A vocabulary-size test of controlled productive ability

Batia Laufer *English Department, University of Haifa*
Paul Nation *English Language Institute, Victoria University of Wellington*

It is important in the design of the vocabulary component of a teaching program that teachers are able to discover the state of their learners' vocabulary knowledge. It is also important that researchers can draw on a variety of vocabulary measures to investigate the nature of vocabulary growth. This study focuses on a controlled production measure of vocabulary consisting of items from five frequency levels, and using a completion item type like the following.

The garden was full of fra____ flowers.

The controlled-production vocabulary-levels test was found to be reliable, valid (in that the levels distinguished between different proficiency groups) and practical. There was a satisfactory degree of equivalence between two equivalent forms of the test.

I Vocabulary testing

Vocabulary knowledge is considered by both first-language and second-language researchers to be of great significance in language competence (Grabe, 1991; Frederiksen, 1982) and vocabulary testing is now receiving the attention it deserves, with studies of the construct validity of some vocabulary tests (Chapelle, 1994; Perkins and Linville, 1987), examination of the effectiveness of particular item types (Henning, 1991; Laufer and Nation, 1995), and a comprehensive examination of the field of vocabulary testing in preparation (Read, forthcoming). The present study attempts to contribute to this knowledge.

This increased interest can be interpreted to mean that there is considerable value in gaining knowledge about specific parts of language learners' proficiency because it can be used effectively for diagnostic,

Address for correspondence: Paul Nation, English Language Institute, Victoria University of Wellington, PO Box 600, Wellington, New Zealand; e-mail: Paul.Nation@vuw.ac.nz

Language Testing 1999 16 (1) 33–51 0265-5322(99)LT160OA

placement and curriculum-design purposes. For example, the Vocabulary Levels Test (Nation, 1983; 1990) has proved to be useful in helping teachers to determine the kind of attention they should be giving to vocabulary for particular groups of learners. This is an important decision in terms of the cost-effectiveness of the use of class time, because high-frequency vocabulary development requires a different program from low-frequency vocabulary development. Following is a sample of three items tested by the Vocabulary Levels Test. The learners have to match three of the six words on the left with the meanings given on the right.

1	business		
2	clock	6	part of a house
3	horse	3	animal with four legs
4	pencil	4	something used for writing
5	shoe		
6	wall		

Similarly, Meara's Eurocentres Vocabulary Size Test (Meara and Buxton, 1987) has been very useful as a quickly administered placement test. Vocabulary tests are useful diagnostic and placement tests largely because they tap very important 'enabling' knowledge and they test a large number of items at one time, allowing reliable decision-making.

An important finding arising from the research on vocabulary-item types is that different item types appear to be tapping different aspects and degrees of vocabulary knowledge (Paul *et al*., 1990). It has long been recognised that there are many dimensions to 'knowing a word' (Richards, 1976; Nation, 1990), and many degrees of knowledge. The receptive/productive distinction is the best known of these. Other aspects of vocabulary knowledge include, for example, collocations, associations, use in context and related meanings. In order to gain a rounded picture of a learner's vocabulary knowledge, it is necessary to have a range of vocabulary measures to draw on.

A variety of vocabulary measures is useful diagnostically to see if particular aspects of vocabulary knowledge are being neglected in a course. For example, learners who gain a high score in a Vocabulary Levels Test, but do not use the full richness of their vocabulary in writing, may need encouragement and well designed tasks to help them draw more readily on what they know.

There is also considerable value in having a range of well-thought-out vocabulary measures to draw on in research on vocabulary learning. Joe (1994), for example, in her study of vocabulary learning from

retelling tasks, used several measures of the same vocabulary and was thus able to measure degrees of strength of knowledge of the words in the study and relate this to the conditions under which they were learned.

II Vocabulary frequency levels

The test format employed in this study was used previously in an examination of lexical richness in writing (Laufer and Nation, 1995). The main idea behind the Vocabulary Levels Test (Nation, 1983; 1990) is that it is useful to view the vocabulary of English (and indeed any language) as consisting of a series of levels based on frequency of occurrence. For example, these levels could consist of groups of 1000 words made up of the most frequent 1000 words of the language, the next most frequent 1000 words of the language and so on. There are several compelling reasons why it is useful to view vocabulary in this way.

First, even a glance at the relative frequency of use of different words shows that there are striking differences between them. The word *the* accounts for 7% of the running words in written texts. The most frequent 10 words account for around 25% of the running words in spoken and written use. The most frequent 1000 words account for around 75% of the running words in formal written texts and around 84% of informal spoken use. By contrast, the tenth 1000 most frequent words account for much less than 1% of the running words in a text.

