

Unknown Vocabulary Density and Reading Comprehension

Marcella Hu Hsueh-chao and Paul Nation

LALS, Victoria University of Wellington, New Zealand

The purpose of this study is to see what percentage coverage of text is needed for unassisted reading for pleasure, where learners are able to read without the interruption of looking up words. The study examines the effect of three densities of unknown vocabulary on two measures of reading comprehension, a multiple-choice test and a cued written recall test. Where 80% of the running words in the text were familiar to the readers, none gained adequate comprehension on either of the measures. Where 90% and 95% of the tokens in the text were familiar to the readers some gained adequate comprehension but most did not. A simple regression analysis of the data showed a predictable relationship between the density of unknown words and degree of comprehension. It seems that around 98% coverage of vocabulary is needed for learners to gain unassisted comprehension of a fiction text.

VOCABULARY KNOWLEDGE AND COMPREHENSION

The relationship between vocabulary knowledge and reading comprehension is complex and dynamic. One way of looking at it is to divide it up into two major directions of effect - the effect of vocabulary knowledge on reading comprehension (which is the main focus of this paper) and the effect of reading comprehension on vocabulary knowledge or growth.

Chall (1987) suggests that these two effects achieve prominence at different times for young native speakers of English. When they begin to learn how to read, native speakers' vocabulary knowledge supports their reading comprehension. That is, typically they work with texts that contain only known vocabulary. As native speakers begin school with a vocabulary approaching 5,000 word families this is not difficult to arrange. After three or four years of learning to read, the relationship changes. Having gained control of many of the skills of reading, reading can become a means of vocabulary growth. That is, the learner learns new vocabulary through reading words that have not been met elsewhere.

Researchers have suggested several models to describe the relationship between vocabulary knowledge and reading comprehension. The factors involved in these models involve language knowledge (of which vocabulary knowledge is a part),

PAUL NATION teaches in the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. He has taught in Indonesia, Thailand, the United States, Finland, and Japan. His specialist interests are language teaching methodology and vocabulary learning. He has just completed a draft of a new book called *Learning Vocabulary in Another Language*.

MARCELLA HU HSUEH-CHAO was a successful candidate for an MA in Applied Linguistics at Victoria University of Wellington. The research reported here was conducted as a part of her MA. She teaches English in Taiwan.

knowledge of the world (sometimes called background knowledge) and skill in language use (of which reading comprehension is one result).

A large number of different types of studies have shown the strong statistical relationship between vocabulary and language use. The causal relationships, however, are not so clear. Anderson and Freebody (1981) and Nation (1993) distinguish three views that are reflected in research.

The instrumentalist view sees vocabulary knowledge as being a major prerequisite and causative factor in comprehension. Good vocabulary knowledge supports good comprehension. Nagy (1991, personal communication) sees a need to be cautious about adopting this viewpoint in that it can lead to a "reductionist" view of reading, i.e. if you know the vocabulary then that is about all you need to be able to read. Diagrammatically, the instrumentalist view can be represented as follows.

vocabulary knowledge → reading comprehension

The aptitude view sees vocabulary knowledge as one of many outcomes of having a good brain. Good reading comprehension is also one of these outcomes. Other outcomes might include skill at non-verbal puzzles and the ability to understand oral explanations. Diagrammatically the aptitude view can be represented as follows.

mental aptitude → large vocabulary
 → good reading comprehension
 → etc.

The knowledge view sees vocabulary as an indicator of good world knowledge. This world knowledge supports reading comprehension because the reader must bring as much information to the text as the reader expects to get from it. It is difficult to read about astrophysics if you know nothing about it. Diagrammatically, the knowledge view can be represented as follows.

knowledge and experience → reading comprehension
 ↓ ↗
 vocabulary knowledge

Anderson and Freebody (1981) point out that no serious scholar holds any of these positions to the exclusion of the others and different relationships exist at different stages of vocabulary growth and skill development.

Mezynski (1983) suggests a fourth view. The access view of the relationship between vocabulary knowledge and language use, like the instrumental view, sees vocabulary as having a causal relationship with comprehension provided that the vocabulary can be easily accessed. Access can be improved through practice. This access can involve several factors including fluency of lexical access, speed of coping with affixed forms, and speed of word recognition.

For non-native-speakers of a language, the relationships are even more complicated. These complications can arise from the learners already being able to read in their first language, the use of a different writing system in the second language compared with the first language (say for Japanese learners of English), and the common situation of beginning to read the second language with virtually no vocabulary knowledge in the second language.

DENSITY AND COVERAGE

There have been several studies which attempt to determine the amount of vocabulary needed by a second language learner in order to be able to read with reasonable comprehension and without lack of vocabulary knowledge becoming a major burden. One approach to this has been to take a commonsense view of the issue and to see how the density of unknown vocabulary and vocabulary size are related in various kinds of texts. This approach makes assumptions about desirable and undesirable densities. Table 1 provides the basis for these assumptions.

% text coverage	Density of unfamiliar and familiar tokens	Number of text lines per 1 unfamiliar word
99	1 in 100	10
98	1 in 50	5
97	1 in 33	3.3
96	1 in 25	2.5
95	1 in 20	2
90	1 in 10	1
80	1 in 5	0.5

Table 1: The number of unfamiliar tokens per 100 tokens and the number of lines of text containing one unfamiliar word.

Table 1 shows that if learners have an 80% coverage of the running words or tokens in a text, then one in every five running words is likely to be an unknown word. This is the same as there being two unknown words per line, if a line contains on average about ten words. A density of two unknown words per line, particularly two unknown content words, would make reading very difficult and would probably result in low levels of comprehension. One of the most significant features of Table 1 is the way that each percentage change in coverage makes a greater change in the density of unknown words around the 95% coverage point. With 95% coverage, one in twenty running words is unknown. With 96% coverage, this drops to one in twenty-five, and with 97% coverage to one in thirty-three. By contrast, moving from 80% coverage to 81% coverage makes very little difference in the density of unknown vocabulary.

LANGUAGE THRESHOLDS

These features of coverage have led several researchers to consider that there may be a threshold where vocabulary knowledge becomes sufficient for adequate comprehension. If the learner is on one side of the threshold, vocabulary knowledge is not sufficient for adequate comprehension. If the learner is on the other side, then the learner knows enough vocabulary, other things being equal, to gain adequate comprehension of the text.

Before we look more closely at this idea of a language knowledge threshold, let us look at how text coverage and vocabulary size are related. The assumption which lies behind this discussion is that vocabulary learning is strongly affected by word frequency. That is, words which occur frequently in the language tend to be learned before words that occur less frequently. There is plenty of evidence to show that this is largely true. For example, Read (1988) in a study of the Vocabulary Levels Test (Nation 1983 and 1990) found that second language learners' scores on the various levels of the test decreased from the high frequency levels to the lower frequency levels. That is, the test had strong implicational scaling. Laufer and Nation (1999) found a similar effect for productive knowledge.

