International Journal of Linguistics and Translation Studies
Volume 2, Issue 2, 2021
Homepage: http://ijlts.org/index.php/ijlts/index
DOI: https://doi.org/10.36892/ijlts.v2i2.131

Word frequency, Range and Lexical diversity: Picking out Changes in Lexical Proficiency among University Learners in an EFL Context

Brahim Ait Hammou
Moulay Ismail University – Meknes
hammou76@gmail.com
Mohammed Larouz
Moulay Ismail University – Meknes
m.larouz@umi.ac.ma
Mustapha Fagroud
Moulay Ismail University – Meknes
mfagroud@hotmail.com

ARTICLE HISTORY

Received: 28 /02/ 2021

Accepted: 13/ 04 /2021

KEYWORDS
lexical knowledge, sophistication, diversity, range, EFL

Abstract
While frequency of occurrence and lexical diversity are common measures of lexical proficiency, very few studies have examined range along with these two other lexical indices, especially in EFL contexts. Using a sample of 529 undergraduate EFL learners' essays, the present study examined the development of three indices of lexical proficiency (word frequency, range and lexical diversity) across 3 different proficiency levels. No similar study has been conducted so far in the Moroccan context, which makes this study a valuable contribution to a better understanding of university learners' productive lexical knowledge. Unexpectedly, although frequency is an established measure which can successfully show differences in learners' proficiency, the results of this study indicate that at the productive level, the range of content words and lexical diversity might better reflect differences among learners of different proficiency levels. While word frequency was stable among the three groups, indices of range and lexical diversity reflected better lexical knowledge in higher proficiency levels. We concluded that in EFL contexts, range and diversity might be better measures of proficiency. This suggests that there should be more focus on developing these two aspects in EFL classes.

1. INTRODUCTION

1.1. LITERATURE REVIEW

1.1.1. Lexical Sophistication

Lexical sophistication is an aspect of lexical proficiency. Milton (2009, p.131) defines lexical sophistication as a calculation of the proportion of infrequent words in a text. Read (2000) uses vocabulary 'rareness' to refer to lexical sophistication. He defines lexical sophistication as a selection of low-frequency words that are appropriate to the topic and style of the writing. Similarly, Kim, et al (2018, p.2) use lexical sophistication to 'refer to the learners' use of sophisticated and advanced words'.

There are two main procedures to compute the frequency of words in a text. Crossley, et al., (2013) distinguish between band-based and count-based indices of lexical sophistication.

Band-based indices calculate frequency based on bands of 1000 words each. One of the earliest band-based measures of lexical sophistication is Laufer and Nation's (1995) Lexical Frequency Profile (LFP). The profiler measures productive vocabulary use in writing by analyzing the vocabulary deployed in a text in terms of the most frequent and the least frequent vocabulary and it provides percentages of each category. The count-based indices calculate word frequency based on the frequency of word occurrences as they occur in a reference corpus. Unlike the band-based measure which provides proportions of words which belong to frequency bands, the count-based measure provides an average word frequency score for the whole text based on the frequencies of each individual word in the target reference corpus rather than grouping the words into frequency bands.

Crossley, et al., (2013) conclude that count-based indices can pick up small changes in learners' productions compared to band-based measures and they are, therefore, closer to the word's real frequency in the corpus. In their study, Crossley, Cobb and McNamara maintain that the count-based indices showed the highest accuracy and Laufer and Nation's (1995) LFP showed the second highest accuracy in predicting the proficiency level of the learners. The study found that all the frequency indices indicated that as L2 learners advance in proficiency level, they begin to use less frequent words, with the native participants using the least number of frequent words.

McNamara, et al., (2009) showed that word frequency demonstrated the largest difference between high and low-proficiency essays. Similarly, the researchers found that word-frequency is a significant predictor of essay ratings. Comparing L1 with L2 learners' lexical differences in writing, Crossley and McNamara (2009) also concluded that L2 writers significantly used more frequent words in their writing than L1 writers. In another study, Crossley and McNamara (2010) analyzed cohesion and lexical features which could predict writing quality. The researchers concluded that lexical variables, including word frequency, accounted for the vast majority of the variance in students' scores in writing. A similar conclusion was found by Crossley, Kyle, Allen, Guo and McNamara (2014).These results indicate that the use of low-frequency vocabulary is an indicator of higher quality writing.

With the exception the study of Zyad, et al., (2017) which used the traditional LFP of Laufer and Nation (1995), no other study examined the lexical sophistication of Moroccan university students' writing. However, because Zyad et al.'s study used the LFP, their results may not reflect learners' actual productive vocabulary knowledge since Laufer and Nation's (1995) LFP relies on word families and also on an old reference corpus (West's General Service Word List, 1953). Because language is constantly changing, especially at the level of lexis and how words are used, we need to rely on more modern and updated corpora which can better mirror how language is used by today's learners.

1.1.2. Lexical range

In spite of the established tradition in corpus linguistic research wherein frequency has been widely adopted as a tool to measure the development of second language learners' lexical knowledge, there has been many calls highlighting the bias which might result from the sole reliance on word frequency in studying learner's productive lexical knowledge. A word might be really frequent in a certain corpus but its frequency is unbalanced as most of its occurrences are only in one part of the target corpus. The issue we are referring to here is called lexical dispersion, range or contextual diversity in studies of corpus linguistics. Gries (2008) postulated that "the most frequently used statistic in corpus linguistics is the frequency of occurrence of some linguistic variable. […] However, even though this is apparently not recognized much in the field, frequencies of (co-)occurrence may sometimes be incredibly misleading" (pp. 403-404).

