), 24-49. https://doi.org/10.1093/applin/amm017

Johnson, W. (1944). Studies in language behavior: A program of research. *Psychological Monographs*, *56*(2), 1-15. https://doi.org/10.1037/H0093508

Kalantar, O. S. (2024). Attaining C1-level scores in TOEFL iBT: From challenges to needs. *International Journal of Language Testing*, *14*(1), 99-113. https://doi.org/10.22034/ijlt.2023.414966.1287

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal*, *102*(1), 120-141. https://doi.org/10.1111/modl.12447

Kim, H., & Lu, X. (2024). L2 English speaking syntactic complexity: Data preprocessing issues, reliability of automated analysis, and the effects of proficiency, L1 background, and topic. *The Modern Language Journal*, *108*(1), 127-146. https://doi.org/10.1111/modl.12907

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York City, NY: Guilford Press.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine-grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Doctoral dissertation, Georgia State University]. Open Access Institutional Repository of Georgia State University. https://scholarworks.gsu.edu/alesl_diss/35/

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757-786. https://doi.org/10.1002/tesq.194

Kyle, K., & Crossley, S. A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, *34*, 12-24. https://doi.org/10.1016/j.jslw.2016.10.003

Kyle, K., & Crossley, S. A. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, *34*(4), 513-535. https://doi.org/10.1177/0265532217712554

Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods*, *50*(3), 1030-1046. https://doi.org/10.3758/s13428-017-0924-4

Kyle, K., Crossley, S. A., & Jarvis, S. (2020). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, *18*(2), 154-170. https://doi.org/10.1080/15434303.2020.1844205

Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, *43*(4), 781-812. https://doi.org/10.1017/S0272263120000546

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307-322. https://doi.org/10.1093/applin/16.3.307

Lieven, E. V. M. (1978). Conversations between mothers and young children: Individual differences and their possible implication for the study of child language learning. In N. Waterson & C. E. Snow (Eds.), *The development of communication* (pp. 173–187). Hoboken, NJ: Wiley.

Ling, C. Y., & Hari, J. (2019). Cohesive devices and writing quality: Singapore learner corpus research. *Corpus Linguistics Research*, *5*, 29-53. https://doi.org/10.18659/CLR.2019.5.02

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, *96*(2), 190-208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

Maas, H. D. (1972). On the relationship between vocabulary and the length of a text. *Zeitschrift für Literaturwissenschaft und Linguistik*, *2*(8), 73.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.

McCarthy, M., & Carter, R. (2014). *Language as discourse: Perspectives for language teaching* (2nd ed.). London: Routledge.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, *24*(4), 459-488. https://doi.org/10.1177/0265532207080767

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381-392. https://doi.org/10.3758/BRM.42.2.381

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. In C. Bazerman, R. Krut, K. Lunsford, S. Null, P. Rogers, & A. Stansell (Eds.), *Handbook of research on writing: History, society, school, individual, text* (pp. 163–177). London: Routledge.

Mendoza Ramos, A., & Martinez, J. A. (2023). The retrofit of an English language placement test used for large-scale assessments in higher education. *International Journal of Language Testing, 13*(1), 139-165. https://doi.org/10.22034/ijlt.2022.354919.1184

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4). http://doi.org/10.1093/applin/amp044

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*(1), 117-134. https://doi.org/10.1177/0267658314536435

Park, S. (2022). Syntactic complexity in a learner written corpus and L2 speaking quality: Suggestions for distinguishing L2 speaking proficiency. *Journal of Language and Linguistic Studies*, *18*(1), 361-371. https://eric.ed.gov/?id=EJ1325786

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Sadri Mirdamadi, F., & De Jong, N. H. (2015). The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research*, *31*(1), 105-116. https://doi.org/10.1177/0267658314554498

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. https://doi.org/10.1177/1362168808089921

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510-532. https://doi.org/10.1093/applin/amp047

Souzandehfar, M. (2024). New perspectives on IELTS authenticity: An evaluation of the speaking module. *International Journal of Language Testing*, *14*(1), 34-55. https://doi.org/10.22034/IJLT.2023.409599.1272

