

○ L2 VOCABULARY KNOWLEDGE IN AND OUT OF CONTEXT: IS IT THE SAME FOR READING AND LISTENING?

Hilde van Zeeland University of Nottingham
aexhv@nottingham.ac.uk

The vast majority of second language (L2) vocabulary research focuses on learners' knowledge of isolated word forms. However, it is unclear to what extent this knowledge can be used as an indicator of knowledge in context (i.e. reading and listening). This study aims to shed light on this issue by comparing ESL learners' knowledge of the meaning of isolated words ('decontextual knowledge') with their knowledge of the same words in both reading and listening ('contextual knowledge'). Decontextual knowledge was measured in a free recall interview. Contextual knowledge was measured through a task in which participants paraphrased sentences containing the target items from both a written and spoken narrative. Results showed that learners' decontextual and contextual knowledge agreed in 65% of the cases. This indicates a considerable gap between the two, and emphasises that scores on decontextualised vocabulary test should not be used as predictors of learners' vocabulary knowledge in context. In addition, learners demonstrated significantly better knowledge of word meaning in the reading than listening mode, which may be due to processing difficulties in listening as well as better inferencing opportunities in reading. Two additional factors found to affect both decontextual and contextual knowledge are word frequency and learners' vocabulary size.

KEY WORDS: L2 vocabulary knowledge, L2 reading, L2 listening, vocabulary knowledge in context

INTRODUCTION

Vocabulary knowledge can be conceptualised and measured in many different ways (Hirsh, 2010; Milton, 2009; Read, 2000; Schmitt, 2010). Examples of commonly used test formats are word completion tasks (Laufer & Nation, 1999), checklists (Meara, 1992), and multiple-choice (MC) tests, such as the Vocabulary Size Test (VST) (Nation, 2008; Nation & Gu, 2007) and Vocabulary Levels Test (VLT) (Nation, 1983, 1990; Schmitt, Schmitt & Clapham, 2001).

Although these tests are certainly informative, they are limited in the sense that they only measure learners' knowledge of isolated words out of context ('decontextual knowledge'), which is different from the knowledge required in the contexts of reading and listening ('contextual knowledge'). Although research has found that learners' decontextual vocabulary knowledge (i.e. size) affects their reading and listening comprehension (Qian, 1999; Stæhr, 2008), it appears that no research has explored the potential gap between these

two. How much of learners' decontextual knowledge is generally used in context? And does this depend on the context mode, i.e. reading or listening? As long as learners' scores on decontextualised vocabulary tests continue to be used as predictors of their reading and listening ability (e.g. Nation, 2006), these are crucial questions. This study aims to shed light on this issue.

VOCABULARY KNOWLEDGE IN AND OUT OF CONTEXT

In a review of general language assessment, Schoonen (2011) observed that hardly any research focuses on communicative language ability, or even on performance tasks such as understanding meaning through reading and listening. This awareness of the importance of investigating L2 knowledge in context is also growing in the field of vocabulary (Nation, 2007; Read, 2004, 2007; Read & Chapelle, 2001). As Read (2000) argues, decontextualised vocabulary tests ignore important dimensions of lexical knowledge, such as learners' ability to exploit their knowledge for communicative purposes, interpret words in a given context, and compensate for lack of vocabulary knowledge. In this line of thinking, Read and Chapelle (2001) have argued for a so-called 'interactionist' approach to assessment, in which traditional trait-tests (e.g. Yes/No and MC), are complemented by tests assessing learners' ability to deploy their lexical knowledge in specific contexts of use: 'in listening and reading tasks that would mean including questions which measure contextual understanding of lexical items in the text' (Read, 2007, p. 120).

Any assessment of contextual vocabulary knowledge should recognise that learners' knowledge likely differs between the reading and the listening context. Milton and Hopkins (2006) created a phonological version of the *X-Lex* (a written test of the Yes/No format) referred to as *Aural Lex*, and compared the orthographic and phonological vocabulary size of Arabic and Greek learners of English. Results revealed that, among the Greek learners, the orthographic vocabulary size was larger than the phonological vocabulary size, suggesting that written vocabulary size tests may overestimate learners' phonological knowledge.

While Milton and Hopkins' (2006) study revealed that learners' knowledge of isolated words differs between the written and spoken mode, it did not explore learners' relative *contextual* vocabulary knowledge in the two modes. Especially in the spoken mode, learners' contextual knowledge is likely to differ from their knowledge of isolated words. L2 listeners generally experience difficulties in recognising individual lexical items in a stream of speech (Field, 2008; Goh, 2000; Vandergrift, 2011), and often have problems distinguishing similar lexical units due to sound perception difficulties (Broersma & Cutler, 2008; Weber & Cutler, 2004). This raises the question to what degree known vocabulary is actually recognised in the listening process (Read, 2007).

