| $\mid$ ISSN: 1112-685X | IV : 15 $\quad$ I Issue: 01 | /June 2022 |
| :--- | :--- | :--- | :--- |

# Vocabulary Profiling of Third Year Students of English in Critical Essays 

## تصنيفمفــرداتطـــلابالسنتالثالثـتّللغــتّ <br> 

Karima TAYAA *
Batna 2 University
Batna / Algeria
Karima.tavaa@univ-tebessa.dz

Souhila HELLALET
Batna 2 University
Batna / Algeria
S.helalet@yahoo.com

Received: 15/08/2021 Accepted :24/01/2022
Published:10/06/2022


#### Abstract

: This study was set out to investigate VocabProfile (VP) use to assess student's productive vocabulary size in critical essays. Third-year students at the department of English were selected randomly to analyse learners 'critical essays. Thus, 10 writing samples build from written assignments, and then student essays were analysed using (VP)to generate student's vocabulary profiles according to word families, type, and tokens ratio and match it with available vocabulary frequency lists. The overall profile shows that words used by the student mostly belonged to the first ( $\mathrm{K}-1$ ) with vocabulary coverage ( $77.20 \%$ ), and then followed by AWL with vocabulary coverage ( $10.90 \%$ ). The second (K-2) is the least frequently used word ( $5 \%$ ) indicating that it is below the necessary level ( $95 \%$ ) in the written essay. Keywords: Lexical Frequency Profile (LFP), VocabProfile (VP), productive vocabulary size, critical essays.

في هذه الدراسة قمنا Vocab Profile (VP) باستعمال لتقييم حجم المفردات الإنتاجية للطالب  المقالات النقدية للمتعلمين. وبالتالي، تم الحصول على 10 نصوص كتابية من واجبات منزلية، بعد ذلك


[^0]```
تم تحليل مقالات الطلاب بإدخالهم الي برنامج (VP) لإنشاء ملف تعريف مفردات وفقًا لعائلات الكلمات
```



```
التي يستخدمها الطالب تنتمي في الغالب إلى الأول (K-1) مع تغطية المفردات (77.20٪)، ثم يتبعها AWL (K) مردا (K)
مع تغطية المفردات (10.90٪). الكلمة الثانية (K-2) هي أقل الكلمات استخدامًا (5٪) تشير إلى أهنا أقل
                                    من المستوى الضروري (95٪) في المقال المكتوب:
الكلمات الرئيسيت: ملف تعريف التردد المعجمي (LFP)، ملف تعريف (Vocab (VP)، حجم المفردات
                                    الإنتاجية، المقالات النقدية.
```


## Introduction :

The purpose of written language assessments is routinely associated with measuring manifold dimensions of learner writing ability. Hence, the overall writing proficiency can be assessed by testing language aspects, this may include grammar, style, spelling, organisation, content, and vocabulary. Vocabulary as the main descriptor in written production, represents a building block that constitutes meaning in the process of language learning and it has particularly been proved as focal language ability in language learning skills or as this was stated that "language learning is largely lexical learning" (Gass \& Selinker, 2008)

Therefore, vocabulary size and range contribute to overall judgment of accuracy and writing quality. For instance, Engber (1995) argued that "the lexicon is an integral component in both the construction and interpretation of meaningful written text". For that reason, many researchers (Astika, 1993; McCarthy\& Carter ,1997) who investigated the different scoring criteria of writing rubrics to understand the degree of association between vocabulary and writing quality, found a positive correlation between the vocabulary bands and the overall scores. For example, Laufer and Nation (1995) through two studies found correlations between lexical measures and more holistic measures of quality in written text.

