Vol. 10 (2019) University of Reading ISSN 2040-3461
LANGUAGE STUDIES WORKING PAPERS
Editors: R. Leonard and B. Garrido

Using the YARC Secondary with adult Arabic L1 learners of English: an exploration of L2 learners' reading comprehension and their ability to learn new words

Jeanine Treffers-Daller, Reham Alkhudiry and Jacqueline Laws

The York Assessment of Reading for Comprehension Secondary Test (YARC) was originally designed by Stothard, Snowling, Clarke, Barmby and Hulme (2010) for secondary school students in the UK between the ages of 11 and 16 (both monolinguals and learners of English as an Additional language). Since tests that were developed for a particular group are not necessarily suitable for other groups, the aim of this research was to find out to what extent the YARC Secondary reading comprehension test could be used to measure reading comprehension among Arabic L1 adult L2 learners of English. As vocabulary knowledge has often been found to be a key predictor of reading, we administered two offline vocabulary tests (measuring depth and size of word knowledge) in addition to the YARC Secondary to 30 native English-speaking and 28 Saudi Arabian learners of English attending university courses in their home country. The findings showed that YARC scores correlated significantly with scores on the vocabulary tests, and that scores on the Single Word Reading Test (SWRT) for adult L1 Arabic learners fell within the range for children in the UK aged 11 years and three months. Evidence of the suitability of the YARC texts was also obtained through an analysis of students' ability to retain meaning and use (in a sentence) of non-words encountered in the YARC texts. We conclude that the YARC Secondary can be used with some groups of adult L2 learners of English, provided adaptations are made to the reading comprehension texts to ensure the vocabulary levels are appropriate for the target group.

1. Introduction

Reading skills are a crucially important resource for academic success, both in children and adults (Harrington & Roche, 2014). In a recent study among international students in Higher Education (HE) in the UK, Trenkic and Warmington (2017) show that non-native speakers achieve lower grades for their university course because of their lower language and literacy levels in English, despite having similar non-verbal cognitive skills. Other researchers have also found that L2 language proficiency is a strong predictor of L2 reading comprehension, even stronger than L1 reading ability (Tsai, Ernst & Talley, 2010; Jiang, 2011; Jeon &Yamashita, 2014). This serves to illustrate the importance of improving language proficiency, not only at university level but also at lower levels of education, where learners with English as an Additional Language (EAL) have often been found to achieve lower learning outcomes (Hutchinson, 2018; Strand & Hessel, 2018). As English-medium instruction is offered in many countries, there are also many L2 learners of English who learn through the medium of English outside the UK. Many of these learners do not have the required language and literacy skills to benefit from instruction delivered through the medium of English (see Dearden, 2014).

One key variable that affects students' reading ability is their vocabulary knowledge (Qian, 2002; Zhang & Yang, 2016). By comparison with native speakers, many L2 learners have smaller vocabularies (Nation & Waring, 1997). According to Laufer (as cited in Schmitt, Jiang & Grabe, 2011) the vocabulary size of high school and university-educated second language learners of English often ranges from 1,000-4,000-word families. As one needs to have a vocabulary of around 8,000-9,000 words to read a novel or a newspaper article (Nation, 2006), it is clear that reading can be very challenging for many L2 learners, particularly because

readers need to know 98% of the words in a text if they are to achieve good comprehension, that is achieve 70% correct answers on a comprehension test (Schmitt, Jiang & Grabe, 2011).

To be able to give advice regarding the kinds of interventions that are needed to help L2 learners of English improve their reading skills, tests which measure this are urgently needed. Unfortunately, there are few – if any – tests that are suitable for adults (Ready, Chaudry, Schatz & Strazzullo, 2012) and even fewer for adult L2 learners. The current study therefore sets out to investigate whether the YARC Secondary can be used to measure reading in adult L2 learners.

We are of course fully aware of the fact that a test which was developed for secondary school students cannot simply be assumed to be suitable for a different target group. Schmitt, Nation and Kremmel (2019) argue that no language test can be valid for all learners, for different purposes, and in all contexts, and the same can be said about tests of reading. The authors also argue that developing new tests is extremely challenging, as rigorous procedures need to be followed to develop and validate any test. An alternative option is to investigate to what extent existing tests can be adapted for new groups. In a recent study, Treffers-Daller and Huang (in press) administered the York Assessment of Reading for Comprehension Secondary Test (henceforth the YARC), designed by Stothard, Snowling, Clarke, Barmby and Hulme (2010) to a group of Chinese L1 adult learners of English in a concurrent validation study of the Chinese Test for English Majors, band IV. They found that the reading comprehension tasks were challenging, although the reading fluency and the word recognition tasks were of acceptable levels of difficulty. They concluded that the YARC could indeed be used with this group of learners, but also note that this does not mean it is suitable for other groups of adult L2 learners. The current study aims to further explore the suitability of the YARC for adult L2 learners in a study among Arabic L1 learners. As this group has been reported to have small vocabularies (Alsaif, 2011; Larson, 2017), they could be expected to experience problems similar to those of the Chinese L1 learners. For this reason, we first ascertain whether learners' vocabulary knowledge falls within the vocabulary levels used in the YARC texts. Second, as learning new words crucially depends on text comprehension (Grabe, 2009), we decided to also focus on students' ability to learn the form and meaning of non-words which replaced real words in the YARC reading comprehension texts.

