

## Lexical Richness in Research Articles Written by Iranian and Foreign Writers

Payman Debash<sup>1</sup>, Seyed Foad Ebrahimi <sup>\*2</sup>

<sup>1</sup>English Department, Bushehr Branch, Islamic Azad University, Bushehr, Iran

<sup>2</sup>English Department, Shadegan Branch, Islamic Azad University, Shadegan, Iran

DOI: [10.30495/LCT.2021.1945054.1049](https://doi.org/10.30495/LCT.2021.1945054.1049)

**Received:** 15/11/2021

**Revised:** 18/12/2021

**Accepted:** 23/12/2021

### Abstract

This study aimed at investigating lexical richness in the articles written by Iranian and foreign scholars. To this end, a quantitative descriptive design was used. As the corpus of this study, 34 English Language Teaching (ELT) papers were randomly selected from different ISI and Academic-Research journals. 17 papers were written by Iranian researchers and 17 by foreign researchers. The journals from which the papers written by foreign writers were selected are EAP Journal and System. The journals from which the papers written by Iranian writers were selected are Journal of Research in Applied Linguistics and Journal of Teaching Language Skills. Analysis of the corpus was done based on the framework proposed by Lei and Yang (2020), wherein lexical richness was measured in three dimensions including lexical diversity, lexical density, and lexical sophistication. To analyze the data, descriptive as well as inferential statistics were used. The findings revealed that the median score of the lexical diversity, lexical density and lexical sophistication of the papers written by the foreign researchers were higher than those of the papers written by the Iranian researchers. To examine whether this difference is significant, Mann-Whitney U test was run and the results confirmed that there is a significant difference between the three dimensions (i.e., lexical diversity, lexical density and lexical sophistication) of the lexical richness of the papers written by Iranian and foreign scholars. The findings of this study could be considered by EFL learners/teachers, curriculum planners and researchers.

**Keywords:** Lexical richness, Lexical diversity, Lexical density, Lexical sophistication

### 1. Introduction

As an integral component of academic writing, lexical richness plays an important role in writing because it can help learners to write higher

\* Corresponding Author's E-mail address: [seyedfoade@gmail.com](mailto:seyedfoade@gmail.com)



This work is licensed under a [Creative Commons Attribution](https://creativecommons.org/licenses/by-sa/4.0/)

quality written texts (Maher Salah, 2008, as cited in Anjomshoa & Zamanian, 2014). Lexical richness is an umbrella term employed to explain the general influence of more particular systems involving lexical variation, lexical density, and lexical sophistication (Nation & Webb, 2011). As put by Jarvis (2017), the effective condition of lexical richness would be encapsulated in at least six elements: volume, variability, evenness, rarity, disparity, and dispersion. Read's (2000) pattern of lexical richness consists of lexical variation, lexical sophistication (utilization of low-repetition vocabulary), lexical compactness (the proportion of fulfilled vocabularies to operational vocabularies), and the repetition of lexical mistakes. The first two of these evaluations (i.e., lexical variation and lexical sophistication) are the most typically employed in examinations of lexical richness (Malvern & Richards, 2013).

Vocabulary knowledge operationalization occurs through lexical richness (Read, 2000). Lexical richness consists of three measures including lexical diversity, lexical density and lexical sophistication (Lei & Yang, 2020). Djwandono (2015) defined lexical richness as the existence of different words in a text, and suggested that it should be measured through type-token ratio (TTR). Lexical richness is more particularly regarded as the analogue of the lexical description (Djwandono, 2015).

The importance of lexical richness is so that some scholars (e.g., Mehrpour, Razmjoo, & Kian, 2011) have mentioned that even if students can successfully decode and fluently read, lexical richness plays a significant role in writing academically. According to Snow (2002), lexical richness influences learners' writing mechanism and written productions. Accordingly, Snow (2002) continues that if readers' level of vocabulary knowledge is low, they may find the essay too complex to write.