As of Fairly recently, computer-based analysis of written texts can provide feasible and solid grounds to rate the writing quality and make it possible for teachers and researchers to target learners' lexical needs, study lexical acquisition, and even provides them with an interactive resource to learn and to assess vocabulary knowledge in a piece of writing. Before (Laufer \& Nation 1995) the expansion of lexical frequency profile and VP, (Nation \& Heatley, 1996 \& Cobb,2002) -it was
an acutely hard and intricate task to measure students 'vocabulary size in written text. VP is an empirical database that represents an advanced online software accessible via the worldwide web that helps to analyze the level of vocabulary knowledge of a language user in terms of the vocabulary level. Part from (Laufer 1995; Nation 2001), VP was used by several other authors, including Morrison (1996) ${ }^{8}$ and adapted by Horst, Cobb, \& Nicolae, (2005), and Meara, Lightbown, and Halter (1997) For example, Sapa-asa (2006) analyzed the vocabulary through comparing between secondary school and university student essays. She found that university students use higher words from base 3 more than secondary students. This study was conducted to examine the vocabulary profile of third-year students in English critical essays. In the light of the focal points are highlighted, this research sought to answer the following questions:

- What is the productive vocabulary size of third-year students in critical essays?
- What percentage of the words in student essays the productive vocabulary size cover?


## 1. Literature Review

### 1.1. Vocabulary Profiling

Vocabulary profiling entails providing details about productive use of vocabulary in written text in terms of frequent bands of vocabulary (Laufer,1994; Nation, 2001) Nation's research agenda offers much insightful computational text analysis. VP is a simple online adaptation of Nation and Hwang's vocabulary assessment instrument developed by (Cobb, 2002) to analyse text and perform lexical analysis and lexical richness of these analysed texts. It was inspired by the RANGE program (Heatly, Nation, and Coxhead) incorporates the General Service List (GSL) of English Words, the Academic Word List (AWL), and the British National Corpus High-Frequency Word List (BNC HFWL). VocabProfile exits in website format Lextutor (www.lextutor.ca) helps to categorise the vocabulary used, it has two versions (1) VocabProfile BNC takes the text as raw input and describes range and diversity of vocabulary according to high and low frequency use and (VocabProfile Classic incorporate or compare targeted texts based on related word frequency list through breaking down the text into words families that
belongs to each band, types, and tokens and four frequency bands such as:

- Band 1 the most common 1,000 words in English (1-1,000)
- Band 2 the next most common 1,000 words in English (1,001-2,000)
- AWL the Academic Word List, is an updated version of the original "University Word List," (Coxhead, 2000) which contains 3,100 words belonging to 570 -word families which appear frequently in a wide range of academic textbooks
- Not in the lists less frequent words

The use of those bands is a fairly common practice in vocabulary studies. Thus, the end up of calculating process shows percentages that might look like bellow (Smith,2005, p.440)

## Band 1: 86.5\% Band 2: 7.0\% AWL: 3.5\% Nil: 3\%

Laufer and Nation (1995) introduced the VP as a valid instrument to assess lexical richness in written text using type-token ratio analysis to identify a number of words that appeared in the text. Laufer and Nation (1995) took a further step and contended that VP might be used to meter the productive vocabulary size of non-native speakers. They conducted (1995) that learners typically present different levels of proficiency and accordingly text provided a predictable function of learners' vocabulary profile. Consequently, emphasis on the following six strength of VP (Laufer \& Nation, 1995)

- Is a reliable and valid measure of lexical use in writing
- Provides similar stable results for two pieces of writing by one person
- Discriminates between learners at different proficiency levels
- Correlates with an independent measure of vocabulary knowledge
- Is a useful diagnostic test
- Is a sensitive research tool

However, Meara (2005) stated that VP has several drawbacks: Initially, related to how the text is processed, this means how the way errors counted (ignore or correct them) and how the proper nouns counted. another problem handled is that the VP analyses the text liable on discrete bands. Finally, a low percentage does not necessarily entail a few words in the text. In addition, Meara (2005) also points out that the VP does not work well with very low-level learners because the percentage of words that go beyond the first 2000 words of English is very low. Beforehand, Meara (1993) subjected the BBC course to VP
analysis. The findings revealed that learners' words belong to band 1 (1000 words) and she determined that the results would not be changed even when the course content moved from beginner to intermediate. Interestingly, she found that using the French language translated book of TinTin, learners were exposed to rich and varied vocabulary. Cobb \& Horst, (2005) and Cobb, \& Morris (2004) studies-based VP analysis of trainees' texts at Concordia, and teacher training establishments in Vietnam. The results demonstrated that success in a TESL program seems better predicted by the proportion of lexis drawn from post-1000 frequency zones than by other predictors. Therefore, VP analysis is regarded as part of an admission test for TESL trainees. Finally, Morris (2004) slightly modified version of Nation's profiler, developed by Cobb (2002), was used to establish a vocabulary profile for each of 140 Canadian TESL trainees. The findings determined Significant correlations were found between academic performance in the training program and K-1 (negative correlation), AWL (positive correlation), and AWL (positive correlation).