Before explaining the design of our study, in Section 2 we will briefly define the key constructs under investigation (reading comprehension and vocabulary knowledge), the ways in which these can be measured. In section 3 we will outline the research questions and methods for the current study. In section 4 the methodology is described. Sections 5 and 6 offer an overview of the results and the discussion. We conclude with a summary of the main findings and an outlook towards the future.

2. Literature Review

2.1 Defining the Construct of Reading Comprehension and its measurement

Reading comprehension is defined as an interactive activity between the reader's knowledge and the characteristics of the written input such as genre and structure (Schelling, Aarnoutse & Leeuwe, 2006). For the purposes of the current study, we will assume that reading comprehension crucially depends on two variables: decoding and linguistic comprehension skills, as proposed by the Simple View of Reading (SVR) (Gough & Tunmer, 1986; Hoover & Gough, 1990). *Decoding* means readers' ability to recognise words, that is to make a link between the printed word and the appropriate entry in the mental lexicon. After decoding the words, readers must draw upon their oral language comprehension skills to understand what is written. This process, which is referred to as *linguistic comprehension*, involves a number of

language skills, such as the ability to derive lexical semantic information, and sentence and discourse level interpretations. Studies based on this model propose that decoding and language comprehension account for between 40% and 70% of the variance in reading comprehension (Cutting & Scarborough, 2006; Joshi & Aaron, 2000).

While there are a number of measures of reading that can be used with adults, there are only a few tests that are suitable for adult L2 learners, particularly those with lower levels of vocabulary. The National Adult Reading Test (NART, Nelson, 1984) contains a large number of low frequency words and is therefore unlikely to be suitable for L2 learners (see also Treffers-Daller & Huang, in press). The same is true for the York Adult Assessment Battery-Revised (YAA-R: Warmington, Stothard & Snowling, 2013), which was developed for university students and measures reading accuracy, speed and comprehension. Another possibility would have been to use one of the Cambridge proficiency tests, such as the Cambridge Preliminary English Test (PET), which is designed for B1 level of CEFR. According to Cambridge English Language Assessment (2014), the PET is particularly suitable for language learners who need to use English in everyday communicative situations when travelling in an English-speaking country. This, however, is not the situation of most classroom learners of English. In addition, the PET does not provide in-depth information regarding decoding and comprehension, which are key to the process of reading.

An alternative option is to use the YARC, which is based on the SVR model (Gough & Tunmer, 1986), and was developed for 11-16-year-olds. While the majority of the students in the standardization sample were native speakers of English, according to the manual the YARC can be used with EAL students. Stothard et al. (2010, p. 114) report that EAL students participated in a test to ensure that the sample was representative of UK school students. Their standard scores on the SWRT, used to select the appropriate level of difficulty of the YARC Passage Reading test for the participants, were on average 3 points lower than those of native speakers, while the reading comprehension standard scores were between 6 and 10 points lower. Reading rate scores were between 3 and 7 points lower and reading fluency scores on average 6 points lower. Although the YARC can therefore be used with bilingual children from the same age group, it remains to be seen whether it is also suitable for adult L2 learners. Treffers-Daller and Huang (in press) provide some evidence that the YARC can indeed be used with some adult L2 learners, although students' levels of achievement were relatively low by comparison with those of secondary school students in the UK. To the best of our knowledge, there are no further studies in which the YARC is used with adult L2 learners.

2.2 Vocabulary knowledge (depth and size) and their measurement

Word knowledge involves more than knowing about the link between form and meaning of a word. Nation (2001) provides the most widely used comprehensive model for defining word knowledge and suggests that knowing a word also entails knowing about the ways in which words are used. It is not only vocabulary size that matters (how many form-meaning links one knows) but also vocabulary depth, that is how well words are known (Milton, 2009; Qian, 1998). Both vocabulary size and vocabulary depth have been argued to be important for reading comprehension. Hsueh-chao & Nation (2000), for example, point to the importance of vocabulary size in order to achieve success in reading comprehension. Conversely, Rashidi and Khosravi (2010) found that depth of vocabulary knowledge of 38 Iranian university students explained more unique variance (69%) in reading comprehension performance than vocabulary size (55%). In another study, Zhang and Yang (2016) revealed that vocabulary depth of L2 learners of Chinese explained unique variance in reading comprehension over and above vocabulary size.

It can be difficult to interpret the results from different studies into the contribution of vocabulary to reading, because the role of vocabulary depends on the ways in which it is measured, and different studies use different measuring tools. One of the most widely used tests of written vocabulary knowledge is Nation and Beglar's (2007) Vocabulary Size Test (VST), that can be used to assess both L1 and L2 learners' written receptive vocabulary size. Since it has a multiple-choice format, learners are provided with the orienting effect of the context, as well as the cuing effect of the options (Nation, 2012). Further details about the validity of the test can be found in Beglar (2010).