Since lexical richness is an indicator of writing proficiency, it should be measured for the purpose of assessment of writers' writing ability. Thus, the volume of research done on lexical richness in general and comparison of lexical richness in writings of different groups of writers with different language backgrounds is scarce. Among the studies conducted on this issue, some scholars (e.g., Daller & Xue, 2009; Read, 2000) have investigated various components of lexical richness, such as lexical variation (i.e., kind-representation proportion), lexical variability (e.g., kind of various vocabulary kind but concentrated on lexical vocabulary), lexical sophistication (i.e., the proportion of advanced words), average word length, and lexical density (i.e., the dimensions of

lexical vocabularies in the whole writing). Grant and Ginther (2000) analyzed a sample of 90 Test of Written English (TWE) papers, written at three degrees of accomplishment as described by TWE classifications to examine paper length, lexical particularity, lexical characteristics, grammatical constructions, and clause level characteristics. Programmed tagging was employed to recognize outlines in the test-takers' writing capability, and it disclosed that a mixture of all of these features discovered L2 writing accomplishment degree. Kim (2014) examined text length, lexical complication and grammatical complication from the Yonsei English Learner Collection (e.g., educational writing) to recognize the best indicators of L2 pupils' writing competency. The results showed that higher competency learners produced longer passage, utilized more diversity of vocabulary (i.e., variation), and demonstrated the capability to employ more vocabularies in a statement (i.e., grammatical complication).

Gregori-Signesca and Clavel-Arroitia (2015) explored lexical density and lexical diversity in the writings of beginner and advanced university students. According to the measures of lexical richness obtained by software, it was found that higher-proficiency students' writings were richer than lower-proficiency students' productions lexically. The researcher attributed this difference to the role of language proficiency.

In Djiwandono's (2016) study, the lexical richness of senior students' writings was compared to their instructors' to academic papers. With the help of text analyzer program, it was revealed that academic papers of lecturers were in a higher level of lexical richness than writings of students. It was therefore concluded that proficiency in English is at play in determining the lexical richness of written works.

Gungor and Yayli (2016) made inquiries about the relationship between lexical description and reading comprehension among a category of Turkish EFL trainees. The participants of this research consisted of 178 college students that used a vocabulary list in relation to the lexical components of two various descriptive passages. The reading comprehension degree was estimated by means of two reading comprehension examinations. The results of this research disclosed that their passage-oriented lexical knowledge relatively correlated with reading comprehension. The correlation was comparatively direct correlation. In addition, the discoveries demonstrated that the 98% lexical description is essential for trainees to realize educational passages, and this description actually relates to nearly the most repetitive 8000 lexical categories according to the same researches.

Ibrahim and Esa (2019) compared the lexical richness of the essays by entry-level university students and those by third-year university students to see if the two groups of written productions differ statistically. As shown by the results, the mean lexical richness of third year university students' essays was statistically higher than that of entry level university students' texts. Therefore, it was suggested that lexical richness is a parameter which is affected by language proficiency level and writing skill.

As is seen in the reviewed literature, the volume of research on lexical richness is not rich. In this scarcity, the researcher found no study on lexical richness in the articles written by Iranian and foreign scholars. This topic was investigated in this study under the following research questions:

1. What are the lexical richness levels of the papers written by Iranian and foreign scholars?
2. Is there a significant difference between the lexical richness levels of the papers written by Iranian and foreign scholars?

## **2. Methodology**

### **2.1 Design**

In this study, quantitative descriptive design was used. To this end, statistical procedures including both descriptive and inferential statistics (i.e., Mann-Whitney U test) were utilized.

### **2.2 Corpus**

As the corpus of this study, 34 English Language Teaching (ELT) papers were randomly selected from different ISI and Academic-Research journals. 17 papers were written by Iranian researchers and 17 by foreign researchers. The journals from which the papers written by foreign writers were selected included EAP Journal and System. Additionally, the papers written by the Iranian writers were selected from Journal of Research in Applied Linguistics and Journal of Teaching Language Skills. The number of words of each paper was in a range from 5000 to 6000.

