Language Learning & Technology
ISSN 1094-3501 2023, Volume 27, Issue 1
pp. 1–21

ARTICLE

The use of lexical complexity for assessing difficulty in instructional videos

Emad A. Alghamdi, King Abdulaziz University
Paul Gruba, The University of Melbourne
Ahmed Masrai, King Abdulaziz Military Academy
Eduardo Velloso, The University of Melbourne

Abstract

Although measures of lexical complexity are well established for printed texts, there is currently no equivalent work for videos. This study, therefore, aims to investigate whether existing lexical complexity measures can be extended to predict second language (L2) learners' judgment of video difficulty. Using a corpus of 320 instructional videos, regression models were developed for explaining and predicting difficulty using indices of lexical sophistication, density, and diversity. Results of the study confirm key dimensions of lexical complexity in estimates of video difficulty. In particular, lexical frequency indices accounted for the largest variance in the assessment of video difficulty (R^2 = .45). We conclude with implications for CALL and suggest areas of further research.

Keywords: *Lexical Complexity, Video Difficulty, Language Learning, Lexical Frequency*

Language(s) Learned in This Study: *English*

APA Citation: Alghamdi, E. A., Gruba, P., Masrai, A., & Velloso, E. (2023). The use of lexical complexity for assessing difficulty in instructional videos. *Language Learning & Technology*, *27*(1), 1–21. https://hdl.handle.net/10125/73524

Introduction

The use of digital video materials has grown in tandem with the rise of computer-assisted language learning (CALL). Extant research has shown that videos provide a catalyst for facilitating incidental vocabulary learning (Rodgers, 2018; Teng & Teng, 2020; Wong & Samudra, 2019), improving listening comprehension (Aldukhayel, 2021; Cross, 2011; Winke et al., 2010), and learning new grammatical structures (Lee & Révész, 2018). However, the difficulty of a video must be appropriately aligned with a learner's ability in order to be effective (Alghamdi, 2021; Fulcher, 1997; Goh, 2000). If a video is too difficult, for example, learners may become frustrated; if a video is too easy, learners may also lose interest. Therefore, video difficulty assessment would help to inform the selections made by CALL teachers, material developers, and researchers.

Of the many factors that can negatively impact second language (L2) learners' understanding of audio and printed texts, lexical complexity has the most adverse effect (e.g., Brunfaut & Révész, 2015; Révész & Brunfaut, 2013). Because successful understanding of spoken texts is contingent on an ability to recognize the vocabulary of an utterance from the flow of speech (Cheng & Matthews, 2018; Goh, 2000; Rost, 2011; van Zeeland & Schmitt, 2013), the present study focuses on the contribution of lexical complexity as a focal element in the assessment of L2 video difficulty. Following a review of key concepts, the study seeks to determine key lexical indices that may help explain and predict language learners' judgment of video difficulty. The study concludes with suggestions for further research in CALL.

Lexical Complexity and Text Difficulty

As a well-established concept, lexical complexity has been used for over a hundred years as a means to assess text difficulty (Dale & Tyler, 1934; Lively & Pressey, 1923). There is ample evidence that text becomes more difficult to read or listen to when it contains a large number of infrequent vocabulary (Milton, 2009). Accordingly, measures of lexical complexity (e.g., lexical frequency and sophistication) have been used in the development of numerous readability formulas, including, for example, the Flesch Reading Ease formula (Kincaid et al., 1975).

Lexical complexity has also been the locus of extensive research in second language acquisition (SLA) literature, and its impact on L2 proficiency is widely acknowledged (De Clercq, 2015). L2 researchers have interrogated, for example, the relationship between lexical complexity and writing quality (e.g., Crossley & McNamara, 2012), vocabulary learning (e.g., Hashimoto & Egbert, 2019), and comprehension of written and spoken text (e.g., Brunfaut & Révész, 2015; Crossley et al., 2008; Révész & Brunfaut, 2013). Although a plethora of lexical complexity research has focused on the measurement of L2 learners' language production, little research has used lexical indices for L2 language reception (e.g., Brunfaut & Révész, 2015; Yoon et al., 2016), and none, to the best of our knowledge, have directly investigated whether lexical complexity indices can explain and predict L2 learners' perception of video difficulty.

The unique and transient characteristics of listening make comprehending video a demanding task for language learners. Simply knowing a word in its orthographic or printed form, for example, does not guarantee it will be recognized when presented aurally (Milton et al., 2010). Perhaps mistakenly, visual elements have been assumed to aid text comprehension as their presence may take full advantage of human processing capacities, as suggested by the Dual Coding Theory (Clark & Paivio, 1991; Paivio, 1990) and the Cognitive Theory of Multimedia Learning (Mayer, 2009). Researchers have also suggested that visual elements provide 'contextual cues' that may assist learners in identifying new words, for example, or otherwise reinforce the knowledge of words that are partially known (Peters, 2019; Peters & Webb, 2018; Webb & Rodgers, 2009). However, when an extensive amount of both verbal and visual elements is presented, the presence of these elements may 'compete and collaborate' in ways that cause learners a range of comprehension difficulties (Gruba, 2006; Kirby, 1993).

Video Complexity and Difficulty

Despite its prominence in language research, the construct of 'difficulty' remains elusive as it is often used alongside or in place of terms such as 'complexity,' 'comprehensibility,' 'learnability,' 'teachability,' and 'accessibility' (Housen, 2014; Housen & Simoens, 2016). Moreover, complexity and difficulty have been operationalized differently in the literature. In learning a language feature, for example, difficulty is often perceived as problematic, but some degree of difficulty is beneficial (Suzuki et al., 2019). In language production research, complexity is a desirable and sought-after indicator of proficiency, but receptively complex language can reduce readability and intelligibility. In the field of text difficulty, Mesmer et al. (2012) asserted that text difficulty and complexity are two distinct constructs, hence treating them as synonymous "conflates causes with effects" (p. 236). To put it differently, the complexity of a text or feature always implies an independent or predictor variable, whereas the difficulty of a text or feature implies a dependent or criterion variable (Mesmer et al., 2012).

Against this background, in this study, we use the term *video difficulty* to refer to learners' perception of difficulty due to "*an increase in the attentional and cognitive resources demanded for successful understanding of videotext*" (Alghamdi, 2021, p. 43). On the other hand, we use the term 'video complexity' to refer to the characteristics of individual videos, such as lexical sophistication and sentence length, that can be analyzed and manipulated.

