Vocabulary Size and Use: Lexical Richness in L2 Written Production

BATIA LAUFER and PAUL NATION

University of Haifa   *Victoria University of Wellington*

This article shows that if there is some control over genre then there will be a close correspondence between the vocabulary size of intermediate learners as reflected in their writing and a more direct measure of vocabulary size. The study proposes a new measure of lexical richness, the Lexical Frequency Profile, which looks at the proportion of high frequency general service and academic words in learners' writing. The study shows that it is possible to obtain a reliable measure of lexical richness which is stable across two pieces of writing by the same learners. It also discriminates between learners of different proficiency levels. For learners of English as a second language, the Lexical Frequency Profile is seen as being a measure of how vocabulary size is reflected in use. In this study, it was found that the Lexical Frequency Profile correlates well with an independent measure of vocabulary size. This reliable and valid measure of lexical richness in writing will be useful for determining the factors that affect judgements of quality in writing and will be useful for examining how vocabulary growth is related to vocabulary use.

1. BACKGROUND

One of the major determinants of the vocabulary used in written production is the vocabulary size of the writer, particularly if the writer is a second language learner with a relatively small vocabulary compared with native speakers. Measures of lexical richness attempt to quantify the degree to which a writer is using a varied and large vocabulary. There has been interest in such measures for two reasons—they can be used to help distinguish some of the factors that affect the quality of a piece of writing, and they can be used to examine the relationship between vocabulary knowledge and vocabulary use. Let us look at each of these applications.

1.1 Lexical richness and the quality of writing

A well-written composition, among other things, makes effective use of vocabulary. This need not be reflected in a rich vocabulary, but a well-used rich vocabulary is likely to have a positive effect on the reader. Engber (1993), like Linnarud (1986), examined the relationship between various lexical measures of a piece of writing and holistic scores of writing quality. Engber found a correlation of .43 between lexical variation including errors, and a holistic measure of quality, and a correlation of .57 between lexical variation minus errors, and a holistic measure of quality. These correlations are substantial,

considering that lexical richness is only one of a variety of factors that affect the overall quality of a piece of writing. Engber's study suggests that it is worth helping and encouraging learners to bring their vocabulary knowledge into active use in writing.

There are many factors besides vocabulary size that could affect lexical richness in writing. These could include familiarity with the topic, skill in writing, and communicative purpose. This means, for example, that a change of topic could result in a marked change in lexical richness. If lexical richness is strongly affected by factors other than vocabulary size, and these factors cannot be controlled, it may prove impossible to obtain reliable measures of richness, making the measures of little use to researchers or teachers. It thus needs to be shown that it is possible to obtain reliable measures of lexical richness from different pieces of writing by the same learner. ·

1.2 *Vocabulary size and vocabulary use*

Vocabulary is not usually learned for its own sake. An important aim of a vocabulary program is to bring learners' vocabulary knowledge into communicative use. Where learners are in a situation where there are demands upon them to make use of what they know, we would expect to see a relationship between direct measures of learners' vocabulary size and the richness of vocabulary in their language production. Laufer (1991) has shown that development in lexical richness in writing can be measured over a period of 14 weeks and 28 weeks, particularly with learners below average proficiency. This shows that it is possible to isolate this factor of lexical richness and observe it change. Although it is clear from Laufer's study that there is a change in the range of vocabulary used, it is not clear if the change in lexical richness is the result of learning new vocabulary, activating previously known vocabulary, or being able to give more attention to vocabulary as a result of becoming more proficient in the writing skill. There is therefore a need to see if lexical richness in writing is related to more direct measures of vocabulary size.

Arnaud (1984) found a correlation of .36 between lexical variation and performance on a productive translation test from L1 to L2. In a subsequent study (Arnaud 1992), he found a correlation of .51 between a measure of lexical quality and performance on a vocabulary test, half of which involved translation (L2 into L1) and half of which involved multiple choice matching of L2 nouns with a picture. He found a correlation of .64 between the lexical quality measures on two sets of compositions by the same learners. This was an indication of the reliability of the measure using test–retest.

The present study surveys measures of lexical richness and proposes a new measure with the goal of determining how well this new measure, the Lexical Frequency Profile (LFP), reflects the vocabulary size of the learner as determined by an independent test. If the LFP can be shown to bear a relationship to vocabulary size, then it also has value as an indicator of quality of vocabulary use in that it can show the extent to which writers are making the fullest use of their available vocabulary knowledge.