### 1.2. Measuring Productive Vocabulary Size

Nation (2001) outlines four diverse approaches to deciding how to count words such as tokens, types, lemmas, word families (Milton, 2009; Nation, 2001; Schmitt, 2000). According to Nation (2001), tokens or running words refer to every word that appears in a spoken or written text. If the same word occurs more than once, each occurrence is counted. For instance, in the sentence: "Taken from his books and his private letters", we can find eight tokens even when we have a word repeated. Contrary, Types, on other hand, consist of the number of different words or "types" that are present in an utterance. If we consider the example above, we have just seven types. Lemmas consist of the headwords, the most frequent infections, and reduced forms. Inflections consider plurals, third-person singular present tense, past tense, past participle, progressive aspect, comparative, superlative, and possessive (Nation 2001). For instance, the verb write includes writes, wrote, and writing; but not write, which is a noun and not a verb.

Additionally, Laufer and Nation (1995) introduce four common types of lexical variety in written text. Lexical Originality (LO) is the first type, which measures the percentage of words in a piece of writing that is used by the writer's performance compared to another writer.it is calculated in the bellow form (Laufer \& Nation, 1995)
$\mathrm{LO}=\frac{\text { number of tokens unique to one writer }}{\text { total number of tokens }}$
The second type is Lexical Density (LD), which measures the percentage of lexical words (nouns, verbs, adjectives, adverbs) in a piece of writing. Thus, the calculation form as follow: (Laufer \& Nation, 1995) $\mathrm{LD}=\frac{\text { number of Lexical tokens }}{\text { total number of tokens }}$

Lexical sophistication (LS) is the next type that measures the percentage of "advanced" words in a text. Therefore, the main calculation is based on dependent advanced words that exist in the text. (Laufer \& Nation, 1995)

$$
\mathrm{LS}=\frac{\text { number of advanced tokens }}{\text { total number of tokens }}
$$

The last type of lexical variation in text production is lexical richness, Lexical Variation (LV) measures the type/token ratio in a text, which means the ratio of the number of different words in a text (types) and the total running number of words (tokens) in a text. (Laufer \& Nation, 1995) ${ }^{5}$

$$
\mathrm{LS}=\frac{\text { number of types }}{\text { total number of tokens }}
$$

## 2. . Research Methods

### 2.1. Participants

This study involved a sample of third-year students of English who were selected randomly in the second semester from a total number ( 87 students) from three groups. It was convenient and easy to access the groups since I was tutoring the written expression module in the second semester.

### 2.2. Text Corpus Used in This Study

To run this study, students' essays were used and is aimed at measuring the productive vocabulary size of third-year university students' writing essay. Vocabulary or word choice plays a critically significant descriptor in writing an essay. During five weeks (approximately 20 hours) of instruction, I emphasized tutoring students the focal techniques, strategies, and guidelines to write different types of analytical essays including critical essays. The session was segmented as follows: one hour for instructions and one hour for whole-class
practice and then the students were asked to write assignments related to the suggested topic.

The topic was selected judiciously focused on the main aspect of the text material which is authenticity to reflect learners' experience:
"Distance learning-classroom platform - during Covid-19 global pandemic and its effect on learner's achievement". Write a critical analytical essay about the main theme of this statement.

### 2.3. Instrument

The student essays entered into VocabProfile (classic edition) analyser program was used for this study to reveal the numbers and percentages of words and word families in student essays coming from each of the word lists to measure of productive size in writing. VP was used as an instrument to analyse 10 written samples, compare vocabulary that existed in each essay, and provides a range or distribution figure (how many texts the word occurs in), a headword frequency figure (the total number of times the actual headword type appears in all the texts), a family frequency figure (the total number of times the word and its family members occur in all the texts), and a frequency figure for each of the texts the word occurs in.