### **2.3 Framework**

Analysis of the corpus was done based on the framework proposed by Lei and Yang (2020), wherein lexical richness was measured in three dimensions including lexical diversity, lexical density and lexical sophistication. In this framework, the ratio of new words (type) to the total number of all words (token) was taken as the measure of lexical diversity.

Lexical density was measured by Lexical Density Calculator. Lexical sophistication was calculated through the Lexical Frequency Profile (LFP) (Laufer & Nation, 1995), using the software AntWordProfiler.

#### **2.4 Procedure**

At the beginning, the researcher randomly collected 34 ELT papers from a variety of ELT-related ISI and Academic-Research journals including EAP Journal, System, US-China Foreign Language, Communication and Linguistics Studies, Journal of Second Language Writing, International Journal of Curriculum and Instruction, etc. Then, he converted the files into textpad format. Next, the converted files were entered into 'http://www.lexutor/ca/vp' so that their lexical diversity profiles could be determined. This program used type-token ratio (TTR) to measure lexical diversity index. That is, the ratio of new words (type) to the total number of all words (token) was taken as the measure of lexical diversity in this program. Then, lexical density was measured by the software of Lexical Density Calculator designed by Ge (2016). Finally, lexical sophistication was calculated through the Lexical Frequency Profile (LFP) (Laufer & Nation, 1995), using the software AntWordProfiler developed by Anthony (2014). At the end, the measures of lexical diversity, lexical density and lexical sophistication obtained for each paper were entered in SPSS for quantitative statistical tests.

#### **2.5 Data Analysis**

To analyze the data, descriptive statistics and inferential statistics were used. Descriptive statistics included mean rank, sum of ranks, and median. It aimed at determining the mean lexical richness of the papers written by Iranian and foreign researchers. Inferential statistics consisted of Mann-Whitney U test. The aim of this test was to compare the papers written by Iranian and foreign researchers concerning their lexical richness level.

### **3. Results**

Table 1 shows the results of descriptive statistics of the three dimensions of the lexical richness measure including lexical diversity, lexical density and lexical sophistication in the two groups of papers, i.e., one group written by Iranian researchers, and the other group written by foreign researchers.

Table 1. Mean Ranks and Median Scores on Lexical Richness Dimensions of the Papers by Groups

	Group	N	Mean Rank	Sum of Ranks	Median
Lexical diversity	Iranian	17	11.50	230.00	32
	Foreign	17	29.20	590.00	68
Lexical density	Iranian	17	25.81	800.00	14
	Foreign	17	75.333	910.00	70
Lexical sophistication	Iranian	17	6.10	79.00	25
	Foreign	17	9.00	98.00	58

As displayed in Table 1, the foreign group (Mdn = 68, 70, 58) had higher median scores than the Iranian group (Mdn = 32, 14, 25) on three dimensions of the lexical richness measure. This means that on average, the papers written by foreign researchers were higher concerning lexical diversity, lexical density and lexical sophistication in comparison with the papers written by Iranian researchers.

To compare the lexical richness dimensions of the two groups of papers, a non-parametric Mann-Whitney U test was run. Table 2 illustrates the results.

Table 2. Mann-Whitney U Test for the Lexical Richness Dimensions of the Papers by Groups

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)	Exact Sig. [2*(1-tailed Sig.)]
Lexical diversity	20.000	230.000	-4.887	.000	.000
Lexical density	41.000	340.000	-11.000	.000	.000
Lexical sophistication	96.000	660.000	-39.000	.000	.000

\*Note: the level of significance is P<05

The results indicated that the foreign group had a significantly higher median scores on three dimensions of the lexical richness measure including lexical diversity (Z = -4.88, P= .000), lexical density (Z = -11.000, P= .000), and lexical sophistication (Z = -39.000, P= .000), than the Iranian group. Thus, it can be concluded that there is a significant difference between the three dimensions of the lexical richness of the

papers written by Iranian and foreign scholars. In a more specific sense, the papers written by foreign scholars were richer in terms of diversity, density and sophistication.

