Comparing Phonological and Orthographic Vocabulary Size: Do Vocabulary Tests Underestimate the Knowledge of Some Learners

James Milton, Nicola Hopkins

The Canadian Modern Language Review / La revue canadienne des langues vivantes, Volume 63, Number 1, September/septembre 2006, pp. 127-147 (Article)

Published by University of Toronto Press
DOI: https://doi.org/10.1353/cml.2006.0048

Comparing Phonological and Orthographic Vocabulary Size: Do Vocabulary Tests Underestimate the Knowledge of Some Learners

James Milton
Nicola Hopkins

Abstract: Any description of what it means to know a word in a foreign language is likely to include recognition of form, both how a word sounds when heard and what it looks like when written. However, tests of vocabulary knowledge focus almost exclusively on the written form of the word. We have little idea of learners' phonological vocabulary knowledge or how it might interact with orthographic knowledge. It is suspected that written vocabulary tests may underestimate the vocabulary knowledge of native Arabic speakers, who often handle English orthography poorly. This paper reports a comparison of the phonological and orthographic vocabulary sizes of Greek and Arabic native speakers. Results suggest that written tests do not underestimate Arabic speakers' vocabulary size. The two aspects of vocabulary knowledge develop differently with language level. Very proficient learners of EFL are characterized by an orthographic word recognition much greater than their phonological word recognition.

Résumé : Toute définition de ce que cela veut dire de connaître un mot dans une langue étrangère devrait inclure la reconnaissance de sa forme, tant sonore quand on l'entend qu'écrite quand on le voit. Toutefois, les examens visant à évaluer la connaissance du vocabulaire se concentrent presque tous exclusivement sur la forme écrite des mots. Nous en savons peu sur le vocabulaire phonologique des apprenants ou son lien avec le vocabulaire orthographique. On soupçonne que les examens écrits ne reflètent pas les connaissances en matière de vocabulaire des arabophones, qui ont souvent de la difficulté avec l'orthographe anglaise. L'article signale qu'une comparaison de la taille des vocabulaires phonologiques et orthographiques a été effectuée chez des personnes dont la langue maternelle est le grec ou l'arabe, et les résultats obtenus aux examens ont correspondu à la taille du vocabulaire des arabophones. Les deux aspects de la connaissance du vocabulaire se développent différemment selon le niveau de langage. Chez les apprenants très compétents en ALS, la reconnaissance orthographique des mots est supérieure à la reconnaissance phonologique.

© 2006 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63,1 (September/septembre), 127–147

Background

Vocabulary breadth is an area where there appears to be some general agreement in what the construct is and how we measure it. Within Nation's framework of what it means to know a word (Table 1), the concept of vocabulary breadth sits broadly within the two sections entitled Form and Meaning or, more precisely, the Form section and the *form and meaning* subsection of Meaning. If a word can be recognized in some form as a word and, perhaps also, if a meaning can be attached to it, then we can count that word as part of a learner's vocabulary. We can then count these words and form an estimate of a learner's vocabulary size. All other things being equal, we would expect that as a learner learns more and more language, then vocabulary size would tend to increase too. We tend to consider how many words a learner knows, to some extent irrespective of how well these words are known. How well

TABLE 1
What is involved in knowing a word

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What words parts are needed to express meaning?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concepts and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What others words does this word make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of word occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use	R	Where, when, and how often would we meet this word?
		P	Where, when, and how often can we use this word?

Note: R = receptive knowledge, P = productive knowledge
Source: Nation, 2001, 27

a word is known would be part of a different axis of word knowledge, that of vocabulary depth.

To measure this concept of vocabulary breadth or size we have a number of seemingly well-constructed tests that give reliable and, most important of all, believable estimates of vocabulary knowledge and the number of words a learner knows. The scores on these tests tie in with other qualities of foreign-language knowledge such as general reading comprehension, which, because of the impact of coverage, we would expect to be associated with overall vocabulary size.