One option to measure vocabulary depth is to use Read's (1993) Word Associates Test (WAT), which assesses test takers' knowledge of collocations, synonyms and antonyms. Unfortunately, it has been found to be very challenging even for advanced foreign language learners at university level (Bogaards, 2000). Another possibility is to measure vocabulary depth with Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS), which consists of a five-point scale ranging from complete unfamiliarity through word recognition, to the ability to produce the word in a correct sentence with grammatical and semantic accuracy. Although this test seems insensitive to other aspects of word knowledge, such as the variety of word meanings (Wolter, 2005), it nevertheless taps into different aspects of word knowledge and is not too complex even for learners with relatively low proficiency levels.

The VKS has also been used by Treffers-Daller and Huang (in press) to measure vocabulary depth in their study among Chinese L1 learners of English. They found that reading comprehension and reading fluency (accuracy) as measured with the YARC correlated significantly with a measure of vocabulary depth, and that vocabulary depth and fluency together explained around 24% of the variance in reading comprehension. Because it has successfully been used in a range of studies and the vocabulary that is to be included in the test can easily be adjusted to individual studies, this is likely to be a good candidate for measuring vocabulary depth.

3. Overview of the present study

The current study builds on Treffers-Daller and Huang (in press) and aims to evaluate the suitability of the YARC for adult L1 Arabic learners of English. As alternative tests of reading for adults are too complex for the current learners, it is not possible for us to carry out a concurrent validity study by comparing the results from the YARC with data collected from another instrument which purportedly measures the same construct. Because most researchers agree that vocabulary is a good, if not the best predictor of reading (Laufer & Ravenhorst-Kalovski, 2010), we propose to study to what extent L2 learners' YARC scores correlate with scores on well-known tests of vocabulary size and vocabulary depth. In addition, we are interested in finding out to what extent native speakers and Arabic L1 learners of English are able to learn new words encountered in the YARC texts. We assume that only learners who have good comprehension would be able to learn these new words from the texts. This study, thus, sought to answer the following three research questions:

1. What are the vocabulary levels of Arabic L1 adult learners of English who are studying for a degree in the KSA?
2. To what extent are levels of text comprehension among native and Arabic L1 learners of English students sufficient to learn the meaning of unknown words in the YARC texts?
3. To what extent do non-native students' vocabulary depth scores correlate with scores on the YARC (reading comprehension/ reading summarisation questions)? Do students' scores on reading comprehension/ reading summarisation questions explain

variance in students' ability to make lexical gains from reading as measured with the VKS?

4. Method

Fifty-eight adult participants took part in the study: 28 Arabic L1 learners of English from Qassim university in the Kingdom of Saudi Arabia (mean age = 19.2), and 30 native English speakers from undergraduate classes at Reading university in the UK (mean age = 23.9). They were all enrolled on the Preparatory Year Program, (see *Qassim University* website), in the English Language and Translation Department. Originally, 30 L1 Arabic learners of English participated in the study, but two of them were removed because their accuracy scores on the VST were lower than 40%. As Jiang (2012) suggested, learners with high error rates (20% or above) should be removed from the data set because these scores are likely to skew the results.

The York Assessment of Reading for Comprehension (YARC) Secondary Test (Stothard et al., 2010) was employed to test the participants' reading comprehension performance. The YARC consists of two sections: decoding and reading comprehension. The Single Word Reading Test (SWRT) was used to select the appropriate level of difficulty of the YARC Passage Reading test for the participants. It consists of 70 words and assesses learners' ability to decode words. As shown in the test manual (Stothard et al., 2010), the result of reliability scores for the different levels of reading comprehension lies between 0.85 and 0.90, i.e., it ranges from good to excellent.

In the present project, Level 1 of the YARC Passage Reading test was selected, based on L2 learners' raw scores (ranging from 43-51) on the SWRT. The procedure we followed was therefore different from the standard procedure for the YARC, which specifies that reading passages are selected on an individual basis, depending on the student's scores on the SWRT. This was not possible in our case because we were interested in learners' ability to learn the meaning of non-words in the text, as is explained in the next paragraph. For this reason, all learners had to read the same text. As part of the process of selecting the comprehension texts, the topic of YARC passages was carefully considered to ascertain whether they were culturally appropriate for learners from Saudi Arabia. This was the case for the first non-fiction passage which discussed a bird called the "Honey Guide," and the second one was a short fictional story about a "River Girl". The level of difficulty of the reading passages was also investigated using *Vocabprofile* (BNC-20 version), available on the *Compleat Lexical Tutor* website (https://www.lextutor.ca/vp/comp/), which provides information about the frequency layers of the words in texts. This analysis revealed that 4% of text A and 3% of words in text B belonged to frequency layers lower than 4k. To ensure that reading the text would be challenging but not too complicated for the learners, any words from frequency layers below 4k were replaced with higher frequency items. The aim of this replacement was to create texts for which 98% of the words would be known and the only unknown words in the texts would be the target non-words. The target coverage was almost achieved: in total 2.5% of the text (12 words for each passage) was composed of target non-words, and the remaining 97.5% were assumed to be known.

In Passage A: *cenedies,* a target non-word, replaced *humans,* a high-frequency (1k) word, which occurred eight times in the text. Another target non-word, *toroko,* replaced a low-frequency word (4k), namely *sting,* which occurred four times in the text. In Passage B, there were also two target non-words: *sataca,* a non-word, replaced *trip,* a high-frequency word (1k), which appeared in the text four times, and another non-word, *pocoko,* replaced a low-frequency word (3k), namely *suburb,* which occurred eight times in the text.