Measures of Lexical Complexity

Lexical complexity is a multidimensional and multifaceted construct that encompasses three interconnected subconstructs: lexical sophistication, lexical density, and lexical diversity (Kim et al., 2018; Lu, 2012). In the following section, we discuss how each construct has been operationalized in the literature.

Lexical Sophistication

Lexical sophistication, or rareness, refers to the proportion of sophisticated or difficult words in a text (Laufer & Nation, 1995). While there is no standard definition of sophisticated or difficult words (Daller et al., 2007), sophisticated lexis is associated with infrequent or uncommon words that are less likely to be known or recognized by language learners (Ellis, 2002; Gries & Ellis, 2015). The frequency in which a word occurs in language, as noted by Schmitt (2010), "is arguably the single most important characteristic of lexis that researchers must address" (p. 63). Beyond frequency, however, there has been little consensus among researchers on how lexical 'sophistication' should be defined (Daller et al., 2007). However, recent investigations suggest that lexical sophistication is a multilayered construct that includes various dimensions beyond frequency such as orthographic density, hypernymy, and n-gram frequency and association strength (Kim et al., 2018).

Révész and Brunfaut (2013) investigated the impact of spoken text complexity on listening task difficulty. The findings of their research showed that the proportion of function words to lexical density had a strong effect on listening difficulty, as it accounted for 40% of the variability in listening difficulty. In a later study, Brunfaut and Révész (2015) found that formulaic sequences correlate positively with listening task difficulty, where learners showed difficulty in understanding spoken text with formulaic strings. Such a result lends support to the role of lexical sophistication within overall text difficulty. In contrast, other studies found no significant relationship between lexical frequency and spoken text difficulty (e.g., Yanagawa & Green, 2008; Ying-hui, 2006). The current study seeks to further examine this issue.

Lexical Density

The term 'lexical density,' originally coined by Ure (1971), refers to the proportion of content words (i.e., nouns, verbs, adjectives, and some adverbs) relative to the total number of words in a text. Earlier research (e.g., Halliday & Hasan, 1985; Ure, 1971) suggests that spoken texts have approximately 40% lower lexical density than written texts. Lexical density is generally regarded as an index of information packaging (Johansson, 2009). In this regard, content words, as opposed to function words (e.g., prepositions, interjections, pronouns, conjunctions, and count words), carry more information in language. Thus, processing texts with high lexical density typically exerts a greater cognitive load on L2 listeners (Bloomfield et al., 2010). In L2 research, Buck and Tatsuoka (1998) found that listening difficulty increased in line with an increase in the content words pertaining to task completion. Results of such correlation have been mixed. In one study, for example, Révész and Brunfaut (2013) found that lexical density was a key predictor of listening task difficulty, yet another study by Brunfaut and Révész (2015) found lexical density to have an insignificant correlation with listening task difficulty.

Lexical Diversity

Lexical diversity, also known as lexical variation and lexical richness, constitutes the range and variety of vocabulary in a text (Read, 2000; Tweedie & Baayen, 1998). A text that contains a greater diversity of unique words is seen to be more challenging for language learners to decode (Bloomfield et al., 2010). Previous studies have found lexical diversity to be a significant predictor of spoken text difficulty (Révész & Brunfaut, 2013; Rupp et al., 2001). Higher lexical diversity is associated with a more varied use of lexis and measures of lexical diversity, therefore, seeking to assess how many unique words there are in a text. Traditionally, lexical diversity is assessed by counting the number of unique words in a text or calculating the ratio of different words (types) to the total number of words (tokens), referred to as type-token ratio (TTR) (Chotlos, 1944; Templin, 1957). A well-known issue with TTR is its sensitivity to text length (McCarthy & Jarvis, 2007). Specifically, TTR values become less stable with longer texts—as the number of tokens increases, the likelihood of those tokens being unique becomes very low (Heaps, 1978). To address this issue, researchers have devised numerous indices using statistical transformation (e.g., Root TTR) or more advanced methods (e.g., Measure of Lexical Textual Diversity; McCarthy & Jarvis, 2010).

To reiterate, given the instrumental role of vocabulary knowledge and familiarity on text comprehension (Muljani et al., 1998), it is reasonable to suggest that lexical complexity contributes to L2 learners' perception of video difficulty. In the literature, numerous indices have been proposed to assess lexical frequency and sophistication, density, and diversity. These indices were found useful in several areas of research, but little is known if they can prove helpful in explaining and predicting video difficulty. While several studies have explored factors contributing to L2 listening comprehension, little research has examined L2 learners' perception of video difficulty, and the present study seeks to address this gap.

The Current Study

The present study focuses on the contribution of lexical complexity to L2 language learners' perception of video difficulty. Specifically, the study investigates whether indices of lexical complexity and its subordinate constructs can be used to explain and predict the difficulty of instructional video lectures as perceived by L2 language learners. The following research questions guided the present study:

1. How do different indices of lexical sophistication, density, and diversity correlate with perceived video difficulty?
2. To what extent can indices of lexical sophistication, diversity, and density explain perceived video difficulty?
3. How well do lexical complexity regression models predict the difficulty of out-of-sample videos?

Video Corpus

The video corpus analyzed in this study came from the Second Language Video Complexity (SLVC) corpus which contains 320 instructional videos and 320 government advertisements (Alghamdi, 2021). The current study targeted the instructional video sub-corpus. All videos were curated from freely available online courses and tutorials from different scientific disciplines as shown in Table 1.

Table 1

Descriptive Statistics of the Instructional Video Sub-Corpus

Discipline	Number	Duration in Minutes (Mean)	Words (Mean)
Biology	27	255 (9.47)	35K (1302)
Business	13	52 (4.02)	8K (608)
Computer Science	29	165 (5.72)	21K (739)
Economics	28	164 (5.72)	24K (841)
Education	14	145 (10.38)	27K (1912)
English	74	291 (3.94)	43K (584)
History	104	1015 (9.76)	172K (1650)
Mathematics	31	106 (3.42)	16K (525)
Total	320	2197 (6.58)	346K (1020)

The difficulty of the instructional videos, of which the majority (95%) were presented as scripted monologues (see Appendix A for examples of video lectures), were judged by English as Foreign Language (EFL) learners through a rigorous process of rating.