A principal example would be Nation's Vocabulary Levels Test (1983) (revised version by Schmitt, Schmitt, & Clapham, 2001). This test samples the 2,000, 3,000, 5,000, and 10,000 word-frequency ranges and the University Word List, and requires learners to match test words with a correct meaning. A second example would be Meara and Milton's X-Lex (2003), which is also designed as a general placement indicator and samples learner knowledge of each of the first five 1,000 word-frequency bands in English. Learners are required simply to identify the words they know. The test includes pseudo-words, which allow a calculation for guesswork and overestimation to be made, and the learner's score is adjusted on that basis. The scores it provides – estimates of how many words in the most frequent 5,000 in English can be recognized – have been shown to tie in with a wide range of general ability levels from quite low level to very advanced (Table 2).

Perhaps it is not surprising that vocabulary-size measures of this kind should link so well with general language exams and hierarchies of ability such as the Common European Framework for Foreign Languages (Council of Europe, 2001), since vocabulary can be shown to predict performance in a number of specific language skills on which these exams and hierarchies rely. A number of studies – for example, Meara and Buxton (1987) – successfully correlate vocabulary size with scores on reading comprehension tests, academic writing scores, and

TABLE 2
Vocabulary size estimates and exam equivalents

CEF level	Cambridge ESOL Test	Vocab size/5,000
A1	Starters, etc.	< 1,500
A2	KET	1,500–2,500
B1	PET	2,750–3,250
B2	FCE	3,250–3,750
C1	CAE	3,750–4,500
C2	CPE	4,500–5,000

Source: Adapted from Meara & Milton, 2003, 12

grammatical knowledge. Vocabulary size is less successful, it seems, in predicting oral fluency among foreign language learners.

The connection between vocabulary size and overall language performance is not limited exclusively to English. Milton (2006), in a cross-sectional study of French foreign-language learners in the United Kingdom, has also been able to link vocabulary size to a hierarchy of exams and the Common European Framework (CEF); these findings are summarized in Table 3. The French exams referred to in the table are the age 16 exams (General Certificate in Secondary Education) and age 18 (Advanced level) exams, which link to the British National Curriculum. Interestingly, the vocabulary sizes associated with the Common European Framework (CEF) levels are significantly different from Meara and Milton's table of EFL equivalences in Table 2.

But no matter how well and persuasively these tests work, one feature stands out: They rely on testing the written form of words only. The reason is probably that until very recently the technology for the reliable, fair, and consistent delivery of vocabulary tests including sound did not exist. Only the arrival of the personal computer has allowed such a test to be delivered in a standardized way. Nonetheless, since these tests are delivered in writing only, it can be argued that they test only the learner's knowledge of orthographic form rather than a broader construct that Nation describes in Table 1, which includes the phonological form. Even a computer-delivered test such as X-Lex still relies on the written form of words for delivery. It is unlikely that a single test could usefully characterize every aspect of vocabulary knowledge.

Valid and reliable measures of phonological vocabulary size, where they exist, are much less well known than the Vocabulary Levels Test or X-Lex. Fountain developed a vocabulary knowledge test, equivalent to the Vocabulary Levels Test, based on dictation (Fountain & Nation, 2000). Details are provided in Nation (2001, p. 429), but the academic literature otherwise contains little reference to it. We have little or no

TABLE 3
Vocabulary size estimates and equivalents from Milton

CEF level	French exam	Vocab size/5,000
A1		
A2	GCSE Lower	800–1,000
B1	GCSE Higher	800–1,000
B2	'A' level	2,000
C1		
C2	BA in French	3,300

Note: Not all CEF levels have a formal exam in this system

idea how learners might score on a phonological test of vocabulary knowledge, or whether these scores might differ from scores on orthographic tests. If we have assumed anything at all, it would be that the two types of knowledge are likely to be closely connected and that if you can recognize the written form of the word then you will probably recognize it in speech as well. If there is to be a difference, one might suspect that there would be a slight advantage in reading vocabulary items in a test over hearing them. Function words in English, for example, may vary in pronunciation according to their sentence position and the other words around them, and may often disappear completely in natural speech. Words like this might well be much less recognizable, whatever form they are given in aurally, than in writing, where there is only one form.

