Principles guiding vocabulary learning through extensive reading

Paul Nation
Victoria University of Wellington
New Zealand

Extensive reading is one of a range of activities that can be used in a language learning course. Ideally, the choice of activities to go into a course should be guided by principles which are well supported by research. Similarly, the way each of those activities is used should be guided by well-justified principles. In this article, we look at the principles justifying the inclusion of extensive reading in a course, and then look in detail at a set of principles guiding how extensive reading can best be carried out to result in substantial vocabulary learning. Extensive reading can result in a wide range of learning outcomes, but in this article we narrow our focus on vocabulary learning (for similar analyses of a wide range of vocabulary learning activities see Webb and Nation [in preparation]).

Principles Justifying the Inclusion of Extensive Reading in a Language Learning Program

There are three major principles justifying the inclusion of extensive reading in a language learning program, namely the principles of learning conditions, the four strands, and cost/benefit.

Learning Conditions

Vocabulary learning occurs because certain mental conditions are created which encourage learning. Essentially, vocabulary learning depends on the number of meetings with each word and the quality of attention at each meeting (see Table 1). The more meetings, the more likely learning is to occur. The deeper the quality of the meetings, the more likely learning is to occur. The few experiments comparing the effects of the number of meetings (repetitions) with the quality of the meetings suggest that, of the two, quality has the stronger effect (Laufer, in press; Webb, 2005).

The quality of the meetings depends primarily on whether the learners give incidental or deliberate attention to a word. There are a few situations where it is not easy to distinguish incidental attention from deliberate attention, but generally incidental attention occurs when the learner's focus is on some other aspect of communication besides the individual words and phrases. Typically this focus would be on the message being communicated. Deliberate attention occurs when the learner consciously focuses on aspects of knowing a word. Both incidental and deliberate attention have various levels of quality, ranging across noticing a word, retrieval of knowledge gained from previous meetings, meeting or using the word in ways which are different from the previous meetings or use, and elaborating on knowledge of the word beyond

its contextual uses (see Table 1). These levels of quality are largely cumulative in that retrieval also includes noticing, and varied use includes retrieval and noticing. Elaboration certainly includes noticing and may include retrieval if the elaborated words have been met before. Because deliberate elaboration can involve decontextualised instances of words, deliberate elaboration does not necessarily include varied use. Vocabulary learning from extensive reading is primarily affected by repetition and varied meetings.

Table 1: *Extensive reading and vocabulary learning conditions*

Number of meetings	Initial occurrence/Repetition	
Quality of attention	Incidental attention	Deliberate attention
Noticing	Guessing from context	Dictionary look-up
Receptive or productive retrieval	Re-occurrence of words in a text or other texts Repetition through repeated reading of the same text	Putting words met during reading onto word cards and learning using the cards
Varied meetings (receptive) or varied use (productive)	Meeting words in varied contexts	Consulting a concordance while reading as in Lextutor Reading with resources
Receptive or productive elaboration	Applying a dictionary definition to a particular textual context	Applying the dictionary strategy

As proficiency develops, reasonably large quantities of input are needed to meet most of the words at a particular 1000 word level, and to meet them enough times for learning to have a chance to occur (Nation, 2014). However, as we shall see later in this article, extensive reading can provide sufficient repetitions.

Besides repetition, extensive reading usefully sets up the condition of varied meetings. When a word occurs again in an extensive reading text such as a graded reader, it typically occurs in a context which is different from the previous contexts in which it occurred. This finding can be easily confirmed by doing a concordance search on a graded reader for words which are likely to be new at the level of that reader, for example, the level 3 words in a level 3 reader. Thus, each new meeting with a word during extensive reading is highly likely to enrich knowledge of that word through its varied contexts as well as strengthen knowledge through repetition.

In addition to repetition and varied meetings, extensive reading also provides opportunities for deliberate learning through looking up the meanings of words in a dictionary. This look-up can be done with a hard-copy dictionary, or with the increasing use of electronic readers and tablets look-up can be done simply by touching a word. The ease of such electronic access makes it much more likely that learners will look up words and thus add a deliberate element to vocabulary learning during extensive reading. Dictionary look-up greatly increases the chances of vocabulary learning (Hulstijn, Hollander, & Greidanus, 1996; Knight, 1994; Laufer & Hill, 2000; Peters, 2007).

Some advocates of extensive reading actively discourage dictionary use on the grounds that it takes time away from reading (Luppescu & Day, 1993) and that it discourages guessing from context. The vocabulary learning benefits of dictionary look-up outweigh this concern, particularly if learners are given training and encouragement to guess from context, and thus see the dictionary as a way of confirming a guess rather than replacing a guess. Electronic look-up is now so speedy that it takes very little time away from reading.