Participants and Procedures

Following a university IRB ethics approval, 322 Saudi male learners of English ($M_{\_age}$ = 20.61 years, *SD* = .92) were asked to assess the difficulty of 320 instructional videos. English classes were part of a compulsory, first-year program in which students take preparatory courses before enrolling in their major. Placement in the English courses was determined by the Cambridge Language Placement test, and the students' overall proficiency was found to be at the B1 level of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001). The volunteer participants were randomly assigned two sets of five videos and asked to watch and then rate the difficulty of each video. The participants watched the videos with no subtitles or captions, and they were allowed to take a break between the two sets.

In this study, we adapted a five-question Likert-scale that was originally developed by Yoon et al. (2016) for assessing difficulty in spoken texts. The use of a rating scale is a common practice in readability and listenability research (e.g., Kotani et al., 2014; Yoon et al.). Items to judge difficulty were adapted to suit our purposes and were translated into the participants' native language (see Appendix B for English and Arabic versions of the scale). The modified scale contains five items that ask the participants to rate their overall understanding and remembering of video content, lexical comprehension, speech rate, and the difficulty of video lectures in relation to their language ability. The scale items were designed to be combined into a single composite score (our dependent variable) during analysis.

The modified scale was piloted with 24 non-participating students using a sample of 28 videos. The participants of the pilot study were drawn from the same population from which the participants for the main study were recruited. After watching each video, they were asked to rate the difficulty of the videos. Bivariate correlation analyses were then conducted on the participants' responses. The results indicated that all five questions significantly correlated with each other, with Pearson values ranging from $r = .33$ to .80 ($p < .001$). The moderate-to-strong inter-correlations among the instrument items suggest that one construct is being measured: the overall perception of video difficulty. Furthermore, a Cronbach's alpha analysis was conducted to assess the reliability of the rating instrument. The result showed that there is a satisfactory internal consistency among the instrument items ($\alpha = .81$) above the suggested benchmark value of .70 (Nunnally, 1978). Post hoc analysis on all participants' responses showed a higher reliability score ($n = 2826$; $\alpha = .86$).

Extracting and Computing Lexical Complexity

To extract and measure the indices of lexical complexity, videos were first transcribed using Microsoft's Azure Cognitive Services (Microsoft, 2021), and the automatic transcriptions were manually checked. Mistakes were then corrected where needed. Next, the official Stanford NLP Python package (Qi et al., 2019) was used to lemmatize, tokenize, and extract POS tags from the transcripts. A Python script was developed to compute indices related to lexical frequency and density. Finally, the Tool for the Automatic Analysis of Lexical Sophistication (TAALED; Kyle et al., 2021) was used to compute indices related to lexical diversity. In total, we extracted and computed 75 indices related to lexical sophistication, diversity, and density (see Appendix C for a description of each index).

Lexical Frequency and Sophistication Indices

A Python script was developed to calculate the proportions of all words (AW), function words (FW), and content words (CW) belonging to the most frequent word families in the K1, K2, K3, and K4 word family bands, off-list words, and words within the 2K, 3K, 4K, 5K, 6K, 7K, and 8K frequency-bands of the BNC-COCA lists (Nation, 2012). We also calculated the proportion of academic words in the transcripts that appear in the Academic Word List (AWL; Coxhead, 2000), Academic Vocabulary List (AVL; Gardner & Davies, 2014), and Academic Spoken Word List (ASWL; Dang et al., 2017). Lexical sophistication was operationalized as lexical items that are not within the first 2000 most frequent lemmas of the BNC-COCA word lists (Lu, 2012).

Lexical Diversity Indices

Approaches to estimate lexical diversity regardless of text length have flourished. In this study, the advice of McCarthy and Jarvis (2010) who recommend the use of multiple indices to better capture all aspects of lexical diversity was followed. In particular, we used TAALED to measure lexical diversity using conventional indices—for example, TTR, Root TTR, Corrected TTR, and logarithmic TTR—and more advanced indices—for example, Measure of Textual Lexical Diversity (MLTD; McCarthy, 2005), HD-D (McCarthy & Jarvis, 2007), Moving-Average TTR (MATTR; Covington & McFall, 2010), and Mass Index (Maas, 1972). These indices were computed for AWs, CWs, and FWs.

Lexical Density Indices

Our calculations of lexical density indices were based on the proportion of content words (CW) to running words in the transcripts. Following Lu (2012), we operationalized content words as nouns; adjectives; verbs (excluding modal verbs, auxiliary verbs, *be*, and *have*); and adverbs with an adjectival base, including those that can function as both an adjective and adverb (e.g., *fast*) and those formed by attaching the *–ly* suffix to an adjectival root (e.g., *particularly*). The proportions of nouns (including proper nouns) and verbs to all running words were also calculated.

Data Analysis

After extracting all lexical complexity features, we checked each feature for normality using histograms and removed features that had low variance. To assess the strength of association between each lexical complexity feature and the participant judgments, we conducted Pearson correlation analysis after scaling the features in our dataset. The results showed that 45 lexical complexity features had a significant correlation with video difficulty above the threshold of $r = .25$ (Plonsky & Oswald, 2014). To prepare the dataset for regression modeling, we shuffled the entire dataset and created both training and testing sets through a 67/33 split approach (Han et al., 2011); the training set then consisted of 214 videos, and the testing set included 106 videos.

To reduce the number of features and avoid overfitting our regression models, we applied the least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996) on the training set. In statistics and machine learning, LASSO is commonly used as a variable selection technique because it penalizes the coefficients of the regression variables and eliminates irrelevant variables that are not associated with the response variable. Through tuning the penalty parameter λ, the strength of the penalty can be adjusted. As the parameter λ increases, more coefficients are shrunk to zero and eliminated (Fonti, 2017). Using the training set, we tuned the parameter λ for the LASSO model using a 10-fold cross-validation technique. Finally, the features selected by LASSO were used as an input in a regression model. The best regression model was then applied to the testing set to determine its generalizability to unseen videos. The prediction performance of the trained model was evaluated based on the Pearson correlation coefficients (between true and predicted video difficulty scores) and Root Mean Square Error (RMSE). One advantage of RMSE is that units of the RMSE are the same as the original units of the target value that is being predicted. The Python machine learning library Scikit-learn (Pedregosa et al., 2012) was used for model building and evaluation.

Results

RQ1: How Do Different Measures of Lexical Sophistication, Density, and Diversity Correlate with Perceived Video Difficulty?