However, a small body of research in vocabulary learning suggests that this general idea – that knowledge of word orthography and of phonology are likely to be closely linked – may not hold true for all learners. In particular, this would appear true for native speakers of Arabic. It has already been noted that vocabulary-size tests that predict reading, writing, and grammar scores so well do not seem to predict oral fluency in English. Nowhere is this distinction more obvious than among native Arabic speakers. Milton (2005) has noted that these learners can be highly cogent in speech even when vocabulary-size tests suggest their knowledge of English is very limited. One explanation is that the vocabulary knowledge these learners have is phonological in form, and the tests we use, presenting words orthographically for recognition, do not allow these learners to show us what they do know. The usual tests we use, which appear to work so well with most learners, may be systematically underestimating the vocabulary knowledge of Arabic learners in particular.

The thinking behind this idea is explained by Ryan (1997) and by Milton (2005). It has its roots in the Arabic writing system, where all Arabic words are based on three radicals. These are three consonant sounds that, because of the close sound/symbol correspondence of Arabic writing, are shown as three written letters. Every word within the same word family is a development of these three radicals. There may be extra sounds and letters added, but the same three consonants always appear, and they are always in same order. It appears from studies of eye movements (Randall, 1990) that readers in Arabic home in on these consonants in particular, presumably because this method provides a quick route to word meaning and is thus a very efficient way of decoding writing in Arabic. But while this process may work well in Arabic, it is likely to be a less efficient way of decoding writing in

English. Efficient readers of English use other strategies and often have quite different eye movements as they seek to recognize whole word shapes or – as in the 'bathtub' effect – examine word beginnings and endings (Randall, 1990). Concentration on only the consonants of English is likely to be misleading for the reader, since consonants in English are less likely to suggest the meaning of a word than in Arabic. A flavour of the type of confusion that results when the consonants are predominantly addressed in reading is provided by Ryan (1997, p. 186) using examples of Arabic speakers writing in English:

> We get water from deep *wheels* (wells).
> You get upstairs in a *left* (lift).
> Goods are carried on a *fright* (freight) train.
> He went to *present* (prison) for the crime.

Ryan points out that in each case the consonants remain relatively unaffected while the vowels are frequently omitted, mis-positioned, or substituted for each other. The result, for those of us who read English fluently, is a confusion of meanings. Not only do Arabic learners suffer problems in recognizing the quality of vowel sounds in English through negative transfer, but they may also suffer difficulties in recognizing the quality of written vowels in English because of their habits of processing written text. Haynes (1984) calls this type of confusion among Arabic speakers 'vowel blindness.' Randall's work suggests that Arabic speakers can and do transfer their eye-movement habits from Arabic and try to read English in the same way, and this, together with the nature of the errors these learners produce, suggests that such learners may have a very real problem reading the words presented to them in vocabulary-size tests.

This difficulty ought to reduce the validity of the usual vocabulary-size tests with learners with this kind of background. In a recognition test such as X-Lex, this obstacle may work in two ways. One is that they may not recognize real words when they are presented in the test, even if they are known. Ryan's examples suggest that words such as *well*, *freight*, and *prison* are known in some form to the learners, presumably in phonological form, but that the written form is much less accessible. Learners with this problem seem likely to under-report their knowledge as a consequence. A second problem, which is potentially more damaging, lies with the use of false words. Test-takers frequently resort to sounding out the words they read in vocabulary tests, and where vowels are so flexible in interpretation to these learners, it seems inevitable that false words can be construed as words that they do

recognize phonologically. The identification of false words as real words, called false alarms, is heavily penalized in checklist-style tests. Al-Hazemi (1993) reports comparatively high false alarm rates among the Saudi learners he tested using English checklist-style tests, although his conclusion is that the tests are valid and reliable nonetheless.

In principle, then, Arabic speakers might under-represent their knowledge on vocabulary tests that are in written format. They might well do rather better, again in principle, if these tests were presented aurally or included an aural element. However, since we do not have standardized measures of phonological vocabulary knowledge to run alongside the orthographic tests we have, this possibility is speculation.

Aims and objectives

The principal aim of this paper, therefore, is to report a study that addresses the issues and questions surrounding learners' phonological – as distinct from their orthographic – word knowledge. It will attempt to answer three specific questions:

1. Is it possible to design and construct an aural test of vocabulary (Aural Lex) that is the same in the selection of test items and in checklist format as an orthographic test of vocabulary (X-Lex), and can give believable estimates of phonological vocabulary size?
2. Using such a test alongside orthographic tests, does it appear that learners' phonological knowledge is related to – and comparable in size to – their orthographic knowledge?
3. Do native Arabic speakers differ from other learners, who do not have the Arabic script problem, in the way their orthographic and phonological vocabulary knowledge compare?