Taş, T., & Khan, Ö. (2020). On the models of communicative competence. *Proceedings of GLOBETSonline: International Conference on Education, Technology and Science* (pp. 86–96). http://www.globets.org/files/Globetsonline_proceedings.pdf

Yazdani, S. (2018). Syntactic complexity in Iranian learners' English writing and speaking. *Journal on English as a Foreign Language*, *8*(1), 75-96. https://doi.org/10.23971/jefl.v8i1.718

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, *31*(2), 236-259. https://doi.org/10.1093/applin/amp024

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, *47*, 100505. https://doi.org/10.1016/j.asw.2020.100505

Appendix A
List of Indices Initially Selected for the Study

Table A1

Indices Initially Selected for the Study

Variables	NLP tools	Indices	Detailed label
Lexical sophistication	TAALES	Contextual distinctiveness	eat_types
		Contextual distinctiveness	eat_tokens
		Contextual distinctiveness	USF
		Contextual distinctiveness	McD_CD
		Contextual distinctiveness	Sem_D
		Word neighbor information	Ortho_N
		Word neighbor information	Phono_N
		Word neighbor information	Phono_N_H
		Word neighbor information	OG_N
		Word neighbor information	OG_N_H
		Word neighbor information	Freq_N
		Word neighbor information	Freq_N_P
		Word neighbor information	Freq_N_PH
		Word neighbor information	Freq_N_OG
		Word neighbor information	Freq_N_OGH
		Word neighbor information	OLD
		Word neighbor information	OLDF
		Word neighbor information	PLD
		Word neighbor information	PLDF
		Word recognition norms	LD_Mean_RT
		Word recognition norms	LD_Mean_RT_Zscore
		Word recognition norms	LD_Mean_RT_SD
		Word recognition norms	LD_Mean_Accuracy
		Word recognition norms	WN_Mean_RT
		Word recognition norms	WN_Zscore
		Word recognition norms	WN_SD
		Word recognition norms	WN_Mean_Accuracy
		Contextual distinctiveness	lsa_average_top_three_cosine
		Contextual distinctiveness	lsa_max_similarity_cosine
		Contextual distinctiveness	lsa_average_all_cosine
		Age of acquisition	aoe_inverse_average
		Age of acquisition	aoe_inverse_linear_regression_slope
		Age of acquisition	aoe_index_above_threshold_40
		Age of acquisition	aoe_inflection_point_polynomial
		Semantic network	content_poly
		Semantic network	poly_noun
		Semantic network	poly_verb
		Semantic network	poly_adj
		Semantic network	poly_adv
		Semantic network	hyper_noun_S1_P1
		Semantic network	hyper_noun_Sav_P1
		Semantic network	hyper_noun_Sav_Pav
		Semantic network	hyper_verb_S1_P1
		Semantic network	hyper_verb_Sav_P1
		Semantic network	hyper_verb_Sav_Pav
		Semantic network	hyper_verb_noun_s1_p1
		Semantic network	hyper_verb_noun_Sav_P1