To answer the first research question, correlation analysis was performed (see Table 2). The results showed that 45 lexical complexity features have a significant relationship with the participants' ratings as the strength of association ranges from $r = .26$ to .64.

Table 2

Highly Correlated Lexical Complexity Indices with Participants' Ratings of Instructional Video Difficulty

	Feature	M	SD	*r*
Lexical Frequency	BNC-COCA K3 (CW)	7.45	3.76	.64
	BNC-COCA K1 (CW)	68.26	10.32	-.60
	BNC-COCA K1-3 (CW)	63.93	6.08	.52
	Academic Word List (AW)	7.96	3.33	.51
Lexical Density	Lexical Density (Type)	.65	.08	.55
	Lexical Density (Token)	.43	.04	.26
Lexical Diversity	MASS TTR (AW)	.06	.01	-.51
	MATTR (CW)	.74	.10	.53
	HD-D (CW)	.82	.12	.46
	Noun Variation	.21	.05	.39

Note. Correlation is significant at the.001 level (2-tailed).

Results in Table 3 show notable correlations between some aspects of lexical complexity and difficulty rating of instructional videos. More specifically, the lexical frequency of K1 items was strongly, but negatively, correlated with difficulty ($r = -.60$, $p < .001$), whereas the lexical frequency of K3 items showed a strong association with the rating of video difficulty ($r = .64$, $p < .001$). These results suggest that the higher the proportion of K1 words in the videos the easier the participants rated them, and the higher the proportion of K3 words the more difficult videos were rated. Additionally, content words in K1 to K3 appeared to moderately correlate with video difficulty rating ($r = .52$, $p < .001$). A moderate correlation ($r = .51$, $p < .001$) was also found between academic words and the difficulty of the videos.

Regarding lexical density, the results, as expected, indicated that the presence of different words (type) in a text contributes to video difficulty. A significant correlation is observed between lexical density (type) and video difficulty ($r = .55$, $p < .001$). Concerning lexical diversity, the MAAS TTR measure showed a negative significant correlation of academic words with text difficulty ($r = -.51$, $p < .001$). The MATTR measure, on the other hand, indicated a positive significant correlation between content words and video difficulty ($r = .53$, $p < .001$). Similarly, the HD-D measure showed that content words are also associated with difficulty ($r = .46$, $p < .001$). Finally, the results show that the noun variation feature of lexical diversity correlated with video difficulty ($r = .39$, $p < .001$). In sum, lexical complexity, collectively, appears to be a determinant factor in the perception of video difficulty.

RQ2: To What Extent Can Measures of Lexical Sophistication, Diversity, and Density Explain Perceived Video Difficulty?

To answer our second research question, multiple regression analyses were conducted to explore if the selected lexical complexity features could significantly explain the variance in the participants' ratings of video difficulty. After removing non-significant features (e.g., features with coefficients exceeding the common alpha level of .05) from the training set, the regression model yielded a significant model, $F(3, 210) = 52.70$, $p < .001$, $r = .66$, $R^2 = .43$. The model included three lexical complexity features: *BNC-COCA K3 (CW)*, *Verb Variation*, and *MASS TTR (FW)*. Collectively, these three features explain 43% of the variance in the participants' judgments of video difficulty. Table 3 shows the beta weights for each index in the training model.

Table 3

Results of Multiple Regression Analysis Using Lexical Complexity Indices as Predictors of L2 Video Difficulty

Predictors	***B***	**SE**	**t**	**Sig.**
BNC-COCA K3 (CW)	1.988	.205	9.674	.000
Verb Variation	-.470	.208	-2.264	.025
MASS TTR (FW)	-.875	.212	-4.134	.000

Note. Constant term = 16.850.

To investigate how lexical sophistication, density, and diversity each contribute to video difficulty, three regression analyses were conducted between the participants' rating of difficulty and significant indices from each of the three dimensions of lexical complexity. The regression analyses yielded a significant lexical frequency model, $F(2, 211) = 70.66$, $p < .001$, $R^2 = .40$, that included two indices: *BNC-COCA K3 (CW)* and *ASWL Level 1*. Four lexical diversity indices were included in the lexical diversity model, $F(4, 209) = 30.15$, $p < .001$ to show that they collectively explain 37% of the variance in the rating of difficulty. Finally, the diversity model was significant $F(1, 212) = 85.20$, $p < .001$. It included a single predictor, *Lexical Density (Type)*, which explained 29% of video difficulty. Additional information about these models, including the beta weights for each predictor used in the training regression models, are shown in Table 4.

Table 4

Summary of Different Lexical Complexity Models and Their Predictors

Predictors	***B***	**SE**	**t**	**Sig.**
Lexical Frequency Model				
BNC-COCA K3 (CW)	2.039	.217	9.399	.001
ASWL Level 1	-.580	.217	-2.676	.008
Lexical Density Model				
Lexical Density (Type)	1.980	.214	9.231	.001
Lexical Diversity Model				
Verb Variation	-.840	.218	-3.854	.001
Corrected Verb	.935	.287	3.265	.001
MATTR 50 (AW)	1.965	.352	5.584	.001
MSTTR 50 (FW)	-0.984	.277	-3.554	.001

Table 5

A Comparison of the Prediction Performances of the Four Regression Models on the Testing Set

Regression Models	Training Set			Testing Set		
	r	R^2	RMSE	*r*	R^2	RMSE
Lexical Frequency Model	.63	.40	2.86	.68	.45	2.60
Lexical Density Model	.54	.29	2.75	.59	.33	2.88
Lexical Diversity Model	.60	.37	2.94	.62	.33	2.92
Lexical Complexity Model (All)	.66	.43	.279	.65	.41	2.71

RQ3: How Well Do Lexical Complexity Regression Models Predict Video Difficulty?

To investigate the generalizability of our regression model, the regression beta coefficients and constant term from the training model were used to predict video difficulty in the test set (*n* = 106 video texts). The findings show that the regression model yielded a correlation of .65 and RMSE of 2.71 and that a collection of selected verbal complexity features explained 41% of the variance in judgments of difficulty. We also developed three regression models using lexical frequency, diversity, and density indices.

The fitted lexical frequency, density, diversity, and lexical complexity models were then used to predict the difficulty of videos in an unseen (held-out) testing set.

As presented in Table 5, the lexical frequency model outperformed the other models in predicting the participants' rating of difficulty in videos.