Four target non-words were carefully created to comply with phonotactic rules in Arabic to suit Arabic learners of English, on the basis of the CVCVCV pattern, as in the Arabic word

kataba “to write”, because consonant clusters are not allowed in syllabic-initial position in Modern Standard Arabic (Hamdi, Ghazali, & Barkat-Defradas, 2005). In order to eliminate any potential matching of these non-words to either Arabic or English words, they were tested by educated English speakers and Arabic speakers in a pilot study.

Four reading comprehension questions were selected from the thirteen YARC comprehension questions for each text, on the basis of their relevance to the key ideas in the passage, and the target non-words. The summarisation question required participants to summarise the content of the passage they had just read, as specified in the YARC test manual (Stothard et al., 2010). The manual offers detailed information with examples regarding acceptable and unacceptable answers. We strictly followed the instructions in the manual for scoring the summarization and comprehension questions.

The Vocabulary Size Test (VST) (Nation & Beglar, 2007) was used to examine the participants' written receptive vocabulary size. This multiple-choice test presents each target word within a short non-defining sentence. The complete set of words belong to 14 different frequency bands, 10 items from each frequency level, totalling 140 items. Beglar (2010) showed that even when comparing learners with different proficiency levels, comparing male participants with females, and comparing the 70-item version of the test with the 140-item version, the VST performs consistently and reliably (reported as > 0.96). Only the first eight frequency layers of the VST were used, as administering words from lower frequency levels was deemed potentially demotivating given the expected low vocabulary levels. Items from the first eight layers include: *saw* (1k), *maintain* (2k), *soldier* (3k), *compound* (4k), *deficit* (5k), *devious* (6k), *olives* (7k) and *erratic* (8k). The reliability was high at Cronbach’s alpha = .96.

In addition to the VST, the Vocabulary Knowledge Scale (VKS) (Brown, 2008), modified from Wesche and Paribakht (1996), was used to assess learning and retention of the meaning and use of the four target non-words. The test-retest procedure revealed reliability measures of above 0.80. Paribakht and Wesche (1997) describe the VKS as a practical measurement that can be used with any words and is useful for research focused on recognition and use of words. The VKS comprises a four-point scale to capture specific stages in the initial development of core knowledge. The scale ranges from the unknown word category, which represents total unfamiliarity, and partial knowledge (form) to the known word category which includes receptive word knowledge (meaning) and productive word knowledge (use). The participants took the VKS for all four target non-words from Passages A and B, as well as for six real words (*beauty, tissue, weather, bicycle, minute, candle*) which did not occur in the texts. All these real words were assumed to be known as they belonged to the 4k band or to higher frequency bands. Participants were awarded two points in the known word category: one point for receptive word knowledge if a synonym or translation of the target word was accurate and one point for productive word knowledge (meaning and use recall) when the use of a target word was both grammatically and semantically accurate. Spelling mistakes were ignored as the test was not designed to measure spelling.

Figure 1 shows the order of tasks carried out by Arabic L1 learners of English and native speakers of English.

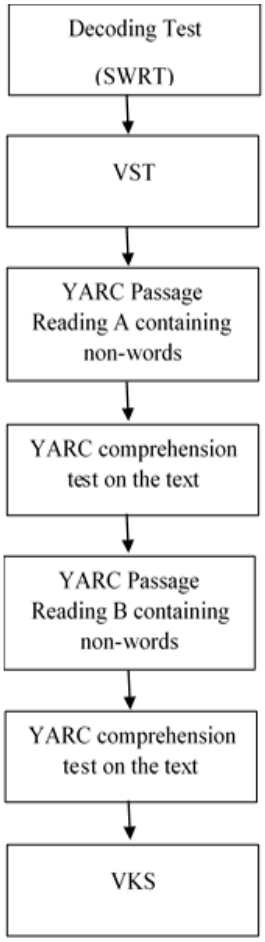

Figure 1. Sequence of tasks carried out by the L1 and L2 groups

Ethical approval to carry out the study was obtained from the Ethics Committee in the Department of English Language and Applied Linguistics prior to data collection.

5. Results

We first present the students' scores on the VST (RQ1). Table 1 shows that the L2 group obtained a score of 38.64 on average on the VST, which means that students knew approximately 3,800 words receptively. Unsurprisingly, the L1 group obtained higher mean scores on the VST than the L2 learners. The difference in accuracy between group means was significant according to the *t*-test (t =15.57, df =56, $p < 0.001$; Eta Squared = 0.812).

	L1			**L2**		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
VST (Max = 80)	73.20 (91.5%)	3.09	66-78	38.64 (48.3%)	11.74	20-64

Table 1: Mean accuracy scores, standard deviations and range for the VST

The group means and standard deviations per frequency band are presented in Figure 2. This reveals that the L2 learners achieved the highest score (8.18 points) for the first layer (1,000) and the lowest scores for the last four bands (5,000 to 8,000), for which their performance is below 4.00, i.e., close to chance level. As shown in the repeated measures mixed model (ANOVA), the main effect of VST levels was significant ($F(7, 392) = 46.37$, $p < 0.001$; Eta Squared = 0.45). The main effect of group was also significant ($F(1, 56) = 242.42$, $p < 0.001$;

Eta Squared = 0.81). In addition, there was a significant interaction between VST level and group ($F(7, 392) = 18.53$, $p < 0.001$; Eta Squared = 0.25). This means that the scores on the different VST levels dropped faster for the L2 than the L1 group. Thus, from the results obtained, it seems that L2 participants' vocabulary size was commensurate with the level of difficulty of the YARC reading passages, since, as mentioned in section 4, all words in both passages belonged to the 4k band or to higher frequency bands.