		Semantic network	hyper_verb_noun_Sav_Pav
Lexical diversity	TAALED	Maas' index	maas_ttr_aw
		Moving average type token ratio	mattr50_aw
		Mean segmental type token ratio	msttr50_aw
		Distribution diversity index	hdd42_aw
		Textual lexical diversity (version 1)	mtld_original_aw
		Textual lexical diversity (version 2)	mtld_ma_bi_aw
		Textual lexical diversity (version 3)	mtld_ma_wrap_aw
Syntactic complexity	TAASSC	Noun phrase complexity	av_nominal_deps
		Noun phrase complexity	av_nsubj_deps
		Noun phrase complexity	av_nsubj_pass_deps
		Noun phrase complexity	av_agents_deps
		Noun phrase complexity	av_dobj_deps
		Noun phrase complexity	av_pobj_deps
		Noun phrase complexity	av_iobj_deps
		Noun phrase complexity	av_ncomp_deps
		Noun phrase complexity	nominal_deps_stdev
		Noun phrase complexity	nsubj_stdev
		Noun phrase complexity	nsubj_pass_stdev
		Noun phrase complexity	agents_stdev
		Noun phrase complexity	dobj_stdev
		Noun phrase complexity	pobj_stdev
		Noun phrase complexity	iobj_stdev
		Noun phrase complexity	ncomp_stdev
		Noun phrase complexity	det_nsubj_deps_struct
		Noun phrase complexity	amod_nsubj_deps_struct
		Noun phrase complexity	prep_nsubj_deps_struct
		Noun phrase complexity	poss_nsubj_deps_struct
		Noun phrase complexity	vmod_nsubj_deps_struct
		Noun phrase complexity	rcmod_nsubj_deps_struct
		Noun phrase complexity	advmod_nsubj_deps_struct
		Noun phrase complexity	conj_and_nsubj_deps_struct
		Noun phrase complexity	conj_or_nsubj_deps_struct
		Noun phrase complexity	det_dobj_deps_struct
		Noun phrase complexity	amod_dobj_deps_struct
		Noun phrase complexity	prep_dobj_deps_struct
		Noun phrase complexity	poss_dobj_deps_struct
		Noun phrase complexity	vmod_dobj_deps_struct
		Noun phrase complexity	rcmod_dobj_deps_struct
		Noun phrase complexity	advmod_dobj_deps_struct
		Noun phrase complexity	conj_and_dobj_deps_struct
		Noun phrase complexity	conj_or_dobj_deps_struct
		Noun phrase complexity	det_pobj_deps_struct
		Noun phrase complexity	amod_pobj_deps_struct
		Noun phrase complexity	prep_pobj_deps_struct
		Noun phrase complexity	poss_pobj_deps_struct
		Noun phrase complexity	vmod_pobj_deps_struct
		Noun phrase complexity	rcmod_pobj_deps_struct
		Noun phrase complexity	advmod_pobj_deps_struct

		Noun phrase complexity	conj_and_pobj_deps_struct
		Noun phrase complexity	conj_or_pobj_deps_struct
		Noun phrase complexity	det_iobj_deps_struct
		Noun phrase complexity	amod_iobj_deps_struct
		Noun phrase complexity	prep_iobj_deps_struct
		Noun phrase complexity	poss_iobj_deps_struct
		Noun phrase complexity	vmod_iobj_deps_struct
		Noun phrase complexity	rcmod_iobj_deps_struct
		Noun phrase complexity	advmod_iobj_deps_struct
		Noun phrase complexity	conj_and_iobj_deps_struct
		Noun phrase complexity	conj_or_iobj_deps_struct
		Clause complexity	cl_av_deps
		Clause complexity	cl_ndeps_std_dev
		Clause complexity	acomp_per_cl
		Clause complexity	advcl_per_cl
		Clause complexity	agent_per_cl
		Clause complexity	cc_per_cl
		Clause complexity	ccomp_per_cl
		Clause complexity	conj_per_cl
		Clause complexity	csubj_per_cl
		Clause complexity	csubjpass_per_cl
		Clause complexity	dep_per_cl
		Clause complexity	discourse_per_cl
		Clause complexity	dobj_per_cl
		Clause complexity	expl_per_cl
		Clause complexity	iobj_per_cl
		Clause complexity	mark_per_cl
		Clause complexity	ncomp_per_cl
		Clause complexity	neg_per_cl
		Clause complexity	nsubj_per_cl
		Clause complexity	nsubjpass_per_cl
		Clause complexity	parataxis_per_cl
		Clause complexity	pcomp_per_cl
		Clause complexity	prep_per_cl
		Clause complexity	prepc_per_cl
		Clause complexity	prt_per_cl
		Clause complexity	tmod_per_cl
		Clause complexity	xcomp_per_cl
		Clause complexity	xsubj_per_cl
		Clause complexity	advmod_per_cl
		Clause complexity	aux_per_cl
		Clause complexity	auxpass_per_cl
		Clause complexity	modal_per_cl
Syntactic sophistication	TAASSC	Average lemma frequency	acad_av_lemma_freq
		Average construction frequency	acad_av_construction_freq
		Average lemma construction frequency	acad_av_lemma_construction_freq
		Average approximate collostructional strength	acad_av_approx_collexeme
		Faith score	acad_av_faith_verb_cue
		Faith score	acad_av_faith_const_cue
		Delta P score	acad_av_delta_p_verb_cue