#### **4. Discussion**

The findings revealed that the median score of the lexical diversity, lexical density and lexical sophistication of the papers written by the foreign researchers were higher than those of the papers written by the Iranian researchers. To examine whether this difference is significant, Mann-Whitney U test was run and the results confirmed that there is a significant difference between the three dimensions (i.e., lexical diversity, lexical density and lexical sophistication) of the lexical richness of the papers written by Iranian and foreign scholars.

These findings are reasonably referring to the arguments which say that lexical richness is under the direct effect of writing proficiency (Breeze, 2008; Ghaddesy, 1989). In fact, the belief is that those with higher writing proficiency can write more lexically rich papers than those with lower language proficiency (Douglas, 2012). Previous literature confirms superiority of foreign writers in terms of linguistic complexity. Knowingly, lexical richness is among different dimensions of linguistic complexity. Therefore, the fact that foreign papers were richer on the aspect of lexical richness seems reasonable in the light of the previous studies.

Justifying the findings is not that much difficult if the previous findings are referred to, arguing that context of language learning and language use (i.e., native context, EFL context, ESL context) exert huge effects on the amount and quality of learned/acquired input. The role of environment and context cannot be missed in this regard. Last but not least, the role of different dimensions of knowledge in the amount of lexical richness of papers written by different writers can be referred to as a potential factor effective in this regard.

In line with the present study, Lei and Yang (2020) compared lexical richness in the papers of native and non-native speakers of English and found that native speakers have produced texts with higher lexical richness. These findings also give support to the previously mentioned strong correlation between writing proficiency and lexical richness. In another study on the same topic, Gregori-Signes and Clavel-Arroitia (2015) found that those writers whose writing ability is higher write texts which are richer lexically when they are matched with the essays written by

writers with a lower writing ability. Ibrahim and Esa (2019) also reached the same finding in their study.

Additionally, consistent with this study, it was found by Azadnia, Lotfi and Biria (2019) that writings of native speakers are richer than those written by Iranian students. Another study whose results are similar to those of this study is the one by Qi (2014) wherein it was proved that writing complexity is more prominent in native speakers' writings compared with non-native speakers' written productions. The same findings were obtained in some other studies including those by Lu (2010) and Lu and Ai (2013).

## **5. Conclusion**

The researcher tends to conclude, based on the results, that lexical richness is a measure which is significantly and richly correlated with writing ability. Therefore, as a prerequisite to enhancing lexical richness, writers should make efforts to improve their general writing ability. However, the potential mediating variables including learning styles, writing styles, verbal intelligence, personality types, topical knowledge, grammar knowledge and other kinds of knowledge whose roles are considered as key ones but mentioning all of them are beyond the patience of this paper should not be neglected in this relationship.

Moreover, it is concluded that in the context of Iran, more emphasis should be put on lexical stock of higher education students if they want to become closer to foreign writers, at least as far as lexical richness is concerned. The mission of English teachers teaching at higher education levels is of importance and significance in this respect. They can motivate students to read more complex texts and learn and retain more new vocabularies as preliminary stages to write more richly.

Last but not least, it is concluded that foreign writers, as expected, are more developed than Iranian writers since contrary to Iran, English is not regarded as a foreign language in many foreign countries and accordingly, many foreign writers have been exposed to English language from their childhood. The results may also be attributed to cultural difference between foreign and Iranian researchers.

The findings have some implications for different groups of stakeholders including EFL learners, EFL teachers, curriculum planners and researchers. EFL learners become more conscious of the dimensions of lexical richness and how they can implement them in their writing to enhance the quality of their writing. EFL teachers can take advantage of the findings by trying to help their students enhance their knowledge of

dimensions of lexical richness. As well, curriculum planners take useful insights from the findings and apply them in planning future EFL curricula, at least as far as lexical richness is concerned in writing courses. Finally, researchers in the field can take useful insights from the findings and increase the lexical richness of the papers they write, in terms of lexical diversity, lexical density and lexical sophistication.