The relationship between text coverage and vocabulary size is strongly affected by the kind of text that is looked at. The high frequency words of English (West 1953) provide poorer coverage of academic text than they do of fiction, for example. Table 2 gives coverage figures for various kinds of text.

	1st 1000	2nd 1000	Academic vocabulary	Others
Conversation	84.3%	6.0%	1.9%	7.8%
Novels for younger readers	84.8%	5.9%		9.3%
Fiction	82.3%	5.1%	1.7%	10.9%
Academic	71.4%	4.7%	10.0%	13.9%

"Academic" includes the words from the Academic Word List, a sub-technical vocabulary (Coxhead, 2000).

"Others" includes low frequency words and proper names

Table 2: Vocabulary size and text coverage

In the present study, we will look at fiction texts. There are several reasons for this. Firstly, as Hirsh and Nation (1992) show, these texts are likely to be the most accessible unsimplified texts for second language learners in terms of vocabulary load. Secondly, the vast majority of simplified graded reading texts are fiction texts. Series of graded readers, like the Oxford Bookworms series, provide large quantities of very readable fiction texts at various vocabulary levels. Thirdly, graded readers and unsimplified novels do not usually require as much background knowledge as

academic texts or newspapers usually do. Thus, fiction texts are likely to represent some of the most favourable conditions for reading.

Table 2 indicates that the most common 2000 words of English provide quite high coverage of fiction. But knowing the most frequent 2000 words still does not provide enough coverage for reading to be comfortable. To read novels and have coverage of 95-98% of the running words, a learner would need a vocabulary size of around 5,000 word families (Hirsh and Nation 1992), and this assumes that proper nouns, which typically account for 4-5% of the running words (Johansson and Hofland 1989; Francis and Kučera 1982) are counted as known words that do not need to be learned before reading a novel.

WHAT IS THE OPTIMUM COVERAGE LEVEL?

We have looked at the commonsense evidence for having a high coverage of the vocabulary in a text in order to gain adequate comprehension, but this is not the only evidence available. There has been experimental research which looks at this.

There has been a continuing interest in whether there is a language knowledge threshold which marks the boundary between not having enough language knowledge for successful language use and having enough language knowledge for successful use. There are at least two ways of defining what a threshold is.

- 1 One way is to see a threshold as an all-or-nothing phenomenon. If a learner has not crossed the threshold, then adequate comprehension is not possible. If the learner has crossed the threshold, then, other things being equal, comprehension is possible for all learners. This is the strong view of a threshold and the one that corresponds to its traditional meaning.
- 2 Another way is to see a threshold as a probabilistic boundary. That is, if a learner has not crossed the threshold, the chances of comprehending adequately are low. If the learner has crossed the threshold, the chances are on the side of the learner gaining adequate comprehension. This second definition of a threshold is the way that Laufer uses the term.

Laufer and Sim (1985) used comprehension questions and interviews with learners to determine a threshold score where learners could be said to be able to comprehend an English for Academic Purposes text in the First Certificate in English exam. They concluded that a score of 65-70% was the minimum necessary. As a result of interviews which involved seeing how learners dealt with the text, and supplying needed items to see how comprehension was affected, Laufer and Sim determined that the most pressing need of the foreign language learner was vocabulary, then subject matter knowledge, and then syntactic structure.

Laufer (1989) then went a step further to see what percentage of word tokens (running words) needed to be understood in order to ensure "reasonable" reading

comprehension of the text. Laufer set reasonable comprehension as a score of 55% or more. Percentage of word tokens known was found by getting learners to underline unfamiliar words in the text and adjusting this score by the number of words mistakenly said to be known as determined by a translation test. This is a much more demanding test than a multiple-choice recognition test. The rest was converted to a percentage of total word tokens in the text which were known.

This calculation should be expressed in 2 stages:

1 The number of words known in the text is:

Total words in text - [words reported as unknown + words reported as unknown x (number of discrepancies x 100/40)].

2 The coverage is [No. of words known in the text x 100/ Total no. of words in text].

In the first stage of the calculation, the number of discrepancies x 100/40 is "the bluff index" i.e. what percentage of vocabulary is not reported as unknown. When multiplied by the reported number of words, we get the number of words that was not reported as unknown but should have been. Therefore to calculate how many words in text are **really** unknown, we add the number of reported words to those that should have been reported. Then to see how many words were known, this number is subtracted from the number of words in the text.

The second stage converts the number of known words into the percentage of the total number of words in the text and this is the coverage.

Let us take an example. Say a person reports 20 unknown words, then in the translation test we find 10 discrepancies (that is, ten words said to be known, but which were not translated correctly). So the bluff index is $10 \times 100/40 = 25\%$. This means he should have reported as unknown $20 + 20 \times 25/100 = 25$. If the text has 200 words, then the number of words known is $200 - 25 = 175$. The coverage is $175 \times 100/200 = 87.5\%$ (Laufer, personal communication).

Laufer found that the group that scored 95% and above on the vocabulary measure had a significantly higher number of successful readers (scores of 55% and above on a reading test) than those scoring below 95%. The 90% level did not result in significant differences between those above and below. A comparison of the 95% and above group with the 90-94% group revealed a significant difference in comprehension scores. In this study, Laufer does not justify the 55% threshold of comprehension (it does not agree with the 65-70% threshold determined in the Laufer and Sim (1985) study) except to say it is the lowest passing grade in the Haifa University system.

The next step is to determine what vocabulary size (number of word types, lemmas, or families) will provide 95% coverage of academic text. Laufer (1989) accepts Ostyn and Godin's (1985: 353) evidence that the 4,839 words (types? lemmas? families?) in the Dutch school books that they had written provide 95-98% coverage of three randomly chosen newspaper clippings. There are several problems with this. First, evidence from frequency studies of Dutch is being applied to English. Second, the unit of counting is not specified - is it types, lemmas, or families? Third, newspapers are not academic text, and fourth, three newspaper clippings make a very, very small corpus.

Corpus studies of English can provide better estimates. Such studies show that the number of word families needed to cover a set percentage, say 95%, of the tokens in a text depends on (1) the type of text - is it a novel, newspaper, academic text, spoken informal conversation etc., (2) the length of the text, and (3) homogeneity of the text - is it on the same topic and by the same writer?

Laufer (1992) in a further study looked at the relationship between reading comprehension score (as measured by an English test produced in the Netherlands or the English subtest of the Israeli entrance examination) and vocabulary size, as measured by the Vocabulary Levels Test (Nation 1983) or the Eurocentre Vocabulary Test (Meara and Jones 1990). The minimal vocabulary level where there were more 'readers' than non-readers (56% in the reading comprehension test) was 3,000 word families.