Second, there is a very large number of words in English (Goulden *et al*., 1990) and it is far beyond the goals of any language course to give attention to anything but a very small proportion of these words. It is, thus, necessary to choose carefully what words to focus on.

For these reasons a distinction is made between the high frequency words of the language, as represented by the most frequent 2000 words (Nation and Hwang, 1995), and the large number of low frequency words of the language. This distinction is essentially a cost–benefit distinction. The cost is the time and effort to teach and learn the words. The benefit is the number of opportunities to use the words, as represented by the frequency of the words. All things being equal, words should be learned roughly in order of their frequency of occurrence, with high frequency words being learned first. To avoid the ludicrous results of a strict interpretation of this guideline, such as, for example, learning *the* before *hello*, words are grouped into frequency bands, such as the first 1000, second 1000 and so on.

A study of frequency figures shows that there is a very rapid drop in the frequency of such bands. Table 1 is based on the lemmatised

Table 1 Frequency band and percentage of text coverage

Frequency level	Cumulative coverage (%)	Coverage (%)
1st 1000	72.0	72.0
2nd 1000	79.7	7.7
3rd 1000	84.0	4.3
4th 1000	86.8	2.8
5th 1000	88.7	1.9
6th 1000	89.9	1.2

figures for the Brown corpus (Francis and Kučera, 1982) which contains a variety of text types and registers. It presents an example of the decreasing coverage by successive word frequency bands.

From a teaching point of view, in courses not focused on well defined areas of use, only the words in the most frequent 2000 words of English (the high frequency words) deserve individual attention. Beyond that level, that is the words in the third, fourth, fifth 1000 levels and onwards, teaching attention should be directed more towards strategies for learning and coping with these words, such as guessing from context, memorisation techniques and procedures, and the learning of word parts. Learners should continue to expand their vocabulary, but teachers should not necessarily continue to teach particular words directly. The reason for this distinction between high- and low-frequency words is primarily one of cost–benefit. The time spent on teaching any one low-frequency word is not justified by the small amount of benefit that learners get from knowing it, that is the number of opportunities to meet or use the word. The dividing line between high- and low-frequency words can be drawn using several criteria: frequency, coverage of the text, size of the high frequency group, overlap between various word counts, and the starting point of specialized vocabularies (Nation and Hwang, 1995). The criteria tend to agree in indicating that the 2000-word level is the most suitable place to make this arbitrary distinction.

This important distinction between high- and low-frequency words then makes it necessary for teachers to know what stage their learners are at in their vocabulary development. This was the motivation behind the construction of the original Vocabulary Levels Test, and now its productive version.

III A test of controlled productive ability

Productive vocabulary ability is not a yes/no phenomenon, but implies degrees of knowledge. For example, a learner may be able

to provide a sentence with an infrequent word when required to do so by the teacher, but be reluctant to use it when left to his own devices, as in a composition writing task and choose to use a simpler, more frequent word of a similar meaning. Such reluctance is often a result of uncertainty about the word's usage. Put differently, lack of confidence is a reflection of imperfect knowledge. We refer to the ability to use a word at one's free will as free productive ability. This type of knowledge is measured by the Lexical Frequency Profile (Laufer and Nation, 1995). We use the term 'controlled productive ability' for the ability to use a word when compelled to do so by a teacher or researcher, whether in an unconstrained context such as a sentence-writing task, or in a constrained context such as a fill-in task where a sentence context is provided and the missing target word has to be supplied. It is the latter format, with modifications described below, that we used for the present test.

For each item, a meaningful sentence context is presented and the first letters of the target item are provided. The first letters prevent the test-takers from filling in another word which would be semantically appropriate in the given context but which comes from a different frequency level. Here is an example eliciting the word 'episodes'.

The book covers a series of isolated epis____ from history.

The test format bears some resemblance to the C-test (Klein-Braley and Raatz, 1984; Klein-Braley, 1985), although for vocabulary-sampling purposes in this study it is not used in a paragraph but a sentence, and the cues are not always half a word. The number of letters for each word was decided on by the elimination of possible alternatives to the tested word. The C-test uses the first half of a word with the smaller number of letters being provided if the word has an odd number of letters (e.g., the first two letters are provided if the word contains five letters). Because our test was a test of productive vocabulary ability, it was thought better to provide the minimal number of letters that would disambiguate the cue. If two letters could start two possible words in the given sentence, an additional letter was added to eliminate this possibility. The size of the underlined space at the end of the incomplete word is no indication of the number of letters needed to complete it.