Discussion

Clearly, the use of videos is increasing throughout CALL research, teaching, and assessment designs. At present, however, researchers lack an empirical basis on which to establish video difficulty in line with indices for printed texts that have long benefitted reading research. As part of a long-term development effort, the purpose of the current study was to investigate whether the use of existing lexical complexity measures can be extended to assess difficulty in L2 videos. Specifically, the study examined the relationship between lexical sophistication, density, and diversity and perceived video difficulty of 320 pedagogical videos (346,000 tokens) that were rated by 322 B1 level participants. The results of the first research question, which targeted the relationship between lexical frequency, diversity, and density indices and video difficulty, showed that many of the lexical complexity features are significantly correlated with video difficulty, meeting or exceeding the *r*-value criterion of ± .10. In particular, lexical diversity had the largest number of indices (*n* = 13) to have moderate correlation, meeting or exceeding the *r*-value criterion of ± .40, with video difficulty, followed by lexical frequency (*n* = 12), and lexical density (*n* = 2). These results support findings from some previous studies (e.g., Brunfaut & Révész, 2015; Révész & Brunfaut, 2013) that lexical complexity is associated with L2 spoken text difficulty.

The second research question queried the extent to which measures of lexical frequency, diversity, and density can explain the difficulty in L2 videos. To answer this question, three regression models were developed using indices from each of three aspects of lexical complexity. The predictive performances of these models were evaluated on a held-out testing set. The results showed that the regression models yielded slightly different outcomes, with the lexical frequency model having the highest explanatory power (R^2 = .45). Two indices were included in the lexical frequency model: K3 content words (a generic frequency index) and ASWL level 1 (an academic frequency index). Our results indicated that the higher the proportion of K3 content words in videos, the more difficult the videos were perceived by our participants. This finding is in line with word frequency models (e.g., Milton, 2009), where lexical frequency plays an important role in a text difficulty and perceived video difficulty (Alghamdi et al, 2022). This finding accords with previous results suggesting that knowledge of the most frequent 2,000–3,000 words is required for effective comprehension of spoken discourse (Matthews, 2018; Matthews & Cheng, 2015; van Zeeland & Schmitt, 2013). Recently, Matthews and Cheng (2015) found knowledge of words in the 2,001–3,000 frequency range alone was able to predict 52% of the variance in L2 listening comprehension scores. As for the second feature, the presence of words from the first level of the ASWL in videos seems to facilitate video understanding. This level contains general high-frequency words from the most frequent 1,000 BNC/COCA word families (e.g., *alright*, *know*, and *ask*).

Finally, the third question investigated the generalizability of lexical complexity regression models on unseen or held-out videos. The results demonstrated overall that there are slight differences between the regression models' performances in training vs. testing sets. A slight decrease was expected due to the small size of the test set ($n = 106$) compared to the training set ($n = 214$). Secondly, the lexical frequency model was the most predictive, outperforming a model with lexical frequency, density, and diversity features included. This finding suggests the superiority of lexical frequency for predicting video difficulty. In line with Schmitt and Schmitt (2014), our results suggest that knowledge of words beyond the most frequent 2,000-word level is needed to attain an adequate degree of text understanding.

In summary, we conclude that lexical complexity indices help to explain and predict video difficulty. Particularly, lexical frequency indices, such as the proportion of content words from the 2,000–3,000-word families and academic spoken words, were found to explain a reasonable variance of intermediate language learners' perception of difficulty. Our results accord with earlier findings that suggest that learners must develop aural knowledge of words up to and including those within the K3 level to comprehend spoken materials (Matthews, 2018).

Limitations and Directions for Future Research

The present study has limitations. Because video difficulty was modeled solely as a function of lexical complexity, a key limitation of our work is that the relative contributions of spoken language, such as acoustic, phonological, syntactic, discoursal, and visual complexity features, were not considered in the present study. Additionally, our results may be limited by the fact that our participants were all male students, and their native language was Arabic. Future studies may seek to include different language backgrounds, genders, and ages to offer further insights into the relationship between lexical complexity and video difficulty. Moreover, the participants in the present study were at the B1 level; thus, to overcome this limitation, informants from various proficiency levels need to be addressed in future research. Further, although our corpus contains various topics and production styles, the videos that we selected may not be representative of the instructional video genre. Though lexical complexity is crucial, we have yet to explore other factors that CALL instructors use in selecting videos for pedagogical purposes. Further work may progress through greater alignment with contemporary SLA research, particularly with a focus on multimodality, to bring together a coherent view of the role of video complexity in relation to the ongoing development of CALL materials, teaching, and assessments (Douglas Fir Group, 2016). One step towards this goal is to harness the recent advances in artificial intelligence, for example, through developing and utilizing more sophisticated video analysis tools (Alghamdi et al, 2021).

Conclusion

As audio-visual materials become a valuable and widely used source in educational contexts, it is particularly important to develop methods through which we can determine the level of difficulty in such materials for use in language teaching, learning, and research. In this study, we examined the relative contribution of a wide range of lexical frequency, density, and diversity indices to video difficulty using a corpus of 320 instructional videos. The findings of regression analyses showed that the three dimensions of lexical complexity slightly vary in their explanatory and predictive powers, with a frequency-based regression model yielding the best result in explaining and predicting difficulty in the videos. In particular, the lexical frequency model predicted about 45% of the variance in video difficulty, outperforming regression models developed using lexical density indices (R^2 = .33) and lexical diversity indices (R^2 =.33) as well as models developed using a set of indices from all three dimensions of lexical complexity (R^2 =.41). The notable predictive value of lexical frequency to the model provides evidence of the importance and utility of lexical frequency lists as an indicator of difficulty in pedagogical videos that are needed to foster CALL listening materials, teaching practices, and test development.

Acknowledgements

This research work was funded by Institutional Fund Projects under grant no. (IFPFP-265-22). Therefore, the authors gratefully acknowledge technical and financial support from Ministry of Education and Deanship of Scientific Research (DSR), King Abdulaziz University (KAU), Jeddah, Saudi Arabia.

References

Aldukhayel, D. (2021). The effects of captions on L2 learners' comprehension of vlogs. *Language Learning & Technology*, *25*(2), 178–191. http://hdl.handle.net/10125/73439

Alghamdi, E. (2021). *Automated video difficulty assessment* [Unpublished doctoral thesis]. University of Melbourne.