Another vocabulary-related difference between the two modes, which is likely related to the relative difficulty of word recognition in listening, is that listeners tend to rely less on

vocabulary than readers do in text comprehension (Lund, 1991; Park, 2004; Reves & Levine, 1988). In fact, various studies have found that learners' vocabulary size is more closely related to their reading than to their listening comprehension. For example, Mecartty (2000) found that lexical knowledge explained a larger amount of variance in reading than listening comprehension, and Stæhr (2008) found that vocabulary size correlated more strongly with reading (.83) than with listening abilities (.69). Similar findings have been reported in studies on lexical inferencing. These have indicated that readers are likely to adhere to local linguistic clues such as collocations and morphology in deriving the meaning of unknown vocabulary (Haynes, 1993; Huckin & Bloch, 1993; Wesche & Parikbakht, 2010), while listeners use mainly global information (Cai & Lee, 2010). Cai and Lee (2010, p. 18) believe that listeners focus less on local clues because 'words easily fade away, making refocusing and checking impossible'.

AIMS AND RESEARCH QUESTIONS

The first aim of this study is to compare decontextual vocabulary knowledge with the ability to recognise and comprehend vocabulary in listening and reading contexts. This should indicate in how far decontextualised vocabulary tests, which are most commonly used in research and pedagogy, provide an accurate indication of learners' vocabulary knowledge in context. The second aim is to shed light on learners' relative contextual knowledge in the reading and listening mode. As the literature review explained, the contexts of reading and listening have been found to differ in a number of vocabulary-related aspects. However, hardly any research has directly compared vocabulary knowledge in the two modes.

The study aims to answer the following research questions:

- i. What is the relationship between L2 learners' decontextual and contextual knowledge of word meaning?
- ii. Is there a difference between learners' contextual vocabulary knowledge in reading and listening?
- iii. Are decontextual and contextual vocabulary knowledge related to participants' vocabulary size and/or to words' frequency?

METHODOLOGY

PARTICIPANTS

Thirty postgraduate students from a university in the United Kingdom participated in the study. As students are required to reach TOEFL iBT 100 or IELTS 6.0 to be accepted to a Postgraduate course at the university, it can be taken with confidence that all participants were of intermediate to advanced level of English. The mean age was 26.1, and 12 participants were male and 18 female. The pool included learners from 20 different L1s.

TARGET ITEMS

In order to explore how different degrees of decontextual knowledge related to the ability to comprehend word meaning in context, it was necessary to have a range of vocabulary items: some that were likely to be known, and some that were unlikely to be known. As frequency can be expected to be a good predictor of knowledge (Ellis, 2002), items were extracted from a range of frequency levels from the Corpus of Contemporary American English (COCA). In addition, as participants' contextual knowledge would be measured from written and spoken narrative context, a second criterion was that the items needed to be suitable for use in two short narratives. Table 1 provides an overview of the ten target items selected, together with their frequency data retrieved from the COCA.

Table 1. The ten target items ranked according to their frequency in the COCA

Item	Frequency rank	COCA
		Freq. per million
poor	1	116.50
patience	2	13.88
authentic	3	9.87
to weep	4	7.76
candid	5	5.07
to shudder	6	4.28
upbeat	7	4.00
whim	8	2.72
to rove	9	0.43
malign	10	0.21

TEXTS

Two short narratives were written to measure participants' knowledge of word meaning in the reading and listening context (see Appendix). Both texts contained the ten target items. As learners need to know at least 95% of the words in a written or spoken text for general comprehension (Hu & Nation, 2000; van Zeeland & Schmitt, 2012), not more than 5% of the words in each text were target items, while all other words were within the 2,000 frequency band. The Vocabulary Profile on the *Compleat Lexical Tutor* website (Cobb, n.d.) was used to ensure this.

It was important that the twenty narrative contexts in which the target items occurred provided equal opportunities to infer their meaning. To this end, 36 native speakers of English were given a written or spoken version of the two narratives, in which the target items were replaced by blanks and bleeps. They were asked to write down the words that they thought were missing. Contexts which enabled many native speakers to guess the target

items’ meanings correctly were rewritten until all twenty contexts provided minimal chances of correct guessing. To further ensure that participants’ overall comprehension of the two stories would lead to equal inferencing opportunities, the 36 native speakers were asked to rate the degree of difficulty of both narratives on a scale from 1 to 10 (1 = difficult, 10 = easy), and to rank the two narratives in difficulty if they felt their difficulty differed enough to do so. In the rating task, the mean difficulty of Narrative 1 was 8.05 and that of Narrative 2 was 8.03. In the ranking task, most participants (24/36) chose not to rank the stories in difficulty, and of those who did, 5 felt Narrative 1 was easier, and 7 felt Narrative 2 was easier. These results show that the two narratives were of equal difficulty.

MEASUREMENT INSTRUMENTS

1. Decontextual knowledge: Interview

In order to get a good idea of learners’ knowledge of the ten target items, a free recall interview was used. The researcher uttered the target item while showing the written item with its part of speech on a word card. From these spoken and written prompts, participants were asked to recall the meaning of the word. Their responses were scored according to three degrees of knowledge: one point for full knowledge, a half point for partial knowledge, and zero points for no knowledge. Table 2 provides examples from the interview data in which participants demonstrated full, partial and no knowledge of the items *to weep* and *to shudder*. A second researcher was asked to rate the interview responses of a third (10/30) of the participants. Scorings by the two raters corresponded in 91 of the 100 cases, indicating an inter-rater reliability of 91%. All 9 mismatch cases were partial versus full or no knowledge, meaning that there were no cases of absolute disagreement (full versus no knowledge) between the two raters.