The overall structure of the test is modelled on the Vocabulary Levels Test (Nation, 1983; 1990). The test samples 18 items at each of the 2000, 3000, 5000, University Word List (UWL), and 10 000 word levels. Test Version A uses the items from the original Levels Test. Three parallel test versions were devised using the items from the three parallel versions of the Levels Test, which were made by Norbert Schmitt.

Two studies were conducted, one to check the reliability and validity of one version of the test, and one to check the equivalence of four parallel forms of the test

IV Study 1: Validation of the test

1 Research question and hypothesis

One kind of evidence for the validity of the test is to see if it distinguishes among different levels of language proficiency since vocabulary size forms a part of language proficiency. Learners at a higher level of language knowledge know more words. Therefore, our research question was whether there would be a significant difference among groups at several language-proficiency levels, in the total score on the test, and in the scores at each vocabulary frequency level: 2000, 3000, 5000, UWL, 10 000. The question was researched with Test Version A (published in Laufer and Nation, 1995). The receptive version of this test, the Vocabulary Levels Test, shows a substantial degree of implicational scaling from one frequency level to the next (Read, 1988).

2 Subjects

The subjects were four groups of foreign learners at different proficiency levels of English as a foreign language: high school 10th graders (n = 24), 11th graders (n = 23), 12th graders (n = 18) and 1st year university students in the English department (n = 14). The 10th graders had studied English for 5 years (4–5 hours a week), the 11th graders – 6 years, 12th graders – 7 years. Since English was studied as a foreign, not second, language and class work was the main source of input, the class grade was a fair indication of language proficiency.

3 Procedure

First, three native speakers were asked to retrieve the tested items. One of the researchers was sitting with them. Whenever someone had a difficulty with the retrieval, the sentence context was modified in the first instance. If that did not help, an additional letter for the target item was provided. The modified versions of the tests were given to seven additional native speakers. This time all the items could be retrieved by six or more of the seven informants.

The controlled productive ability test was given to the four groups. The grading was in terms of correct/incorrect for each item. Minor

Table 2 Reliabilities for each level of test version A

Level	Reliability
2000 level	.77
3000 level	.81
UWL	.84
5000 level	.84
10 000 level	.90

spelling mistakes were not marked as incorrect, and grammatical mistakes were also ignored. Each learner was given 6 scores: a score for the number of correct items at each of the 2000, 3000, UWL, 5000, and 10 000 levels and for the total score of correctly retrieved items.

4 Analysis and results

The entire test version A for all subjects had an internal consistency of .86 using the Kuder–Richardson formula KR21. Table 2 contains the reliabilities for the five levels in Test Version A. Sets of ANOVAs with Duncan post-hoc tests were performed on each test level and the total test scores.

By looking across each row in Table 3, we can see how the scores on the test change as general proficiency increases. For example, the total score on all five levels of the test grows from 21.7 out of 90 for the 10th grade students, to 33.4 out of 90 for the 11th grade students,

Table 3 Mean scores and F-tests for four proficiency level groups on the five levels and the total score of the original productive levels test

	10th grade (n = 24)	11th grade (n = 23)	12th grade (n = 18)	University (n = 14)	F-test
2000 level	11.8	15.0	16.2	17.0	17.9 p=.0001
3000 level	6.3	9.3	10.8	14.9	21.2 p=.0001
UWL level	2.6	5.3	7.4	12.6	34.6 p=.0001
5000 level	1.0	3.9	4.7	7.4	12.6 p=.0001
10 000 level	0.0	0.0	0.9	3.8	13.6 p=.0001
Total	21.7	33.4	40.1	55.8	32.6 p=.0001

to 40.1 for the 12th grade students, and to 55.8 for the university students. This increase is present not only in the totals for each proficiency level but also for each level of the test for each group. For example, at the 2000-word level, the 10th graders' score is 11.8, the 11th graders' 15, the 12th graders' 16.2 and the university students' 17.

Similarly there is a decrease in score for each group at each of the levels of the test. For example, at the 2000 level the 10th graders score 11.8, at the 3000 level 6.3, at the UWL level 2.6, at the 5000 level 1.0 and at the 10 000 level 0.0. The patterning is remarkably consistent in both directions.