Laufer has approached the vocabulary threshold question from several directions, by looking at the relative contributions of vocabulary, grammar and background knowledge to reading comprehension, by looking at vocabulary coverage and reading comprehension, and by looking at vocabulary size and reading comprehension. Her main interest has been in determining the minimal language proficiency level where teachers can usefully switch from concentrating on language development to the development and transfer of reading skills. Her studies have consistently shown the 3,000 word family level to be a minimum for the reading of unsimplified text.

Holley (1973) looked at the effect of the density of unknown words on vocabulary learning, comprehension, and learners' perception of difficulty and enjoyment. Time taken to read the text was also recorded. The subjects were learners of German as a foreign language. Seven versions of the same 750 word text were prepared, each with a different density of unknown words. The densities ranged from 1 unknown word in 150 (99.3% coverage) to 1 unknown word in 15 (93.3% coverage). Subjects were allowed 30 to 40 minutes to read the text and study the words. This represents a reading rate of around twenty words per minute which is very slow. The meanings of words were provided in a glossary that accompanied the text. The learners did

not have to guess from context and had plenty of time to study the words as well as read the text. Holley found that in terms of raw number of words learned, the number kept increasing as the density increased. This is not surprising as there are more words that can be learned as density increases (from 5 words to 50) and there was plenty of time to study them with their given meanings. Holley's experiment then probably did not necessarily measure the effect of unknown word density on learning and comprehension, but may have been measuring rote learning, and comprehension of text containing known and glossed items. There was no significant difference between the densities for comprehension. Holley's densities of unknown words were all light, especially when viewed in terms of the percentage coverage of known words (93.3% - 99.3%). Heavier densities of one unknown word in five (80% coverage) and one unknown word in ten (90% coverage) need to be looked at.

Research so far has not been able to provide a clear guideline about the optimal density of unknown words.

THE STUDY

This study attempts to answer the following research questions.

- 1 Will different densities of unknown words result in differences in comprehension? In particular, as the number of unknown words increases, will comprehension decline?
- 2 Is there a vocabulary coverage level which acts as a threshold between adequate and inadequate comprehension of a fiction text?

THE SUBJECTS

The sixty-six subjects were all adults attending a pre-university English course in an English speaking country. Most were preparing for undergraduate or post-graduate study in an English speaking university. They were chosen on the basis of their Vocabulary Levels Test scores (Nation 1983, 1990) from among the most proficient learners on the course. They came from a wide variety of language backgrounds including Thai, Chinese, Ni-Vanuatu, Indonesian, Japanese, Korean, Vietnamese and German.

THE TEXT

The text chosen for the study was a story about a person being tricked by a caller about an escape from a mental institution. It was 673 words long and was told in the first person with some dialogue. Four versions of the text were prepared by making two kinds of adjustments to the text. Firstly, the low frequency words in the text were replaced by nonsense words. In the 95% coverage version, 5% of the running words were replaced by nonsense words (32 words). In the 90% coverage version, 10% of the running words were replaced by nonsense words, and in the

80% version, 20% of the running words were replaced. The words to be replaced were chosen by referring to Francis and Kučera's (1982) frequency list and cross-checking with West's (1953) *General Service List of English Words* and the *Longman Dictionary of Contemporary English*. The words were replaced in reverse order of their frequency, with the least frequent words being replaced in the 95% version, and then the next most frequent words in the 90% version and so on. The fourth version of the text was a 100% version which contained no nonsense words (see Appendix). It was decided to replace the low frequency words with nonsense words rather than blank spaces because blank spaces are more distracting. Nonsense words were used rather than pretested low frequency words because learning can occur from pretesting, and more importantly, it would be difficult, if not impossible, to find suitable words that were not known by any of the learners and keep the density constant. Learners differ not only in vocabulary size but in the particular words they know (Saragi, Nation and Meister 1978). The Appendix contains the 80% version of the text. Where a word to be replaced by a nonsense word occurred more than once in the text, it was replaced by the same nonsense word. Repetitions of words were counted as separate tokens when calculating percentage coverage.

The second adjustment to make the four versions was simplification. In the 95%, 90% and 80% versions, all of the words which were not replaced by nonsense words had to be within the most frequent 2000 words of English. So, any words outside the 2000 word level were replaced by words or phrases within the 2000 level. This was to ensure that there were no vocabulary difficulties except for the nonsense words. In the 80% version, some of the replaced words were lower frequency items within the most frequent 2000 words. All of the words in the 100% version were similarly brought within the 2000 level.

All the learners taking part in the experiment were tested using the Vocabulary Levels Test. Any learner not receiving a score of more than 14 out of 18 at the 2000 word level as well as a score higher than 10 out of 18 at the 3000 level was not included in the experiment. Most of the subjects were well above these minimum scores. This selection was done to ensure that the learners taking part in the experiment would not have any vocabulary difficulties with the text except for the nonsense words. Table 3 summarises the data on the four versions of the text. The text contained 14 tokens which were proper nouns and exclamations. These were left unchanged in the text and were counted as familiar words.

Coverage	Number of nonsense words	Density of unknown words
100%	0	0
95%	32	1 in 20
90%	63	1 in 10
80%	127	1 in 5

Table 3: Number of words replaced by nonsense words in each of the four versions of the text.

Because the words replaced by nonsense words were chosen on the basis of their general frequency in English, each version of the text realistically represents the vocabulary problems that learners of differing vocabulary sizes would face when reading the text. The least frequent words in the original text *asylum*, *gurgling*, *lunatic*, *mystified*, *shrubbery*, and *quavering* occurred only once in the 1,000,000 token corpus used by Francis and Kučera (1982).

THE COMPREHENSION TESTS

Two comprehension tests were prepared - a multiple choice test and a cued written recall test. The texts were taken away before the learners sat the tests, so they could not refer back to them to match up any vocabulary clues to the answers.

THE MULTIPLE CHOICE TEST

The multiple choice test contained fourteen items (see Appendix). Each item had four choices. All items were written within the first 2000 words of English and covered the main ideas of the story. Four items dealt with explicit information and ten items dealt with implicit information. A native speaker divided the passage into idea units and classified these into four roughly equally sized groups - those that were in the top 25% idea units in terms of importance in the story, the top 50%, the top 75%, and the least important 25%. All of the idea units in the top 50% were covered by the multiple choice questions

Seven native speakers were also given the questions to answer without reading the text. They all scored less than five correct out of fourteen. This showed that the questions were not too easy and required knowledge of the text. Then the native speakers read the 100% version of the text and answered the questions again. Most got almost all questions correct. This showed that the questions were not too tricky.

THE CUED WRITTEN RECALL TEST

The multiple choice questions ensured that all the main ideas in the story and a good range of implicit items were tested. The cued written recall test gave learners the opportunity to show their understanding of the main ideas and the supporting detail. The test consisted of a set of pronominal questions that gave away little of the detail of the story and its vocabulary, but which covered all the steps of the story in chronological order. There were twenty-seven questions in total. Here are the first five.

