

AN ANALYSIS OF THE VOCABULARY OF NCEA

David Wallace

Victoria University of Wellington

Introduction and background

The National Certificate of Educational Achievement (NCEA) was introduced to Year 11 students in New Zealand high schools in 2002. It involved a number of changes to the past methods of assessment, most notably the implementation of an achievement based assessment system and the introduction or expansion of internal assessment to many Year 11 subjects. This has been followed by the implementation of NCEA level two into Year 12 in 2003 and the scheduled implementation of level 3 in 2004. The marking schedule for each assessment grades the students' results into four levels of achievement: not achieved, achieved, achieved with merit, and achieved with excellence. Currently most subjects are approximately 50% internally assessed, although there are some subjects, for example Physical Education, which are 100% internally assessed. The three levels have, to some extent, been designed to scaffold student learning as they move through the achievement standards at each level. NCEA has also been adopted by New Zealand universities as not only a requirement for university entrance but as a measure of language competence. Achievement in reading and writing at level two in English has been adopted as a language requirement for university entrance.

While the exact assessment conditions can be adjusted to suit the assessment, many of the internals and most of the externals are completed under examination conditions. These conditions include separation of desks, no talking and importantly first and second language dictionaries are not usually permitted. Therefore, the difficulty of vocabulary in these assessments becomes a significant factor in determining whether NESB (non-English speaking background) students achieve in the assessment and the level of that achievement. Furthermore, for the ESOL (English to speakers of other languages) teacher it becomes critical to determine the vocabulary that the student will need to know in order to achieve in any particular task.

How much vocabulary is necessary to comprehend different texts?

Obviously vocabulary knowledge is not the only determining factor in the comprehension of a language. An understanding of grammar is important, along with a background knowledge of the culture, and the skill of guessing meanings from context, among other things. However, in a study of ESOL reading comprehension, Laufer and Sim (1985) determined that knowledge of the vocabulary was in fact the most important requirement for comprehension, followed by subject matter knowledge and finally grammar. Thus, NESB students with a good knowledge of the subject matter may not achieve to their true potential in NCEA assessments because of inadequacies in their known vocabulary, much of which is subject-related.

A problem for many ESL teachers is how to target vocabulary learning in order to best prepare students for the rigours of internal and external assessments. In a study comparing L2 vocabulary scores and reading comprehension of unsimplified texts, Laufer (1992) found that vocabulary knowledge is actually a very good predictor of NESB students' reading comprehension scores. Furthermore, Laufer (*ibid*) states that

3000 word families (about 4,800 lexical items) constitutes a lexical threshold where there were more than half of the learners could reach the required 56% target to pass a reading comprehension task using an unsimplified text.

Using a fiction text, Hu and Nation (2000) found that on average learners' comprehension improved in a predictable way as known vocabulary increased. Furthermore, while the comprehension scores varied widely, only a few students with an understanding of less than 95% of the vocabulary (word tokens) were able to gain reasonable comprehension. Hu and Nation (ibid) determined that a threshold of 98% knowledge of vocabulary was necessary to gain adequate unassisted comprehension of a fiction text.

Other types of text may require a greater knowledge of vocabulary. Table 1 demonstrates that while a knowledge of the first 2000 words of English, taken from the General Service List (West, 1953), may be enough to understand about 87% of fiction, it will only allow the learner to comprehend about 80% of newspapers and 76% of academic texts.

Can a specialised vocabulary be developed to improve students' comprehension?

Nation and Hwang (1995) demonstrate that while a general service vocabulary of up to 2000 words gives adequate comprehension, beyond that level students learning English for specific purposes (ESP) will get a better return by learning a vocabulary targeted to that purpose. In high schools, students can obtain a wide range of word lists, glossaries and subject dictionaries that can assist them to master the technical vocabulary of their chosen disciplines. Technical word lists have been found to assist ESP learners. Ward (1999) developed a list of 3000 words which included technical and academic words essential to the understanding of engineering. He found that in a typical introductory engineering text the 3000 word list covered 97.8% of the running words with the first 1000 covering 92.6%. However, it may not always be appropriate for an ESL teacher to focus on a highly technical or specialised vocabulary where a class may have a more general focus.

Cross-discipline word lists have been developed to assist students to improve their comprehension of texts found in university study. Table 1 below demonstrates that knowledge of academic words is vital for the comprehension of academic texts.