	Delta P score	acad_av_delta_p_const_cue
	Average lemma frequency	acad_av_lemma_freq_type
	Average construction frequency	acad_av_construction_freq_type
	Average lemma construction frequency	acad_av_lemma_construction_freq_type
	Average approximate collostructional strength	acad_av_approx_collexeme_type
	Faith score	acad_av_faith_verb_cue_type
	Faith score	acad_av_faith_const_cue_type
	Delta P score	acad_av_delta_p_verb_cue_type
	Delta P score	acad_av_delta_p_const_cue_type
	Collostruction ratio	acad_collexeme_ratio
	Collostruction ratio	acad_collexeme_ratio_type
	Lemma type-token ratio	acad_lemma_ttr
	Construction type-token ratio	acad_construction_ttr
	Lemma construction type-token ratio	acad_lemma_construction_ttr
	Lemmas in text in reference corpus	acad_lemma_attested
	Constructions in text in reference corpus	acad_construction_attested
	Lemmas and constructions in text in reference corpus	acad_lemma_construction_attested
	Average lemma frequency	news_av_lemma_freq
	Average construction frequency	news_av_construction_freq
	Average lemma construction frequency	news_av_lemma_construction_freq
	Average approximate collostructional strength	news_av_approx_collexeme
	Faith score	news_av_faith_verb_cue
	Faith score	news_av_faith_const_cue
	Delta P score	news_av_delta_p_verb_cue
	Delta P score	news_av_delta_p_const_cue
	Average lemma frequency	news_av_lemma_freq_type
	Average construction frequency	news_av_construction_freq_type
	Average lemma construction frequency	news_av_lemma_construction_freq_type
	Average approximate collostructional strength	news_av_approx_collexeme_type
	Faith score	news_av_faith_verb_cue_type
	Faith score	news_av_faith_const_cue_type
	Delta P score	news_av_delta_p_verb_cue_type
	Delta P score	news_av_delta_p_const_cue_type
	Collostruction ratio	news_collexeme_ratio
	Collostruction ratio	news_collexeme_ratio_type
	Lemma type-token ratio	news_lemma_ttr
	Construction type-token ratio	news_construction_ttr
	Lemma construction type-token ratio	news_lemma_construction_ttr
	Lemmas in text in reference corpus	news_lemma_attested

	Constructions in text in reference corpus	news_construction_attested
	Lemmas and constructions in text in reference corpus	news_lemma_construction_attested
	Average lemma frequency	mag_av_lemma_freq
	Average construction frequency	mag_av_construction_freq
	Average lemma construction frequency	mag_av_lemma_construction_freq
	Average approximate collostructional strength	mag_av_approx_collexeme
	Faith score	mag_av_faith_verb_cue
	Faith score	mag_av_faith_const_cue
	Delta P score	mag_av_delta_p_verb_cue
	Delta P score	mag_av_delta_p_const_cue
	Average lemma frequency	mag_av_lemma_freq_type
	Average construction frequency	mag_av_construction_freq_type
	Average lemma construction frequency	mag_av_lemma_construction_freq_type
	Average approximate collostructional strength	mag_av_approx_collexeme_type
	Faith score	mag_av_faith_verb_cue_type
	Faith score	mag_av_faith_const_cue_type
	Delta P score	mag_av_delta_p_verb_cue_type
	Delta P score	mag_av_delta_p_const_cue_type
	Collostruction ratio	mag_collexeme_ratio
	Collostruction ratio	mag_collexeme_ratio_type
	Lemma type-token ratio	mag_lemma_ttr
	Construction type-token ratio	mag_construction_ttr
	Lemma construction type-token ratio	mag_lemma_construction_ttr
	Lemmas in text in reference corpus	mag_lemma_attested
	Constructions in text in reference corpus	mag_construction_attested
	Lemmas and constructions in text in reference corpus	mag_lemma_construction_attested
	Average lemma frequency	fic_av_lemma_freq
	Average construction frequency	fic_av_construction_freq
	Average lemma construction frequency	fic_av_lemma_construction_freq
	Average approximate collostructional strength	fic_av_approx_collexeme
	Faith score	fic_av_faith_verb_cue
	Faith score	fic_av_faith_const_cue
	Delta P score	fic_av_delta_p_verb_cue
	Delta P score	fic_av_delta_p_const_cue
	Average lemma frequency	fic_av_lemma_freq_type
	Average construction frequency	fic_av_construction_freq_type
	Average lemma construction frequency	fic_av_lemma_construction_freq_type
	Average approximate collostructional strength	fic_av_approx_collexeme_type