Who is telling the story?

What was happening at the very beginning of the story?

What happened next?

What were her feelings when she spoke on the telephone?

Why did she feel this way?

A cued written recall test was used instead of a free oral recall, because pilot testing indicated that second language learners were likely to produce more if there was not the time pressure of oral recall and if they were prompted by questions. The aim of the cued written recall was to get learners to produce as much as they possibly could about the story, no matter how much they had understood.

SCORING THE MEASURES

The multiple choice questions were scored by giving one mark for each correct answer. The cued written recall was scored by weighting each idea unit according to its ranking in the top one quarter (4 points), top second quarter (3 points), third quarter (2 points), and bottom quarter (1 point) of the idea units. Where the answer involved several idea units, the points for each idea unit were added to each other. So each question's score depended on the number of idea units in the answer and the value of each idea unit. Learners were not penalized for spelling and grammatical errors if the meaning seemed clear. An inter-rater reliability check was performed on the scoring of the cued written recall. This involved an initial training session and then separate scoring of the learners' answers. An agreement of 95% was reached.

The number of words each learner wrote in the cued written recall was also counted to see if vocabulary density affected the quantity of words produced.

PREDICTION

After reading the text and before seeing the comprehension measures the learners were asked to predict what their score on the comprehension test was likely to be. This was done for two reasons, firstly to see how accurately learners could gauge their level of comprehension, and secondly to see if different versions of the text had different effects on learners' feelings about their level of comprehension. The prediction was done using a fifteen-point scale and a percentage scale.

THE PROCEDURE

Each learner was randomly assigned to one of four conditions - 100%, 95%, 90%, 80% text coverage and read the appropriate version of the text. Then without seeing the questions they were asked to predict what they thought they would get on a multiple choice test and what percentage of the text they considered they had understood.

The text and the predictions were collected and the learners were all given the same multiple choice test. When that was completed, it was collected and the learners were then given the same cued written recall test. The time to read the text and answer the questions was not limited and each learner indicated when they were

ready for the next part of the experiment. The time taken to read the text, and predict and answer the questions ranged from 23 minutes to 48 minutes.

Sixty-six learners were involved in the experiment, seventeen reading the 100% version, seventeen the 95% version, sixteen the 90% version and sixteen the 80% version.

The learners were informed of the purpose of the experiment, and how the results would be used and were given the opportunity to not participate or to withdraw at any time they wished. None chose to withdraw.

RESULTS

THE MULTIPLE-CHOICE QUESTIONS

There are various ways of deciding what can be considered as adequate comprehension of the text. These include the following.

- 1 One way is simply to make an arbitrary decision allowing for a degree of human error. Thus fourteen correct out of the fourteen multiple-choice questions is obviously adequate, and allowing for human error then twelve or thirteen correct out of fourteen should also be acceptable.
- 2 A second way to determine adequate comprehension is to use the scores of native speakers on the 100% version as the standard.
- 3 Another way is to use the non-native speaking subjects' scores on the 100% version as the standard of adequate comprehension.
- 4 All of the preceding ways are simply quantity measures. That is, they add up the number of questions answered correctly. It is possible to add a quality dimension to the decision by using the ranking of importance of the idea units. To achieve adequate comprehension using this criterion, learners should get most of the questions based on the top 50% of the idea units correct. Twelve of the fourteen questions were based on these idea units (questions 6 and 11 were not).
- 5 A combination of the quality and quantity measures is to give weighted scores for each question, with three points for a question focusing on an idea unit in the top 25% of idea units, two points for a question in the top 50% but not in the top 25%, and one point for a question in the top 75% but not in the top 50%. There were no multiple choice questions focusing on idea units in the bottom 25% of idea units.

All of these measures yielded results that were so similar that only one of them, setting an arbitrary level of at least twelve correct out of fourteen, will be discussed in detail.

Table 4 shows that fifteen of the seventeen subjects reading the 100% version got at least twelve correct out of the fourteen questions. That is, most subjects (15 out of 17) had adequate comprehension.

	100	95	90	80
14	2	2		
13	3	3	3	
12	10	1	1	
11	1	3	3	
10	1	2	1	2
9		2	4	
8		1		1
7			1	2
6		1	1	4
5		1	1	3
4		1	1	4
Total subjects	17	17	16	16
Average score	12.24	10.18	9.5	6.06
s.d.	0.97	3.04	2.82	1.94

Table 4: Number of subjects gaining various scores on the four versions of the text

Six of the subjects reading the 95% version gained adequate comprehension, but eleven subjects had a score lower than twelve. The results indicate that 95% text coverage is not adequate for most learners to gain good comprehension. With the 90% version, only four got twelve or higher. Twelve subjects did not reach the criterion of twelve correct. In the results of the 95% and 90% versions, the spread of scores from four to thirteen is quite similar. None of the subjects' reading the 80% version gained adequate comprehension. The scores on the 100% version cluster near the top of the comprehension scale. The scores on the 80% version cluster near the bottom and this is partly reflected in the standard deviations.

WRITTEN RECALL QUESTIONS

Table 5 gives the range of scores on the cued written recall measure for subjects reading the various versions of the text. The highest possible score is 124. No-one gained that score. One subject gained a score in the 100-109 range. That subject read the 100% version. Seven students gained a score in the 80-89 point range, of whom five read the 100% version and two read the 95% version.

Score range	100%	95%	90%	80%
100-109	1			
90-99	2			
80-89	5	2		
70-79	2	5	3	
60-69	5	3	2	
50-59	2	2	4	
40-49		2	4	1
30-39		3	2	5
20-29				5
10-19			1	4
0-9				1
Total subjects	17	17	16	16
Mean score	77.17	61	51.31	24.6
s.d.	14.49	18.2	15.97	11.2

Table 5: Number of subjects on the four different versions scoring within certain score ranges on the written recall measure

The spread of scores is somewhat similar to that of the multiple choice questions. If the boundary between adequate and inadequate comprehension is set around 70 out of 124 (near the average of the subjects reading the 100% version), then most of the learners reading the 100% coverage version gained adequate comprehension, a minority of those reading the 95% version did, as did fewer of those reading the 90% version. No learners reading the 80% version gained adequate comprehension.

TESTING THE DATA

The data from the two comprehension measures was tested using a simple regression model in SPLUS. The shortcomings of applying a regression model and interpreting the p values produced are (1) the variables are discrete not continuous, (2) coverage of known words (the predictor variable) has an allowed maximum, and (c) there is non-homogeneity of variance. However, both analyses broadly speaking are consistent with each other. The analysis of the residuals from the model indicates reasonable fit. The model shows that the average score depends on the coverage of known words according to a function which is estimated $a+b$ (coverage), where a = the intercept and b = the slope coefficient. Fitting the data to a model is a very stringent test of the relationship between coverage of known words and comprehension. Instead of comparing the four coverages for effects on reading comprehension, it places them in a fixed order and looks for a predictable relationship between word coverage and comprehension. If the data fits the model well, it is then possible to use the model to estimate the effect of coverages other than those measured in the experiment. Figures 1 and 2 are the plots for the two comprehension measures.