During extensive reading, vocabulary gains occur through guessing from context and through dictionary look-up. Guessing from context also includes gaining knowledge of unfamiliar family members of previously known words. In addition, as Pigada and Schmitt (2006) have shown, guessing from context can strengthen and enrich knowledge of partially known words. Guessing from context involves drawing on a wide variety of clues including information from the linguistic context (both the immediate context and the wider context), knowledge gained from earlier parts of the text, knowledge of the world and particular subject areas, common-sense, and morphological clues. It is likely that noticing is a factor affecting learning from guessing through context, and words that are consciously guessed are likely to be better retained than words which are not consciously noticed.

Where a word is repeated, the occurrences of the word after the first meeting provide an opportunity for a combination of guessing from context clues and retrieval of the meaning of the word gained from previous meetings. This retrieval is likely to have strong positive effects on learning, but its success is dependent on the ability of the reader to see the connection between the present context and previous contexts. There is some evidence that learners differ in this skill (van Daalen-Kapteijns & Elshout-Mohr, 1981). Research on glossing has shown the importance of retrieval (Barcroft, 2007). Better retention was found for words that were not glossed every time they occurred, but which were glossed on the first meeting and then not glossed on every subsequent meeting, allowing retrieval to occur. Glossing removes the need for retrieval to occur because the meaning is provided.

Where a word occurs several times within the same text, or even in different texts, in the vast majority of the cases the linguistic contexts are not the same. This occurrence of varied contexts is likely to be the strongest quality factor affecting vocabulary learning from guessing from context. This of course underlines the importance of quantity of input in learning from guessing from context, because the larger the amount of input, the greater the opportunity for repetition and meeting the word in varied contexts.

The Four Strands

The principle of the four strands says that a well-balanced language course should involve equal amounts of (a) learning through comprehensible listening and reading input, (b) learning through pushed spoken and written output, (c) deliberate language-focused learning, and (d) learning through fluency development in each of the four skills of listening, speaking, reading and writing (Nation, 2007; Nation, 2013a). Each strand provides different kinds of opportunities for learning and the combination of these opportunities sets up ideal conditions for vocabulary learning. Extensive reading fits into the meaning-focused input strand and also the fluency development strand. It should make up around half of the meaning-focused input strand (the other half

involves extensive listening), and one-quarter of the fluency development strand (the other three-quarters involve the three skills of listening, speaking, and writing). Because each strand should take up one-quarter of the course time, the reading fluency development part of the fluency strand should take up one-sixteenth of the total course time. In total, extensive reading should make up around three-sixteenths of the course time.

In a well-designed extensive reading course around two thirds of this time should be spent on reading material which contains a small proportion of unknown words (around 2% of the running words). Around one-third of the time should be spent reading very easy material containing little or no unknown words with a focus on reading for fluency. While there is no direct research justification for these proportions, applying the four strands principle provides a rational, justifiable way of deciding how much time should be given to each kind of activity in a course.

Cost/Benefit Analysis

The cost/benefit principle says that time spent on an activity should be justified by the learning benefits it brings. In other words, if just under one-quarter of the course time is spent on extensive reading, what research evidence is there that it results in substantial learning? The classic study on learning from extensive reading is the book flood experiment in Fiji by Elley and Mangubhai (1981a) in rural Fijian primary schools, although this did not include measures of vocabulary knowledge. Elley and Mangubhai found that by devoting three quarters of the time on the English course to extensive reading (less than three hours per week) the learners in the experimental group made the equivalent of fourteen months progress over the nine months of the course and these gains were still maintained a year later (Elley & Mangubhai, 1981b).

Much smaller scale experiments focusing on vocabulary gains from reading have shown that a reasonable number of words are learnt incidentally to various strengths of knowledge (Brown, Waring & Donkaewbua, 2008; Horst, 2005; Pigada & Schmitt, 2006; Pulido, 2004; Waring & Takaki, 2003). If only delayed posttest and recall measures are used, then the vocabulary gains from extensive reading are small (Waring & Takaki, 2003). If a range of immediate posttest vocabulary measures are used including sensitive measures, then the vocabulary gains are more substantial considering that they are the result of incidental learning.

Considering the Elley and Mangubhai study and the vocabulary-focused studies, it is reasonable to suggest that the single most significant change that a teacher could make to a language learning course would be to include a substantial extensive reading program. The three principles we have looked at (conditions, four strands, cost/benefit) clearly and strongly support the inclusion of an extensive reading program in a language learning course. Not only does extensive reading set up powerful vocabulary learning conditions, it provides a range of opportunities for learning, and delivers learning results that justify its inclusion.