Table 1: The vocabulary requirements of different texts

	Academic texts (from Coxhead, 2000)	Newspapers (from Nation, 2001)	Fiction (from Nation, 2001)
First 1000	71.4	75.6	82.3
Second 1000	4.7	4.7	5.1
AWL	10.0	3.9	1.7
Not in Lists	13.9	15.7	10.9

The Academic Word List (AWL), compiled by Coxhead (2000), is a specialised vocabulary of 570 word families outside of the General Service List (GSL) that make up

approximately 10% of academic texts. Word families in the AWL include the headword and its closely related inflected and derived forms even if the part of speech is different. A good knowledge of the word families found in the AWL vocabulary would be a great advantage for NESB high school students and should be essential for students undertaking university study. From the studies in this section, it would seem essential for ESL teachers to be aware of the particular needs of their students in order to be able to guide them in learning the vocabulary that will give them the best return for their efforts.

The purposes of this study

This study deals with the analysis of level one and level two NCEA achievement standards to determine how much vocabulary is required to understand the texts used in these assessment tasks. The specific research questions are:

How large a vocabulary is needed to be familiar with most of the vocabulary in NCEA assessments?

Does the vocabulary of NCEA occur with equal frequency over the four major disciplines, Arts, Science, Fine Arts and Technology, and Commerce?

Are there some subjects that have higher vocabulary demands than others?

Method

Selection of words for the NCEA corpus

The New Zealand National Qualifications Framework is administered by the New Zealand Qualifications Authority (NZQA) and all NCEA assessments are coordinated and maintained by this government authority. Sample NCEA internal assessments are available from the teachers' resource web site Te Kete Ipurangi (<http://www.tki.org.nz>). The 2002 NCEA level one examinations and the NZQA approved level one and two practice external assessments were also studied. These can be downloaded from NZQA (<http://www.nzqa.govt.nz>). For the purposes of this research it was decided to use just Year 11 and 12 assessments. This was because they had all been moderated by NZQA, and had been subjected to at least two years of trialling and implementation in schools.

After downloading the files, the teacher's instructions, marking schedules, generic NZQA reference information and foreign language words were removed to leave only the text that would be presented to a student studying the Year 11 subject. The assessments were then classified into four main disciplines: Arts, Sciences, Fine Arts and Technology, and Commerce. Table 2 below shows the range of assessments that were obtained. The number of internal assessments is larger than externals because the Ministry of Education has, over the past few years, developed a range of internal assessments to act as exemplars for the development of suitable tasks. On the other hand, there was usually only one assessment or assessment exemplar available at NZQA for each externally assessed Achievement Standard.

Table 2: The range and number of NCEA assessments analysed in this study.

Discipline	No. of subjects	No. of internal assessments	No. of external assessments	Total assessments	Total no. of words
Arts	20	172	98	270	243,724
Sciences	8	173	114	287	165,754
Fine Arts and Technology	6	99	44	143	91,023
Commerce	2	28	29	57	37,513

Analysis of words for the NCEA corpus

The corpus was then analysed by using the Range computer programme. This programme records the frequency of each word and calculates the proportion of words occurring in a text that match base-lists of selected word families. In the first part of this study, the NCEA corpus was analysed by using three base-lists containing the AWL and the word families for the first and second 1000 words of the GSL.

Results

Figure 1 illustrates the results obtained by the NCEA corpus analysis.