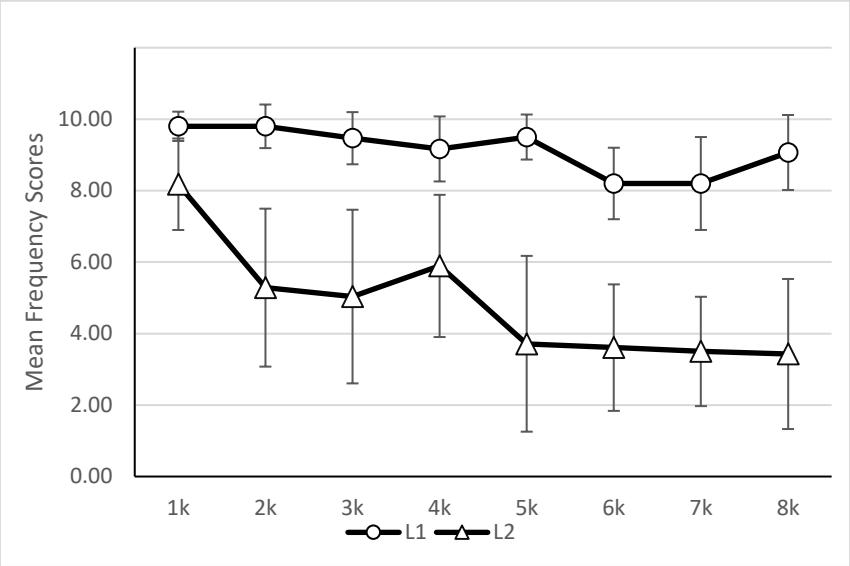

Figure 2. Mean accuracy scores and standard deviations on Nation and Beglar's Vocabulary Size Test (Levels 1-8)

Table 2 presents the overall VKS results for the real-word and target non-word categories. The mean accuracy scores of the VKS for both groups nearly reached the maximum of 12 points in the real-word category; however, both groups obtained low scores in the VKS target non-word category, as mean scores were below 1 for retention of meaning and ability to use the target non-word in a sentence. As can be seen in Table 2, L2 learners had learned the meaning of the target non-words and were able to use them in a sentence (productive level) 4.88% of the time. This indicates that at the productive level, approximately one in twenty of the unknown words was learned. Thus, the rates of learning and retention seemed quite low.

	L1		**L2**		**Sig/Diff**
	Mean	**SD**	**Mean**	**SD**	
VKS target non-words (Max = 4x2=8)	0.73 (9.17%)	1.60	0.39 (4.88%)	1.10	NS
VKS (receptive word knowledge-meaning- of the target non- words) (Max 1x4=4)	0.47 (11.75%)	0.86	0.21 (5.25%)	0.57	NS
VKS (using the target non-words in sentences) (Max 1x4=4)	0.30 (7.5%)	0.80	0.18 (4.5%)	0.55	NS
VKS real words (Max 6x2=12)	11.97 (99.17%)	0.18	11.75 (97.92%)	0.65	NS

Table 2: Mean accuracy scores and standard deviations for the VKS

With respect to RQ2, it is important to note firstly that the mean accuracy score obtained in the SWRT by the L2 participants was 47.21 (67.44%), as shown in Table 3. L1 participants did not take this part of the test as it was assumed they knew all these words. Stothard et al. (2010) report that the SWRT scores can be used to provide a reading age score for children whose first

language is English. Based on L2 participants' SWRT scores, adult L1 Arabic students have an average reading age which is equivalent to a native speaker age of 11 years and three months. Considering that the YARC is suitable for students between the ages of 11 and 16, this means that the performance of the L1 Arabic learners of English was within the range of scores expected for this test.

	L1(N=30)		L2 (N=28)		SWRT	Sig/Diff
	Mean	**SD**	**Mean**	**SD**	**(Year)**	
SWRT (Max = 70)	_____	____	47.21 (67.44%)	2.97	11.03	___
RCQ (Max = 8)	5.43 (67.88%)	1.41	6.32 (79%)	1.02	____	0.01
RSQ (Max = 17)	11.90 (70%)	3.07	11.57 (68.06%)	2.77	____	NS

Table 3: Mean accuracy scores and standard deviations on the YARC Secondary Test

Accuracy scores for the eight reading comprehension questions (RCQ) are presented in Table 3. For the analysis of these results non-parametric tests were used because the data were not normally distributed. The mean accuracy score for L1 participants was significantly lower than that achieved by the L2 participants ($U = 258.5$, $p < 0.01$, d = 0.79). This was somewhat unexpected because, as one reviewer points out, it would have been easier for the L1 to realise that these were non-existent words. We believe that the comparatively high scores of the L2 group may be due to the fact that, as learners of English, they were used to taking language and literacy tests prior to the experimental session, whereas the L1 group did not have experience with such tests. In other words, the L2 learners' success may be the result of a practice effect, as another reviewer suggests. There were no significant differences in accuracy, however, between the two groups in the reading summarisation question (RSQ). As summarisation is a skill widely used in education, it is perhaps not unexpected that both groups performed in similar ways on this part of the test.