Table 2. Examples of knowledge scoring in the free meaning recall interview			
Item	Full knowledge (1 point)	Partial knowledge (0.5 point)	No knowledge (0 points)
<i>to weep</i>	to cry, to shed tears	Is it to cry or something similar? [hesitant]	[no answer given], to clean with a brush
<i>to shudder</i>	to shiver, to shake	To make a movement with your body	[no answer given]

2. Contextual knowledge: reading and listening

In order to test knowledge of word meaning in the reading and listening context, participants were asked to paraphrase ten sentences which contained the target items from both a written and spoken narrative. A sheet with paraphrasing examples was provided so that participants knew what was expected from them. In the reading task, participants were given the narrative on a piece of paper (see Appendix). It was divided into ten paragraphs, and the last sentence

of each was in bold. Participants were given a separate sheet on which to write a paraphrase of each of the ten bold sentences. Where they did not know how to paraphrase a sentence, they were asked to select the option ‘I don’t know how to paraphrase this’. There was no time limit to finishing this task.

A recording was made of both narratives for the listening exercise. Both were made by the same British male speaker to ensure similarity between the two passages. Participants were given another paraphrasing sheet, and listened to the recording of the text that was not used in the reading exercise. The recording paused automatically after each paragraph, and participants were asked to paraphrase the last sentence they had heard. Once they had completed their paraphrase, they could press the *play* button on the computer screen for the recording to continue. This way there was no time pressure, and the paraphrasing task was the same for the written and spoken mode. Responses were scored as correct (1 point) or incorrect (0 points). Again, the responses of 10 participants were scored by the second rater. An inter-rater agreement of 93% was achieved.

Participants also completed the VLT (levels 2,000, 3,000, 5,000 and 10,000) in order to get an indication of their vocabulary size.

PROCEDURE

Tests were completed in a one-to-one session with the researcher in the order below. The session lasted approximately 45 minutes.

- i. Interview;
- ii. 2,000 and 3,000 level of the VLT;
- iii. Paraphrasing from reading and listening (counterbalanced);
- iv. 5,000 and 10,000 level of the VLT.

To ensure that no potential difference between the two narratives interfered with the modality effect, each narrative was used in the reading and listening exercise by half of the participants. To further minimise any bias towards one of the two modalities, the order in which participants carried out the reading and listening task was counterbalanced. This led to the organisation illustrated in Table 3.

Table 3. Sequencing of the reading and listening of the two stories			
Group 1 (N = 8)	Group 2 (N = 7)	Group 3 (N = 8)	Group 4 (N = 7)
Reading: Story 1	Reading: Story 2	Listening: Story 1	Listening: Story 2
↓	↓	↓	↓
Listening: Story 2	Listening: Story 1	Reading: Story 2	Reading: Story 1

RESULTS

RELATIVE KNOWLEDGE OF WORD MEANING ON THE THREE TESTS

Table 4 shows the mean scores (and SDs) achieved on the three tests. As can be seen, participants managed to recall the meaning of almost half of the items on the interview (this includes the 0.5 scores for partial knowledge). Their ability to recall the items’ meaning from reading and listening was similar, with around 5 of the 10 target items recalled correctly.

Table 4. Mean score (SD) on the three tests (Max = 10)

Interview	Reading	Listening
4.6	5.4	4.5
(1.5)	(1.61)	(1.72)

As Shapiro-Wilk tests revealed that the data was not normally distributed, non-parametric tests are used in this analysis. A Friedman test revealed the difference in scores on the three tests to be significant (χ^2 (2, n = 30), 19.918, p < .001). Wilcoxon Signed Rank tests indicated that participants scored highest on reading, with these scores being significantly higher than those on the interview (z = -3.600, p < .001, r = 0.47) and listening (z = - 3.407, p < .01, r = 0.44). There was no significant difference in scores on the interview and listening tests (Figure 1).