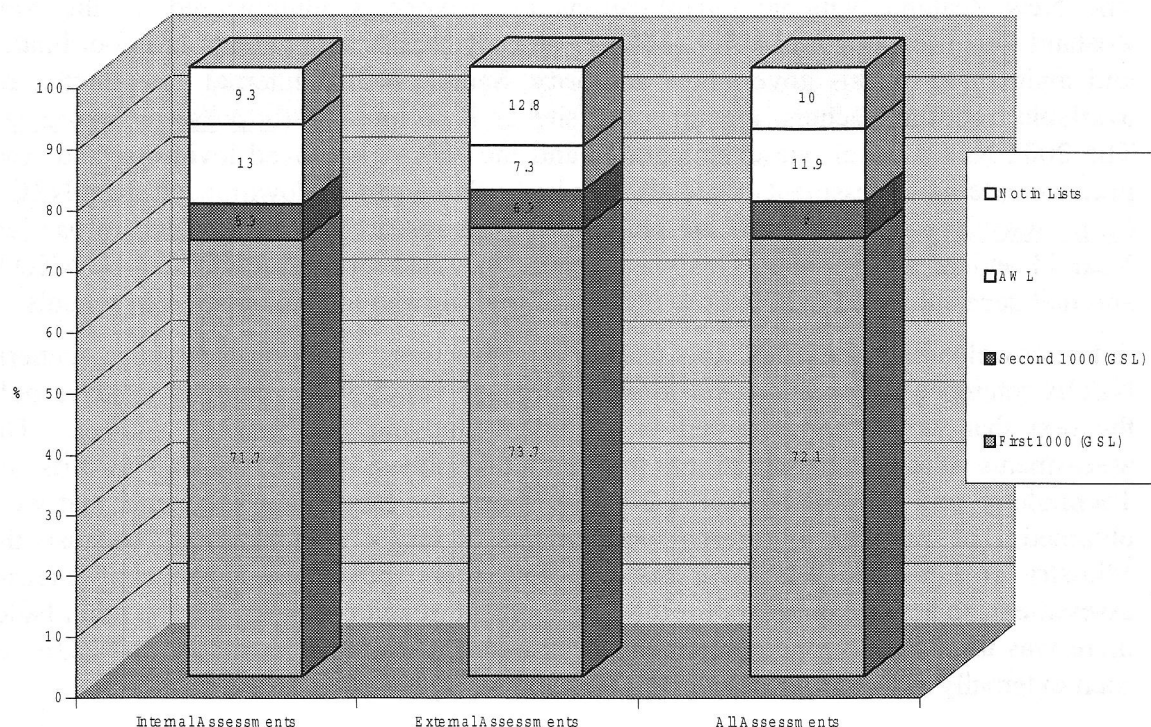


Figure 1: The coverage of the NCEA corpus by the General Service List and the Academic Word List

The internal assessments were firstly compared to the externals. In general the internals drew much more heavily from words found in the AWL with fewer words from the first 2000. However, the external assessments had slightly more words from outside the three lists. When the whole corpus was analysed it was found that 78.1% of the words (tokens) came from the GSL, 11.9% from the AWL and 10.0% of the words came from outside of the three lists.

Coxhead (2000) found that the AWL made up about 10% of the words in academic texts. The NCEA corpus included even more (11.9%), but had fewer words not found in the lists and correspondingly more from the GSL. While readability of a text is determined by a great range of factors, one could tentatively conclude from the data in Table 3 that NCEA assessments are slightly more difficult to read than newspapers and more difficult again than general fiction. This would be further substantiated if the words not found in the lists of newspapers and fiction had many more proper nouns, such as the names of people and places, than the NCEA corpus.

Table 3: A comparison of the NCEA results with different texts

	NCEA Assessments	Academic Texts (from Coxhead, 2000)	Newspapers (from Nation, 2001)	Fiction (from Nation, 2001)
First 1000	72.1	71.4	75.6	82.3
Second 1000	6.0	4.7	4.7	5.1
AWL	11.9	10.0	3.9	1.7
Not in Lists	10.0	13.9	15.7	10.9

1. How large a vocabulary is needed to be familiar with most of the vocabulary in NCEA assessments?

As demonstrated in Figure 1, even with knowledge of the GSL and the AWL, English language students will still only understand 90% of the tokens found in NCEA assessments. This is well below the 98% required for an adequate comprehension (Hu & Nation, 2000). In my experience, students completing Year 11 subjects and requiring ESOL support typically have a vocabulary of between 1000-2000 words. Obviously this lack of vocabulary knowledge will make it difficult for these students to do well in NCEA assessments. It becomes critical then to analyse the words found in the assessments from outside the GSL to find out if a more focussed vocabulary list can be created to enable the students to improve their comprehension.

2. Does the vocabulary of NCEA occur with equal frequency over the four major disciplines?

When analysed in terms of the different disciplines, the Arts and the Sciences are quite similar in their vocabulary requirements (refer to Figure 2), with 70-75% of the vocabulary coming from the first 1000 words and 9-15% from the AWL. Fine Arts and Technology had the highest number of words from the AWL (15.3%) and Commerce had a higher percentage of words outside the three lists (11.3%).