Principles Guiding the Running of an Extensive Reading Program

In addition to principles that justify the inclusion of extensive reading in a language course, there are also principles that can guide the way an extensive reading program is run. These principles

are ranked in order of importance and cover comprehensible input, quantity of input, opportunities for learning, and maximising learning conditions.

Comprehensible Input

Extensive reading simply involves the learners quietly reading books which are at the right level for them. Ideally each learner would be reading a different book of their own choice, and they would be interested in what they are reading and be gaining enjoyment from the reading.

Because extensive reading involves reading texts which are at the right level for the reader, it is essential for low and intermediate proficiency learners to use graded readers. Graded readers are books written within carefully controlled vocabulary levels. The main effect of the control is to exclude words which are well beyond the learners' current level. Graded readers typically go up to the 3000 word level. The mid-frequency readers (Nation & Anthony, 2013) go from the 4000 to 8000 word levels. The development of mid-frequency reading texts is an attempt to make material at the right level of difficulty available even for learners of high proficiency, although Uden, Schmitt & Schmitt (2014) suggest that these may not be necessary. Nonetheless, successful guessing requires contexts that do not contain a high density of unknown words. Books written for young native speakers of English typically use a vocabulary which is much larger than the vocabulary size of foreign language learners. This is because young native speakers already know thousands of words. A seven-year-old native speaker for example knows at least 5000 words. For this reason, specially prepared graded readers are much more accessible for foreign language learners than books written for young native speakers. Every major ELT publisher has their series or several series of graded readers. The Extensive Reading Foundation runs an annual competition to find the best graded readers published that year, and the results of the competition can be found on the Extensive Reading Foundation website.

If a text contains too many unknown words, then it is likely to be difficult for learners to comprehend and they will thus have difficulty in guessing the unfamiliar words that they meet in the text. Research has suggested that learners need to know around 98% of the running words in a text for vocabulary not to be a major issue in comprehension (Hu & Nation, 2000; Schmitt, Jiang, & Grabe, 2011). There is growing debate over whether text coverage is a sensitive enough factor for determining the difficulty of texts.

Another aspect of unsimplified texts that is often over-looked is the number of different unknown words in a text. For a learner with a vocabulary size of 5000 words, the average unsimplified novel will contain over 2000 different words beyond this level, the vast majority of which will occur only once in the novel (see Nation 2014 for an example). Even with speedy electronic look-up as in Kindle this can be a discouragingly high number of look-ups. Simplified or adapted texts get rid of this unproductive burden.

Quantity of Input

The most important way that vocabulary learning from extensive reading can be increased is to do a lot of extensive reading. The minimum amount of reading should be around one graded reader every two weeks (Nation & Wang, 1999). This figure is regardless of the proficiency level

of the learners, because as the level of graded readers increases the length of the graded readers also increase. So, the graded readers written for beginning learners of English are only a few hundred words long, while the graded readers for intermediate learners are several thousand words long. One way to support large amounts of reading is to get the learners to do a speed reading course. Speed reading courses simply involve doing around 20 timed readings followed by comprehension questions. The aim of such a course is to bring learners' reading speed up to around 200 to 300 words per minute. Such courses are usually very successful (Tran, 2012). Table 2 (from Nation, 2014) provides not only weekly time requirements, but also daily (5 days a week) time requirements for extensive reading. It assumes the goal of learning around 1000 word families a year by meeting each word around 12 times.

Table 2: *Amount of reading input and time needed to meet the word families in each of the most frequent nine 1000 word families 12 times*

1000 word list level	Amount to read (tokens)	Time needed in one year for reading per week (and per day) at a reading speed of 150 words per minute
2nd 1000	200,000	33 minutes per week (7 minutes per day)
3rd 1000	300,000	50 minutes (10 minutes per day)
4th 1000	500,000	1 hour 23 minutes (17 minutes per day)
5th 1000	1,000,000	2 hours 47 minutes (33 minutes per day)
6th 1000	1,500,000	4 hours 10 minutes (50 minutes per day)
7th 1000	2,000,000	5 hours 33 minutes (1 hour 7 minutes per day)
8th 1000	2,500,000	6 hours 57 minutes (1 hour 23 minutes per day)
9th 1000	3,000,000	8 hours 20 minutes (1 hour 40 minutes per day)

Note. The per-week figure is based on forty weeks. The daily rate is based on 5 days per week.