2. SOME MEASURES OF LEXICAL RICHNESS

The most popular measures used in the description of the productive lexicon are lexical originality (LO), lexical density (LD), lexical sophistication (LS), and lexical variation (LV). Other less frequently employed measures are semantic variation (Mendelsohn 1981), lexical quality (Arnaud 1984, 1992), T-unit length and error free T-unit length (Cohen 1989). We will define each measure and will question its adequacy as a diagnostic and research tool.

Let us start with Lexical Originality which is the percentage of words in a given piece of writing that are used by one particular writer and no one else in the group:

$$LO = \frac{\text{Number of tokens unique to one writer} \times 100}{\text{Total number of tokens}}$$

The Lexical Originality index measures the learner's performance relative to the group in which the composition was written. If the group changes, the index changes too. In other words, the LO of a particular composition is unstable, since it is defined not only by the composition in question, but by the group factor. This detracts from its reliability.

Lexical Density is defined as the percentage of lexical words in the text, i.e. nouns, verbs, adjectives, adverbs:

$$LD = \frac{\text{Number of lexical tokens} \times 100}{\text{Total number of tokens}}$$

Since lexical words are the words which primarily convey information, a text is considered 'dense' if it contains many lexical words relative to the total number of words, i.e. lexical and functional words. However, we could argue that the Lexical Density index does not necessarily measure lexis, since it depends on the syntactic and cohesive properties of the composition. Fewer function words in a composition may reflect more subordinate clauses, participial phrases and ellipsis, all of which are not lexical but structural characteristics of a composition. As the LD measure is influenced by the number of function words, this affects its validity.

Lexical Sophistication is the percentage of 'advanced' words in the text:

$$LS = \frac{\text{Number of advanced tokens} \times 100}{\text{Total number of lexical tokens}}$$

What is labelled as 'advanced' would depend on the researcher's definition. To decide what vocabulary is advanced, it is necessary to take the learner's level into consideration. Thus, the lexis in the lexical syllabus of the last two school grades could be considered advanced for school students, but not necessarily for university students. The lexis of the two last school years may not be the same in different countries with different educational objectives and different amounts of instruction. Here lies the weakness of the LS measure. The same piece of writing may be analysed differently in terms of LS, depending on how

'advanced' vocabulary is defined. This makes the measure unstable. Second, as LS is determined by the researcher's definition of advanced or sophisticated words, its uses are limited. If one is evaluating a lexical syllabus in a country and checks how many advanced words (e.g. those met in the last two years of high school) have been learnt, then the LS measure may well be adequate for the purpose. If, however, one wants to conduct comparative research of groups of learners from different educational systems, then a more standardized definition of 'advanced', or 'sophisticated' vocabulary is needed.

Lexical Variation is the type/token ratio, i.e. the ratio in per cent between the different words in the text and the total number of running words:

$$LV = \frac{\text{Number of types} \times 100}{\text{Number of tokens}}$$

The type/token ratio has been shown to be unstable for short texts and can be affected by differences in text length; even more sophisticated formulas have been shown to be unsuitable for short texts like learners' essays, and one solution is to use samples of equal lengths. Even if we solve the problem of the sensitivity of Lexical Variation to essay length by drawing a fixed number of words, two problems remain unsolved. First, LV is dependent on the definition of a word. If derivatives are considered to be different words, as is usually done in this measure, LV will be higher than if a word family is considered to be one word. If each word form is counted as a different word, high lexical variation does not necessarily indicate rich vocabulary, in terms of the number of word families used. A learner who used many derived forms of a few families would not be distinguished from a learner who used a lot of different families.

LV also does not distinguish what kinds of words are used. In a composition of 300 tokens, for example, 200 word types could be used by someone who knows 2,000 words, or someone who knows 5,000 words. The 200 word types in the two compositions may be from different frequency levels, even though the number of word types is identical. What a high type/token ratio may show is how well a learner can express himself with the vocabulary he knows, not what types of words he knows. This is because LV distinguishes only between the different words used in a composition, but not between the quality of the different words as defined by their rarity. Since similar LV indices can reflect different vocabulary sizes in terms of lexical richness, this detracts from the validity of LV.

Other less used measures of lexical richness have weaknesses. Semantic Variation (Mendelsohn 1981), which is the number of types per topic, depends on how a topic, and a sub-topic is defined and where the line is drawn between elaboration of a topic or mere repetition of ideas. These decisions are difficult to standardize.