### 2.4. Procedures and Data Analysis

After gathering the required written samples in which text length range from 200 to 400 words, all student samples were typed into documents and then converted into txt computer files to be analyzed in VocabProfile classic version by providing several words in the text, Type/Token Ratio. Therefore, we observed how many words of band list K1, band list K2, and band list AWL occurred in each group. Some words do not belong to any of these 3 bases off lists (are excluded to describe and analyse in this study). The qualitative analysis of the productive vocabulary size is used and the individual essay profiles of high-, average- and low-scoring participants from each participant and corpus are presented to answer the second question. Therefore, highfrequency words in a text a rich in vocabulary use in the text.

## 3. Findings and Discussion

### 3.1. Findings

After gathering corpus from 10 students who completed the essay on a given topic, we built up a profile of vocabulary input for selected
samples of third students studying at semesters two. Table 1 shows a thorough description and evaluation of word frequency of $\mathrm{K}-1, \mathrm{~K}-2$, AWL in terms of word families, tokens, and types (other 'off-list words are excluded in this list).

Table 1: The Overall Vocabulary Frequent Words

|  | Families | Types | Tokens | Percent |
| :--- | :---: | :---: | :---: | :---: |
| K1 Words (1-1000): | 263 | 370 | $\mathbf{1 8 0 5}$ | $\mathbf{7 7 . 7 3 \%}$ |
| K2 Words (1001-2000): | 53 | 65 | $\mathbf{1 3 1}$ | $\mathbf{5 . 6 4 \%}$ |
| AWL Words: | 94 | 128 | $\mathbf{2 5 3}$ | $\mathbf{1 0 . 9 0 \%}$ |
| Off-List Words: | $?$ | 82 | $\mathbf{1 3 3}$ | $\mathbf{5 . 7 3 \%}$ |
|  | 410 | 645 | 2322 | $100 \%$ |

Table 1 shows that a large majority of the vocabulary consists of K1 words totaling 77.73 \% then followed by The AWL cover the 10,90 $\%$ and K2 words with a much smaller percentage of $5,64 \%$ only. Students' essay-writing heavily relies on the 2,000 most frequently used words rather than an academically suitable vocabulary. Therefore, the students used words from frequency levels of AWL more than the K2 frequency level.

Table2: The Productive Vocabulary Text Profile Gathered from 10 Writing Samples

|  | Families-Tokens-Types (Percentage Coverage \%) |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathbf{1}$ | $\mathbf{2}$ | $\mathbf{3}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{6}$ | $\mathbf{7}$ | $\mathbf{8}$ | $\mathbf{9}$ | $\mathbf{1 0}$ |  |
| K1 | 72.45 | 79.3 | 68.8 | 81.9 | 75.0 | 76.92 | 75.9 | 80.8 | 80.71 | 78.76 |  |
| K2 | 5.26 | 5.8 | 7.7 | 4.8 | 6.5 | 7.14 | 4.3 | 5.1 | 3.86 | 6.19 |  |
| AWL | 14.24 | 9.78 | 13 | 7.89 | 13.5 | 12.64 | 12.7 | 10.66 | 9.00 | 9.44 |  |

Table 3 shows student's vocabulary profiles reveal the scores of all of the participants. All the students receive the highest score in K-1 ranged from 72 to $81 \%$. The lowest score with the maximum score is 7 $\%$. average scores were respectively recorded in K-2. Finally, the analysis of student essays revealed that they use mostly words belong to the third base AWL with word coverage between 9 to $14 \%$ and only one corpus 4 below this word coverage ( $7 \%$ ).