The third research question focused on correlations between reading comprehension as measured with the RCQs and RSQ and students' VKS scores which measured success in learning the non-words. As can be seen in Table 4, a slightly stronger correlation was obtained between the VKS target non-words and the RSQ ($r_s = 0.496$, $p < 0.01$) than between the VKS target non-words and RCQ scores ($r_s = 0.474$, $p < 0.05$). Significant positive correlations were also obtained between the VKS (total scores) and RSQ scores ($r_s = 0.465$, $p < 0.05$), and the VKS (total scores) and RCQ scores ($r_s = 0.452$, $p < 0.05$). Thus, as expected, L2 learners who score highly on the RCQ and the RSQ also obtain high scores on the VKS. The current findings therefore suggest that there is a positive relationship between reading comprehension performance and students' ability to make lexical gains during reading.

	VKS (target non-words)	**VKS (total scores)**	**RCQ (total scores)**	**RSQ (total scores)**
VKS (target non-words)	_____			
VKS (total scores)	_____	_____		
RCQ (total scores)	**0.474***	**0.452***	_____	
RSQ (total scores)	**0.496****	**0.465***	**0.611****	_____

Table 4: Spearman Correlations between VKS, RCQ and RSQ for learners of English, * $p<0.05$, ** $p<0.01$

In a next step, we carried out various regression analyses to investigate whether learners' reading comprehension and summarisation scores could predict the VKS scores. In the first model, VKS non-word scores were the dependent variable and the RCQ scores the predictor. This model revealed that the RCQ measure was a significant predictor of the VKS non-word scores ($\beta = 0.44$, $F(1, 26) = 6.39$, $p = 0.018$, $p < 0.05$). The overall model was significant, and the adjusted R^2 value was 0.197. This means that 19.7% of the variance in the VKS target non-word scores was accounted for by RCQ scores. Second, we investigated to what extent the RSQ could predict students' scores on the non-words. This was indeed the case ($\beta = 0.43$, $F(1, 26) = 6.05$, $p = 0.021$, $p < 0.05$). Again, the overall model was significant ($R^2 = 0.189$). This means that 18.9% of the variance in VKS non-word scores was accounted for by the summarisation scores.

Finally, using a hierarchical regression model, we regressed the VKS non-word scores on both the RCQ and the RSQ scores. In a first model we entered the RCQ in a first step and the RSQ in a second step, and found the multicollinearity statistics were within acceptable limits. The resulting model was significant ($F(1,26)=6.394$, $p =0.018$), and the adjusted R^2 was 0.167. While after adding RSQ the explained variance appeared to increase slightly ($R^2 = 0.178$), the change in R^2 was not significant. The β for RCQ (0.285) was slightly higher than for RSQ (0.259), but the differences were very small. In a second model, we reversed the order of entry of the two predictor variables and found that the results were almost identical. Adding RCQ in the second step after RSQ did not lead to a significant change in R^2. Thus, we conclude that the two variables are important predictors of the VKS scores but one does not explain additional variance over and above the variance explained by the other.

6. Discussion

In this study, we first investigated learners' vocabulary size. The L1 Arabic learners' scores on the VST suggest that this group of L2 learners had a receptive vocabulary of just under 4,000 words. As pointed out by Gyllstad, Vilkaitė and Schmitt (2015) it is possible that the VST overestimates test takers' vocabulary sizes by up to 26% because the VST is a multiple-choice test. Students' ability to guess words might have distorted the results, and actual vocabulary sizes could therefore be considerably lower. However, the results from the SWRT indicate that students were able to read 47.21 out of the 70 items on this test (67.44%). As the first half of the SWRT consists mainly of words up to the 3k level (Treffers-Daller & Huang, in press), and students were able to accurately read far more than the first 35 words on the test, it is clear that they knew (at least the form of) a considerable number of words beyond the 3k level. The combined evidence from the VST and the SWRT therefore makes it likely that students knew on average no less than 3,000 words and possibly up to 4,000 words, which means that their performance was comparable to that of many L2 learners in Larson (2017). On the SWRT the Arabic students outperformed the learners in Treffers-Daller and Huang (in press). The Chinese students in the latter study were able to read just over 50% of the items correctly on the SWRT.