Table 2 shows that from the 4th 1000 level on, the increase required in the amount of reading is 500,000 words per year. From the 7th 1000 level on, over an hour a day five days a week, forty weeks of the year would need to be devoted to reading. This is a lot, but it assumes that this quantity of input is coming only through reading. Spoken sources are of course possible but these provide less intensive input. It takes around two hours to watch a typical 10,000 token movie (a rate of around 83 words per minute, or just over half of a slow reading rate of 150 words per minute). Nonetheless, an hour to an hour and forty minutes five times a week at this proficiency level is possible.

Opportunities for Learning

We have already looked at how an extensive reading course should include the two strands of learning from comprehensible input where a small amount of vocabulary (no more than 2% of the running words) is outside the learners' current knowledge, and learning through fluency development where all the vocabulary is familiar and the focus is on reading quickly. Fluency development reading can involve re-reading previously read texts, and reading new texts that are well within the learners' knowledge. Re-reading provides valuable repetition, while reading new easy texts enriches previously met vocabulary.

There is also value in linking extensive reading to the meaning-focused output strand where learners talk or write about what they have read. Nation (2013b) has a whole chapter on linked skill activities which include this chaining of different skills with the same content focus. Linked

skill activities set up ideal conditions for vocabulary learning and are very similar to content-based instruction.

Maximise the Effects of the Learning Conditions

An important way of improving vocabulary learning from extensive reading is to combine extensive reading with deliberate learning. If learners confirm the meaning of a word, for example by looking it up in a dictionary after they have guessed it from context clues, this greatly increases learning (Fraser, 1999; Mondria, 2003). Mondria also found that there was no significant difference between deliberate learning after guessing and deliberate learning with no guessing. Guessing in itself does not seem to have any special qualities that enhance learning. However, repeated opportunities to meet a word in varied contexts may provide the opportunity for retrieval and enriching knowledge of the word.

Dictionary use applies the condition of deliberate noticing, particularly when it occurs with the first meeting of a word. There are two major kinds of receptive dictionary use. One kind of use involves simply gaining quick access to the meaning or some other information about a word. This is by far the most common kind of dictionary use. The other kind of use is focused on learning and can be considered a vocabulary learning strategy. The dictionary use strategy involves using a dictionary to help remember a word, and is probably best used with words that are already partly known. When applying the dictionary use strategy, the learner looks through the various senses of the word in order to find out the core meaning, that is, the meaning that runs through all of the senses. The learner also looks at the form of the word and at entries which are near the entry for the word to find words that share the same word stem. If the dictionary contains easily accessible and comprehensible etymological information, then the learner also looks at this to enrich knowledge of the word. Most learner dictionaries contain plenty of examples, and part of the dictionary use strategy can involve looking at these examples to gain some idea of the use of the word and what words it collocates with. The dictionary use strategy is a deliberate learning strategy, and learners need training to use it well.

Lower proficiency learners need to make use of bilingual dictionaries. This is because in order to use a monolingual dictionary, a learner needs to have a vocabulary size of around 2000 to 3000 words in order to understand the definitions. Dictionary look-up results in vocabulary learning (Bruton, 2007; Hulstijn, Hollander, & Greidanus, 1996; Knight, 1994; Laufer & Hill, 2000; Peters, 2007). Most studies show that learners do not use dictionaries well, and that the amount and quality of dictionary look-up depends on the saliency of words in the text, and learners' view of the goal and importance of the reading task (Peters, 2007).

Dictionaries have long been accepted as an essential tool in learning another language, and with the growth of electronic media and electronic devices there have been enormous changes in the availability and accessibility of dictionaries. Electronic readers and electronic reading apps now come with their attached dictionary for easy access while reading. Look-up can occur quickly with just the touch of a finger. In addition, the quality of dictionaries especially learners' dictionaries has continued to improve. This traditional tool has a very modern face.

Utilizing deliberate learning also means that unknown words found during extensive reading

should be immediately put on word cards with their translation on the other side so that they can become the focus of deliberate study later. If the learners are reading graded readers, then almost every word in the books they read will be a useful word for them and would be well worth learning. Dictionary use will help in the making of word cards and will also help in remembering the words through focused noticing.

If the learners are reading unsimplified texts, then there is value in doing narrow reading. Narrow reading involves reading within a very limited topic area. This narrow content focus dramatically reduces the number of different words that the learners will meet, generally to around 50% of the different words they would meet if they were reading the same amount on a wide variety of topics (Sutarsyah, Nation, & Kennedy, 1994).

Re-reading books that have already been read before is a way of increasing reading fluency. It also has the positive effect of allowing repetition and retrieval of previously met words. The re-reading should probably be within a few weeks of the first reading of the book so that receptive retrieval of vocabulary is likely to be successful.