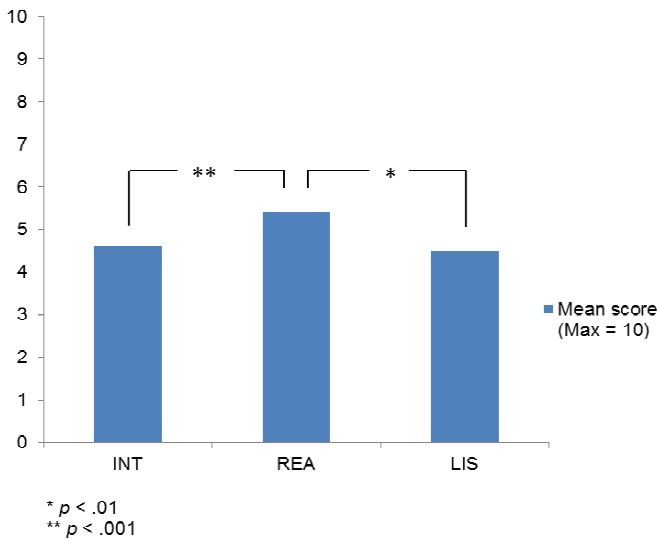


Figure 1. Mean scores on the three tests with significant differences indicated

DECONTEXTUAL AND CONTEXTUAL KNOWLEDGE

Table 5 provides the degree of knowledge demonstrated on the interview together with the numbers of correct scores on the reading and listening tasks. Out of all 300 cases (10 items x 30 participants), 162 (54.0%) were scored correct in reading and 135 (45.0%) in listening, showing the advantage of reading over listening (Table 4). In most cases, participants had either full knowledge in the interview and correct recall in both modes (90, 30.0%) or no knowledge and incorrect recall in both modes (106, 35.3%). Decontextual knowledge as found in the interview, at least in the cases of full or no knowledge, thus relates to the ability to comprehend word meaning in both modes in 65.3% of the cases.

Table 5. Cases of knowledge on the interview and correct/incorrect recall from reading and listening

	REA and LIS correct	REA and LIS incorrect	REA correct, LIS incorrect	LIS correct, REA incorrect	Total
INT full	90	6	19	5	120
INT partial	17	14	5	2	38
INT unknown	16	106	15	5	142
Total	123	126	39	12	300

An interesting follow-up question is how learners' contextual knowledge relates to their degree of decontextual knowledge (full, partial, or no knowledge). Of the 120 cases of full decontextual knowledge (Table 5, Figure 2), participants recalled the meaning from reading in 109 cases (90.8%) and from listening in 95 cases (79.2%). In 90 (75.0%) of the cases they managed to recall it correctly from both modes. Thus, while full decontextual knowledge agreed with both context modes in three-fourths of the cases, it was not uncommon for learners to have difficulty with recalling words from context even if they knew these words. Of the 38 partial knowledge cases (Table 5, Figure 3), participants managed to recall the meaning from reading in 22 (57.9%) and from listening in 19 cases (50.0%), while they managed to recall it from both in 17 (44.7%). This shows that there were many cases in which the narrative context aided meaning recall, especially in the written mode. Context also aided recall of unknown items (Table 5): while 106 of the 142 cases (74.6%) were not recalled correctly in either mode, in 16 cases (11.3%), participants managed to do so in both. Again, reading has a small advantage over listening, suggesting that participants were slightly better able to infer the meaning of unknown vocabulary from reading.

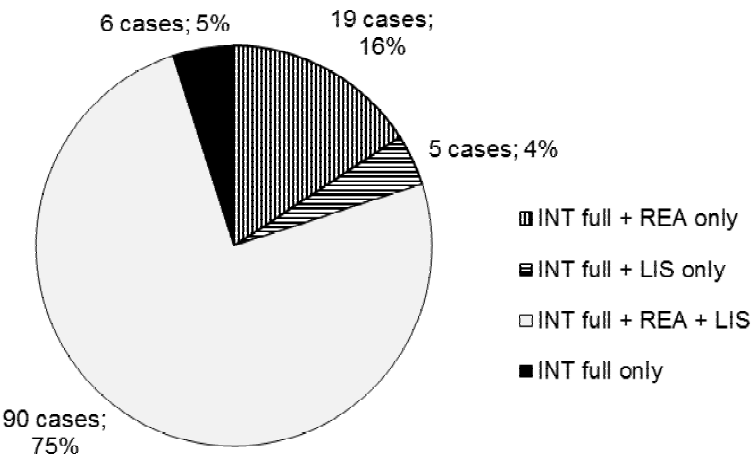


Figure 2. The number of cases and percentages in which participants showed full knowledge of items' meaning in the interview (total = 120) and scored correctly in the reading and listening tasks (based on data from Table 5)

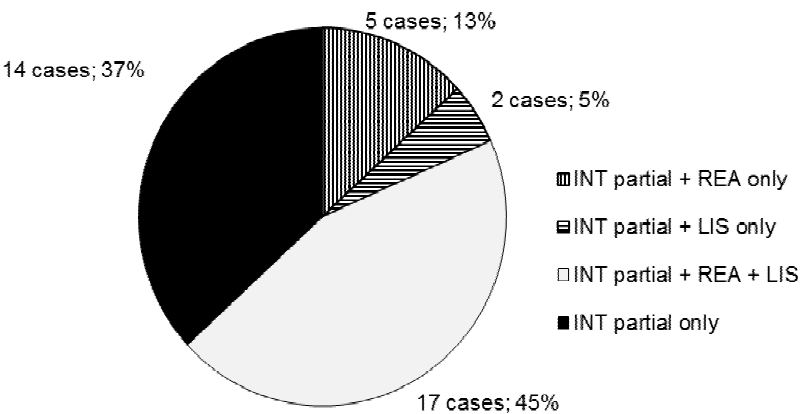


Figure 3. The number of cases and percentages in which participants showed partial knowledge of items' meaning in the interview (total = 38) and scored correctly in the reading and listening tasks (based on data from Table 5)

So can learners' decontextual knowledge be taken as an indication of their contextual knowledge? These results reveal a fair degree of agreement between the two. Where learners had either full or no knowledge of a word, they were also successful or not successful in comprehending it in context in about 75% of the cases. If the partial knowledge cases are included, there is an agreement between decontextual and contextual knowledge in two-thirds (65.3%) of the cases. However, for the purpose of vocabulary assessment, this is clearly not enough. It means that a decontextualised vocabulary test fails to provide an accurate indication of learners' contextual knowledge in about one-third of the cases. Moreover, results suggested that the degree to which a decontextualised test accurately predicts contextual knowledge depends on the context mode. This difference between the reading and listening modes is analysed in more detail in the following section.