The Lexical Quality (LQ) formula (Arnaud 1992) is the sum of the number of types and 'rare' words minus twice the number of lexical errors. As stated earlier, 'rare' words are defined differently in different systems. The inclusion of errors in the measure of lexical quality was motivated by the effect produced on

the reader, but it is problematic if one is interested essentially in the writer's vocabulary knowledge. Firstly, no differentiation has been made between error types and error tokens. A learner who repeated the same error 5 times and a learner who made 5 different errors were 'penalized' identically in the calculation of their lexical quality. Secondly, it is often hard to distinguish between lexical and non-lexical errors. For example, *they are responsible education' can be interpreted as a grammatical error—omission of a preposition 'for', a lexical error—not knowing the collocation of 'responsible', or a global syntactic error—incompatibility between subject and subject complement. Third, if we are interested in lexical richness, it is sensible to regard words that are not used correctly as not being a part of the learner's lexical knowledge and assess vocabulary size without them rather than subtracting them from the existing vocabulary as the LQ formula does.

T-unit length and error free T-unit length, which are sometimes used as indicators of lexical richness, are problematic for several reasons. Firstly, the measure, which is a main clause together with subordinate clauses, takes into account a syntactic property of writing—subordination. Secondly, a definition of a main clause, like that of topics and sub-topics, is not necessarily objective and therefore the measure is not objective either. Thirdly, the length of the T-unit may reflect verbosity rather than lexical richness. Lastly, both main clause and subordination are sometimes hard to identify in the writing of low-proficiency learners. The errors make it hard to determine where one unit starts and another begins and also whether subordination is the appropriate way of expressing the learner's intention or not. In other words, the calculation of T-unit length involves the subjective interpretation of the reader. It is because of the above limitations of the various measures that we propose a different measure of lexical richness of written production—the Lexical Frequency Profile (LFP).

3. LEXICAL FREQUENCY PROFILE
3.1 *What is the LFP?*
The LFP shows the percentage of words a learner uses at different vocabulary frequency levels in her writing—or, put differently, the relative proportion of words from different frequency levels. We suggest two different LFP measures: one for less proficient students, the other for advanced students. Since the less proficient learner is less likely to use rare vocabulary than the advanced one, the main distinction, in the case of the less proficient learners, should therefore be between the first 1,000 most frequent words, the second 1,000, and any other vocabulary. For more advanced learners, a finer distinction should be made above the basic vocabulary. Therefore, the profile could look at the total number of word types of the second 1,000 most frequent words, the academic vocabulary (UWL—University Word List: see below), and the less frequent words, i.e. words that are not in the first 1,000 most frequent words and not in any of the above two lists. The three categories of second 1,000, UWL, and words not in any lists would constitute the 100 per cent of the count. The LFP is

calculated as follows. Let us imagine a composition of an intermediate learner which consists of 200 word families. Among the 200, 150 belong to the first 1,000 most frequent words, 20 to the second 1,000, 20 to the UWL, and 10 are not in any list. To calculate the LFP, we convert these numbers (the number of word families at each frequency level) into percentages out of the total of 200 word families. The LFP of the composition is therefore 75%–10%–10%–5%. · The UWL is a list of 836 word families containing vocabulary that is not in the first 2,000 words of English, but which is frequent and wide range across a variety of written academic texts from a variety of disciplines. The list can be found in Nation (1990) and in Xue and Nation (1984).

The entire calculation is done by a computer program which compares vocabulary lists against a text that has been typed in (without lemmatization) to see what words in the text are and are not in the lists and to see what percentage of the items in the text are covered by the lists. The VocabProfile package consists of the program itself and three accompanying word lists. The program compares the words in a text in ASCII format with the words in the word lists. It compares the words in a text and lists the words from the text in types and families marks the words in the text and lists the words from the text in types and families according to the list they occurred in. It also provides frequency and coverage data. The words in the lists that accompany the program are arranged under head words with derived forms listed below them indented by a TAB, for example:

```
push
    pushed
    pushes
    pushing
```

A word is defined in the program as a base form with its inflected and derived forms, i.e. a word family. The program can calculate the LFP on the basis of word tokens, word types, or word families. It is the latter calculation that we consider more revealing as an indication of lexical richness, because it uses a definition of what should be counted as a word which most closely matches how learners view words. That is, learners at the levels of proficiency involved in this study have no difficulty in seeing that *happy, happiness, happyish, happily,* and *unhappy* are closely related. The levels described in Bauer and Nation (1993) were used in this study, with each word family being at level 3 on their scale. Level 3 includes inflections and the affixes *-able, -er, -ish, -less, -ly, -ness, -th, -y, non-,* and *un-*. So if *governable* or *ungovernable* occurred in a text, they would be counted as part of the word family *govern. Government* would be counted as a different word family because the suffix *-ment* is beyond level 3 on Bauer and Nation's scale.