Monteiro, et al., (2018, p.4) maintain that "contextual diversity, also known as range and dispersion, is a measure that indicates the number of unique contexts in which linguistic

items appear." This indicates that range is simply a count of the different contexts rather than the different times in a which a word appears. Gries (2020) states that range is the simplest measure of dispersion and it refers to the number of corpus parts in which a word is attested, irrespective of the size of each part. In the present study, we will use *range* and *contextual diversity* for this aspect of lexical knowledge.

Adelman, et al., (2006) believe that a word which is used in a variety of contexts will likely to be a frequent one, which facilitates its learning compared to a word which is less frequent and/or more contextually restricted. This suggests that words which are more restricted in their contextual use are more difficult to learn since they are rarely encountered. This conclusion is outlined by McDonald and Shillcock (2006) who maintain that infrequent words are very rarely encountered as their contextual behavior is more constrained. Citing Ellis (2002a, 2002b), Monteiro, et al., (2018) believe that contextually diverse words are learned faster. Having a large contextual diversity makes encountering the word an easy task for the language learner. In their own study, Perea, et al., (2013) observed that contextual diversity is the main determinant of word identification times in young readers. This study showed that compared to providing a list of words all at once, the number of contexts in which a word occurs facilitates both learning and recognition.

In assessment, Kyle and Crossley (2015) reported that the index of range in their study explained 26% of the variance in the scores which are assigned by human raters to lexical proficiency. Similarly, in another study Kyle and Crossley (2016) showed that range explained 17% of the human scores of the TOEFL independent writing samples. Monteiro, Crossley and Kyle (2018) examined the linguistic variables which contributed to variance in the TOEFL writing scores. The results of the study concluded that the range index of content words was a significant predictor and that this index alone contributed to explaining 17% of the L2 holistic writing score. A similar conclusion was outlined by Guo, et al., (2013) and also by Crossley, et al., (2014). Both studies examined range among other lexical sophistication indices and linguistic micro-features which predict L2 writing proficiency, and they both agreed on the conclusion that the essays which are scored high by humans had greater vocabulary range while the essays which obtained lower scores are characterized by words which are more frequent and more general in their contextual distribution. In a more recent study, Kim, et al (2018) also examined the multi-dimensional nature of lexical proficiency and they concluded that 'words which are used in fewer contexts' are among the aspects which human raters consider as features of quality writing. These studies, taken together, show that contextual diversity is an important aspect of lexical proficiency and it should be taken into account in studies which examine the development of productive vocabulary knowledge.

1.1.3. Lexical diversity

McCarthy & Jarvis (2010) define lexical diversity (LD) as "the range of different words used in a text, with a greater range indicating a higher diversity" (p. 381). For Jarvis (2013, p. 3) LD is the opposite of repetition, a text which is not lexically diverse is characterized by the repetition of words which are stated earlier. In this respect, González (2017, p. 3) states that "[…] lexical diversity measures a text's ability to use a number of different words in order to limit repetition". In a lexically diverse text one would expect to find more word types than tokens. In principle, this means that LD is calculated by counting the number of different words in relation to the total number of words.

Various measures have been put forward to compute lexical diversity. Earlier indices such as the Number of Different Words (NDW) and the Type-Token Ratio (Richards and Malvern, 1997) were reported to yield unreliable results due to their being vulnerable to differences in text length (Malvern, et al.,, 2004). A number of mathematical transformations of the type-token ratio were advanced to account for the unreliability of diversity scores due to

variations in text length. These include the Index of Guiraud (Guiraud, 1954), which was found to be affected by differences in text length too (Hess, et al., 1986).

Other sophisticated measures of LD have been developed. Richards and Malvern (1997) developed the *D* index. However, in a study comparing *D* to 13 other LD measures, McCarthy & Jarvis (2007) concluded that "all 14 LD measures significantly correlated with text length. That is, none, including *vocd* [the computer program which is used to compute *D*], overcame the projection problem of text length dependency" (p. 480). McCarthy and Jarvis (2007, 2010) developed the HD-D index. However, like all the other previous measures, it significantly varied as an effect of text length. Both McCarthy and Jarvis (2010) and Treffers-Daller (2013) concluded that both *D* and HD-D correlated equally strongly with the number of tokens.

McCarthy (2005) developed the Measure of Textual Lexical Diversity (MTLD). MTLD is an index that is widely used in studies of LD nowadays thanks to its apparently high resistance to variations in text length. Jarvis (2013) commented on MTLD saying that "there is only one measure of variability I am aware of that does not vary as a function of volume (text length) or evenness, and this is the MTLD measure developed by McCarthy" (p. 28). Treffers-Daller (2013), however, reported that MTLD is also affected by text length.

Covington and McFall (2010) introduced the Moving Average Type-Token Ratio (MATTR). In a study comparing *D*, Maas index, MTLD and MATTR, Fergadiotis, Wright and Green (2015) showed that the validity of *D* was not as strong as that of MTLD and MATTR. Similarly, the researchers concluded that the results of Maas should not be interpreted as valid LD measures. In contrast, MTLD and MATTR were found to be strong measures of LD.

In a recent study, Vidal and Jarvis (2018) introduced a new measure which is expected to overcome the calculation problems of MTLD, as highlighted by Treffers-Daller (2013) and McCarthy and Jarvis (2010). The new measure, MTLD-W (or MTLD Wrap Around) combines the MTLD measurement procedure and the moving-window approach of MATTR. Vidal and Jarvis maintain that MTLD-W is characterized by more stability and it also reduces the overestimation or underestimation which is done by MTLD in estimating LD. However, many studies reported that MTLD-W too is sensitive to text-length.

To validate D, MTLD and HD-D Treffers-Daller (2013) investigated LD in two levels of L2 learners of French and native speakers. The study showed that there was a significant difference in the number of types and tokens produced by level 1 students and the native speakers, with native speakers producing the largest number of types and tokens. Using a proficiency C-test, Treffers-Daller concluded that HD-D and D correlated most strongly with the C-test while MTLD and Maas correlated only less strongly.