As the F-test results in Table 3 indicate, the differences between the four groups of learners for the total scores and scores at individual frequency levels were significant. To check specifically which groups differed from one another, Duncan post-hoc tests were carried out. The Duncan's groupings show that on the total score, except for the 11th and the 12th grades, the groups are significantly different from one another. With regard to the 11th and the 12th grades, even though the difference in vocabulary size is not significant, there is nevertheless progress from a total score of 33.4 to 40.1. The lack of statistical significance between these two groups may be explained by the specific educational conditions of the 12th grade. A lot of teaching time is spent on preparation for the matriculation exam, i.e., revision of material rather than new material. It is also possible that, with larger numbers of subjects in these two groups, statistical significance would be reached in spite of the small difference in mean scores. The test as a whole distinguishes among most of the proficiency groups.

At the 10 000-word level, only the university students' score is significantly different from the others. This is not surprising as the 10th and 11th graders did not score at all at this level which represents low-frequency vocabulary beyond their knowledge, and the 12th graders' score was very low. At the 5000 and 3000 levels as with the total scores, the 11th and 12th graders are not significantly different from each other. These two groups are significantly different at these levels from the 10th graders and university students.

At the 2000 word level, there is no significant difference between the university students and the 12th graders. Both groups have scores that show mastery of the words at this high-frequency level. The 11th graders are not significantly different from the 12th graders but are significantly different from the university students. The 11th graders are approaching mastery of the high frequency words. The 10th graders have significantly different scores from the other three groups with mastery of about two-thirds (11.8 out of 18) of the 2nd 1000 most frequent words. The most interesting subtest is the university word

list which discriminates among all the four proficiency levels. Apparently, there is a gradual and significant increase throughout the high school and the university in the knowledge of academic words.

These results clearly show the gradual mastery of the successive frequency levels of the test as proficiency increases, indicating that it is a valid measure of vocabulary growth.

5 Practicality

The Productive Vocabulary Levels Test is a very practical instrument. It is easy to administer and can be completed in a short time. It is easy to mark as there is only one correct word for each item and each answer is marked as correct or incorrect. A whole test can be fitted on to three pages and although the test sheets are not reusable, it is economical to duplicate. It could be computerised but some allowance would need to be made for scoring minor misspellings.

The test is easy to interpret. Each level represents 1000 words, except the UWL level which represents a list of 836 words. A learner's percentage score on a level is a very rough indication of the number of words known at that level (for example, 9 out of 18 equals 50%; and this would roughly equal 500 out of 1000 words). Deciding whether a learner has satisfactory mastery of a level is a matter of judgement and depends what level is being considered, but is probably around 15 or 16 out of 18 (85% or 90%) for the 2000-word level, indicating that less than 150 words at that level are not readily available for productive use.

V Study 2: The equivalence of four parallel versions

In addition to the original test (Version A), three additional test versions were made up, each version using different items from the same frequency levels. The existence of parallel test versions can be useful in projects where we would like to measure vocabulary growth in test/retest situations to eliminate the memory effect of the items. Having produced the additional versions, it was necessary to check whether they would correlate highly with one another when administered to the same learners.

1 Subjects and procedure

Four groups of learners were selected at different proficiency levels for the four test levels: 2000, 3000, 5000, UWL. That is, one group of learners sat four versions of the 2000-word level, another group sat four versions of the 3000-word level and so on. We tried not to

have a situation where the selected test would be too easy or too difficult for a particular group, i.e., we avoided results with almost all correct or incorrect answers. If this had been the case, then the correlations among the scores of parallel versions would certainly have been high, but not revealing. All they would have revealed is either complete knowledge, or total ignorance of the words at the tested level. The 10 000-word level was not tested simply because, among our foreign learners, we do not have learners with a good enough knowledge of the words at that level. The subjects in this study were different from the subjects in Study 1. In Study 1, each subject took the entire test consisting of five frequency levels, with 18 items at each level. In Study 2, each subject took four versions of only one of the frequency levels.

Pearson correlations between the four test versions were calculated for each of the four tested vocabulary-frequency levels.

2 Results and discussion

Table 4 contains the reliabilities (KR21) for each of the levels. Because different groups of students sat different levels, it is not possible to provide total reliability figures for the four test versions. Moreover, different numbers of students sat different levels of the tests. However, total reliability figures are available from separate testing for Form A (.86) and Form C (.91) on KR21.

The difference in the reliabilities for the various levels and the whole tests (Form A and Form C) is because there are 90 items in the whole test and only 18 in each of the five levels. The reliabilities for the 5000 level are low because of the small number of subjects (18) and the homogeneity of that group.

Table 5 shows the correlations between four levels of the four versions of the Productive Vocabulary Levels Test. For example, the 2000-word level sections of test Versions A and B correlated .82 with each other which was significant at the .0001 level.