Eventually, it is worth noting that this study suffered from some limitations in terms of sampling, setting, corpus, and rating, etc. Thus, it is recommended to be removed in the future studies so that the findings can be generalized and extrapolated with higher confidence. It is hoped that the findings contribute to production of richer papers by Iranian writers in terms of lexicon.

**Funding:** This research received no external funding from any agency.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Anjomshoa, L., & Zamanian, M. (2014). The effect of vocabulary knowledge on reading comprehension of Iranian EFL learners in Kerman Azad University. *International Journal on Studies in English Language and Literature (IJSELL)*, 2(5), 90-95.
- Anthony, L. (2014). *AntWordProfiler*. Tokyo, Japan: Waseda University. Version 1.4.0. Retrieved from <https://www.laurenceanthony.net/software>.
- Azadnia, M., Lotfi, A.R., & Biria, R. (2019). A study of syntactic complexity via Coh-Metrix: Similarities and differences of Ph.D. dissertations written by Iranian University students and English native speakers. *Research in English Language Pedagogy (RELP)*, 7(2), 232-254.
- Breeze, R. (2008). Researching simplicity and sophistication in student writing. *International Journal of English Studies*, 8(1), 51-66.
- Daller, M. H., & Xue, H. (2009). Vocabulary knowledge and academic success: A study of Chinese students in UK higher education. In B. Richards, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller, (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 179-193). London: Palgrave Macmillan.
- Djiwandono, P.I. (2016). Lexical richness in academic papers: a comparison between students' and lecturers' essays. *Indonesian Journal of Applied Linguistics*, 5(2), 209-216.

- Douglas, R. S. (2012). *Non-native English-speaking students at university: Lexical richness and academic success* (Unpublished dissertation). Universitas Calgary.
- Ge, X. (2016). *Lexical density calculator*. Retrieved from <http://gexiaoshuai.top/>.
- Ghaddesy, M. (1989). The use of vocabulary and collocations in the writing of primary school students in Singapore. In Nation, P. & Carter, R. (eds). *Vocabulary acquisition* (pp. 106 – 117).
- Gungor and Yayli (2016). The interplay between text-based vocabulary size and reading comprehension of Turkish EFL learners. *Educational Sciences: Theory and Practice*, 16(4), 1171-1188.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123-145.
- Gregori-Signesa, C., & Clavel-Arroitia, B. (2015). Analyzing lexical density and lexical diversity in university students' written discourse. *Procedia - Social and Behavioral Sciences*, 198, 546-556.
- Ibrahim, E.H.E., & Esa, Z. (2019). A comparison of lexical richness in L2 written productions. *Ijet*, 14(20), 174-181.
- Jarvis, S. (2017). Grounding lexical diversity in human judgments. *Language Testing*, 34(4), 537–553.
- Kim, J.Y. (2014). Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study. *English Teaching*, 69(4), 27-51.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Lei, S., & Yang, R. (2020). Lexical richness in research articles: corpus-based comparative study among advanced Chinese learners of English, English native beginner students and experts. *Journal of English for Academic Purposes*, 47, 1-9.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International journal of corpus linguistics*, 15(4), 474-496.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International journal of corpus linguistics*, 15(4), 474-496.
- Lu, X., & Ai, H. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. *Studies in Corpus Linguistics*, 59, 249-264.

- Malvern, D., & Richards, B. (2013). *Lexical diversity and language development: Quantification and assessment*. Basingstoke, UK: Palgrave Macmillan.
- Mehrpour, S., Razmjoo, S., & Kian, P. (2011). The relationship between depth and breadth of vocabulary knowledge and reading comprehension among Iranian EFL learners. *Journal of English Language Teaching and Learning*, 22(53), 140-147.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.
- Qi, D. (2014). *Syntactic complexity of EFL, ESL and ENL: Evidence of the international corpus network of Asian learners of English* (Master's thesis). National University of Singapore, Singapore.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica: The RAND Corporation.