More recently, Vidal and Jarvis (2018) compared MTLD, MTLD-W and MATTR, investigating LD in the writings of Spanish university students of English. The authors reported a strong correlation between MTLD and MTLD-W which shows that the two indices measure the same construct. Similar to Treffers-Daller (2013), Vidal and Jarvis (2018) reported that both MTLD and MTLD-W were affected by text length unlike MATTR whose values remained consistent. Concerning the development of the target groups' lexical diversity, the study showed that there was no significant difference between 1st and 3rd year students whether the scores of MTLD, MTLD-W or MATTR are used, which indicates that students did not increase their productive lexical diversity scores after three years of instruction. However, MTLD and MTLD-W were able to predict learners' proficiency as measured by Oxford Placement Test. These two measures correlated significantly highly with the scores of the test, while MATTR correlated with the test only moderately. Interestingly, the quality of the participants' essays and their lexical diversity correlated somewhat significantly, with MTLD and MATTR showing the highest correlations. This further supports the conclusions of et al., (2015) about the two indices as reported above and also in Fergadiotis, Wright and West (2013). In a very recent study, Zenker and Kyle (2021) examined the stability of a variety of LD indices and they concluded that MATTR is stable across a variety of text lengths. What can be retained from

this study is that lexical diversity may not always increase as proficiency increases. Learners might improve their overall language proficiency without improving the ability to use a rich and varied repertoire of words in their written texts. These studies also indicate that MTLD and MATTR are more reliable measures of LD.

1.2. Significance of the study and research questions

The present study contributes to the efforts of the researchers to better understand the development of the productive vocabulary knowledge with respect to different proficiency levels by examining which of these three dimensions (word frequency, word range or lexical diversity) can pick out changes in language proficiency from a grade level perspective. Similarly, being conducted in a foreign language (FL) context, the present study sheds light on a context in which no such similar study has been conducted so far: The Moroccan EFL context. Also, no study has examined Moroccan university learners' productive vocabulary knowledge at the level of its diversity and range. This makes this study of valuable contribution to understanding the effect of the interaction of the three indices on lexical development. Similarly, this study uses MATTR as measure of lexical diversity. This index has been reported to be resistant to the effects of text length and there are very few studies which used it in the analysis of LD. Hence, the present research will also serve as an attempt to further validate the ability of MATTR to significantly differentiate between different proficiency levels. The present study is guided by the following research questions:

1. Do Moroccan university learners of different proficiency levels differ in the sophistication of the words they deploy in their writing?
2. Does word range show any improvement as an effect of proficiency level?
3. Do Moroccan university learners of different proficiency levels significantly enhance the diversity of the words they use in their writing?

2. METHODS

2.1. Sample and participants

Following a quantitative approach, this study uses as its corpus written essays from a total number of 529 university students who belong to three different intact undergraduate groups. Our target groups are semester 1 (S1) (N =167), semester 3 (S3) (N =188) and semester 5 (S5) (N =174) students majoring in EFL at the School of Arts and Humanities in Meknes. In this study we take grade level as an indicator of language proficiency. The informants are from 9 different groups conveniently chosen based on their and their teachers' voluntary involvement. Table 1 provides the related statistics for each group:

Table 1.
The number of participants by proficiency level

Proficiency Level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	S1	167	31,6	31,6	31,6
	S3	188	35,5	35,5	67,1
	S5	174	32,9	32,9	100,0
	Total	529	100,0	100,0	

2.2. Data collection and data analysis

To collect data for this study, we administered a writing test at the beginning of the academic year. Three target groups were asked to write an essay of approximately 300 words during their officially-scheduled classes. The participants were asked to write their essays in response to a unified topic. Participants were all given a maximum duration of one hour to complete their writing. The question that was used as a prompt for the essay writing is as follows:

> *Young people study because they want to have a job when they grow up.*
> *What are the other reasons for which young people go to school?*

Before processing the data for lexical indices, we corrected the spelling and grammar mistakes. We also changed the abbreviations and the contracted forms to their complete forms. No attempt was made at changing or correcting wrong words. Similarly, we deleted proper nouns such as America, Moroccan, etc. in order not to bias the frequency counts since these words might be extremely frequent both in learner corpus and in the reference corpus. After these procedures, all the texts were saved under a *.txt* format. For the analysis of lexical sophistication and range, the corpus of each of the three groups was processed using the open-source computer tool TAALES (Kyle and Crossley, 2014; Kyle, et al., 2018). The reference corpus for our data is the academic sub-corpus of COCA. We opted for using only the academic section of the corpus since our data is collected as written essays, and we think that the language which is elicited is mainly academic. The frequency and range scores from TAALES are generated for word tokens rather than word types. Each token in a text receives a frequency and a range score based on the reference corpus. These individual token scores are used to compute individual text mean scores and this is done by dividing the sum of the frequency scores by the total number of the tokens that received a score in the text. These are used to generate the whole target group's mean frequency and range score. For the analysis of lexical diversity, we used the open-source computer program TAALED (Kyle, et al., 2020). The present study used the MATTR statistic which was first introduced by Covington and McFall (2010). In the present study, lexical diversity is measured in learner texts based on word lemmas. A lemma includes a headword and its most immediate inflections within the same part of speech category. Lexical sophistication and range are examined with reference to both content and function words. Previous studies of lexical richness examined the learners' corpus from a unitary perspective focusing on the whole words together. The differences between the functions and distribution of function and content words is likely to engender important differences in the production of each of these two categories of words, which makes their examination separately worthwhile. For lexical diversity, we analyzed the whole words together.