In general, the correlations are moderate to high and are significant. The lower correlations at the 5000 level may be due to the small

Table 4 Reliabilities for the levels in each of the four test versions

Level	Form A	Form B	Form C	Form D
2000 level	.51	.67	.80	.67
3000 level	.50	.39	.47	.56
UWL	.72	.63	.61	.78
5000 level	.61	.38	.04	.02

Table 5 Correlations between four versions of the Productive Vocabulary Levels Test at four of the five frequency levels in the tests

	A/B	A/C	A/D	B/C	B/D	C/D
2000 level (n = 45)	.82*	.82*	.78*	.83*	.81*	.77*
3000 level (n = 36)	.71*	.70*	.82*	.82*	.71*	.80*
UWL level (n = 33)	.75*	.80*	.84*	.83*	.76*	.80*
5000 level (n = 18)	.72 (p = .004)	.83*	.69 (p = .003)	.49 (p = .1)	.77 (p = .003)	.67 (p = .006)

Note: *significant at .0001 level.

number of subjects, but is most likely the result of patchy, unsystematic knowledge at this level which is at the edge of most of the learners' low-frequency vocabulary growth.

Because the tests are designed as diagnostic tests, another way to check the equivalence of the four forms is to see if they lead to the same decision regarding individuals who sit the tests. That is, if a criterion score is set at the 2000 level, do all four forms of the test always put an individual on the same side of the criterion score at that level? We focus on the 2000 level because this is the dividing line between high-frequency and low-frequency vocabulary. Table 6 has the results for the 2000-word level on the four forms. Fifteen out of 18 is a preferred criterion, but the criterion was set at 12 out of 18 because so few learners (only 6 out of 45) gained a score of 15 or more on the tests.

Table 6 shows that 62% of the learners (28 out of 45) were on the same side of the criterion on all four forms, 91% (28 plus 13 out of 45) were on the same side of the criterion on three out of the four forms and 95% (28 plus 13 plus 4 out of 45) were on the same side of the criterion on two out of the four forms. When the criterion score was set at 9, which was the mean of the means of all four versions, by coincidence the results were exactly the same as for the criterion

Table 6 Number of each individual's scores in four, three and two versions of the productive levels test on the same side of the criterion (n = 45)

Criterion	All four versions	Three out of four	Two out of four
12	28	13	4
9 (grand mean)	28	13	4

of 12. In terms of decision-making, the tests show a high degree of equivalence.

Unfortunately, although the four versions correlated well enough with each other and led to similar decision-making, the means were not similar enough. Sets of paired t-tests and ANOVAs with repeated measures were carried out and, on the basis of these, pairs of tests at each level were chosen that were not significantly different and had a good correlation (see Table 7). For diagnostic purposes, any of the four versions could be used, while for test/retest purposes, the two new parallel versions are recommended.

All correlations are significant at the $p = .0001$ level. Table 7 shows that the two parallel versions (see Appendix 1 and Appendix 2) are Version C at all levels, and a test made up of the 2000-level items from Version B, the 3000 and UWL items from Version D, and the 5000 level items from Version A. Further analysis of Version C showed that it had a reliability of .91 on KR21, and discriminated between learners of different proficiency levels.

VI Conclusion

The Productive Vocabulary Levels Test is a reliable, valid and practical measure of vocabulary growth. It is an additional quantitative measure which enables us to research some important issues in vocabulary acquisition.

Paul *et al.* (1990: 1) conclude that for vocabulary testing 'the choice of test format depends on the type of information desired'. The Productive Vocabulary Levels Test provides a useful addition to a range of largely receptive measures that have been used previously. The three formats investigated by Paul *et al.* were multiple-choice, interview and yes/no. These are all receptive measures in that the word form was provided and the learners were tested on their knowledge of the meaning. Similarly, Joe (1994) used three receptive measures (a sensitive multiple-choice test, a more demanding multiple-choice and an interview) to determine the strength of vocabulary

Table 7 Two equivalent forms with similar means and a good correlation at each level

Level	2000 B/C	3000 C/D	5000 A/C	UWL C/D
Means	6.7/6.3	3.8/3.9	3.7/3.5	5.1/5.7
Standard deviations	3.3/3.3	2.3/2.6	2.3/1.7	2.9/3.8
Correlations	.83	.80	.82	.80

knowledge. The Productive Vocabulary Levels Test allows researchers to investigate other aspects of vocabulary knowledge and thus look more effectively at breadth of vocabulary knowledge.