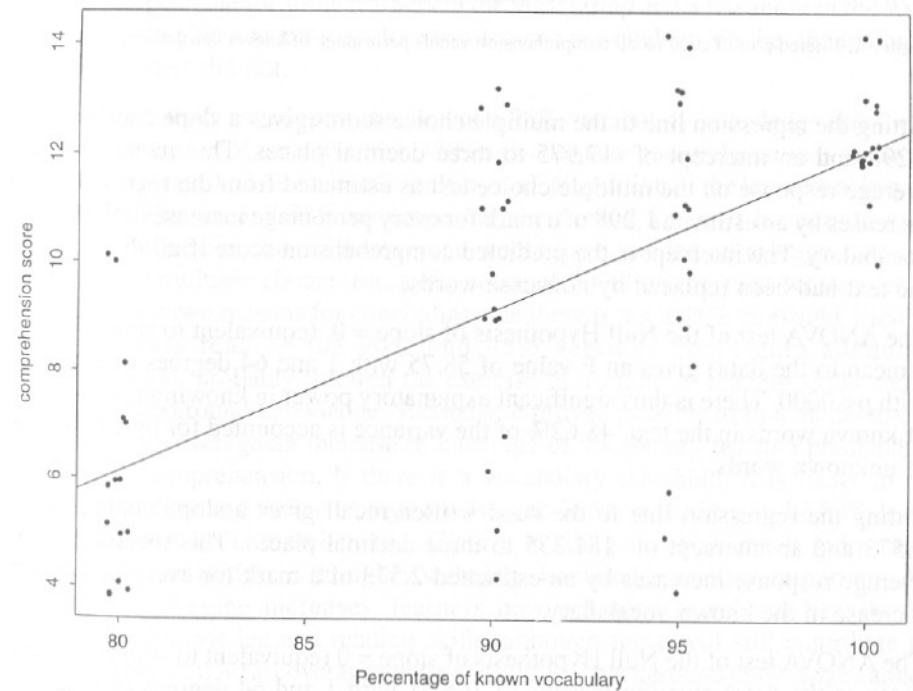


Figure 1: Jittered plot of multiple-choice comprehension versus percentage of known vocabulary.

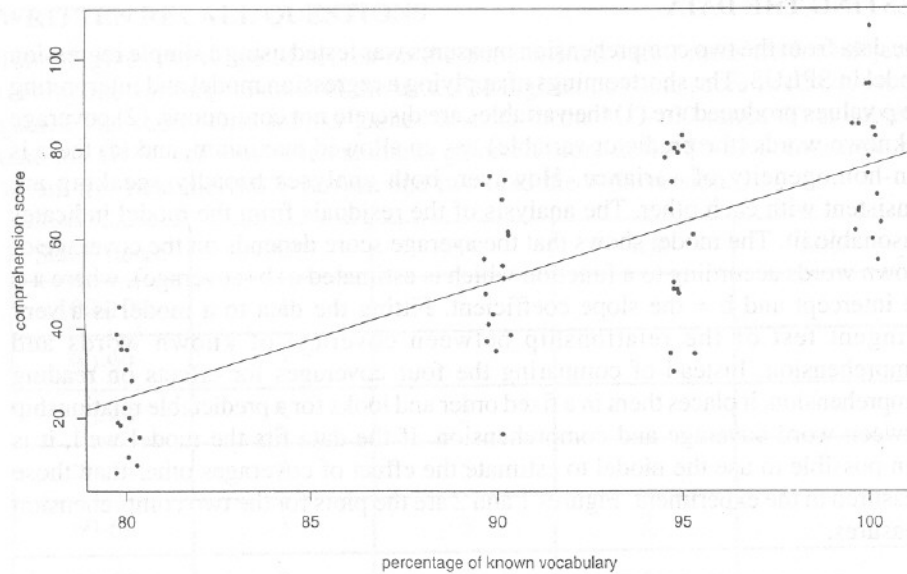


Figure 2: Jittered plot of cued recall comprehension versus percentage of known vocabulary.

Fitting the regression line to the multiple choice scores gives a slope coefficient of 0.298 and an intercept of -17.975 to three decimal places. This means that the average response on the multiple choice test as estimated from the regression line increases by an estimated .298 of a mark for every percentage increase in the known vocabulary. The intercept is the predicted comprehension score if all the words in the text had been replaced by nonsense words.

The ANOVA test of the Null Hypothesis of slope = 0, (equivalent to simply fitting a mean to the data) gives an F value of 58.75 with 1 and 64 degrees of freedom, with $p < .0000$. There is thus significant explanatory power in knowing the coverage of known words in the text. 48.62% of the variance is accounted for by the density of unknown words.

Fitting the regression line to the cued written recall gives a slope coefficient of 2.573 and an intercept of -181.335 to three decimal places. This means that the average response increases by an estimated 2.573 of a mark for every percentage increase in the known vocabulary.

The ANOVA test of the Null Hypothesis of slope = 0 (equivalent to simply fitting a mean to the data) gives an F value of 105.31 with 1 and 64 degrees of freedom, with $p < .0000$. Once again the coverage by known words (or alternatively, density of unknown words) has substantial explanatory power. 62.18% of the variance is

accounted for by the density of unknown words. The models for the multiple choice and cued recall questions show that each percentage increase in the coverage of known words makes a predictable difference in comprehension. Whether this difference is large enough to be of interest to a learner or teacher is a matter of judgement. The two comprehension measures correlated with each other at 0.84 ($p > .0000$).

The results for both the multiple choice and written recall questions agree on the following points.

- 1 On average, learners' comprehension scores increase to a predictable degree as the coverage of known words increases.
- 2 No learners reading the 80% coverage version of the text gained adequate comprehension. All learners in this group gained uniformly low comprehension scores.
- 3 The range of scores of learners in the 95% and 90% coverage groups was wide.
- 4 It was possible for some learners in the 90% group and a few more in the 95% group to gain adequate or close to adequate comprehension, but the majority of learners did not.