3. RESULTS

3.1. Lexical sophistication

This section examines the sophistication of learners' productive knowledge of content words. To start with, the descriptive results on Table 2 indicate that there are some differences between the three groups in the sophistication of the lexical items they produced in their writing. For content words, S1 (mean = 1167.52; sd. = 245.52) had a lower content word frequency mean compared to S3 (mean = 1175.17; sd. = 237.21). For S5, the mean frequency of the content words they produced in their writing is 1159.34 (sd. = 216.99) which is lower than the means of both S1 and S3, which may indicate that S5 learners produced more sophisticated (i.e. less frequent) lexical items compared to the other two proficiency groups. After checking the normality of the variable (Figure 1), we conducted an analysis of variance and the results indicate that the difference between the three groups is *not* statistically significant ($p.> .05$) (Table 2).

For function words, we observe from the very high means that the learners of the three groups used highly frequent function words. Similar to what was observed for the frequency of content words, S1 learners had a mean score of 17609.04 (sd. = 3417.47) which is higher than that of S3 (mean = 17491.21; sd. = 2948.15) and lower than the mean of S5 (mean = 17831.56; sd. 2848.66). Surprisingly, S5 had an even higher mean compared to both S1 and S3. It appears from these descriptive results that after spending two years at university, learners still rely on the use of high frequency function words. Using Q-Q plots (Figure 2), we checked the normality of the variable and the plot shows that the data is normally distributed. The analysis of variance (Table 2) shows that the difference between the three groups is *not* statistically significant for function words as well (*p.*> .05).

Figure 1
The Q-Q normality plot for the sophistication of content words

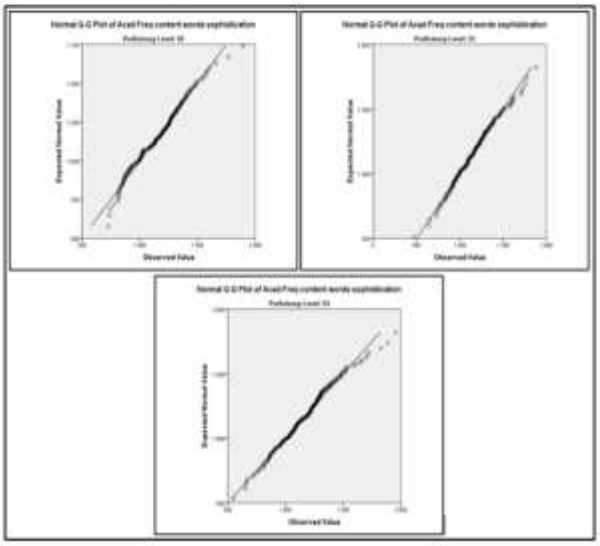

Figure 2
The Q-Q normality plot for the sophistication of function words

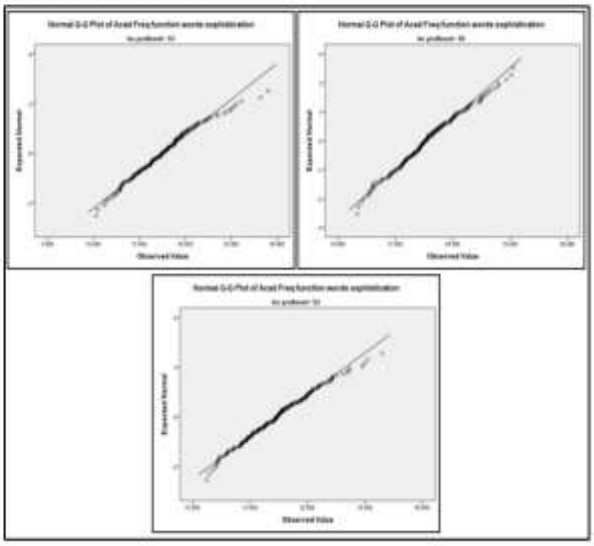

Table 2.
Anova test for the difference in (content and function) word frequency means between the three proficiency groups

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Acad Freq content words sophistication	Between Groups	22665,293	2	11332,646	,208	,812
	Within Groups	28674736,63	526	54514,708		
	Total	28697401,92	528			
Acad Freq function words sophistication	Between Groups	15372772,45	2	7686386,224	,814	,444
	Within Groups	4967943249	526	9444759,028		
	Total	4983316021	528			

To further check this similarity which we have demonstrated statistically, we plotted all the words on frequency bands using the computer program *Lextutor* which relies on frequency lists from Brown's (Francis and Kučera, 1964) over 1-million-word corpus in the following table (Table 3).

Table 3.
Profiling learners' vocabulary on frequency bands

S1			
Total word count : 25241 tokens			
Frequency band*	**Tokens**	**Percent**	**Cumulative %**
k-01	23144	91.80	91.80
k-02	1341	5.32	97.12
k-03	464	1.84	98.96
k-off list	54	0.21	99.98
S3			
Total word count: 29429 tokens			
Frequency band	**Tokens**	**Percent**	**Cumulative %**
k-01	26545	90.37	90.37
k-02	1747	5.95	96.32
k-03	768	2.62	98.94
k-off list	62	0.21	99.97
S5			
Total word count: 30230 tokens			
Level	**Tokens**	**Percent**	**Cumulative %**
k-01	26966	89.33	89.33
k-02	1957	6.48	95.81
k-03	883	2.93	98.74
k-off	62	0.21	99.98

*The frequency bands which do not appear on the table are deleted as they contain less than 1% of words each

The plotting (Table 3) supports our previous results about the similarity of learners' profiles in terms of the frequency of the vocabulary they produced in their writings. Irrespective of their proficiency level, the majority of learners' productive vocabulary belongs to the first frequency band. For S1 students, 91.8% of their vocabulary belongs to K1. A very similar percentage of S3 and S5 learners' vocabulary belongs to K1 (90.37% and 89.33%, respectively). We also noticed that the learners of the three groups produced a very similar amount of their vocabulary which belongs to the second band (K2) with S1 having 5.32% of their vocabulary in this band, 5.95% for S3 and 6.48% for S5. For the three groups, the remaining vocabulary spreads over the other bands. These results, along with the non-significant difference in the analysis of variance, indicate that at the productive level, after spending two years at university, the learners seem to rely more on the use of highly-frequent vocabulary items in their writing.