CONTEXTUAL KNOWLEDGE IN READING AND LISTENING

Figure 4 presents the relative correct meaning recall from reading and listening. The vast majority of the 300 cases show agreement between the two modes: in 41.0% participants demonstrated knowledge of an item in both, and in 42.0% they failed to do so in both. In substantially fewer cases participants knew an item in only one of the two. This was more often reading (13.0%) than listening (4.0%), which is reminiscent of the small advantage of reading over listening reported in the previous section.¹

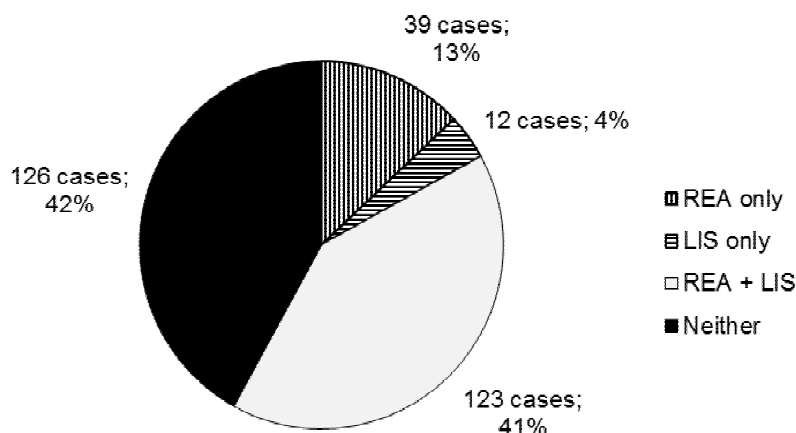


Figure 4. The number of cases and percentages of correct and incorrect scores in the reading and/or listening task (total = 300) (based on data from Table 5)

Why would learners have more difficulty with recalling word meaning from listening than from reading? More information may be provided by an analysis of the cases in which participants recalled the meaning from reading but not from listening. Table 6 provides the number of such cases for each item, together with participants' knowledge of these items in the interview. There were a total of 24 occurrences in which learners had partial or full knowledge of a word, but were only able to recall its meaning correctly from reading. A closer look reveals that a considerable number of these occurrences come from the same participants. Of the 19 cases of full knowledge, 5 were from the same participants (3 by participant 15 and 2 by participant 3). If the occurrences in which an item was partially known are included, making a total of 24, 11 occurrences come from the same participants. This may indicate that these particular learners experienced difficulty with correctly identifying (partially) known lexical items in listening. This explanation is in agreement with the recurrent research finding that lexical segmentation and word recognition are prime obstacles for L2 listeners (e.g. Field, 2008; Goh, 2000). Interestingly, a similar analysis was carried out for the cases where (partially) known items were recalled from listening but not from reading, and no such participant-related pattern was found.

There are also certain item-related factors that may have contributed to the advantage of recall from reading over listening. Two words that were unknown but commonly recalled correctly from reading but not listening were *to rove* and *candid* (Table 6). This suggests that the meaning of these items may have been easier to infer from reading than from listening. A look at the narratives reveals that co-textual information was required to infer the meaning of both items. For example, in the case of *candid*, the narratives refer to behaviour which is explained elsewhere in the text: *But after the police had been so candid in explaining this...* (Appendix, paragraph 9) and *I thought she wasn't being very candid with me*. While readers could search the text to infer the meaning of these words, listeners could not.

Table 6. The number of times the meaning of each item was recalled correctly from reading but not from listening, together with the number of times each item was known fully, known partially or unknown on the interview. The participant number (1-30) is provided in brackets.

Item	Recalled from REA only	Knowledge in INT		
		Full	Partial	Unknown
poor	3	3 (7/24/27)		
patience	8	8 (2/11/12/15/18/19/21/22)		
authentic	1	1 (15)		
to weep	5	3 (3/5/26)		2 (24/25)
candid	8	3 (3/4/15)	3 (19/28/30)	2 (26/27)
to shudder	3			3 (17/18/20)
upbeat	1	1 (30)		
whim	1			1 (29)
to rove	5		1 (21)	4 (8/13/24/30)
malign	4		1 (16)	3 (15/26/28)
Total	39	19	5	15