Faith score	fic_av_faith_verb_cue_type
Faith score	fic_av_faith_const_cue_type
Delta P score	fic_av_delta_p_verb_cue_type
Delta P score	fic_av_delta_p_const_cue_type
Collostruction ratio	fic_collexeme_ratio
Collostruction ratio	fic_collexeme_ratio_type
Lemma type-token ratio	fic_lemma_ttr
Construction type-token ratio	fic_construction_ttr
Lemma construction type-token ratio	fic_lemma_construction_ttr
Lemmas in text in reference corpus	fic_lemma_attested
Constructions in text in reference corpus	fic_construction_attested
Lemmas and constructions in text in reference corpus	fic_lemma_construction_attested
Average lemma frequency	all_av_lemma_freq
Average construction frequency	all_av_construction_freq
Average lemma construction frequency	all_av_lemma_construction_freq
Average approximate collostructional strength	all_av_approx_collexeme
Faith score	all_av_faith_verb_cue
Faith score	all_av_faith_const_cue
Delta P score	all_av_delta_p_verb_cue
Delta P score	all_av_delta_p_const_cue
Average lemma construction frequency	all_av_lemma_construction_freq_log
Average lemma frequency	all_av_lemma_freq_type
Average construction frequency	all_av_construction_freq_type
Average construction frequency	all_av_lemma_construction_freq_type
Average approximate collostructional strength	all_av_approx_collexeme_type
Faith score	all_av_faith_verb_cue_type
Faith score	all_av_faith_const_cue_type
Delta P score	all_av_delta_p_verb_cue_type
Delta P score	all_av_delta_p_const_cue_type
Collostruction ratio	all_collexeme_ratio
Collostruction ratio	all_collexeme_ratio_type
Lemma type-token ratio	all_lemma_ttr
Construction type-token ratio	all_construction_ttr
Lemma construction type-token ratio	all_lemma_construction_ttr
Lemmas in text in reference corpus	all_lemma_attested
Constructions in text in reference corpus	all_construction_attested
Lemmas and constructions in text in reference corpus	all_lemma_construction_attested
Average lemma frequency	all_av_lemma_freq_stdev
Average construction frequency	all_av_construction_freq_stdev