This research does not support the idea of a 95% vocabulary knowledge threshold for comprehension of narrative text. Although adequate comprehension is possible with this coverage by known words, only a small proportion of the learners achieved it. It seems that around 98% coverage may be needed for most learners to gain adequate comprehension. 98% coverage would have yielded an average score of 11.53 on the multiple choice test, and a score of 70.82 on the cued written recall test. There are two reasons for concluding that there is not a 95% threshold. Firstly, some learners reading the 90% and 95% versions could gain adequate comprehension, probably through the exercise of a range of reading skills and a wealth of background knowledge. Secondly, it proved possible to fit a model to the data where on average an increase in coverage of vocabulary led to a predictable increase in comprehension. If there is a vocabulary threshold, it is likely to be between the 80% and 90% coverage levels as no learners reading at the 80% level achieved adequate comprehension. Clearly, reading involves much more than vocabulary knowledge, although vocabulary knowledge is very important. As vocabulary coverage increases, learners may not need to rely so heavily on background knowledge and reading skills, although these will still contribute to comprehension. At 80% coverage however, skills and background knowledge cannot usually make up for lack of vocabulary knowledge.

LEARNERS' PREDICTION OF PERFORMANCE ON THE COMPREHENSION MEASURES

Predicted score	100%	95%	90%	80%
14	2			
13	4	2		1
12	4	3	1	
11	2	2	1	
10	1	1		
9	2	4	4	2
8	2	1	3	1
7		3	5	
6				3
5		1		5
4			1	2
3			1	1
2				1
Total	17	17	16	16

Table 6: Number of learners' self-predicted comprehension score for the various versions

Table 6 shows that two learners reading the 100% version predicted that they would get all fourteen multiple choice comprehension questions correct. Four learners predicted that they would get one wrong.

The pattern of the predictions generally reflects the comprehension scores, with predicted comprehension dropping from one coverage level to the next. The learners' impressions of the difficulty of the texts were clearly affected by the density of unknown words.

Table 7 shows how accurate the learners were in predicting their multiple choice comprehension score. Five of the seventeen learners reading the 100% version accurately predicted their score. Another five were only one point out of fourteen above or below the actual score.

	100%	95%	90%	80%	Total
Same score	5	3	3	1	12
1 point away	5	5	2	6	18
2 points away	4	2	5	2	13
3 points away	1	3	2	4	10
4 points away	1	1	2	1	5
5 points away	1	2	2	1	6
6 points away					
7 points away		1			1
8 points away					
9 points away				1	1

Table 7: The accuracy of learners' predictions of multiple choice comprehension score across the four versions

Two-thirds (43) of the learners predicted their score within two points of their actual score, and four-fifths predicted their score within three points. Learners show a reasonable degree of accuracy in gauging their score without having seen the comprehension questions. This does not seem to be unduly affected by the version of the text they read, although it is a little easier for those reading the 100% version who have no unknown vocabulary to feel uncertain about.

As could be expected, the prediction of the number correct on the multiple choice and the prediction of the percentage of the text understood correlated highly with each other (0.88, $p > .0000$). As Table 8 shows there were moderate correlations between the subjects' prediction of their comprehension and their subsequent actual comprehension scores.

	Prediction of number correct	Prediction of percentage understood
Multiple choice comprehension score	0.62*	0.67*
Cued recall score	0.67*	0.68*

* $p > .0000$

Table 8: Correlations between predicted scores and actual scores for the two measures.

The prediction measures show that learners see texts containing more unknown words as more difficult to comprehend and their prediction of their level of comprehension roughly agrees with their actual comprehension scores. Both performance and impression of difficulty are affected by the density of unknown words.

CONCLUSION

This study shows that the density of unknown words has a marked effect on text comprehension. The text used in this study was a fiction text with a strong chronological story line and was thus not a "difficult" text. Other text types, particularly newspapers and academic texts, would place greater demands on the reader. However, even with this reasonably easy text, most learners would need around 98% coverage to gain adequate unassisted comprehension of the text. This provides experimental support for the position taken by Hirsh and Nation (1992), namely that learners need to have around 98% coverage of the words in the text to be able to read for pleasure. It was found that there was a predictable relationship between density of unknown words and comprehension. As density of unknown words increases, comprehension drops. A few learners reading the 95% and 90% coverage versions seemed able to overcome some of the obstacles posed by the unknown words and gain adequate or close to adequate comprehension, but they were in a minority.

This conclusion must not be interpreted as saying that with 98% coverage of the vocabulary no other skills or knowledge are needed to gain adequate comprehension. All of the subjects in this study were readers in their first language, had considerable knowledge of English grammar, were experienced in reading English, and brought considerable background knowledge to their reading. These all contribute to their skill in comprehending text and account for some learners reading the 95% and 90% versions getting high scores. However, as readability studies show, vocabulary knowledge is a critical component in reading.

If the reading text used in this study was accompanied by a glossary as in the Holley (1973) study, then with this glossary and more time, learners would have achieved much higher levels of comprehension. The purpose of this study however was to see what percentage coverage of text was needed for unassisted reading for pleasure, where learners could read without interrupting the reading to look up words.

Teachers should keep the findings of this study in mind when guiding learners in choosing books for extensive reading, and when using meaning focused input activities in class. The findings agree closely with a study of native speakers by Carver (1994: 432) who concluded "when the material being read is relatively easy, then close to 0% of the words will be unknown, ... when the material is relatively

hard then around 2% or more of the words will be unknown, ... and when the difficulty level of the material is approximately equal to the ability level of the individual, then around 1% of the words will be unknown".

The results of this study have an important message for teachers of English. We can divide the reading done by EFL learners into three types:

- 1 Intensive reading where learners work through texts containing unfamiliar language features, but where learners are assisted by teacher guidance and discussion or by the use of dictionaries, glossaries or elaboration in the text.
- 2 Extensive reading for language growth where learners read texts containing some unfamiliar language features but cope with these with only minor interruptions to understanding the message of the text.
- 3 Extensive reading for developing fluency in reading skills where learners read texts containing no unknown language features and where the reading should not be interrupted by the need to deal with unfamiliar language features.

The research described in this study has looked at the second type of reading, extensive reading for language growth. The results suggest that for this kind of reading learners need to know at least 98% of the running words in the text. That is, the density of unknown words should be around one in fifty. This is the same figure arrived at by West (1955:21) and Hirsh and Nation (1992) simply by considering the amount of interruption to the flow of reading.

This study has taken a somewhat extreme position. That is, the learners in the study read the text without access to a dictionary or glossary. Further research is needed to see how learners cope with the unknown words. Are most unknown words easily guessed at the 98% coverage level? What proportion can be guessed from 95% coverage? When learners have access to dictionaries, what density of unknown words makes reading a long text too burdensome, so that enjoyment and a feeling of achievement disappears?

Perhaps the most useful application of this study will come in guiding learners in the selection of graded readers. If learners are to read graded readers for pleasure, picking up some new vocabulary as they read, then they need to choose texts that do not have too heavy a vocabulary load. If the vocabulary load is greater than one unknown word in fifty running words, then extensive reading has become intensive reading. These two types of reading practise some similar skills and some different skills. Teachers need to ensure that extensive reading truly is extensive reading, so that the unique skills that it develops have a chance to be practised.