Table 3: Individual VocabProfile of Participants (10 written samples)

|  | Word frequency | Families | Types | Tokens |
| :---: | :---: | :---: | :---: | :---: |
| 1 | K1 | 95 | 111 | 234 |
|  | K2 | 12 | 13 | 17 |
|  | AWL | 37 | 41 | 46 |
| 2 | K1 | 87 | 103 | 219 |
|  | K2 | 11 | 14 | 16 |
|  | AWL | 20 | 22 | 27 |
| 3 | K1 | 58 | 64 | 106 |
|  | K2 | 9 | 10 | 12 |
|  | AWL | 17 | 18 | 20 |
| 4 | K1 | 85 | 105 | 218 |
|  | K2 | 13 | 13 | 13 |
|  | AWL | 18 | 19 | 21 |
| 5 | K1 | 66 | 74 | 150 |
|  | K2 | 11 | 11 | 13 |
|  | AWL | 21 | 22 | 27 |
| 6 | K1 | 62 | 68 | 140 |
|  | K2 | 10 | 10 | 13 |
|  | AWL | 21 | 22 | 23 |
| 7 | K1 | 75 | 85 | 220 |
|  | K2 | 11 | 12 | 14 |
|  | AWL | 22 | 25 | 29 |
| 8 | K1 | 92 | 114 | 251 |
|  | K2 | 10 | 10 | 12 |
|  | AWL | 23 | 25 | 28 |
| 9 | K1 | 91 | 117 | 267 |
|  | K2 | 15 | 17 | 21 |
|  | AWL | 24 | 27 | 32 |
| 10 | K1 | 85 | 110 | 230 |
|  | K2 | 11 | 12 | 14 |
|  | AWL | 22 | 25 | 28 |

Word families are one of the most telling criteria that are presented in table 3. The number of word families obtained concerning lists one and two show similar numbers (word families average between 95- and 60 ) for K-1 and (word families average between 10-to-17-) for K-2. However, if we take a look at the Academic word list, some differences
can be observed: third-year students seem to use more words belonging to list three in their third essays more than the second list (K-2): they go from an average of 20 to an average of 24 . There is therefore a meaningful evolution between the use of K-2 to AWL for all student essays. The results also indicate that only one essay includes more word families on average (37) in this type in their essay.

Table 3 also presents a type and token analysis; a token is any occurrence of a word form in the text or number of words in the text. As we can see that maximum running words exist in K1, which makes up (267) of the total words in the text. Then minimum words are in K2 that makeup between ( 12 to 21 ) of the running words in total words in the text. Here 21 to 27 words belong to AWL and It has been noted again that corpus 1 includes words that belong to this list more than other corpora not that represent 46 running words of the whole words. Students used K1 and AWL more than K-2 in their words.

Moreover, type is any word form that occurs once regardless of how many more times it might occur. Among 10 essays, three essays record-high average words belong to K1 makes up between 111 to 117 of the totals. In turn, word forms belong to K2 makeup between 10-to17 only-word forms in the text. Remarkably, the last corpus records average word form more than other forms ( 17 words). The maximum word forms are in AWL three making up 27-word forms.

### 3.2. Discussion

To summarise, the main research questions of this article were:

- What is the productive vocabulary size of third-year students in critical essays?
- What percentage of the words in student essays the productive vocabulary size cover?
After profiling all student essays using VP, the first thousand covered approximately $77.73 \%$ from 2322 words in written texts which indicates that K-1 covers the most words used by participants. It is evident that students use basic simple or "core" words belong to 1000 words in their essay. Admittedly, the results obtained allow us to conclude that students depend heavily on the K-1 and AWL more than $\mathrm{K}-2$. The results consistently acquainted with previous related studies that treated this issue (Douglas,2010; Breeze,2008) Nation (2001) asserts that the first 2000 most frequent words are domineering to language learning. The same view was supported by Milton (2009) who contends
that the instructional activities focus on the words that belong to K-1 and represent around $80 \%$ of a text coverage, while the next 2000 frequent words account for a mere $8 \%$ coverage.

Our result also proved that third-year students used vocabulary from AWL coverage more than word coverage belongs to K-2. Coxhead's (2000) academic words in the AWL are assumed to reach nearly $10 \%$ coverage of the whole words in general academic texts.in this regard, it is worth noting that although K-2 routinely indicates the lowest percentage, however, Breeze (2008) suggested that thus it would a lower percentage of words belongs to K-2 does not indicate poorer vocabulary "because it might merely indicate that the student is using a range of words that are far more sophisticated, or perhaps more specialized, than those contained in Nation's second thousand".