The second aim of the study was to establish whether the reading passages from the YARC were suitable for the L2 learners in the current study. If we assume that the learners had the vocabulary levels described in the previous paragraph, the reading passages from the YARC were indeed appropriate for these learners, but only after words from frequency layers lower than the 4k level had been replaced. This meant that students could be expected to know almost 98% of the words in the stories and would be able to guess the meaning of the two non-words in the text. The results on the reading comprehension questions and the summarisation questions, where students gave between 68% and 79% correct answers, suggest that was indeed the case: students' comprehension levels were therefore close to or exceeded the 70% comprehension levels mentioned in Schmitt et al. (2011) as the target for adequate

comprehension. Their ability to guess the meaning of the non-words in the text, and to use these in a sentence was, however, limited. Respondents from both groups obtained scores between 4.5% and 12% for different parts of the VKS related to non-words (see Table 2). In general, they appeared to be slightly better at recalling the meaning than the form, and least good at using the word accurately in a sentence. These results seem low, by comparison with those reported in Godfroid, Ahn, Ballard, Cui, Johnston, Lee, Sarkar and Yoon (2018), who report scores between 30% accuracy for form and meaning recognition and 13% accuracy for meaning recall. It is important to note, first of all, in relatively short texts, such as those from the YARC, frequency of occurrence is necessarily low. Students in the current study had had relatively few opportunities to encounter the non-words, as they occurred only four or eight times. According to Pellicer-Sánchez and Schmitt (2010), readers often need ten occurrences for any learning gains to occur. Results differ between studies, however: in the study of Godfroid et al., many words had very low frequencies (below 4 occurrences in the text). Second, the L1 and L2 group were not significantly different from each other in their ability to recall the non-words. In that respect, the results of the current study are comparable to those of Godfroid et al., who also found no differences between native and non-native speakers.

The third aim of the study was to establish to what extent students' reading comprehension/ summarisation skills correlated with their scores on the VKS, and whether the former explained any variance in the latter. The correlations show that there was indeed a positive relationship between reading comprehension performance and students' ability to make lexical gains during reading. While the data therefore support the claim that higher comprehension leads to better word learning, both the L1 and L2 participants seemed to struggle with learning the non-words. Thus, despite the fact that the learners were familiar with 98% of the words in the text, and levels of comprehension were generally high, this does not automatically mean that meanings and forms of novel words will be learned.

7. Conclusion

This study set out to investigate the suitability of the YARC Secondary Test for L1 Arabic adult learners of English. This aim was pursued, first of all, through an analysis of students' scores on the Single Word Reading test (SWRT), which is a component of the YARC that measures decoding. Second, we looked at the vocabulary in the YARC texts, and the percentage of words from these texts that the target group was likely to know given their vocabulary knowledge. Third, we analysed students' ability to learn the form, meaning and use of non-words that replaced existing words in the YARC texts.

We found, firstly, that students' scores on the SWRT were close to the lower end of the range of scores for children aged between 11 and 16, as shown in the manual. We subsequently investigated to what extent the corresponding reading comprehension texts from the YARC were appropriate for the target group. Using Nation and Beglar's (2010) Vocabulary Size Test we found that students knew on average just under 4,000 words, and that the texts from the YARC needed to be simplified slightly to ensure the 98% coverage would be reached. A modified version of Paribakht and Wesche's (1996) Vocabulary Knowledge Scale (VKS) was used to assess students' ability to make lexical gains. While students' scores on the VKS were low, moderately strong correlations were found between students' reading comprehension and summarisation scores on the one hand and their scores on the VKS on the other. Regression analyses also revealed that the comprehension and summarisation scores explained 19-20% of the variance in VKS scores. Thus, levels of comprehension were sufficient for some students to make at least some lexical gains, but across the board lexical gains were limited.

Overall, we conclude that the SWRT component and the reading comprehension texts in the YARC were suitable for the target group once the reading texts had been simplified slightly to

accommodate the learners' vocabulary levels. Our findings confirm those of Treffers-Daller and Huang (in press) who arrived at the same conclusions in a study among Chinese L1 adult L2 learners of English. This does not mean that the YARC can be used with any adult L2 learners, regardless of their language proficiency. We recommend that the vocabulary levels of potential test takers be carefully checked and the suitability for the target group investigated in any future studies among adult L2 learners, because readers need to know 98% of the words to be able to understand the texts. As the vocabulary in reading comprehension tests for native speakers is often too complex for learners, it is useful for studies which focus on adult L2 learners to consider using (an adapted version of) the YARC Secondary.

References

Alsaif, A. (2011). *Investigating vocabulary input and explaining vocabulary uptake among EFL learners in Saudi Arabia.* Unpublished PhD thesis, Swansea University.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101-118.

Bogaards, P. (2000). Testing L2 vocabulary at a high level: The case of the Euralex French tests. *Applied Linguistics, 21*,490-516.

Brown, D. (2008). Using a modified version of the Vocabulary Knowledge Scale to aid vocabulary development. *The Language Teacher, 32*(12), 84-105.

Cambridge English Language Assessment (2014). *The Preliminary English Test (PET)*. Cambridge: Cambridge University Press.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10,* 277-299.

Dearden, J. (2014). *English as a medium of instruction–a growing global phenomenon: Phase 1*. London: British Council.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A. & Yoon, H.-Y. (2018). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism, Language and Cognition 21*(3), 563-584.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015) Assessing vocabulary size through multiple-choice formats. *International Journal of Applied Linguistics, 166*(2), 278-306.

Hamdi, R., Ghazali, S., & Barkat-Defradas, M., (2005). Syllable Structure in Spoken Arabic: A Comparative Investigation. In: Proceedings of the *International Conference of Speech, Communication, and Technology*, Lisbon, Portugal, 4-5 September 2005, 2245-2248.