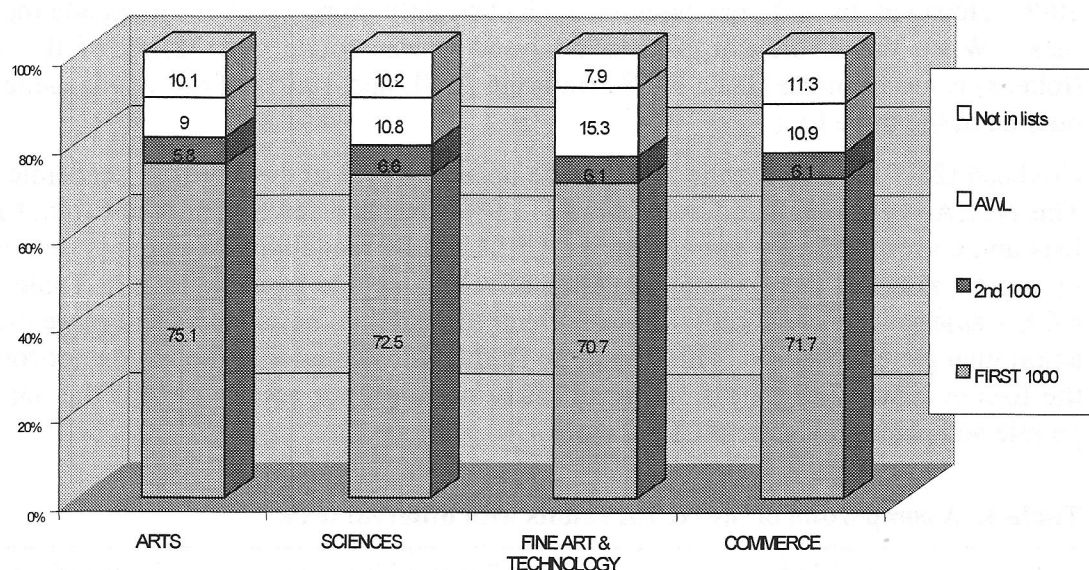


Figure 2: NCEA vocabulary by discipline

3. Are there some subjects that require a higher vocabulary than others?

When the Year 11 subjects were analysed individually some subjects had a much higher percentage of words outside the GSL (as shown in Figure 3 below). Graphics, for example, drew 18.5% of its total words from the AWL and 9.9% came from outside the three lists. This was followed by Technology with 18.2% AWL and 4.1% outside. Biology and Visual Art had the highest percentage of words drawn from outside the lists, 12.5% and 12.4% respectively.

An analysis of these particular subject areas found that in the subjects with the highest use of AWL vocabulary, for example Graphics, a large proportion of the academic language was used to give instructions with precision. Examples of this kind of language include, *requirements*, *specification*, *evaluate*, *appropriate* and *process*. Clearly these texts will be difficult for a typical NESB student to read. However, it should also be noted that unlike most other subjects, Graphics and Technology assessments are typically completed over several classes with teacher instruction, therefore students who do not understand the instructions can obtain assistance. Clearly these texts will be difficult for a typical NESB student to read.