We have looked at two sets of principles that justify and guide the use of extensive reading in a language course. These principles are well supported by research and theory. The practical guidelines derived from the principles are:

1   include an extensive reading program as a part of your language course
2   make sure that learners spend enough time each week on extensive reading, either around 3/16 of the total course time or better still enough time to meet the words often that they need to learn
3   make sure that there are two strands to the extensive reading program – (a) the strand where they read texts at the right level for them (around 2% of the running words are unfamiliar), and (b) the fluency strand where they read easy familiar texts quickly
4   support the fluency development strand by getting learners to do a timed readings course
5   support vocabulary learning from extensive reading by getting the learners to do dictionary look-up, preferably while reading electronic texts
6   support vocabulary learning from extensive reading by getting the learners to note unfamiliar words on word cards for later independent study
7   link some of the extensive reading to extensive listening, and to speaking and writing about what has been read
8   if necessary, provide training in the guessing from context and word card strategies.

Teachers are often reluctant to include a substantial extensive reading program in their language course, largely I think because it does not involve direct teaching and involves the learners sitting quietly and getting on with their reading. Teachers may feel guilty about playing such a passive role. However, a good extensive reading course puts good principles of learning and teaching into practice and both the principles and the practice of extensive reading are supported by research. Teachers should feel a sense of accomplishment and satisfaction in having an extensive reading program as part of their course.

References

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*, 35–56.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*, 136–163.

Bruton, A. (2007). Vocabulary learning from dictionary reference in collaborative EFL translational writing. *System*, *35*, 353–367.

Elley, W. B., & Mangubhai, F. (1981a). *The impact of a book flood in Fiji primary schools*. Wellington: New Zealand Council for Educational Research.

Elley, W. B., & Mangubhai, F. (1981b). The long-term effects of a book flood on children's language growth. *Directions*, *7*, 15–24.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, *21*, 225–241.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: a measurement study. *Canadian Modern Language Review*, *61*, 355–382.

Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430.

Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*, 327–339.

Knight, S. M. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, *78*, 285–299.

Laufer, B. (in press). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*.

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology*, *3*, 58–76.

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries and vocabulary learning. *Language Learning, 43*(2), 263–287.

Mondria, J. A. (2003). The effects of inferring, verifying and memorising on the retention of L2 word meanings. *Studies in Second Language Acquisition*, *25*, 473–499.

Nation, I. S. P. (2007). The four strands. *Innovation in language learning and teaching, 1*, 1–12.

Nation, P. (2013a). *What Should Every EFL Teacher Know?* Seoul: Compass Publishing.

Nation, P. (2013b). *What Should Every ESL Teacher Know?* Seoul: Compass Publishing. (available free in electronic form from www.compasspub.com/ESLTK)

Nation, I.S.P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*, 1–16.

Nation, I. S. P., & Anthony, L. (2013). Mid-frequency readers. *Journal of Extensive Reading*, *1*, 5–16.

Nation, P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, *12*, 355–380.

Peters, E. (2007). Manipulating L2 learners' online dictionary use and its effect on L2 word retention. *Language Learning & Technology*, *11*, 36–58.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*, 1–28.

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *54*, 469–523.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*, 26–43.

Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based study. *RELC Journal, 25*, 34–50.

Tran, Y. T. N. (2012). The effects of a speed reading course and speed transfer to other types of texts. *RELC Journal*, *43*, 23–37.

Uden, J., Schmitt, D., & Schmitt, N. (2014). Jumping from the highest graded readers to ungraded novels: Four case studies. *Reading in a Foreign Language*, *26*, 1–28.

van Daalen-Kapteijns, M. M., & Elshout-Mohr, M. (1981). The acquisition of word meanings as a cognitive learning process. *Journal of Verbal Learning and Verbal Behavior*, *20*, 386–399.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163.

Webb, S. (2005). Receptive and productive vocabulary learning. *Studies in Second Language Acquisition*, *27*, 33–52.

Webb, S., & Nation, I.S.P. (in preparation). *How Vocabulary is Learned*. Oxford: Oxford University Press.

About the Author

Paul Nation is emeritus professor in the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. He writes and researches in the areas of second language vocabulary learning and language teaching methodology. His most recent books on vocabulary are Nation, I.S.P. and Webb, S. (2011) Researching and Analyzing Vocabulary. Boston: Heinle Cengage Learning, and Nation, I.S.P. (2013) Learning Vocabulary in Another Language. Cambridge: Cambridge University Press (second edition). E-mail: Paul.Nation@vuw.ac.nz