We see, therefore, that the qualitative approach used in this study to explores the productive vocabulary size using VP analyser program involves clues regarding FL writers' challenges especially when it is associated with vocabulary size. Consequently, many implications can be derived through this study; firstly, teaching vocabulary does not concentrate on teaching aspects of words rather than on receptive and productive vocabulary use (Webb \& Nation, 2017). Additionally, teachers need to decontextualize vocabulary that belong to different bands, which means creating opportunities for exposure to vocabulary through authentic language instructions.

## 4. Conclusion

In this study, we used VP program to analyse third year student vocabulary profile by exploring in particular type, ratio, and token. The results have shown that students vocabulary level slightly varied at each band, the percentage difference between the number of words found in each frequency band generally increased exponentially K-1 became increasingly less frequent in K-2 and AWL.

## References

Astika,G,G.(1993). Analytical Assessments of Foreign Students' Writing, RELC journal, Vol 24, Issue 1, https://doi.org/10.1177/003368829302400104

Breeze, R.(2008). "Researching Simplicity and Sophistication in Student Writing", International Journal of English Studies, vol.8, no. 1, January,2008,51-66, http://www.um.es/ijes.

Cobb, T .(2002). "Web VocabProfile" http://www.lextutor.ca/vp/], An Adaptation of Heatley, Nation \& Coxhead's Range Heatley, A., Nation, I.S.P. \& Coxhead, A. (2002). RANGE and FREQUENCY programs. Available at http://www.victoria.ac.nz/lals/staff/paulnation.aspx. Conference Proceedings. Tokyo: JALT

Douglas, S. R .(2010). Non-Native English-Speaking Students at University: Lexical Richness and Academic Success (Unpublished doctoral thesis), University of Calgary, Calgary, AB). doi:10.11575/PRISM/16571http://hdl.handle.net/1880/48195 doctoral thesis

Engber, C.A .(1995). "The relationship of lexical proficiency to the quality of ESL compositions", Journal of Second Language Writing, Volume .4, Issue 2, p151,https://doi.org/10.1016/1060-3743(95)900047

Gass, S. M., \& Selinker, Second language acquisition: An introductory course (3rd ed.), New York, NY: Routledge,2008, p. 173.

Laufer, B .(1994). The Lexical Profile of Second Language Writing: Does It Change Over Time?, Vol 25, Issue 2, https://doi.org/10.1177/003368829402500202

Laufer, B., \& Nation, P, "Vocabulary Size and Use: Lexical Richness in L2 Written Production", Applied Linguistics, vol. 16 ,no.3, 1995, ,307-322.

McCarthy, M. and Carter .(1997). Written and spoken vocabulary".Cambridge university press.United Kingdom

Meara, P.(1993). The bilingual lexicon and the teaching of vocabulary. In Schreuder and Weltens,279-297.

Meara, P.(2005). Lexical Frequency Profiles: A Monte Carlo analysis, Applied Linguistics, vol.26, 1-16.

Meara, P., Lightbown, P. and Halter, R.H., .(1993). Classrooms as lexical environments, Language Teaching Research, vol.1, no;1,1997,28-47.

Milton,J,.(2009). Measuring Second Language Vocabulary Acquisition, In: Second Language Acquisition, Multilingual Matter.

Morris, L., \& Cobb, T .(2004). Vocabulary profiles as predictors of the academic performance of teaching English as a second language trainees. System, vol.32, no.1, 75-87.

Morris,L, T ,Cobb,(2003).Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees

Morrison,L .(1996). Talking about words: a study of French as a second language learners' lexical inferencing procedures, Canadian Modern Language Journal 53, no 1,1996,41-75.

Nation, I.S.P .(2001). Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.

Schmitt,N .(2000). Vocabulary in Language Teaching Cambridge University Press, Cambridge.

Smith, R .(2004). The Lexical Frequency Profile: Problems and Uses,2005, In K. Bradford Watts, C. Ikeguchi, \& M. Swanson (Eds.) JALT2004.

System,vol. 32 ,no.1, 75-87
Webb,S \& Nation,P, .(2017). How Vocabulary is Learned, instructed language acquisition


[^0]:    * Corresponding author.