Alghamdi, E., Velloso, E., & Gruba, P. (2021). AUVANA: An automated video analysis tool for visual complexity. https://doi.org/10.31219/osf.io/kj9hx

Alghamdi, E., Gruba, P., & Velloso, E. (2022). The relative contribution of verbal complexity to second language video lectures difficulty assessment. *The Modern Language Journal, 106*(2), 393–410. https://doi.org/10.1111/modl.12773

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult?* The University of Maryland. https://apps.dtic.mil/sti/pdfs/ADA550176.pdf

Brunfaut, T., & Révész, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, *49*(1), 141–168. https://doi.org/10.1002/tesq.168

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, *15*(2), 119–157. https://doi.org/10.1177/026553229801500201

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing, 35*(1), 3–25. https://doi.org/10.1177/0265532216676851

Chotlos, J. W. (1944). IV. A statistical and comparative analysis of individual written language samples. *Psychological Monographs*, *56*(2), 75–111. https://doi.org/10.1037/h0093511

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Press Syndicate of the University of Cambridge.

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

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Cross, J. (2011). Comprehending news videotexts: The influence of the visual content. *Language Learning & Technology*, *15*(2), 44–68. http://dx.doi.org/10125/44251

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, *42*(3), 475–493. https://www.jstor.org/stable/40264479

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

Dale, E., & Tyler, R. W. (1934). A study of the factors influencing the difficulty of reading materials for adults of limited reading ability. *The Library Quarterly*, *4*(3), 384–412. https://www.jstor.org/stable/4302090

Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge University Press.

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959–997. https://doi.org/10.1111/lang.12253

De Clercq, B. (2015). The development of lexical complexity in second language acquisition: A cross-linguistic study of L2 French and English. *EUROSLA Yearbook*, *15*(1), 69–94. https://doi.org/10.1177/0267658316674506

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *Modern Language Journal, 100*, 19–47. https://doi.org/10.1111/modl.12301

Ellis, N. (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

Fonti, V. (2017). *Feature selection using LASSO*. *VU Amsterdam research paper in business analytics*, *30*, 1–25

Fulcher, G. (1997). Text difficulty and accessibility: Reading formulae and expert judgement. *System*, *25*(4), 497–513. https://doi.org/10.1016/S0346-251X(97)00048-1

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3

Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, *65*(S1), 228–255. https://doi.org/10.1111/lang.12119

Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 learning. *Language Learning & Technology*, *10*(2), 77–92. http://dx.doi.org/10125/44062

Halliday, M. A. K., & Hasan, R. (1985). *Language, context and text: Aspects of language in a social semiotic perspective*. Deakin University Press.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Elsevier.

Hashimoto, B. J., & Egbert, J. (2019). More than frequency? Exploring predictors of word difficulty for second language learners. *Language Learning*, *69*(4), 839–872. https://doi.org/10.1111/lang.12353

Heaps, H. (1978). *Information retrieval: Computational and theoretical aspects*. Academic Press.

Housen, A. (2014). Difficulty and complexity of language features and second language instruction. *The encyclopedia of applied linguistics* (pp.1–7). https://doi.org/10.1002/9781405198431.wbeal1443

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, *38*(2), 163–175. https://doi.org/10.1017/S0272263116000176

Johansson, V. (2009). Lexical diversity and lexical density in speech and writing. *Lund University, Department of Linguistics and Phonetics*, *53*.

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

Kincaid, J. P., Fishburne Jr., R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel*. Naval Technical Training Command, Millington TN Research Branch.

Kirby, J. R. (1993). Collaborative and competitive effects of verbal and spatial processes. *Learning and Instruction*, *3*(3), 201–214. https://doi.org/10.1016/0959-4752(93)90004-J

Kotani, K., Ueda, S., Yoshimi, T., & Nanjo, H. (2014). A listenability measuring method for an adaptive computer-assisted language learning and teaching system. *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing* (pp. 387–394).

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). 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

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

Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study. *The Modern Language Journal, 102*(3), 557–577. https://doi.org/10.1111/modl.12503

Lively, B. A., & Pressey, S. L. (1923). A method for measuring the 'vocabulary burden' of textbooks. *Educational Administration and Supervision*, *9*, 389–398.

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

Maas, H. D. (1972). Über den Zusammenhang zwischen Wortschatzumfang und Länge eines Textes [On the relation between vocabulary size and length of a text]. *Zeitschrift für Literaturwissenschaft und Linguistik*, *8*, 73–79.

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, *72*, 23–36. https://doi.org/10.1016/j.system.2017.10.005

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, *52*, 1–13. https://doi.org/10.1016/j.system.2015.04.015

Mayer, R. E. (2009). *Multimedia learning*. Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (Unpublished doctoral dissertation, The University of Memphis).

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

Mesmer, H., Cunningham, J., & Hiebert, E. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading Research Quarterly*, *47*(3), 235–258. https://www.jstor.org/stable/43497518

Microsoft. (2021). *Azure cognitive services* [Computer software]. https://azure.microsoft.com/en-au/services/cognitive-services/

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacon-Beltran, C. Abello-Contesse, M. Torreblanca-Lopez, & M. Lopez-Jimenez (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83–97). Multilingual Matters.

Muljani, D., Koda, K., & Moates, D. R. (1998). The development of word recognition in a second language. *Applied Psycholinguistics*, *19*, 99–113. https://doi.org/10.1017/S0142716400010602

Nation, I. S. P. (2012, September 17). *The BNC/COCA word family lists*. Unpublished paper. http://www.victoria.ac.nz/lals/about/staff/paul-nation.aspx

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.1007/s13398-014-0173-7.2

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Qi, P., Dozat, T., Zhang, Y., & Manning, C. D. (2019). Universal dependency parsing from scratch. *Proceedings of the CoNLL 2018 Shared Task: Multilingual parsing from raw text to universal dependencies*, 160–170. http://arxiv.org/abs/1901.10457

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

Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, *35*(1), 31–65. https://doi.org/10.1017/S0272263112000678

Rodgers, M. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL-International Journal of Applied Linguistics*, *169*(1), 191–211. https://doi.org/10.1075/itl.00012.rod

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Longman.