Using the Productive Vocabulary Levels Test together with the receptive levels test and the Lexical Frequency Profile (a measure of free active vocabulary, see Laufer and Nation, 1995), we can investigate questions such as the following:

1) What developments occur in the different types of vocabulary knowledge over a period of time (for example, receptive, controlled productive, free productive)?
2) How are the different types of vocabulary knowledge related to one another in the same individuals?
3) How do the relationships between the different types of knowledge change over time?
4) How do the different types of knowledge develop in different input conditions and with different teaching methods?

VII References

Chapelle, C. 1994: Are C-tests valid measures for L2 vocabulary research? *Second Language Research* 10, 157–87.

Francis, W.N. and **Kučera, H.** 1982: *Frequency analysis of English usage*. Boston, MA: Houghton Mifflin.

Frederiksen, J.R. 1982: A componential theory of reading skills and their interactions. In Mislevy R.J. (ed.) *Advances in the psychology of human intelligence*, Vol. 1. Hillsdale, NJ: Lawrence Erlbaum.

Goulden, R., **Nation, P**. and **Read, J**. 1990: How large can a receptive vocabulary be? *Applied Linguistics* 11, 341–63.

Grabe, W. 1991: Current developments in second language reading research. *TESOL Quarterly* 25, 375–406.

Henning, G. 1991: *A study of the effects of contextualisation and familiarisation on responses to the TOEFL vocabulary test items*. Princeton, NJ: Educational Testing Service.

Joe, A. 1994: The effects of text-based tasks on incidental vocabulary learning. Unpublished MA thesis. Victoria University of Wellington, New Zealand.

Klein-Braley, C. 1985: A cloze-up on the C-test: a study in the construct validation of authentic tests. *Language Testing* 2, 76–104.

Klein-Braley, C. and **Raatz, U.** 1984: A survey of research on the C-test. *Language Testing* 1, 134–46.

Laufer, B. and **Nation, P.** 1995: Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16, 307–22.

Meara, P. and **Buxton, B.** 1987: An alternative to multiple choice vocabulary tests. *Language Testing* 4, 142–51.

Nation, I.S.P. l983: Testing and teaching vocabulary. *Guidelines* 5, 12–25.

—— 1990. *Teaching and Learning Vocabulary*. Boston, MA: Heinle and Heinle.

Nation, P. and **Hwang K.** 1995: Where would general service vocabulary stop and special purposes vocabulary begin? *System* 23, 35–41.

Paul, P.V., **Stallman, A.C.** and **O'Rourke, J.P.** 1990: Using three test formats to assess good and poor readers' word knowledge. Technical Report No. 509, Center for the Study of Reading, University of Illinois at Urbana-Champaign, IL.

Perkins, K. and **Linville, S.E.** 1987: A construct definition study of a standardized ESL vocabulary test. *Language Testing* 4, 125–41.

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

—— forthcoming: *Testing vocabulary knowledge and use*. Cambridge: Cambridge University Press.

Richards, J.C. l976: The role of vocabulary teaching. *TESOL Quarterly* l0, 77–89.

Appendix 1

One of two equivalent versions of the A LEVELS TEST OF PRODUCTIVE VOCABULARY: Parallel Version 1 (Version C)

Complete the underlined words. The example has been done for you.

He was riding a bic*ycle*.

The 2000-word level

1. I'm glad we had this opp____ to talk.
2. There are a doz____ eggs in the basket.
3. Every working person must pay income t____.
4. The pirates buried the trea____ on a desert island.
5. Her beauty and cha____ had a powerful effect on men.
6. La____ of rain led to a shortage of water in the city.
7. He takes cr____ and sugar in his coffee.
8. The rich man died and left all his we____ to his son.
9. Pup____ must hand in their papers by the end of the week.
10. This sweater is too tight. It needs to be stret____.
11. Ann intro____ her boyfried to her mother.
12. Teenagers often adm____ and worship pop singers.
13. If you blow up that balloon any more it will bur____.
14. In order to be accepted into the university, he had to impr____ his grades.
15. The telegram was deli____ two hours after it had been sent.
16. The differences were so sl____ that they went unnoticed.
17. The dress you're wearing is lov____.
18. He wasn't very popu____ when he was a teenager, but he has many friends now.