3.2. The Range of Learners' Words

We now turn to analyzing learners' produced vocabulary in relation to the range of the words. the results suggest that the mean range of the three groups' produced vocabulary is very similar both for content and function words. For content words, we can observe that S5 learners produced words with a more restricted range (mean = 0.47; sd. = 0.037) compared to both S1 (mean = 0.489; sd. = 0.040) and S3 (mean = 0.481; 0.041). Concerning function words, it can be observed that the texts of the three groups are characterized by the use of function words with higher mean range indices (S1mean = 0.922; sd. = 0.027; S3 mean = 0.924; sd. = 0.028 and S5 mean = 0.926; sd. = 0.027). It is important to note that function words are very likely

to have higher range indices due to their utility in a wider variety of texts and registers. As the Shapiro-Wilk test (Table 4) shows, the variable of word range is not normally distributed ($p.< .05$) for function words while it is normally distributed for content words ($p.>.05$). We, therefore, used an analysis of variance to check the significance of the difference in means between the three groups with respect to the range of content words, and we resorted to Kruskall-Wallis to test the significance of the difference in the mean range of function words.

Table 4
Normality test for the range index

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Proficiency Level	Statistic	df	Sig.	Statistic	df	Sig.
Acad Range content words sophistication	S1	,044	167	,200*	,994	167	,728
	S3	,054	188	,200*	,991	188	,327
	S5	,051	174	,200*	,988	174	,152
Acad Range function words sophistication	S1	,061	167	,200*	,974	167	,003
	S3	,057	188	,200*	,983	188	,019
	S5	,092	174	,001	,951	174	,000

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Table 5.
Anova test for the difference in range means between the three groups

ANOVA

Acad Range content words sophistication

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,027	2	,013	8,307	,000
Within Groups	,847	526	,002		
Total	,874	528			

The results of the analysis of variance (Table 5) show that the difference in the mean content word range between the three groups is statistically significant ($p.< .05$). That is, semester 5 learners' lowest mean range indicates that this group is using more sophisticated vocabulary in terms of its contextual restrictedness. The post-hoc test (Table 6) suggests that compared to S1 and S3, S5 learners demonstrated better knowledge of productive vocabulary at the level of range. More specifically, S5 learners produced vocabulary items which are more restricted in their usage scope.

Table 6.
Tukey HSD for group differences in mean range for content words

Multiple Comparisons

Dependent Variable: Acad Range content words sophistication
Tukey HSD

		Mean Difference (I-J)			95% Confidence Interval	
(I) Proficiency Level	(J) Proficiency Level		Std. Error	Sig.	Lower Bound	Upper Bound
S1	S3	,007971	,004267	,149	-,00206	,01800
	S5	,017673*	,004347	,000	,00746	,02789
S3	S1	-,007971	,004267	,149	-,01800	,00206
	S5	,009702	,004221	,057	-,00022	,01962
S5	S1	-,017673*	,004347	,000	-,02789	-,00746
	S3	-,009702	,004221	,057	-,01962	,00022

*. The mean difference is significant at the 0.05 level.

For the range of function words, the results of the analysis of variance (Table 7) ($p.>.05$) indicate that the mean range of function words for the three groups is not statistically

significant. This shows that learners' use of function words is very similar in terms of their range. This similarity might be due to the limited category of function words and their inevitable use in all types of texts and registers. This is what explains the high and similar mean range which is obtained by each proficiency group.

The results of this section indicate that the three proficiency groups differ significantly in terms of the range of content words, while they show no meaningful difference in the range of function words. The results seem to point towards better lexical performance for more advanced proficiency groups (S5) since they managed to use content words which are more restricted in their range. However, because function words are not specifically related to a certain type of texts and they are required in all contexts, learners show very similar productive knowledge in this category of words.

3.3. Lexical Diversity

Lexical diversity deals with the variety of learners' vocabulary in terms of the number of types (new words) compared to the number of tokens (repetitions of the same word). This section tests the hypothesis that the learners of three proficiency groups (S1, S3 and S5) do not demonstrate significant differences in the diversity of their productive vocabulary. To test this hypothesis, we first describe the performance of the learners before we resort to inferential statistics.

The descriptive statistics show some variability in the data of the three proficiency groups. The mean MATTR indicates that the learners of the three groups are approximately similar to each other in their diversity indices. S1 learners had an average MATTR of 0.72 (sd. = .050). S3 learners had one point higher (mean MATTR = 0.73; sd. = 0.43) and the highest proficiency group, S5, had a mean MATTR of 0.74 (sd. = .042). The similarity of learners' performance is also reflected in the very low standard deviation scores, which reflects the homogeneity of the groups in their lexical diversity performance. To test the significance of the differences, we first checked normality. The Q-Q plots (Figure3) indicate that the MATTR variable is normally distributed for the three groups.