3.2 The methodological advantages of the LFP
The LFP has several advantages over the other measures of lexical richness mentioned above. The LFP provides a more objective tool than Lexical Originality since it is independent of the learner's environment and therefore

does not change with the change of group. Unlike Lexical Density, LFP is largely independent of syntax and text cohesiveness. The measure focuses on lexis proper and is therefore more suitable for assessing lexical richness.

Since the Lexical Sophistication score claims to show how advanced the learner's vocabulary is, it can be argued that it is not very different from the LFP. There are, however, two important differences. First, the LFP provides a more detailed picture of the different types of words (as defined by frequency levels), unlike Lexical Sophistication which distinguishes only between two types of words—frequent and sophisticated. Second, because of different definitions of 'sophisticated' vocabulary, LS is unsuitable for comparing groups from different learning environments. The 'sophisticated' vocabulary of the LFP is defined in terms of word frequency, not in terms of a lexical syllabus or words from reading material. Therefore, it can be used when comparing groups from different educational systems, or groups acquiring language outside the classroom.

The LFP has advantages over Lexical Variation. The LFP will discriminate between subjects who use frequent and less frequent vocabulary, not just between those who can or cannot vary their possibly limited vocabulary.

.The LFP is also more reliable than other less frequent measures as it is free of subjective decisions regarding what is a topic, sub-topic, elaboration, and thematic unit. A word used incorrectly is not considered to be part of the learner's lexicon. Let us now look at how these advantages work out in practice.

4. THE STUDY
4.1 Aims
The study aims at establishing the reliability and the validity of the Lexical Frequency Profile as a measure of lexical richness in free written production. The profile can be considered reliable if it remains stable in different samples of writing produced by the same subject. In other words, if the profile does not change with the change of composition topic, we can claim that any properly collected sample of writing is likely to be representative of a learner's lexical richness in a particular kind of written performance. We expect the LFP to be the same in different samples collected at the same stage of learning. It is expected that at a more advanced level, the samples will be richer in lexis. One measure of validity is the correlation with other established measures of lexical knowledge, like the Vocabulary Levels Test (Nation 1983) in its active and even passive versions, although only an active version was used in this study (see Appendix 2). Even though the original Levels Test measures receptive knowledge, subjects with higher lexical profiles should also score higher on the Levels Test, as one cannot use a word correctly without understanding it, or without being able to provide it in a word-elicitation task. Good and significant correlations between the LFP and the Levels Test will demonstrate concurrent validity. The LFP measure can also be considered valid if it distinguishes beween different levels of language proficiency since lexical richness is a part of language proficiency. Another way of demonstrating the validity of the measure is to see if the LFP can reveal the differences between the proficiency levels.

4.2 Research questions

a. Will there be a significant difference between the LFPs of learners of different language proficiency levels?
b. Will the LFP of the compositions correlate highly with the scores of the same learners on the active version of the Vocabulary Levels Test?
c. Will the LFPs in two sets of compositions written by the same learners correlate highly with each other?
d. Will the percentages of words at each frequency level correlate highly with each other in the two sets of compositions?

Questions a. and b. address the issue of LFP validity; questions c. and d., the issue of LFP reliability.

4.3 Subjects

The subjects were foreign learners of English in New Zealand and Israel. The New Zealand group consisted of 22 learners in Victoria University. They were enrolled, at the time of the experiment, in an English for Academic Purposes course. They were native speakers of Chinese, Japanese, Thai, Samoan, Polish, Malay, Russian. The English proficiency levels of the learners were measured by an in-house placement test and were found to be at a low intermediate level. The Israeli subjects were students in the department of English Language and Literature. They were all graduates of Israeli high schools, which require a level equivalent to the Cambridge First Certificate in English, and had passed an entrance exam to the department, which involved production of a written essay. Of these students, 20 were in their first semester in the department, and 23 had completed two semesters. Altogether, there were 65 subjects. Thus, we had three proficiency levels in the experiment. The lowest proficiency group (group 1) was the New Zealand group, the next (group 2) was the first-semester Israeli group, and the highest one was the end-of-two-semesters group (group 3).

4.4 Procedure

4.4.1 *Data collection.* Two compositions were written by the subjects during class time in one week. The short time intervals between the compositions meant that the language level of the learners would not change to a significant degree. The learners were given one hour to complete each composition. The required length of each composition was 300–350 words which was feasible within the available time. Profiles over 200 words were found to be stable, while those done on less than 200 words were not.