The 3000-world level

1. He has a successful car____ as a lawyer.
2. The thieves threw ac____ in his face and made him blind.

3. To improve the country's economy, the government decided on economic ref____.
4. She wore a beautiful green go____ to the ball.
5. The government tried to protect the country's industry by reducing the imp____ of cheap goods.
6. The children's games were funny at first, but finally got on the parents' ner____.
7. The lawyer gave some wise coun____ to his client.
8. Many people in England mow the la____ of their houses on Sunday morning.
9. The farmer sells the eggs that his he____ lays.
10. Sudden noises at night sca____ me a lot.
11. France was proc____ a republic in the 18th century.
12. Many people are inj____ in road accidents every year.
13. Suddenly he was thru____ into the dark room.
14. He perc____ a light at the end of the tunnel.
15. Children are not independent. They are att____ to their parents.
16. She showed off her sle____ figure in a long narrow dress.
17. She has been changing partners often because she cannot have a sta____ relationship with one person.
18. You must wear a bathing suit on a public beach. You're not allowed to be na____.

The 5000-word level

1. Soldiers usually swear an oa____ of loyalty to their country.
2. The voter placed the ball____ in the box.
3. They keep their valuables in a vau____ at the bank.
4. A bird perched at the window led____.
5. The kitten is playing with a ball of ya____.
6. The thieves have forced an ent____ into the building.
7. The small hill was really a burial mou____.
8. We decided to celebrate New Year's E____ together.
9. The soldier was asked to choose between infantry and cav____.
10. This is a complex problem which is difficult to compr____.
11. The angry crowd sho____ the prisoner as he was leaving the court.
12. Don't pay attention to this rude remark. Just ign____ it.
13. The management held a secret meeting. The issues discussed were not disc____ to the workers.
14. We could hear the sergeant bel____ commands to the troops.
15. The boss got angry with the secretary and it took a lot of tact to soo____ him.
16. We do not have adeq____ information to make a decision.
17. She is not a child, but a mat____ woman. She can make her own decisions.
18. The prisoner was put in soli____ confinement.

The University Word List level

1. There has been a recent tr____ among prosperous families towards a smaller number of children.
2. The ar____ of his office is 25 square meters.
3. Phil____ examines the meaning of life.

4. According to the communist doc____, workers should rule the world.
5. Spending many years together deepened their inti____.
6. He usually read the sport sec____ of the newspaper first.
7. Because of the doctors' strike the cli____ is closed today.
8. There are several misprints on each page of this te____.
9. The suspect had both opportunity and mot____ to commit the murder.
10. They insp____ all products before sending them out to stores.
11. A considerable amount of evidence was accum____ during the investigation.
12. The victim's shirt was satu____ with blood.
13. He is irresponsible. You cannot re____ on him for help.
14. It's impossible to eva____ these results without knowing about the research methods that were used.
15. He finally att____ a position of power in the company.
16. The story tells us about a crime and subs____ punishment.
17. In a hom____ class all students are of a similar proficiency.
18. The urge to survive is inh____ in all creatures.

The 10 000-word level

1. The baby is wet. Her dia____ needs changing.
2. The prisoner was released on par____.
3. Second year University students in the US are called soph____.
4. Her favorite flowers were or____.
5. The insect causes damage to plants by its toxic sec____.
6. The evac____ of the building saved many lives.
7. For many people, wealth is a prospect of unimaginable felic____.
8. She found herself in a pred____ without any hope for a solution.
9. The deac____ helped with the care of the poor of the parish.
10. The hurricane whi____ along the coast.
11. Some coal was still smol____ among the ashes.
12. The dead bodies were muti____ beyond recognition.
13. She was sitting on a balcony and bas____ in the sun.
14. For years waves of invaders pill____ towns along the coast.
15. The rescue attempt could not proceed quickly. It was imp____ by bad weather.
16. I wouldn't hire him. He is unmotivated and indo____.
17. Computers have made typewriters old-fashioned and obs____.
18. Watch out for his wil____ tricks.

Appendix 2

The second of two equivalent versions of the A LEVELS TEST OF PRODUCTIVE VOCABULARY: Parallel Version 2

Complete the underlined words. The example has been done for you.

He was riding a bic*ycle*.

The 2000-word level

1. It is the de____ that counts, not the thought.
2. Plants receive water from the soil through their ro____.

3. The nu____ was helping the doctor in the operation room.
4. Since he is unskilled, he earns low wa____.
5. This year long sk____ are fashionable again.
6. Laws are based upon the principle of jus____.
7. He is walking on the ti____ of his toes.
8. The mechanic had to replace the mo____ of the car.
9. There is a co____ of the original report in the file.
10. They had to cl____ a steep mountain to reach the cabin.
11. The doctor ex____ the patient thoroughly.
12. The house was su____ by a big garden.
13. The railway con____ London with its suburbs.
14. She wan____ aimlessly in the street.
15. The organisers li____ the number of participants to fifty.
16. This work is not up to your usu____ standard.
17. They sat down to eat even though they were not hu____.
18. You must have been very br____ to participate in such a dangerous operation.