Rupp, A. A., Garcia, P., & Jamieson, J. (2001). Combining multiple regression and CART to understand difficulty in second language reading and listening comprehension test items. *International Journal of Testing, 1*, 185–216. https://doi.org/10.1080/15305058.2001.9669470

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484-503. https://doi.org/10.1017/S0261444812000018

Suzuki, Y., Nakata, T., & Dekeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *Modern Language Journal,103*(3), 713–720. https://doi.org/10.1111/modl.12585

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 26–288. https://www.jstor.org/stable/2346178

Templin, M. (1957). *Certain language skills in children: Their development and interrelationships*. The University of Minnesota Press.

Teng, M. F., & Teng, F. (2020). Vocabulary learning through videos: Captions, advance-organizer strategy, and their combination. *Computer Assisted Language Learning*, 1–33. https://doi.org/10.1080/09588221.2020.1720253

Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? Measures in lexical richness in perspective. *Computers and the Humanities*, *32*(5), 323–352. https://doi.org/10.1023/A:1001749303137

Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics: Selected papers of the second international congress of applied linguistics, Cambridge 19*. Cambridge University Press.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407–427. https://doi.org/10.1093/applin/amp010

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65–86. http://dx.doi.org/10125/44203

Wong, K. M., & Samudra, P. G. (2019). L2 vocabulary learning from educational media: extending dual-coding theory to dual-language learners. *Computer Assisted Language Learning*, *34*(8), 1182–1204. https://doi.org/10.1080/09588221.2019.1666150

Yanagawa, K., & Green, A. (2008). To show or not to show: The effects of item stems and answer options on performance on a multiple-choice listening comprehension test. *System*, *36*(1), 107–122. https://doi.org/10.1016/j.system.2007.12.003

Ying-hui, H. (2006). An investigation into the task features affecting EFL listening comprehension test performance. *Asian EFL Journal*, *8*(4), 33–54.

Yoon, S., Cho, Y., & Napolitano, D. (2016). Spoken text difficulty estimation using linguistic features. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications,* (pp. 267–276).

Appendix A. Some Examples of Instructional Videos and Their Difficulty Ratings

Category	Topic	Difficulty	URL
Economics	Investing: Why You Should Diversify	18	shorturl.at/gmrs9
History	World history	22	shorturl.at/cVY12
English	The vowel-shift irregular verb	14	shorturl.at/sY136
Computer Science	A mathematical theory of communication	20	shorturl.at/beNTW

Appendix B. Video Difficulty Rating Scale

How would you rate your understanding of the video?

5) less than 60%
4) 70%
3) 80%
2) 90%
1) 100%

How much of the information in the video can you remember?

5) less than 60%
4) 70%
3) 80%
2) 90%
1) 100%

Estimate the number of words you missed or did not understand

5) more than 10 words
4) 6-10 words
3) 3-5 words
2) 1-2 words
1) none

The speech rate was…

5) fast
4) somewhat fast
3) neither fast nor slow
2) somewhat slow

1) slow

I believe the language of the video is…

5) at much higher level than my language ability

4) at higher level than my language ability

3) neither high nor low

2) at lower level than my language ability

1) at much lower level than my l language ability

Appendix C. Summary of Lexical Complexity Measures Used in the Study

Indices	Type	Formula
BNC-COCA K1 AW	LS	The proportion of words in the text that are among the 1,000 most frequent words in the BNC-COCA list
BNC-COCA k2 AW	LS	The proportion of words in the text that are among the 2,000 most frequent words in the BNC-COCA list
BNC-COCA k3 AW	LS	The proportion of words in the text that are among the 3,000 most frequent words in the BNC-COCA list
BNC-COCA k1-2 AW	LS	The proportion of words in the text that are among the 2,000 most frequent words in the BNC-COCA list
BNC-COCA k1-3 AW	LS	The proportion of words in the text that are among the 3,000 most frequent words in the BNC-COCA list
BNC-COCA k1-4 AW	LS	The proportion of words in the text that are among the 4,000 most frequent words in the BNC-COCA list
BNC-COCA k1-5 AW	LS	The proportion of words in the text that are among the 5,000 most frequent words in the BNC-COCA list
BNC-COCA k1-6 AW	LS	The proportion of words in text that are among the 6,000 most frequent words in the BNC-COCA list
BNC-COCA k1-7 AW	LS	The proportion of words in text that are among the 7,000 most frequent words in the BNC-COCA list
BNC-COCA k1-8 AW	LS	The proportion of words in text that are among the 8,000 most frequent words in the BNC-COCA list
BNC-COCA k1 CW	LS	The proportion of content words in the text that are among the 1,000 most frequent words in the BNC-COCA list
BNC-COCA k2 CW	LS	The proportion of content words in the text that are among the 2,000 most frequent words in the BNC-COCA list
BNC-COCA k3 CW	LS	The proportion of content words in the text that are among the 3,000 most frequent words in the BNC-COCA list
BNC-COCA k1-2 CW	LS	The proportion of content words in the text that are among the 2,000 most frequent words in BNC-COCA list
BNC-COCA k1-3 CW	LS	The proportion of content words in the text that are among the 3,000 most frequent words in the BNC-COCA list
BNC-COCA k1-4 CW	LS	The proportion of content words in the text that are among the 4,000 most frequent words in BNC-COCA list
AWL	LS	Number of words in text that are in the AWL list
AWL Sublist_1	LS	Number of words in text that are in the AWL Sublist_1
AWL Sublist_2	LS	Number of words in text that are in the AWL Sublist_2
AWL Sublist_3	LS	Number of words in text that are in the AWL Sublist_3
AWL Sublist_4	LS	Number of words in text that are in the AWL Sublist_4
AWL Sublist_5	LS	Number of words in text that are in the AWL Sublist_5
AWL Sublist_6	LS	Number of words in text that are in the AWL Sublist_6
AWL Sublist_7	LS	Number of words in text that are in the AWL Sublist_7
AWL Sublist_8	LS	Number of words in text that are in the AWL Sublist_8
AWL Sublist_9	LS	Number of words in text that are in the AWL Sublist_9
AWL Sublist_10	LS	Number of words in text that are in the AWL Sublist_10
ASWL_1	LS	Number of words in text that are in the ASWL Sublist_1