THE ROLE OF VOCABULARY SIZE AND ITEM FREQUENCY

To explore the effect of vocabulary size, the participant pool was divided into three groups based on VLT scores. A careful look at the VLT results showed that participants' scores on the 10,000 frequency were good indicators of their scores on the three lower frequency levels (i.e. learners with relatively high or low scores on this level also scored high or low on the lower levels). This led to the following distinction: Group 1 (large vocabulary size) scored 20 or more on the 10,000 level, Group 2 (medium vocabulary size) scored 11 to 19, and Group 3 (small vocabulary size) scored 10 or less. The three groups differed clearly in their total VLT scores (Max = 120) and in their mean scores on the four frequency levels (Max = 30): Group 1 had a mean total score of 116.30 (range 111-120), and a mean score of 27 or higher on all four frequency levels, so they demonstrated mastery of vocabulary up to the 10,000 band; Group 2 had a mean total score of 102.17 (range 94-107), and a mean score of 27 or higher on the 2,000, the 3,000 and the 5,000 levels, indicating an average vocabulary size of at least 5,000 word families; and Group 3 achieved a mean total score of 83.38 (range 71-94) and a mean score of 27 or higher on only the 2,000 and 3,000 levels.

Table 7 shows that the mean scores on all three tests are higher for learners with a larger vocabulary size than with a smaller size. Kuskall-Wallis tests found that the difference between scores of the three vocabulary size groups was significant for all three tests: interview (χ^2 (2) 17.747, $p < .001$), reading (χ^2 (2) 14.085, $p < .01$), and listening (χ^2 (2) 17.276, $p < .001$). A series of Mann-Whitney U tests revealed that scores on the three tests decreased significantly ($p < .05$) between the three vocabulary size groups (large > medium >

small size), confirming that learners’ decontextual as well as contextual vocabulary knowledge relates to their vocabulary size.

Table 7. Mean scores (SD) (Max = 10) on the three tests achieved by the three vocabulary size groups

Vocabulary size group	INT	REA	LIS
Group 1 (N = 10): large	6.00 (1.13)	6.80 (1.31)	6.00 (1.41)
Group 2 (N = 12): medium	4.54 (1.03)	5.17 (1.19)	4.50 (1.17)
Group 3 (N = 8): small	3.06 (0.73)	4.00 (1.06)	2.62 (0.52)

It was also analysed whether learners’ relative performance on the three tests related to their vocabulary size. Wilcoxon Signed Rank Tests found that participants scored significantly higher on the reading test than on the interview in both the small size group ($z = -2.214, p < .05, r = 0.55$) and the large size group ($z = -2.217, p < .05, r = 0.49$), but not in the medium size group. In addition, although the mean score on the reading test was higher than on the listening test for all three size groups, this difference was significant only in the small size group ($z = -2.456, p < .05, r = 0.61$). The advantage of reading over listening is thus mainly found among learners with a relatively smaller vocabulary size. Why would this be the case? One explanation may be that learners with a smaller size are generally of lower proficiency, and that they are likely to encounter more problems in the processing of spoken language than higher proficiency learners (Vandergrift, 2011). Another explanation is that they lack the resources for more global comprehension strategies such as lexical inferencing, because they tend to focus primarily on linguistic processing (Buck, 2001). This is a plausible explanation. Vocabulary knowledge has been found to affect learners’ inferencing ability in reading (Qian, 2005; Wesche & Paribakht, 2010), and it can be expected to be the same in listening. In this line of thinking, the inferencing success of learners with a smaller vocabulary size may be affected considerably more strongly in the listening than reading mode because listening is particularly problematic for this learner group. Clearly, more research is needed to explain if the difference between the two modes in this learner group is due to processing differences, to inferencing differences, or to an interaction between these and/or other factors.

Next, as frequency is a prime predictor of decontextual vocabulary knowledge and the basis of most standardised tests (e.g. the VLT), it was considered whether learners’ contextual knowledge also related to this factor. Table 8 and Figure 5 show the ten items ordered by their frequency in the COCA together with the mean test score on each item. This shows a clear relationship between items’ frequency and test scores. Spearman rank order correlations revealed strong, positive relationships between frequency data and participants’ scores on the three tests: interview ($\rho = .923, p = .000$), reading ($\rho = .912, p = .000$) and listening ($\rho = .896, p = .000$). It can be taken from this that word frequency is a strong predictor not only of learners’ decontextual vocabulary knowledge as has been shown before (e.g. Schmitt, Schmitt, and Clapham, 2001), but also of their contextual knowledge in both reading and listening.

Table 8. The mean score (Max = 1.0) on all ten items, organised by frequency rank in the COCA

freq. rank	item	Mean test score (Max = 1.0)		
		INT	REA	LIS
1	poor	1.00	0.93	0.90
2	patience	1.00	0.80	0.57
3	authentic	0.82	0.83	0.83
4	to weep	0.65	0.90	0.73
5	candid	0.30	0.47	0.20
6	to shudder	0.23	0.40	0.40
7	upbeat	0.23	0.30	0.40
8	whim	0.15	0.23	0.20
9	to rove	0.02	0.27	0.10
10	malign	0.23	0.27	0.17