REFERENCES

- Anderson, R.C. and Freebody, P. (1981) Vocabulary knowledge. In J.T. Guthrie (ed): 77-117.
- Arnaud, P.J.L. and Béjoint, H. (Eds) (1992) *Vocabulary and Applied Linguistics*. London: Macmillan.
- Carver, R.P. (1994) Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: implications for instruction. *Journal of Reading Behavior*, 26(4), 413-437.
- Chall, J.S. (1987) Two vocabularies for reading: recognition and meaning. In McKeown and Curtis: 7-17.
- Coxhead, A. (2000, in press) A new academic word list. *TESOL Quarterly*, 34.
- Francis, W.N. and Kučera, H. (1982) *Frequency Analysis of English Usage*. Boston: Houghton Mifflin Company.
- Guthrie, J.T. (Ed) (1981) *Comprehension and Teaching: Research Reviews*. Newark: International Reading Association.
- Hill, L.A. and May, D.J. (1978) *Appreciation Pieces for Overseas Students*. London: Oxford University Press.
- Hirsh, D. and Nation, P. (1992) What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.
- Holley, F.M. (1973) A study of vocabulary learning in context: the effect of new-word density in German reading materials. *Foreign Language Annals*, 6, 339-347.
- Johansson, S. and Hofland, K. (1989) *Frequency Analysis of English Vocabulary and Grammar*. Volumes 1&2. Oxford: Clarendon Press.
- Laufer, B. and Sim, D.D. (1985) Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals*, 18(5), 405-411.
- Laufer, B. (1989) What percentage of text-lexis is essential for comprehension? In C. Lauren and M. Nordman (Eds) *Special Language: From Humans Thinking to Thinking Machines* Clevedon: Multilingual Matters.
- Laufer, B. (1992) How much lexis is necessary for reading comprehension? In Arnaud and Bejoint (1992): 126-132.
- Laufer, B. and Yano, Y. (in preparation) Learners' estimations of vocabulary knowledge.

- Laufer, B. and Nation, I.S.P. (1999) A vocabulary size test of controlled productive ability. *Language Testing*, 16(1), 36-55.
- McKeown, M.G. and Curtis, M.E. (eds) (1987) *The Nature of Vocabulary Acquisition*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Meara, P. and Jones, G. (1990) *Eurocentres Vocabulary Size Test. 10KA*. Zurich: Eurocentres.
- Mezynski, K. (1983) Issues concerning the acquisition of knowledge: effects of vocabulary training on reading comprehension. *Review of Educational Research*, 53(2), 253-279.
- Ostyn, P. and Godin, P. (1985) RALEX: an alternative approach to language teaching. *Modern Language Journal*, 69(4), 346-355.
- Nation, I.S.P. (1983) Testing and teaching vocabulary. *Guidelines*, 5(1), 12-25.
- Nation, I.S.P. (1990) *Teaching and Learning Vocabulary*. Rowley, Mass: Newbury House.
- Nation, I.S.P. (1993) Vocabulary size, growth and use. In R. Schreuder and B. Weltens (eds) *The Bilingual Lexicon*. Amsterdam/Philadelphia: John Benjamins. 115-134.
- Read, J. (1988) Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19(2), 12-25.
- Saragi, T., Nation, I. S. P. and Meister, G. F. (1978) Vocabulary learning and reading. *System*, 6(2), 72-78.
- West, M. (1953) *A General Service List of English Words*. London: Longman, Green & Co.
- West, M. (1955) *Learning to Read a Foreign Language*. London: Longman. 2nd ed.

APPENDICES

THE ESCAPED MADMAN (100% VERSION)

I was in the middle of making a cake for dinner when the telephone rang. I washed my hands and went to answer it.

'Hullo,' said an anxious voice when I had picked up the receiver, 'could I speak to Mrs. Scott, please?'

'Speaking,' I answered.

'Mrs. Rosalind Scott?' the voice went on, even more anxiously.

'Yes,' I answered. 'What's the matter?' By now I was getting a bit frightened myself. Thoughts of my husband being in a railway crash, or of one of the children breaking his leg playing football in the school-yard raced through my mind.

'This is the Northfields Hospital,' said the voice. 'One of our patients has escaped, and one of his friends heard him say that he was going to kill you. We thought we'd better warn you.'

'Oh—yes—thank you,' I answered and stopped to think very quickly. 'What's the name of this—er—patient?' I then went on.

There was a pause at the other end, as if someone was searching through a list. Then, 'George Hitchcock is his name,' came the reply. 'He's short and dark-haired and very dangerous.'

'But why does he want to—er—kill me? I've never even heard of him.'

A strange sound came from the telephone and it went dead. Thoughts of people cutting telephone wires before attacking a house flashed through my brain. Would a madman think of doing a thing like that?

I was still holding the receiver when there was a knock at the front-door. I dropped the receiver and raced to the door. I locked it just as the man outside was beginning to open the cover of the letter-box. Then I rushed to the back-door, knocking a small table over on my way and breaking the flower-bowl on it to pieces. The water made the floor smooth and I slipped and crashed into a chair, knocking that over too. However, I reached the back-door, locked it and then went around the house making sure all the windows were shut. They had no bars on them, so it would be easy to break one open.

Meanwhile, the telephone-operator, finding I had not put my telephone down and hearing the crashes and knocking, thought there was a fight going on in my house and telephoned the police. In a few minutes, I heard heavy footsteps advancing up the drive and there was a firm sound on the knocker of the front door.

'Who's that?' I said in a trembling voice.

'Police,' came the answer. 'Is there anything wrong, Lady?'

I thought fast. 'Come round to the window so that I can see you,' I said. The heavy footsteps circled the garden, and two policemen—a police officer and a younger policeman—appeared at the dining-room window. I ran joyfully to the front door to let them in. The mailman was standing a few yards off, looking at me in a way that struck me as unusual.

'Did you catch the madman?' I asked the police-officer.

'What madman?' he asked, clearly completely confused.

'Why, the one that was knocking at my door. He escaped from the hospital, and his name's Hitchcock. George Hitchcock.'

'George Hitchcock?' said the officer, even more confused than before. 'But that's the name of the Governor of the hospital.'

'And it was me that knocked at your door,' said the mailman.

We looked at each other—the two policemen, the mailman and I. Fright, surprise and suspicion followed each other across our faces. Then the police officer said, 'Somebody must have been having a joke with you, lady, and a stupid joke at that.' We all burst into relieved laughter.

THE CULVAZED SELGIAN (80% VERSION)

(Note: the unmarked nonsense words and the nonsense words in bold and italics were all in the 80% version. The nonsense words in bold, and bold and italics were the only nonsense words in the 90% version. The nonsense words in italics were the only nonsense words in the 95% version. In the texts read by the learners none of the nonsense words were highlighted in bold or italics.)

I was in the seplenity of making a *rane* for pestance when the pilution whimmt. I premurred my hands and went to answer it.