The 3000-word level

1. I live in a small apa____ on the second floor.
2. The pro____ of failing the test scared him.
3. Before writing the final version, the student wrote several dra____.
4. It was a cold day. There was a ch____ in the air.
5. The cart is pulled by an o____.
6. Anthropologists study the struc____ of ancient societies.
7. After two years in the Army, he received the rank of lieu____.
8. The statue is made of mar____.
9. Some aristocrats believed that blue blood flowed through their ve____.
10. The secretary assi____ the boss in organizing the course.
11. His beard was too long. He decided to tr____ it.
12. People were whir____ round on the dance floor.
13. He was on his knees, ple____ for mercy.
14. You'll sn____ that branch if you bend it too far.
15. I won't tell anybody. My lips are sea____.
16. Crying is a nor____ response to pain.
17. The Emperor of China was the supr____ ruler of his country.
18. You must be awa____ that very few jobs are available.

The 5000-word level

1. Some people find it difficult to become independent. Instead they prefer to be tied to their mother's ap____ strings.
2. After finishing his degree, he entered upon a new ph____ in his career.
3. The workmen cleaned up the me____ before they left.
4. On Sunday, in his last se____ in Church, the priest spoke against child abuse.
5. I saw them sitting on st____ at the bar drinking beer.
6. Her favorite musical instrument was a tru____.
7. The building is heated by a modern heating appa____.
8. He received many com____ on his dancing skill.
9. People manage to buy houses by raising a mor____ from a bank.

10. At the bottom of a blackboard there is a le____ for chalk.
11. After falling off his bicycle, the boy was covered with bru____.
12. The child was holding a doll in her arms and hu____ it.
13. We'll have to be inventive and de____ a scheme for earning more money.
14. The picture looks nice; the colours bl____ really well.
15. Nuts and vegetables are considered who____ food.
16. The garden was full of fra____ flowers.
17. Many people feel depressed and gl____ about the future of the mankind.
18. He is so depressed that he is cont____ suicide.

The University Word List level

1. I've had my eyes tested and the optician says my vi____ is good.
2. The anom____ of his position is that he is the chairman of the committee, but isn't allowed to vote.
3. In their geography class, the children are doing a special pro____ on North America.
4. In a free country, people can apply for any job. They should not be discriminated against on the basis of colour, age, or s____.
5. A true dem____ should ensure equal rights and opportunities for all citizens.
6. The drug was introduced after medical res____ indisputably proved its effectiveness.
7. These courses should be taken in seq____, not simultaneously.
8. Despite his physical condition, his int____ was unaffected.
9. Governments often cut budgets in times of financial cri____.
10. The job offer sounded interesting at first. But when he realised what it would involve, his excitement subs____ gradually.
11. Research ind____ that men find it easier to give up smoking than women.
12. In a lecture, most of the talking is done by the lecturer. In a seminar, students are expected to part____ in the discussion.
13. The airport is far away. If you want to ens____ that you catch your plane, you have to leave early.
14. It's difficult to ass____ a person's true knowledge by one or two tests.
15. The new manager's job was to res____ the company to its former profitability.
16. Even though the student didn't do well on the midterm exam, he got the highest mark on the fi____.
17. His decision to leave home was not well thought out. It was not based on rat____ considerations.
18. The challenging job required a young, successful and dyn____ candidate.

The 10000-word level

1. The new vic____ was appointed by the bishop.
2. If your lips are sore, try lip sal____, not medicine.
3. Much to his chag____, he was not offered the job.
4. The actors exchanged ban____ with reporters.
5. She wanted to marry nobility: a duke, a baron, or at least a vis____.
6. The floor in the ballroom was a mos____ of pastel colours.
7. She has contributed a lot of money to various charities. She is known for her generosity and bene____.

8. This is an unusual singer with a range of three oct____.
9. A thro____ controls the flow of gas into an engine.
10. Anyone found loo____ bombed houses and shops will be severly punished.
11. The crowd soon disp____ when the police arrived.
12. The wounded man squi____ on the floor in agony.
13. The dog crin____ when it saw the snake.
14. He imme____ himself in a hot bubbly bath forgetting all his troubles for a moment.
15. The approaching storm stam____ the cattle into running wildly.
16. The problem is beginning to assume mam____ proportions.
17. His vind____ behaviour towards the thief was understandable.
18. He was arrested for illi____ trading in drugs.