The topics of the compositions were of a general nature and dealt with controversial issues. None of the topics required expert knowledge of particular subject matter. (The topics are listed in Appendix 1.) The topic of the first composition was identical for all learners. In the other composition, the learners could choose one topic out of three. The choice of topics was offered in order to maximize the learners' interest in the content of their essays. Throughout the experiment, the motivation of the learner was ensured by making the two

compositions part of their regular class work, grading them and, for the Israeli students, including the grades in the final course grade.

The learners in the experiment were also given the 'active version' of the Vocabulary Levels Test which elicited the use of target words at 5 frequency levels in given sentences (the second 1,000 words, the third 1,000, the fifth 1,000, the University Word List, and the tenth 1,000). The target words were equivalent to those in the original Levels Test (Nation 1983). To prevent the possibility of a non-target word being filled in a sentence, the first two or three letters of the desired word were provided (see Appendix 2).

4.4.2 *Data processing.* All compositions were entered into the computer. As the composition length for data analysis was 300 word tokens, only the first 300 running words were entered for compositions which were longer than that. (In practice, most of the compositions were around 300 word tokens long.) When a word was clearly used incorrectly, it was omitted, as it could not be considered as part of the subject's productive lexicon. This did not occur often. If, on the other hand, it was used correctly but misspelled, the error was corrected and the word was considered as familiar to the subject. A wrong derivative of a word was not considered an error since all the derivatives forming one word family have the same frequency. Proper nouns were deleted from the samples.

The following lists were produced for each composition: the first 1,000 most frequent words, the second 1,000, the University Word List, and the 'not-in-the-lists' word list. For each composition, the LFP was calculated, on the basis of the proportion of word families at each of these four levels. The computer program which was developed to do this is an IBM compatible program called VocabProfile and is available free with its accompanying word lists to interested researchers from the second author.

The weakness of the program is in its not distinguishing between homonyms. However, a manual analysis of a sample of essays showed that the average number of homonyms per paper was between 2 and 3. Therefore, the imprecision created by the lack of distinction was minimal.

5. RESULTS AND DISCUSSION

Let us address the issue of validity first and answer the research questions pertinent to it:

a. Will there be a significant difference beween the LFPs of different language proficiency levels?
b. Will the LFP of the compositions correlate highly with the scores of the same learners on the active version of the Vocabulary Levels Test?

Table 1 presents the mean percentages of words at different frequency levels which were used by the three groups of learners. It also shows the results of an ANOVA (comparison of means of the three groups at each frequency level) and the degree of the significance of the F test. The means were considered different

Table 1: Mean percentages and standard deviations of word families at different frequency levels

	1st 1,000		2nd 1,000		UWL		Not in lists	
	Comp1	Comp2	Comp1	Comp2	Comp1	Comp2	Comp1	Comp2
Group 1	86.5	87.5	7.1	7.0	3.2	4.1	3.3	2.8
SD	3.8	5.3	2.0	2.3	1.8	2.5	2.3	1.8
Group 2	79.7	79.4	6.7	6.8	8.1	7.8	5.6	6.6
SD	5.3	4.5	1.7	2.2	2.3	2.3	3.5	3.3
Group 3	77.0	74.0	6.6	5.6	8.1	10.1	7.5	· 8.7
SD	6.1	5.9	2.6	2.5	3.2	2.9	2.9	3.5
F-test	19.35	33.1	0.29	1.89	24.86	27.40	10.46	22.74
p value	.0001	.0001	.75	.16	.0001	.0001	.0001	.0001

from each other when the *p* value, i.e. the significance level of their difference, was not higher than .05.

As can be seen from Table 1, the three proficiency groups of learners were found to be significantly different from each other in the percentage of the first 1,000 most frequent word families. In a *post hoc* analysis of the ANOVA (using the Duncan procedure), the following differences were observed. In composition 1, the first group was different from groups 2 and 3. The subjects in group 1 used considerably more words of the highest frequency. The trend was towards group 2 using more of the highest frequency words than group 3. In composition 2, the three groups were more clearly differentiated with the score of each group being significantly different from those of the other two groups. Group 1 had the largest number of the first 1,000 words, and group 3 the smallest. As for the percentage of the second 1,000 words, the three groups were not significantly different from one another in both compositions, although there was a consistent trend for the less proficient groups to use more of the second 1,000. With regard to the UWL, in the first composition, group 1 used significantly fewer words of this type than the two others. In composition 2, the three groups were different from one another. Group 1 used the smallest number of UWL words, group 3 the largest. The three groups were also different from one another in the use of 'not-in-the-lists' words. This was true for both compositions. These results show the following: the less proficient students make more use of the first 1,000 most frequent words; there is a tendency for the less proficient to make more use of the 'middle level' vocabulary (the second 1,000); the significant differences emerge again with the more sophisticated vocabulary, the UWL and the 'not-in-the-lists' words. These differences are in accordance with the concept of language proficiency which assumes that richer vocabulary is characteristic of better language knowledge. If the LFP has tapped these differences, this is evidence for its validity.