Figure 3.
The Q-Q normality plots for the lexical diversity variable

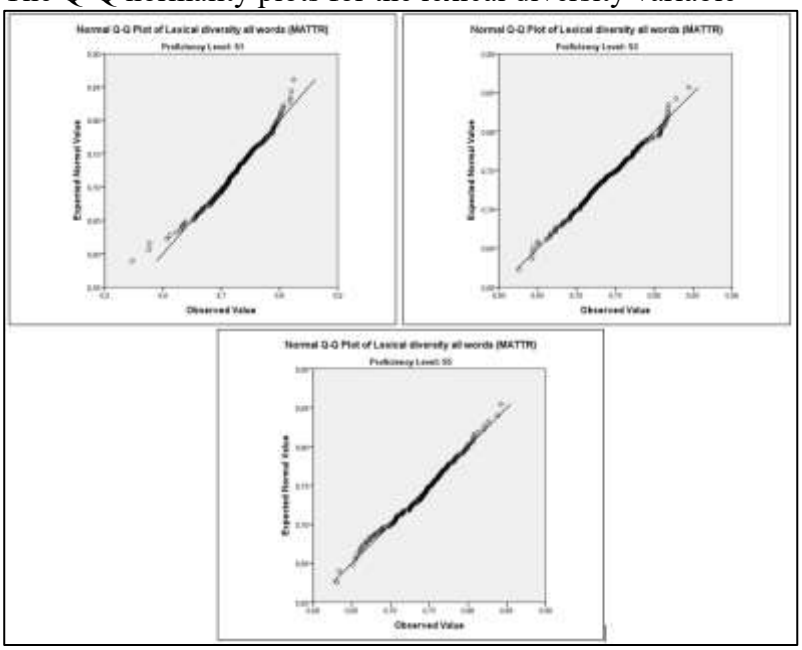

Table 7.
Anova for the difference in lexical diversity means between the three groups

ANOVA

Lexical diversity all words (MATTR)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,025	2	,012	6,029	,003
Within Groups	1,090	526	,002		
Total	1,115	528			

The analysis of variance (Table 7) indicates that the difference in the MATTR means is statistically significant. Our results indicate that the learners' proficiency level affects the production of varied and rich vocabulary in written texts.

Table 8.
Tukey LSD between groups comparison for lexical diversity (MATTR means)

Multiple Comparisons

Dependent Variable: Lexical diversity all words (MATTR)
Tukey HSD

(I) Proficiency Level	(J) Proficiency Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
S1	S3	-,014546*	,004840	,008	-,02592	-,00317
	S5	-,015031*	,004931	,007	-,02662	-,00344
S3	S1	,014546*	,004840	,008	,00317	,02592
	S5	-,000485	,004788	,994	-,01174	,01077
S5	S1	,015031*	,004931	,007	,00344	,02662
	S3	,000485	,004788	,994	-,01077	,01174

*. The mean difference is significant at the 0.05 level.

Using Tukey HSD between groups comparisons (Table 8), we observed that the difference in the MATTR means is statistically significant between university freshmen (S1) and the two other groups, S3 and S5 ($p.< .05$). The difference between the students who spent one year at university (S3) and those who spent two years (S5) is not, however, statistically significant. These results suggest that S1 learners joined university with the ability to produce texts which are characterized by more repetition. However, being exposed to instruction over the first year enabled them to enhance their productive lexical vocabulary, which allowed them to produce texts which are more lexically diverse. This significant improvement in the learners' productive vocabulary is expanded over the second year at university. We observed from the descriptive means that S5 learners had the highest diversity mean, and Table 8 shows that the difference between S1 and S5 is significant, which indicates that over the first two years at university, students significantly improved their vocabulary. However, it seems from the non-significant difference between S3 and S5 that the diversity of learners' productive lexical knowledge slows down in the second year, compared to the first.

4. DISCUSSION

This study examined three indices of lexical richness (word frequency, word range and lexical diversity). First, for word frequency, we could not observe any meaningful difference between the learners of the three proficiency groups. Although the mean frequency scores

showed a decreasing trend between S1 and S5, especially for content words, the frequency mean scores for the three groups are very similar, and we observed that the means are somewhat high both for content and function words, which indicates that learners rely in their written production on the use of frequent lexical items.

In our study, we profiled all the lexical items which are produced by the learners of the three groups using *Lextutor*, and the lexical profiles showed that the majority of the three groups' productive vocabulary belongs to the first frequency band: 91% for S1, 90% for S3 and 89% for S5 learners. The statistical analysis along with the lexical profiles showed that after two years of studying English at university (besides other previous 4 years in the secondary school), learners' written production is still characterized by the use of highly frequent lexical items. This conclusion seems to go counter most trends in previous studies which highlighted that more proficient L2 learners produce less frequent vocabulary (e.g., Laufer and Nation, 1995; Crossley & McNamara, 2009; Guo, et al.,, 2013; Crossley, et al.,, 2016; Crossley & Cai, 2018; Crossley & Kyle, 2018).Similar to our study, however, González (2017), reported that there wasn't any significant difference between different proficiency levels of advanced multilingual student writers and first year monolingual learners of English in the frequency of the produced words. Gonzaléz observed that the effects of lexical frequency declined between more advanced levels.

Our results show that there seems to be very little effect of proficiency on learners' productive vocabulary knowledge at the level of word frequency and that even second year university students (S5) still rely on the use of very frequent words in their writing. Eguchi and Kyle (2020), for example, reported that more proficient learners produced mainly content words which are more, rather than less, frequent. However, unlike content words, the same study concluded that proficient learners produced less frequent function words. Crossley and McNamara (2009) explained that L2 learners' use of more frequent words can be due to the frequent words' easy processing and the ability of the readers to understand them easily. More frequent words are readily available in the input and easily accessed in the learners' lexicon, which makes their production an efficient and economical process in L2 production.

Not all aspects of productive lexical knowledge develop in the same way. Unlike what was highlighted for word frequency, the analysis of word range showed that the three groups' lexical knowledge demonstrates some statistically meaningful differences. First, we observed that the difference in the mean range of content words is significant between the highest proficiency group (S5) and the two other lower proficiency groups (S1 and S3). These results indicate that various aspects of lexical knowledge might take different developmental paths.