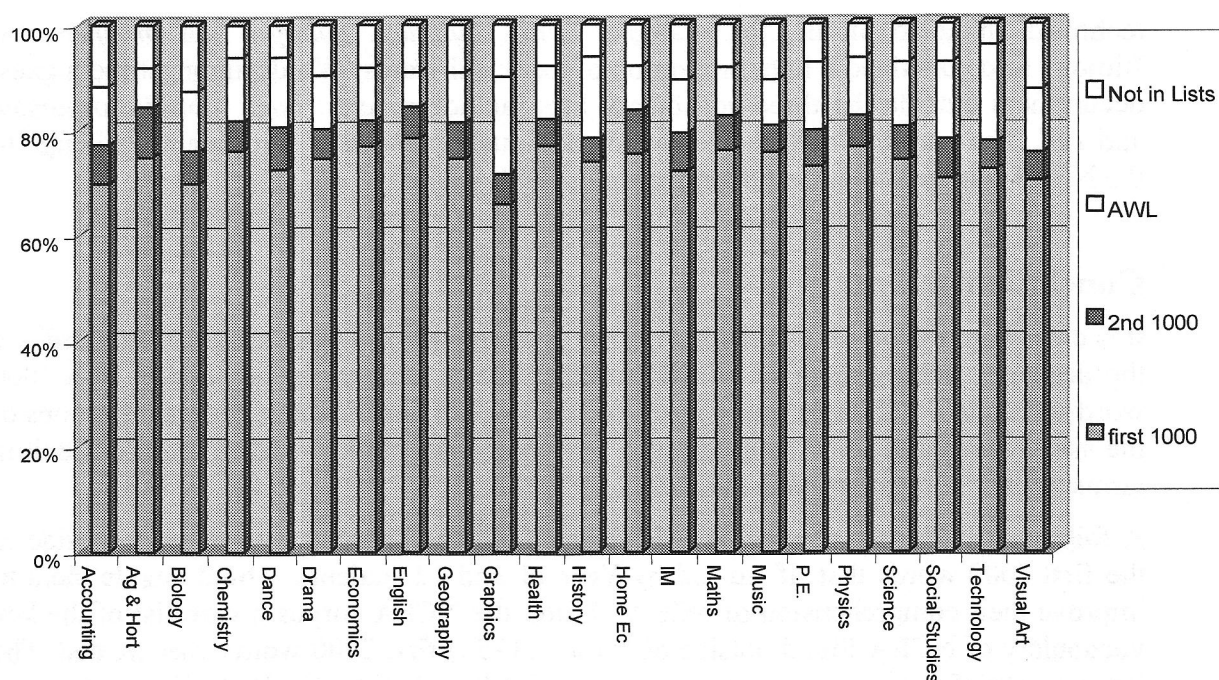


Figure 3: NCEA Vocabulary by subject

Technical Words

As shown in Table 4, many of the words occur in only one of the four disciplines, indeed a very large proportion of these appear in only one subject. Assuming that the proper nouns are known by the student or could be worked out from their context, there is still a large proportion of words that the student is expected to know but are outside the GSL and the AWL.

Table 4: An analysis of the vocabulary found outside the three word lists

Vocabulary	Number of Words	Percentage of Total Words Found Outside Word Lists
Total words found outside word lists	5,379	100%
Proper nouns	447	8.3%
Words found in one discipline only	4,568	84%
Words found in one subject only	4,230	78%

Biology and Visual Art have the largest proportion of words from outside the three lists. The percentage found in Visual Art may be a result of the sample, as four of twenty-three collected internal assessments dealt with Maori art. Thus words such as *waka*, *Maori* and *whakapapa* had a high incidence. However, there was also a large proportion of proper nouns (9.5%) made up of names such as *Cezanne*, *Hotere* and *Angus* along with artistic styles such as *Cubism* and *Expressionism*. Many of the other terms were the

technical language of art, for example, *folio*, *sculpture*, *collage*, and *brushstrokes*. Biology also contained a high proportion of technical words. Words amongst the highest occurrences outside the three vocabulary lists include *species*, *bugs*, *abiotic*, *organism* and *cell*. The learning of technical words obviously presents a significant challenge to the NESB learner and quite a dilemma for the ESL teacher.

Conclusion

It is clear from the present study that NCEA assessments may pose a significant hurdle to the academic achievement of NESB students. Even learners in possession of a 2000 word vocabulary will most likely misunderstand or not understand significant portions of the assessments and some subjects, such as Graphics and Biology, seem more difficult in terms of their vocabulary than others.

A follow up study was conducted to investigate whether there is a vocabulary outside of the first 2000 words that, if studied by Year 11 and 12 students, would enable them to improve their comprehension of NCEA. Using the NCEA corpus a wordlist of the key vocabulary of NCEA found outside of West's (1953) first 2000 words was created. The list and a brief explanation of its creation can be found at the ESOL Online website at: http://english.unitechnology.ac.nz/nedb/resource_exchange/

References

- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- Hu, M. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403-430.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension. In P. J. L. Arnold & H. Bejoint (Eds), *Vocabulary and applied linguistics*. London: Macmillan.
- Laufer, B. & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes. *Foreign Language Annals*, 18, 405-411.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. & Hwang, K. (1995). Where would a general service vocabulary stop and a special purposes vocabulary begin? *System* 23, 35-41.
- Ward, J. (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language* 12, 309-323.
- West, M. (1953). *A general service list of English words*. London: Longman, Green and Co.