Research question b. addressed the concurrent validity of the LFP measure. The Levels Test is the best available measure to use to estimate concurrent validity, as it is based on similar vocabulary levels. Read (1988) showed that it can produce a satisfactory degree of implicational scaling of the levels.

The composite grade on the Active Levels Test was correlated with each component of the LFP. It was necessary to use a composite grade, because the levels test had no levels corresponding to the first 1,000, and 'not-in-the-lists'. The correlations are presented in Table 2. As can be seen from the results,

Table 2: Correlations between the LFP and the Levels Test

	Not in lists		UWL		2nd 1,000		1st 1,000	
	Comp1	Comp2	Comp1	Comp2	Comp1	Comp2	Comp1	Comp2
Levels Test/LFP	.6	.8	.7	.6	.01	.2	−.7	−.7
p value	.0002	.0001	.0001	.001	.9	.3	.0001	.0001

learners who got higher scores on the Levels Test used more of the more sophisticated vocabulary (UWL and 'not-in-the-lists'). There was a negative correlation between the Levels Test and the first 1,000, i.e. the higher one's vocabulary knowledge, the fewer high frequency words were used in a composition. There was no correlation between the Levels Test and the second 1,000, i.e. the 'middle level' vocabulary was used as frequently by learners with low vocabulary size as by those with a larger vocabulary size.

To answer question c., on the reliability of the measure, a within-subject analysis was carried out on the two sets of compositions that each student wrote. The comparison was done first separately for each word frequency level and then for the proportions among the frequency levels. A change was introduced for group 1 (the low-proficiency group). Since the number of 'not-in-the-lists' words was very small, it was added to the UWL words and a composite percentage was compared to the other levels. Table 3 presents the significance values of the various comparisons: matched *t*-tests for the individual levels and MANOVA for comparing the proportions between the levels.

As can be seen from Table 3, groups 1 and 2 exhibited stable profiles in the two compositions (none of the differences was significant), while in group 3 there was a difference in the first 1,000, the UWL, and the proportions. These results suggest that the LFP is stable except for the advanced learners whose vocabulary apparently becomes too varied to remain stable across different samples of writing.

Another way of analysing the same results was tried out following the argument that since the first 1,000 words include almost all the function words and the most basic lexical words, they are not an indication of a developed

Table 3: Comparison of LFP in the two compositions

	1st 1,000 *t*-paired	2nd 1,000 *t*-paired	UWL· *t*-paired	Not in lists *t*-paired	MANOVA *F*
Group 1	0.83 $p = 0.42$	2.01 $p = 0.06$	0.27 $p = 0.79$		1.49 $p = 0.25$
Group 2	0.83 $p = 0.42$	0.47 $p = 0.65$	0.48 $p = 0.64$	1.77 $p = 0.10$	0.65 $p = 0.64$
Group 3	2.70 $p = 0.01$	0.25 $p = 0.81$	3.12 $p = 0.005$	1.59 $p = 0.13$	4.15 $p = 0.009$

lexicon, but a *sine qua non* for written expression. Therefore, the true lexical quality of a piece of writing is determined by the proportion of all the other words at the more advanced frequency levels. The LFP was then treated as a proportion between the second 1,000, the UWL and 'not-in-the-lists' words (for group 1, between the second 1,000 and the combination of UWL and 'not-in-the-lists' words). The results were reanalysed. A total of 100 per cent was taken to be the number of all words except those in the first 1,000. The new profile consisted of the percentage of the words of the second 1,000, the percentage of the UWL words, and the percentage of 'not-in-the-lists' words, each percentage out of the new total of 100 per cent. The same statistical procedures as with the original LFP were carried out. Results are presented in Table 4. In this form of analysis, none of the differences between the two essays is significant, showing that the LFP is stable between two compositions. It may be that this profile is a better measure of lexical richness in general, or that it is better for the advanced students in our sample.