Harrington, M., & Roche, T. (2014). Identifying academically at risk students in an English-as-a-Lingua-Franca university setting. *Journal of English for Academic Purposes, 15*, 37–47.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127-160.

Hsueh-chao, M., & Nation, P. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Hutchinson, J. (2018). *Educational outcomes of children with English as an additional language*. Education Policy Institute.

Jeon, E. H., & Yamashita, J. (2014). L2 Reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160-212.

Jiang, X. (2011). The role of first language literacy and second language proficiency in second language reading comprehension. *The Reading Matrix, 11*(2), 177-190.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. Routledge.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21*(2), 85-97.

Larson, M., (2017). *Thresholds, Text Coverage, Vocabulary Size, and Reading Comprehension in Applied Linguistics*. Published PhD thesis. Victoria University of Wellington.

Laufer, B. & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15-30.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Clevedon, UK: Multilingual Matters.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.
Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/La revue canadienne des langues vivantes*, *63*(1), 59–82.
Nation, P. (2012). Measuring vocabulary size in an uncommonly taught language. Paper presented at the International Conference on *Language Proficiency Testing in the Less Commonly Taught Languages*, Bangkok, Thailand.
Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9-13.
Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge, England: Cambridge University Press.
Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174-200).
Pellicer-Sanchez, A., Schmitt, N. (2010). Incidental Vocabulary Acquisition from an Authentic Novel: Do "Things Fall Apart"? *Reading in a Foreign Language*, *22*(1), 31-55.
Qian, D. D. (1998). *Depth of vocabulary knowledge: Assessing its role in adults' reading comprehension in English as a second language*. (Unpublished doctoral dissertation). University of Toronto, Toronto, Canada.
Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*(3), 513-536.
Rashidi. N., & Khosravi, N. (2010). Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics*, *14*(1), 81-108.
Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*, 355-371.
Ready, R. E., Chaudhry, M. F., Schatz, K. C., Strazzullo, S. (2012). "Passageless" administration of the Nelson-Denny Reading Comprehension Test: Associations with IQ and reading skills. *Journal of Learning Disabilities*, *46*, 377–384.
Schellings, G., Aarnoutse, C., & Leeuwe, J. V. (2006). Third-graders' think-aloud protocols: Types of reading activities in reading an expository text. *Learning and Instruction*, *16*(6), 549-569.
Schmitt, N., Nation, P. & Kremmel, B. (2019). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, *52*(4).
Schmitt, N., Jiang X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26-43.
Stothard, S. E., Snowling, M. J., Clarke, P. J., Barmby, P., & Hulme, C. (2010). *York assessment of reading and comprehension* (2nd ed.). London: GL Assessment.
Strand, S. & Hessel, A. (2018). *English as an Additional Language, proficiency in English and pupils' educational achievement: An analysis of Local Authority data*. The Bell Foundation.
Tsai, Y. R., Ernst, C., & Talley, P. C. (2010). L1 and L2 strategy use in reading comprehension of Chinese EFL readers. *Reading Psychology*, *31*(1), pp.1-29.
Treffers-Daller, J. & Huang, J. (in press). Measuring reading and vocabulary with the Test for English Majors Band 4: a concurrent validity study. To appear in J. Clenton and P. Booth (Eds.). *Vocabulary and the four skills- current issues future concerns*. Taylor and Francis.
Trenkic, D, & Warmington, M. (2017). Language and literacy skills of home and international university students: How different are they, and does it matter? *Bilingualism: Language and Cognition*, *22*(2), 349-365.
Warmington, M., Stothard, S.E., & Snowling, M.J. (2013). Assessing dyslexia in higher education: The York adult assessment battery – revised. *Journal of Research in Special Educational Needs*, *13*, 48–56.
Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*, 13-40.
Wolter, B. (2005). *V-Links: A new approach to assessing depth of word knowledge*. (Unpublished PhD dissertation). University of Wales, Swansea.
Zhang, D., & Yang, X. (2016). Chinese L2 learners' depth of vocabulary knowledge and its role in reading comprehension. *Foreign Language Annals*, *49*(4), 699-715.

Author Bios

Jeanine Treffers-Daller is Professor of Multilingualism in the Department of English Language and Applied Linguistics, University of Reading, UK. She has published widely on the measurement of vocabulary knowledge and use among bilinguals and second language learners and has recently edited a volume with Carmen Silva-Corvalán on *Language dominance in bilinguals: issues of operationalization and measurement*. She is a member of the Editorial Board of the *International Journal of Bilingualism* and of *Bilingualism, Language and Cognition*.

Reham Alkhudiry is an Assistant Professor of Applied Linguistics in the Department of English Language and Translation, Qassim University, SA. Reham was a PhD student in the Department of English Language & Applied Linguistics, University of Reading and was awarded her doctorate in 2018. Her main areas of interest are: second language acquisition, vocabulary learning and teaching, L2 lexical representation and development, assessment of reading comprehension and vocabulary in L2 learners.

Jacqueline Laws is an Honorary Fellow in Linguistics in the Department of English Language and Applied Linguistics, University of Reading. Her research interests relate to constructional approaches to grammar, child language acquisition, cognitive linguistics, corpus linguistics and motion event cognition. One of her current projects on English derivational morphology evaluates the distributional properties of complex words in adult spoken language as a function of register.