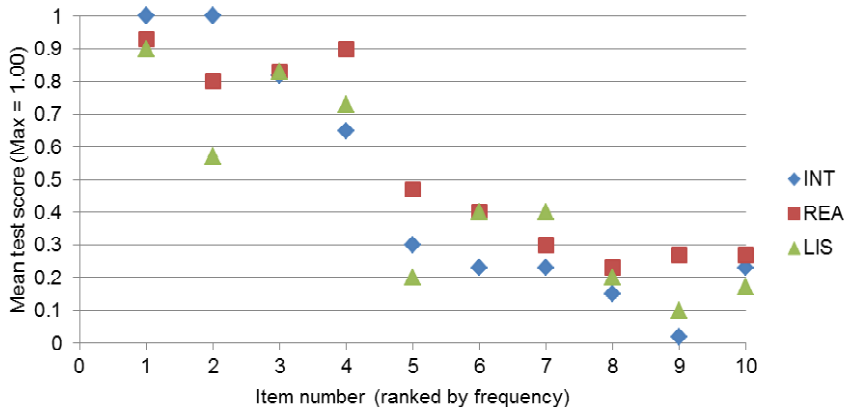


Figure 5. The mean score for each item on the three meaning recall tests, organised by their frequency in the COCA (Item 1 is highest in frequency [poor], Item 10 is lowest in frequency [malign]: see Table 8)

DISCUSSION AND SUGGESTIONS FOR FUTURE RESEARCH

This study explored learners’ knowledge of word meaning in and out of context. The first research question concerned the relationship between decontextual word knowledge (as found in a free recall interview) and knowledge in context. There was approximately 65% overlap between learners’ interview scores and their performance in the reading and listening tasks, showing fair agreement, but also indicating a considerable gap between learners’ decontextual and contextual knowledge. It was not uncommon for learners to have decontextual knowledge

of a word's meaning without having the ability to recall it from context, and vice versa. This finding supports the argument put forward by some researchers (e.g. Read & Chapelle, 2001) that vocabulary knowledge should be tested in context. As argued by Read and Chapelle (2001), discrete decontextualised tests (such as the interview used here, or the more commonly used VST and VLT) are very useful for certain purposes, but the problem arises as soon as such tests are employed without considering what they really tell us about learners' knowledge in communicative contexts of use. There is a serious need for the development of contextualised vocabulary tests that can be used for research and pedagogical purposes. While the paraphrasing task used in this study has proven useful in measuring word knowledge in context, it is a time-consuming task and not appropriate for large-scale assessment.

Another interesting finding is that both full and partial word knowledge was found to be used slightly more often in the reading than the listening mode, while unknown vocabulary was also inferred correctly more often from reading than from listening. One may take these results as confirming the findings on vocabulary size in the written and spoken mode (Milton & Hopkins, 2006), i.e. reflecting greater orthographic than phonological vocabulary knowledge. However, as many correct paraphrasing responses in the reading mode were found to be due to inferencing (i.e. they had been unknown on the interview), the results seem to indicate that the written mode provides better chances to infer the meaning of vocabulary than the listening mode does. This would mean that contextualisation of vocabulary is more useful in written than spoken context. However, these are assumptions, and more research is needed into learners' relative ability to recognise words as well as to infer the meaning of unknown words in the two modes.

Participants' decontextual and contextual vocabulary knowledge was furthermore related to their vocabulary size and to items' overall frequency in the English language. Interestingly, with regard to vocabulary size, the advantage of recall from reading over listening was significant only for the smallest-size group. A possible explanation is that those learners are of lower proficiency and that they have more difficulty with recognising word forms. Moreover, due to these linguistic processing efforts they may fail to use the context to infer the meaning of unknown words (Buck, 2001; Vandergrift, 2011). Items' frequency was found to correlate strongly with scores on all three tests. This indicates that frequency is not only a strong indicator of learners' ability to recall word meaning, but also to recognise and comprehend it in listening and reading contexts. Many standardised tests are based on frequency (e.g. the VST and VLT), and while the results show that such decontextualised tests do not always predict contextual success, they also indicate that the frequency effect does appear to be transferred to context. This finding furthermore implies that frequency constitutes an appropriate basis for the development of tests of contextualised vocabulary knowledge.

It should be emphasised that the decontextualised test used in this study, a free recall interview, gives a highly accurate indication of test-takers' knowledge. It is likely that a

larger gap exists between knowledge found by other decontextualised test formats (e.g. translation and MC tests) and learners' contextual knowledge. This issue should be investigated in future research, particularly because research and pedagogy use mainly such test formats to measure learners' vocabulary knowledge.