Our results show that although learners' productive lexical knowledge might be very similar in terms of frequency, the range of content words shows some statistically significant differences with higher proficiency learners (S5) using words which are more restricted in their use, compared to lower proficiency groups (S1 and S3). Previous studies (e.g., Kyle & Crossley, 2015, 2016; Vögelin, et al., 2019) documented that higher range scores negatively correlate with essay scores, indicating that better quality essays are characterized by words which are restricted in their use across different texts and contexts. More than this, these studies indicated that range is a stronger predictor of holistic essay quality compared to word frequency.

For the range of function words, our results demonstrated that there was no statistically significant difference between the learners of the three proficiency groups. This seems to be a logical conclusion since function words cannot, by their nature, be limited to a type of context or limited to a certain register. Function words are widely used and dispersed across different texts, registers and contexts. This might explain the very high range means which the learners of each of the three groups obtained. This conclusion is similar to what we observed for the performance of learners in relation to the frequency index of function words: for both indices (frequency and range of function words), we observed that the performance of the three groups is very similar and that the differences are not statistically meaningful. We explained this as

being an effect of the nature of function words, which makes them widely and frequently resorted to by writers irrespective of the type of text or the communicative genre they are adopting.

The examination of the range index shows that at early stages of second language learning, the range of content words might be a better indicator of language proficiency compared to word frequency. We observed that after spending two years at university, S5 learners are not statistically different from both S3 and university freshmen (S1) at the index of word frequency, while content word range significantly could distinguish at least between the students with relatively better language proficiency (S5) compared to students who have spent only one year at university (S3).

For lexical diversity, we observed a meaningful difference between the groups. It was shown that university freshmen (S1) significantly differed at the level of lexical diversity with the two upper proficiency groups (S3 and S5). However, the results did not show any significant difference between S3 and S5. The results seem to indicate that lexical diversity can distinguish between university freshmen and more proficient learners. This suggests that the first year of studying English at university provided learners with opportunities to enrich their lexical repertoire with new lexical items, which enabled them to vary the words they deployed in their writing. Learners' productive vocabulary knowledge gets more diverse after spending two years at university, as reflected in the statistically significant difference between S1 and S5. However, because the difference between S3 and S5 learners was not statistically significant, we may maintain that the diversity of learners' vocabulary somewhat slows down during the second year at university.

S5 learners' significant difference with the two lower proficiency levels (S1 and S3) shows that after two years at university, learners demonstrate better productive lexical knowledge. In this respect, previous studies indicated that lexical diversity is an indicator of vocabulary size, and that the number of different words which are produced by a writer is a characteristic of more proficient learners (Baba, 2009; Crossley, et al.,, 2010; Eguchi & Kyle, 2020; Malvern, et al.,, 2004). Our results, therefore, support previous studies in that higher proficiency learners (S3 and S5) produced significantly varied texts, reflecting better productive vocabulary knowledge both in terms of its quality (i.e., diversity) and size.

The results of our study seem to align with findings from previous studies (McNamara, et al., 2009; Gonzalez, 2018) on lexical diversity. These studies analyzed lexical diversity in students' writings and they indicated that lexical diversity is a strong predictor of language proficiency. This shows that lexical diversity improves along with the learners' proficiency in writing. Gonzalez (2018), for instance, showed that monolingual English speakers have higher levels of lexical diversity compared to advanced multilingual learners of English. This result shows that as L2 learners approach native speaker proficiency, they demonstrate better lexical diversity in their writing.

CONCLUSION

The results of the present study showed the word frequency, word range and lexical diversity as measures of lexical proficiency do not adhere to similar developmental paths in EFL contexts. Although previous research (e.g., Nation, 2014; Treffers-Daller & Milton, 2013; Schmitt, 2014; Stæhr, 2008; Vermeer, 2004; Webb, 2008) indicated that there is quick improvement in the frequency of learners' receptive vocabulary knowledge, the productive vocabulary knowledge seems to take some time before it is projected into learners' written productions. Our results suggest that learners might feel comfortable using the words which they know better and which they think their interlocutors might understand because of their high frequency. The results of lexical range indicate that after the first two years at university, learners manage to produce written texts which are characterized by the use of context

restricted vocabulary, as can be seen from semester 5 learners' significantly lower range score. It seems that although the university learners prefer to use high frequency words, more advanced groups tend to produce words which are more context-specific. This same conclusion is observed for the diversity of learners' vocabulary, we observed that S5 learners had a significantly better word diversity performance compared to university freshmen who produced essays which are less lexically-diverse. We explained these results as an indication that like range, lexical diversity is a tacit lexical aspect which requires at least two years of studying English at university before learners' writing is characterized by significant levels of lexical variation.

Pedagogically, taking into account that lexical variation is a quality of higher proficiency groups, it is important to raise learners' awareness to the importance of varying one's vocabulary in writing through the use of newer words. Language teachers are invited to enrich their classes with larger amounts of distinct lexical items in order to expand their learners' lexical knowledge. Similarly, because research has outlined some discrepancy between the amount of vocabulary which learners know and that which they can actually produce, it is of great importance to encourage learners in writing classes to pay more attention to the richness of vocabulary in learners' writing. It is necessary that teachers provide learners with specific feedback on the diversity of vocabulary. Similarly, learners should be encouraged to include more specific vocabulary whenever an alternative to general words can be used.

As previous research concluded that lexical sophistication is a quality both of higher proficiency writing, it is necessary that teachers should focus more on varying the vocabulary to which learners are exposed in terms of its frequency. This can be done by amplifying reading and listening opportunities with the use of graded input which moves learners gradually towards encountering less frequent and also less general lexical items. This requires a careful and research-based selection of texts to which learners are exposed.