		Average lemma construction frequency	all_av_lemma_construction_freq_stdev
		Average approximate collostructional strength	all_av_approx_collexeme_stdev
		Faith score	all_av_faith_verb_cue_stdev
		Faith score	all_av_faith_const_cue_stdev
		Delta P score	all_av_delta_p_verb_cue_stdev
		Delta P score	all_av_delta_p_const_cue_stdev
		Average lemma frequency	acad_av_lemma_freq_stdev
		Average construction frequency	acad_av_construction_freq_stdev
		Average lemma construction frequency	acad_av_lemma_construction_freq_stdev
		Average approximate collostructional strength	acad_av_approx_collexeme_stdev
		Faith score	acad_av_faith_verb_cue_stdev
		Faith score	acad_av_faith_const_cue_stdev
		Delta P score	acad_av_delta_p_verb_cue_stdev
		Delta P score	acad_av_delta_p_const_cue_stdev
		Average lemma frequency	news_av_lemma_freq_stdev
		Average construction frequency	news_av_construction_freq_stdev
		Average lemma construction frequency	news_av_lemma_construction_freq_stdev
		Average approximate collostructional strength	news_av_approx_collexeme_stdev
		Faith score	news_av_faith_verb_cue_stdev
		Faith score	news_av_faith_const_cue_stdev
		Delta P score	news_av_delta_p_verb_cue_stdev
		Delta P score	news_av_delta_p_const_cue_stdev
		Average lemma frequency	mag_av_lemma_freq_stdev
		Average construction frequency	mag_av_construction_freq_stdev
		Average lemma construction frequency	mag_av_lemma_construction_freq_stdev
		Average approximate collostructional strength	mag_av_approx_collexeme_stdev
		Faith score	mag_av_faith_verb_cue_stdev
		Faith score	mag_av_faith_const_cue_stdev
		Delta P score	mag_av_delta_p_verb_cue_stdev
		Delta P score	mag_av_delta_p_const_cue_stdev
		Average lemma frequency	fic_av_lemma_freq_stdev
		Average construction frequency	fic_av_construction_freq_stdev
		Average lemma construction frequency	fic_av_lemma_construction_freq_stdev
		Average approximate collostructional strength	fic_av_approx_collexeme_stdev
		Faith score	fic_av_faith_verb_cue_stdev
		Faith score	fic_av_faith_const_cue_stdev
		Delta P score	fic_av_delta_p_verb_cue_stdev
		Delta P score	fic_av_delta_p_const_cue_stdev
Cohesion	TAACO	Basic connectives	basic_connectives
		Conjunctions	conjunctions
		Disjunctions	disjunctions
		Lexical subordinators	lexical_subordinators

	Coordinating conjunctions	coordinating_conjuncts
	Addition words	addition
	Sentence linking	sentence_linking
	Order words	order
	Reason and purpose words	reason_and_purpose
	Causal connectives	all_causal
	Positive causal connectives	positive_causal
	Opposition words	opposition
	Determiners	determiners
	Demonstratives	all_demonstratives
	Attended demonstratives	attended_demonstratives
	Unattended demonstratives	unattended_demonstratives
	Additive connectives	all_additive
	Logical connectives	all_logical
	Positive logical connectives	positive_logical
	Negative logical connectives	negative_logical
	Temporal connectives	all_temporal
	Positive intentional connectives	positive_intentional
	Positive connectives	all_positive
	Negative connectives	all_negative
	All connectives	all_connective

Appendix B
List of the 17 Indices Confirmed for the Measurement Model

Table B1
Indices Confirmed for the Measurement Model

Variables	Indices	Detailed labels
Lexical sophistication	Contextual distinctiveness	eat_types
	Word recognition norms	WN_Mean_RT
	Age of acquisition/exposure	aoe_index_above_threshold_40
Lexical diversity	Mass's type-token ratio	maas_ttr_aw
	Distribution diversity index	hdd42_aw
	Textual lexical diversity	mtld_ma_wrap_aw
Syntactic complexity	Dependents per direct object	av_dobj_deps
	Dependents per direct object (SD)	dobj_stdev
	Determiners per nominal	det_all_nominal_deps_struct
	Adjectival modifiers per nominal	amod_all_nominal_deps_struct
Syntactic sophistication	Average Delta P score construction-academic	acad_av_delta_p_const_cue_stdev
	Average Faith score construction-news	news_av_faith_const_cue_stdev
	Average Delta P score construction-magazine	mag_av_delta_p_const_cue_stdev
	Average Delta P score construction-fiction	fic_av_delta_p_const_cue_stdev
Cohesion	Basic connectives	basic_connectives
	Conjunctions	conjunctions
	Positive connectives	all_positive