As the computer program easily allows for different lists to be used, we may attempt to produce a different profile for different levels of language proficiency. For example, the LFP for post Cambridge FCE level (the level equivalent to group 3 in our sample) could consist of the second 1,000 words,

Table 4: LFPs of two compositions (without the 1st 1,000)

	2nd 1,000 *t*-paired	UWL *t*-paired	Not in lists *t*-paired	MANOVA *F*
Group 1	2.64 $p = 0.12$	2.64 $p = 0.12$		2.64 $p = 0.12$
Group 2	1.11 $p = 0.3$	0.64 $p = 0.43$	0.07 $p = 0.79$	0.61 $p = 0.55$
Group 3	0.03 $p = 0.86$	1.51 $p = 0.25$	2.83 $p = 0.13$	1.27 $p = 0.33$

the third 1,000, the UWL, and 'not-in-the-lists'; or alternatively, the second 1,000 and third 1,000, the UWL, the fifth 1,000 and 'not-in-the-lists'. When considering the use of other lists, it is necessary to use lists that are based on some objectively determined measure of frequency and range, otherwise the LFP will be open to the criticisms of some of the other lexical richness measures.

6. CONCLUSIONS

The Lexical Frequency Profile has been shown to be a reliable and valid measure of lexical use in writing. It provides similar stable results for two pieces of writing by the same person, and discriminates between learners of different proficiency levels. It correlates well with an independent measure of vocabulary knowledge. Its main strength is that as a measure it focuses directly on lexis, putting aside at least to some degree, the influence of grammar. It has the additional advantage of being almost completely computerized. Except for the preparation of the written texts as computer input, all the analysis and the production of the profile is done by the computer. It provides a detailed picture of vocabulary use over several levels and is thus a useful diagnostic tool as well as a sensitive research tool.

The study also shows that we can reasonably expect learners' vocabulary size as measured by a vocabulary test to be reflected in the learners' productive use of the language. Where this does not happen, we need to look with concern at the opportunities that the learner has to bring language knowledge into use and adjust the language programme accordingly.

REFERENCES

Arnaud, P. J. L. 1984. 'The lexical richness of L2 written productions and the validity of vocabulary tests' in T. Culhane, C. Klein-Braley, and D. K. Stevenson (eds.) 1984: *Practice and Problems in Language Testing.* Colchester: University of Essex Occasional Papers No. 29.

Arnaud, P. J. L. 1992. 'Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate component tests' in P. J. L. Arnaud and H. Béjoint (eds.) 1992: *Vocabulary and Applied Linguistics.* London: Macmillan.

Bauer, L. and I. S. P. Nation. 1993. 'Word families.' *International Journal of Lexicography* 6/3: 253–79.

Cohen, A. D. 1989. 'Attrition in the productive lexicon of two Portuguese third language speakers.' *Studies in Second Language Acquisition* 11: 135–49.

Engber, C. 1993. 'The relationship of lexis to quality in L2 compositions.' Paper delivered at TESOL '93.

Laufer, B. 1991. 'The development of L2 lexis in the expression of the advanced language learner.' *Modern Language Journal* 75/4: 440–8.

Linnarud, M. 1983. 'On lexis: the Swedish learner and the native speaker compared' in K. Sajavaara (ed.) 1983: *Cross Language Analysis and Second Language Acquisition.* Jyväskylä: Jyväskylä Cross Language Studies 11.

Linnarud, M. 1986. *Lexis in Composition.* Lund: Lund Studies in English.

Mendelsohn, D. 1981. 'We should access lexical richness, not only lexical errors.' Paper delivered at TESOL '81.

Nation, I. S. P. 1983. 'Teaching and testing vocabulary.' *Guidelines* 5/1: 12–25.

Nation, I. S. P. 1990. *Teaching and Learning Vocabulary.* New York: Heinle and Heinle.

Read, J. 1988. 'Measuring the vocabulary knowledge of second language learners.' *RELC Journal* 19/2: 12–25.

Xue Guoyi and Nation, I. S. P. 1984. 'A university word list.' *Language Learning and Communication* 3/2: 215–29.

Ure, J. 1971. 'Lexical density and register differentiation' in J. E. Perren and J. L. M. Trim (eds.) 1971: *Application of Linguistics.* Cambridge: Cambridge University Press.

APPENDIX 1

Essay topics

The common topic:

'Should a government be allowed to limit the number of children a family can have?' Discuss this idea considering basic human rights and the danger of population explosion.

Essay two: one topic to be chosen out of three:

'A person cannot be poor and happy, because money is always needed to gain something that is important to that person.' Argue for and against this idea.

'It is always what you do not have as a child that is important to you as an adult.' Agree or disagree with this statement.

'In a free country, industry has the right to develop any product that will sell, and industry can sell it to anyone who can pay for it.' Do you agree with this idea or do you think that the government should be able to control what is produced and sold?

APPENDIX 2

The active version of the Vocabulary Levels Test

Complete the underlined words. The first one has been done for you.