REFERENCES

- Broersma, M. & Cutler, A. (2008). Phantom word activation in L2. *System*, 36, 22–34.
- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Cai, W. & Lee, B. P. H. (2010). Investigating the effect of contextual clues on the processing of unfamiliar words in second language listening comprehension. *Australian Review of Applied Linguistics*, 33, 1–28.
- Cobb, T. (n.d.). *Web Vocabprofile*. Retrieved April 3, 2012, from <http://www.lex tutor.ca/vp/>, an adaptation of Heatley and Nation's (1994) Range.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188.
- Field, J. (2008). Revising segmentation hypotheses in first and second language listening. *System*, 36, 35–51.
- Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28, 55–75.
- Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46–66). Norwood: Ablex.
- Heatley, A. & Nation, P. (1994). *Range*. Victoria University of Wellington, NZ. [Computer program, available at <http://www.vuw.ac.nz/lals/>].
- Hirsh, D. (2010). Researching vocabulary. In B. Paltridge & A. Phakiti, (Eds.), *Continuum companion to research methods in Applied Linguistics* (pp. 222–239). London: Continuum International Publishing Group.
- Huckin, T. & Bloch, J. (1993). Strategies for inferring word meaning in context: a cognitive model. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153–180). Norwood: Ablex.
- Hu, H.-C. M. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13, 403–430.
- Laufer, B. & Nation, P. (1999). A vocabulary–size test of controlled productive vocabulary. *Language Testing*, 16, 33–51.
- Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, 75, 196–204.
- Meara, P. (1992). *EFL vocabulary tests*. Swansea: Center for Applied Language Studies, University of Wales Swansea.

- Mecartty, F. H. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning*, 11, 323–348.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Milton, J. & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review*, 63, 127–147.
- Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12–24.
- Nation, P. (1990). *Teaching and learning vocabulary*. Boston: Heinle and Heinle.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63, 59–81.
- Nation, P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton & J. Treffers-Daller, (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 35–44). Cambridge: Cambridge University Press.
- Nation, P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle.
- Nation, P. & Gu, P. Y. (2007). *Focus on vocabulary*. Sydney: National Centre for English Language Teaching and Research.
- Park, G.-P. (2004). Comparison of L2 listening and reading comprehension by University students learning English in Korea. *Foreign Language Annals*, 37, 448–458.
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, 56, 283–307.
- Qian, D. D. (2005). Demystifying lexical inferencing: The role of aspects of vocabulary knowledge. *TESL Canada Journal*, 22, 34–54.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer, (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 209–228). Amsterdam: John Benjamins.
- Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, 7, 105–125.
- Read, J. & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18, 1–32.
- Reves, T. & Levine, A. (1988). The FL receptive skills: Same or different? *System*, 16, 327–336.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55–88.
- Schoonen, R. (2011). How language ability is assessed. In E. Hinkel, (Ed.), *Handbook of research in second language teaching and learning* (Vol. II) (pp. 701–716). London: Routledge.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36, 139–152.

- Vandergrift, L. (2011). Second language listening: Presage, process, product and pedagogy. In E. Hinkel, (Ed.), *Handbook of research in second language teaching and learning* (Vol. II) (pp. 455-471). London: Routledge.
- van Zeeland, H. & Schmitt, N. (2012). Lexical coverage and L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*. Advance Access published December 18, 2012, doi: 10.1093/applin/ams074.
- Weber, A. & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1–25.
- Wesche, M. B. & Paribakht, T. S. (2010). *Lexical inferencing in a first and second language: Cross-linguistic dimensions*. Clevedon: Multilingual Matters.

APPENDIX

One of the two narratives used in the study. Participants read the story in this format in the reading task.

Something strange happened to me a few weeks ago. I was walking down the street when a man stopped me. I had never seen him before. **He looked very poor and sad.**

He was wearing dirty clothes, had messy hair, and very bad teeth. He asked me for money. I didn't have any money on me, so I said 'I'm sorry, I can't give you any money'. But he didn't believe me. He grabbed my arm and said: 'Give me your money, now'. I saw that he was holding a knife. **I started shuddering.**

I thought of what I could do. What could I give him? What if he would really hurt me? I looked around for help, but there was nobody on the street. The man looked at me, still holding his knife. **I could see that he had no patience.**

I suddenly had an idea: I could give him my golden watch, which was worth a lot of money. It was a very old watch, which my father had once given me. I really didn't want to lose it, but I knew I had no choice. **So I showed it to him and I told him it was authentic.**

He looked at it, took it from me and put it in his pocket, together with his knife. **Without saying anything he ran away, very upbeat.**

Of course I called the police immediately. I told them what had happened and gave them a description of the man. But after doing that, I didn't know what to do. I felt horrible. This watch meant a lot to me, and this man and his knife had really scared me! **I just kept roving all afternoon.**

I finally went home, and I received a phone call from the police. They said my watch had been returned. I couldn't believe it! They also said: 'We understand that you are upset by what happened today. We know the man who took your watch today. **His actions can be quite malign and strange.**

But you should know he has good intentions. At times he can get very confused, and he doesn't know what he's doing. He came to the police station to return your watch this afternoon.' I was incredibly happy. The man realised he had been wrong. **It had been a whim to take my watch.**

I must say that I was very pleased that the police explained everything to me. It made me lose some of my fear. If they hadn't told me about this man's condition, I would have thought of him as a horrible man. **But after the police had been so candid in explaining this, I understood the situation much better.**

So I went to the police station to pick up my watch. As I walked into the building, I saw the man. **He was weeping near the entrance and he didn't seem to recognise me.**

ENDNOTES

- 1 It is worth mentioning that participants did not fill in fewer answers in the listening than reading task: of the 300 test items in each, 29 and 25 were left blank in the reading and listening data.