While this study might be the first to examine these three lexical aspects in the writing of Moroccan tertiary levels using state-of-the-art computer programs, it would be of added value if these aspects are further supplemented with a regression analysis in order to check which among them can positively affect learners' scores in writing. Although this study used quite a large sample and adopted a quasi-longitudinal approach to how the three lexical aspects develop among Moroccan university learners, future research can adopt a small size sample with a longitudinal design. Finally, because, previous studies (e.g., Akki and Larouz, 2021) showed that there is a high correlation between the performance of students in speaking and writing, it would be a valuable contribution to analyze the indices which are examined in this study using learners' spoken rather than written productions.

REFERENCES

Akki, F. & Larouz, M. (2021). The relationship between speaking and writing in descriptive discourse in a Moroccan university EFL context. *International Journal of Linguistics and Translation Studies, 2*(1), 124–133.

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing, 18* (3), 191–208

Cobb, T. *Lextutor*. [Computer program]. Accessed April 13th, 2021 at https://www.lextutor.ca/cite/

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., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, *18*(2), 119–135. https://doi.org/10.1016/j.jslw.2009.02.002

Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, *60*(3), 573–605.

Crossley, S., & McNamara, D. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. Proceedings of the Annual Meeting of the Cognitive Science Society*,* 32*.* Retrieved from https://escholarship.org/uc/item/6n5908qx.

Crossley, Cobb, T., & McNamara, D. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, *41*, 965–981. https://doi.org/10.1016/j.system.2013.08.002

Crossley, S. A., Kyle, K., Allen, L. K., Guo, L., & McNamara, D. S. (2014). Linguistic microfeatures to predict L2 writing proficiency: A case study in automated writing evaluation. *Journal of Writing Assessment, 7*(1). Retrieved from http://www.journalofwritingassessment.org/archives.php?issue=17i.

Crossley, S., Allen, L. K., Snow, E. L., & McNamara, D. S. (2016). Incorporating learning characteristics into automatic essay scoring models: What individual differences and linguistic features tell us about writing quality. *Journal of Educational Data Mining, 8*(2), 1–19.

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

Ellis, N. C. (2002a). 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

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*.

Fergadiotis, G., Wright, H. H., & Green, S. B. (2015). Psychometric evaluation of lexical diversity indices: assessing length effects. *Journal of Speech, Language, and Hearing Research: JSLHR*, *58*(3), 840–852. https://doi.org/10.1044/2015_JSLHR-L-14-0280

Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech-Language Pathology*, *22*(2), 397–408. https://doi.org/10.1044/1058-0360(2013/12-0083)

González, M. C. (2017). The contribution of lexical diversity to college-level writing. *TESOL Journal*, *8*(4), 899–919. https://doi.org/10.1002/tesj.342

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, *13*, 403–437. https://doi.org/10.1075/ijcl.13.4.02gri

Gries, S. T. (2010). Dispersions and adjusted frequencies in corpora: Further explorations. In *Corpus-linguistic applications* (pp. 197–212). Brill Rodopi. https://doi.org/10.1163/9789042028012_014

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, *18*(3), 218–238. https://doi.org/10.1016/j.asw.2013.05.002

Jarvis, S. (2013). In Jarvis, S., & Daller, M. (eds.) (2013). *Vocabulary Knowledge: Human ratings and automated measures*. Amsterdam: John Benjamins.

Jarvis, S., & Daller, M. (2013). *Vocabulary Knowledge: Human ratings and automated measures*. Amsterdam: John Benjamins.

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

Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing. 38*. 10.1016/j.asw.2018.06.004.

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

Kyle, K., Allen, L., Guo, L., & McNamara, D. (2014). Linguistic microfeatures to predict L2 writing proficiency: A case study in Automated Writing Evaluation. *Journal of Writing Assessment*, *7*, 1–16.

Kyle, K., & Crossley, S. (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. (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., & Jarvis, S. (2020). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, *0*(0), 1–17. https://doi.org/10.1080/15434303.2020.1844205

Kyle, K., Crossley, S., & 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

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in l2 written production. *Applied Linguistics*, *16*(3), 307–322.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230511804

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. Doctoral dissertation. Available from ProQuest Dissertations and Theses.

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

McDonald, S., A. & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44*(3), 295–322. https://doi.org/10.1177/00238309010440030101

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2009). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86. https://doi.org/10.1177/0741088309351547

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters.

Monteiro, K. R., Crossley, S. A., & Kyle, K. (2020). In search of new benchmarks: Using L2 lexical frequency and contextual diversity indices to assess second language writing. *Applied Linguistics*, *41*(2), 280–300. https://doi.org/10.1093/applin/amy056

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, *116*(1), 37–44. https://doi.org/10.1016/j.jecp.2012.10.014

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

Richards, B. J., and Malvern, D. D. (1997). *The new Bulmershe papers. Quantifying lexical diversity in the study of language development.* Reading: The University of Reading.

Treffers-Daller, J. (2013) Measuring lexical diversity among L2 learners of French: an exploration of the validity of D, MTLD and HD-D as measures of language ability. In: Jarvis, S. and Daller, M. (eds.) *Vocabulary knowledge: human ratings and automated measures.* Benjamins, Amsterdam, pp. 79-104. ISBN 9789027241887. Accessed at: http://centaur.reading.ac.uk/28712/

Treffers-Daller, J., Parslow, P., & Williams, S. (2018). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, *39*(3), 302–327. https://doi.org/10.1093/applin/amw009

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, *17*(1), 65–83. https://doi.org/10.1177/026553220001700103

Vidal, K., & Jarvis, S. (2018). Effects of English-medium instruction on Spanish students' proficiency and lexical diversity in English. *Language Teaching Research*, *24*(5), 568 587. https://doi.org/10.1177/1362168818817945

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

Zyad, H., Rguibi, S., & Bouziane, A. (2017). *The relationship between L2 writing quality and objective measures of syntactic and lexical complexity*. Proceedings of the 4th International Conference on Cultures and Languages in Contact. pp. 393-410.