The 2,000-word level

He was riding a bicycle.
They will restore the house to its orig_____ state.
Each room has its own priv_____ bath and WC.
The tot_____ number of students at the university is 12,347.
They met to ele_____ a president.
Many companies were manufac_____ computers.
The lakes become ice-free and the snow mel_____.
They managed to steal and hi_____ some knives.
I asked the group to inv_____ her to the party.
She shouted at him for spoi_____ her lovely evening.
You must spend less until your deb_____ are paid.
His mother looked at him with love and pri_____.
The wind roa_____ through the forest.
There was fle_____ and blood everywhere.
She earns a high sal_____ as a lawyer.

The sick child had a very high tempe_____.
The bir_____ of her first child was a difficult time.
My favourite spo_____ is football.
In A.D. 636 an Arab army won a famous vic_____ over another army.

The 3,000-word level

They need to spend less on adminis_____ and more on production.
He saw an ang_____ from Heaven.
The entire he_____ of goats was killed.
Two old men were sitting on a park ben_____ and talking.
She always showed char_____ towards those who needed help.
He had a big house in the Cape Prov_____.
Oh Harold dar_____, I am sorry. I did not mean to upset you.
Judy found herself listening to the last ec_____ of her shoes on the hard floor.
He cut three large sli_____ of bread.
He sat in the shade beneath the pa_____ trees.
He had a crazy sch_____ for perfecting the world.
They get a big thr_____ out of car-racing.
At the beginning of their journey they encoun_____ an English couple.
Nothing illus_____ his selfishness more clearly than his behaviour to his wife.
He took the bag and tos_____ it into the bushes.
Every year she looked forward to her ann_____ holiday.
There is a defi_____ date for the wedding.
His voice was loud and sav_____, and shocked them all to silence.

The 5,000-word level

Some people find it difficult to become independent. Instead they prefer to be tied to their mother's ap_____ strings.
After finishing his degree, he entered upon a new ph_____ in his career.
The workmen cleaned up the me_____ before they left.
On Sunday, in his last se_____ in Church, the priest spoke against child abuse.
I saw them sitting on st_____ at the bar drinking beer.
Her favourite musical instrument was a tru_____.
The building is heated by a modern heating appa_____.
He received many com_____ on his dancing skill.
People manage to buy houses by raising a mor_____ from a bank.
At the bottom of a blackboard there is a le_____ for chalk.
After falling of his bicycle, the boy was covered with bru_____.
The child was holding a doll in her arms and hu_____ it.
We'll have to be inventive and de_____ a scheme for earning more money.
The picture looks nice; the colours bl_____ really well.
Nuts and vegetables are considered who_____ food.
The garden was full of fra_____ flowers.
Many people feel depressed and gl_____ about the future of mankind.

The University Word List level

The afflu_____ of the western world contrasts with the poverty in other parts.
The book covers a series of isolated epis_____ from history.
Farmers are introducing innova_____ that increase the productivity per worker.

They are suffering from a vitamin defic_____.
There is a short term oscill_____ of the share index.
They had other means of acquiring wealth, pres_____, and power.
The parts were arranged in an arrow-head configu_____.
The learners were studying a long piece of written disco_____.
People have proposed all kinds of hypot_____ about what these things are.
The giver prefers to remain anony_____.
The elephant is indig_____ to India.
You'll need a mini_____ deposit of $20,000.
Most towns have taken some eleme_____ civil defence precautions.
The presentation was a series of sta_____ images.
This action was necessary for the ulti_____ success of the revolution.
He had been expe_____ from school for stealing.
The lack of money depressed and frust_____ him.
The money from fruit-picking was a suppl_____ to their regular income.

The 10,000-word level

He wasn't serious about art. He just da_____ in it.
Her parents will never acq_____ to such an unsuitable marriage.
Pack the dresses so that they won't cre_____.
Traditionally, men were expected to nu_____ women and children.
Religious people would never bl_____ against God.
The car sk_____ on the wet road.
The politician delivered an arrogant and pom_____ speech.
The Romans used to hire au_____ troops to help them in their battles.
At the funeral, the family felt depressed and mo_____.
His pu_____ little arms and legs looked pathetic.
A vol_____ person will change moods easily.
The debate was so long and tedious that it seemed int_____.
Drink it all and leave only the dre_____.
A hungry dog will sa_____ at the smell of food.
The girl's clothes and shoes were piled up in a ju_____ on the floor.
Some monks live apart from society in total sec_____.
The enemy suffered heavy cas_____ in the battle.
When the Xmas celebrations and rev_____ ended, there were plenty of drunk people
  everywhere.