ASWL_2	LS	Number of words in text that are in the ASWL Sublist_2
ASWL_3	LS	Number of words in text that are in the ASWL Sublist_3
ASWL_4	LS	Number of words in text that are in the ASWL Sublist_4
LS (CW)	LS	Number of content words in text that are not in the first 2000 most frequent BNC-COCA lists of lemmatized word families
LS_2 (CW)	LS	Number of content words in text that are not in the first 2000 most frequent BNC-COCA lists of lemmatized word families divided by number of types.
LS Noun	LS	Number of nouns in text that are not in the first 2000 most frequent BNC-COCA lists of lemmatized word families
LS Verb	LS	Number of verbs in text that are not in the first 2000 most frequent BNC-COCA lists of lemmatized word families
Lexical Density (Type)	LDen	Percentage of types that are content words
Lexical Density (Token)	LDen	Percentage of tokens that are content words
Verb Variation	LDiv.	Number of verb (types) divided by number of tokens
Adjective Variation	LDiv.	Number of adjective (types) divided by number of tokens
Adverb Variation	LDiv.	Number of adverbs (types) divided by number of tokens
Noun Variation	LDiv.	Number of nouns (types) divided by number of tokens
Noun TTR	LDiv.	Type-token ratio for nouns
Verb TTR	LDiv.	Type-token ratio for verbs
Adj TTR	LDiv.	Type-token ratio for adjectives
TTR (AW)	LDiv.	Type-token ratio for all words
TTR (CW)	LDiv.	Type-token ratio for content words
TTR (FW)	LDiv.	Type-token ratio for function words
Root TTR (AW)	LDiv.	Number of types divided by the square root of the number of tokens
Root TTR (CW)	LDiv.	Number of types divided by the square root of the number of CW tokens
Root TTR (FW)	LDiv.	Number of types divided by the square root of the number of FW tokens
Log TTR (CW)	LDiv.	Log of number of types divided by the log of the number of CW tokens.
Log TTR (FW)	LDiv.	Log of number of types divided by the log of the number of FW tokens.
Log TTR (AW)	LDiv	Log of number of types divided by the log of the number of tokens.
Root TTR (verb)	LDiv.	Number of verb types divided by the square root of the number of tokens
Corrected Verb	LDiv.	Number of verb types divided by the square root of 2 times the number of tokens
Mass TTR (AW)	LDiv.	A transformation of TTR that attempts to fit the value to a logarithmic curve.
Mass TTR (CW)	LDiv.	More complex transformation of TTR (content words only) that attempts to fit the value to a logarithmic curve.
Mass TTR (FW)	LDiv	More complex transformation of TTR (function words only) that attempts to fit the value to a logarithmic curve.
MATTR-50 (AW)	LDiv.	Moving average TTR (50-word window)

MATTR-50 (CW)	LDiv.	Moving average TTR for CW (50-word window)
MATTR-50 (FW)	LDiv.	Moving average TTR for FW (50-word window)
MSTTR-50 (AW)	LDiv.	Mean segmental TTR for CW (50-word non overlapping segments)
MSTTR-50 (CW)	LDiv.	Mean segmental type-token ratio for content words (50-word non-overlapping segments)
MSTTR-50 (FW)	LDiv.	Mean segmental type token ratio for function words (50-word non overlapping segments)
HD-D-42 (AW)	LDiv.	HD-D uses the hypergeometric distribution to calculate the probability of encountering one of its tokens in a random sample of 42 tokens.
HD-D-42 (CW)	LDiv.	HD-D uses the hypergeometric distribution to calculate the probability of encountering one of its tokens in a random sample of 42 tokens (content words only).
HD-D-42 (FW)	LDiv.	HD-D uses the hypergeometric distribution to calculate the probability of encountering one of its tokens in a random sample of 42 tokens (function words only).
MTLD Original (AW)	LDiv.	MTLD is based on the average number of tokens it takes to reach a given TTR value (.720).
MTLD Original (CW)	LDiv.	the average number of content words it takes to reach a given TTR value (.720).
MTLD Original (FW)	LDiv.	the average number of function words it takes to reach a given TTR value (.720).
MTLD MA Wrap (AW)	LDiv.	A version of MTLD (content words only) that takes a moving-average approach to calculate the index. The final factor is calculated by wrapping back to the beginning of the text.
MTLD MA Wrap (CW)	LDiv.	A version of MTLD (content words only) that takes a moving-average approach to calculate the index. The final factor is calculated by wrapping back to the beginning of the text.
MTLD MA Wrap (FW)	LDiv.	A version of MTLD (content words only) that takes a moving-average approach to calculate the index. The final factor is calculated by wrapping back to the beginning of the text.

Note. Descriptions of the lexical diversity indices were taken from the website of TAALED at https://www.linguisticanalysistools.org/taaled.html.

About the Authors

Doctor Emad Alghamdi is Assistant Professor of Computational Linguistics at King Abdulaziz University. He is the founding director of the AI in Language Learning & Assessment Lab. He obtained his PhD from the University of Melbourne where he received the 2022 Chancellor's Prize for Excellence in the PhD Thesis. His research interests lie at the intersection of applied linguistics, human-computer interaction (HCI), and AI. Alghamdi is passionate about developing innovative technologies for language learning, assessment, and research. His recent projects are the Automated Video Analysis (AUVANA) software for extracting, computing, and visualizing visual complexity in videos. He has published numerous studies and book chapters in top venues such as *The Modern Language Journal*, *Language Learning & Technology*, and the *International Journal of Computer-Assisted Language Learning and Teaching* (*IJCALLT*).

E-mail: eaalghamdi@kau.edu.sa

Doctor Paul Gruba is Associate Professor in the School of Languages and Linguistics, University of Melbourne. Originally trained as a journalist, Paul taught ESL in Mali, Japan, and the United States before earning a PhD in second language listening. His research areas include blended learning, language assessment, and program evaluation.

E-mail: p.gruba@unimelb.edu.au

Doctor Ahmed Masrai is Associate Professor of Applied Linguistics at the Institute of Research and Consulting Services at Prince Sattam Bin Abdulaziz University, Saudi Arabia. His research and teaching interests include second language acquisition, lexical studies, and vocabulary testing. This interest has led to publication in a number of prestigious journals and contribution to edited volumes.

E-mail: a.masrai@hotmail.com

Doctor Eduardo Velloso is Senior Lecturer in Human-Computer Interaction at the School of Computing and Information Systems, University of Melbourne. He is interested in the user experiences enabled by novel sensing modalities, the design of new interaction devices and techniques, and the use of AI in interactive systems.

E-mail: eduardo.velloso@unimelb.edu.au