'Hullo,' said a **stergic** voice when I had teladed up the tangerity, 'could I speak to Mrs. Scott, please?'

'Speaking,' I answered.

'Mrs. Rosalind Scott?' the voice went on, even more **stergically**.

'Yes,' I answered. 'What's the matter?' By now I was getting a kineary *sorant* myself. Thoughts of my husband being in a **cubular dract**, or of one of the children racomizing his courney playing **coziment** in the school-curander advelted through my mind.

'This is the Northfields **Slivian** Home,' said the voice. 'One of our glises has culvazed, and one of his **emseleards** heard him say that he was going to aleand you. We thought we'd better gleat you.'

'Oh—yes—**yard** you,' I answered and stopped to dreed my *latments*. 'What's the name of this—er—glise?' I then went on.

There was a fabument at the other end, as if someone was **swanding** through a buncike. Then, 'George Hitchcock is his name,' came the feldination. 'He's short and dark-haired and very unizacy.'

'But why does he want to—er—aleand me? I've never even heard of him.'

A strange *clairy* sound came from the pilution and it went pually. Thoughts of people cutting pilution undriments before rotacizing a house scizzled through my veculion. Would a *selgian* think of doing a thing like that?

I was still holding the tangerity when there was a *crang* at the front-door. I antelased the tangerity and advelted to the door. I rounced it just as the man outside was beginning to open the *rajera* of the letter-box. Then I globerized to the back-door, pertruding a small table over on my way and *higoning* the flower-*puliard* on it to *lacraments*. The water made the *alamuts adianned* and I *strang* and *dracted* into a sabrity, pertruding that over too. However, I reached the back-door, rounced it and then went around the house making sure all the windows were *garnage*. They had no venegades on them, so it would be easy to racomize one open.

Culberously, the pilution-quelson, finding I had not put my pilution down and hearing the *dractments* and *crangs*, thought there was a perbuerty going on in my house and piluted the orcian. In a few minutes, I heard heavy footsteps feunding up the drive and there was a *prined* sound on the pertruder of the front door.

'Who's that?' I said in a *gramonous* voice.

'Orcian,' came the answer. 'Is there anything wrong, *stragar*?'

I thought fast. 'Come round to the window so that I can see you,' I said. The rakacrant footsteps scalanded the *quenery*, and two orcians—an orcian *cheltian* and a *mirocan*—appeared at the *morice*-room window. I ran turnoazantly to the front door to let them in. The *dripler* was *heaking* a few curanders off too, looking at me in a way that privanated me as *utermous*.

'Did you sploud the *vesbian*?' I asked the orcian-*cheltian*.

'What *vesbian*?' he asked, *emartically* completely *poygant*.

'Why, the one that was *cranging* at my door. He escaped from the *selgian tardan*, and his name's Hitchcock. George Hitchcock.'

'George Hitchcock?' said the *cheltian*, even more *poygant* than before. 'But that's the name of the phantropist of the *slivian* home.'

'And it was me that pertruded at your door,' said the *dripler*.

We looked at each other—the orcian-*cheltian*, the *mirocan*, the *dripler* and I. *Sorance*, surprise and *sanctiance nuzriated* each other across our faces. Then the *orcian-cheltian* said, 'Somebody must have been having a *shorenge* with you, *stragar*, and a *prolatic shorenge* at that.' We all *lucay* into *salerized* prinalment.

THE MULTIPLE CHOICE QUESTIONS

The comments in brackets were not seen by the test takers. They indicate whether the question requires implicit or explicit interpretation of the text, and $\frac{1}{4}$ indicates

that the answer was in the top $\frac{1}{4}$ of the ranked information units, $\frac{1}{2}$ = top half, and $\frac{3}{4}$ indicates top three quarters.

1. Mrs. Scott worried at the beginning of the telephone call because (implicit, $\frac{1}{2}$)
 - her husband was in a railway crash.
 - a child broke his leg.
 - the person speaking to her was very worried.
 - the children were not at home.
2. The telephone call told Mrs. Scott that (explicit, $\frac{1}{4}$)
 - someone was going to attack her house.
 - someone wanted to kill her.
 - she had to go the Northfields Home.
 - her husband had escaped.
3. George Hitchcock (implicit, $\frac{1}{4}$)
 - was Mrs. Scott's husband.
 - was talking to Mrs. Scott on the telephone.
 - was a close friend of Mrs. Scott.
 - was supposed to be a crazy person.
4. Mrs. Scott stopped talking on the telephone because (implicit, $\frac{1}{2}$)
 - the wires to the house were cut.
 - she could not hear the other person at the other end.
 - she dropped the phone.
 - someone came to the door.
5. Mrs. Scott ran to the door (implicit, $\frac{1}{4}$)
 - to stop someone getting in.
 - to get the letter that had just arrived.
 - to tell someone about the telephone message.
 - to let her husband in.
6. Some things in Mrs. Scott's house were broken (explicit, $\frac{3}{4}$)
 - by George Hitchcock.
 - by Mrs. Scott.
 - by the postman.
 - by an unknown person
7. The letter-box was being opened by (explicit, $\frac{1}{2}$)
 - the postman.
 - Mrs. Scott.
 - an unknown person.
 - the police.

8. The police came to the house because (explicit, $\frac{1}{4}$)
 - a neighbour called them.
 - George Hitchcock called them.
 - the telephone operator called them.
 - Mrs. Scott called them.
9. The police were called because (implicit, $\frac{1}{4}$)
 - someone was trying to kill Mrs. Scott.
 - there was a lot of noise.
 - the postman became very worried.
 - Mrs. Scott's phone did not work.
10. When the police came (implicit, $\frac{1}{4}$)
 - Mrs. Scott thought they might be George Hitchcock.
 - they could not find the house at first.
 - they could not find Mrs. Scott.
 - they caught George Hitchcock.
11. The postman (implicit, $\frac{3}{4}$)
 - was helping the police.
 - could not understand Mrs. Scott's behavior.
 - was really George Hitchcock.
 - was a friend of Mrs. Scott's.
12. The policemen (explicit, $\frac{1}{4}$)
 - did not know about the madman.
 - did not know George Hitchcock.
 - were worried about the postman.
 - were looking for George Hitchcock.
13. Mrs. Scott (implicit, $\frac{1}{4}$)
 - had been tricked.
 - finally met George Hitchcock.
 - helped catch the madman.
 - did not know what had happened.
14. Everybody laughed because (implicit, $\frac{1}{2}$)
 - Mrs. Scott said something funny.
 - there was no serious problem.
 - the police were at the wrong house.
 - The postman was at the wrong house.

ACKNOWLEDGMENTS

We would like to express our gratitude to Dr Brian Dawkins of the School of Mathematical and Computing Sciences, Victoria University of Wellington for help with the statistical procedures and their interpretation.