Aural Lex design

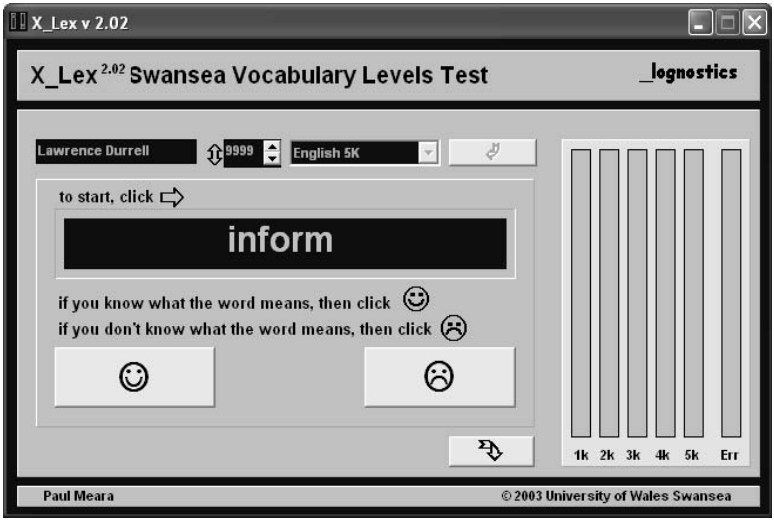
The phonological test for use in this study has been designed to mimic, as precisely as possible, the X-Lex vocabulary-size test in the number and choice of test items and in manner of delivery. The only difference, which is deliberately sought, is that test words will be heard and not seen.

X-Lex tests knowledge of the most frequently occurring 5,000 words in English and estimates overall knowledge of this vocabulary. The word lists it uses are drawn from Hindmarsh (1980) and Nation (1984) and are lemmatized. It is a computer-delivered Yes/No test, which presents learners with 120 words, one by one. Learners have to indicate

whether they know each word. There are 20 randomly selected words from each 1,000 word-frequency band, and a further 20 pseudo words that are designed to look and sound like words in English but are not real English words. The number of Yes responses to these pseudo-words allows the score on the real words to be adjusted for guessing and overestimation of knowledge. There is no time limit to the test, which generally takes 5 or 10 minutes to complete. A learner's vocabulary knowledge is calculated by counting the number of Yes responses to real words and multiplying the result by 50 to give a raw score out of 5,000. The number of Yes responses to pseudo-words is then calculated and multiplied by 250. This figure is deducted from the raw score to give an adjusted score, which thus includes a compensation for guesswork. The adjusted scores are those reported in this study. They are estimations of the learners' passive receptive vocabulary size. A screenshot to illustrate the delivery format is shown in Figure 1.

The phonological test, Aural Lex (Milton & Hopkins, 2005) replicates this format wherever feasible. The test words follow the same selection procedure, are the same in number, and are delivered one by one as in X-Lex. These words are obviously not seen, but the learner presses an on-screen button to hear each word, which is given in citation form. The button may be pressed as many times as needed so the test word can be reheard. Once the learner has made a decision and has pressed the Yes

FIGURE 1
X-Lex test format



or No button, a new test word is loaded and this, in turn, can be heard as many times as is needed. This process continues until all the test words have been heard. The test words include the same sample of pseudo-words, which are non-words both in sound and in writing. The scoring process is identical. This version of the test takes slightly longer than the written version, approximately 10 to 15 minutes on average. A screenshot to illustrate the delivery format is shown in Figure 2.

Experimental design and participants

Aural Lex and X-Lex were administered to 126 learners from a range of EFL levels and were drawn from a mixture of Arabic and Greek native-speaking backgrounds.

The participants included 88 Greek native speakers ranging in age from seven to adult, and in ability from beginners completing their first year of English up to very advanced users of English preparing for the Cambridge proficiency exam. All were students at a language school in Larissa, Greece. Greek native speakers were chosen since they are learners of English for whom vocabulary-size tests generally work well in predicting both overall language level and performance on EFL exams. These learners do not usually display the kind of reading difficulties associated with Arabic speakers but, in learning English, they do have to learn a new orthography like the Arabic speakers.

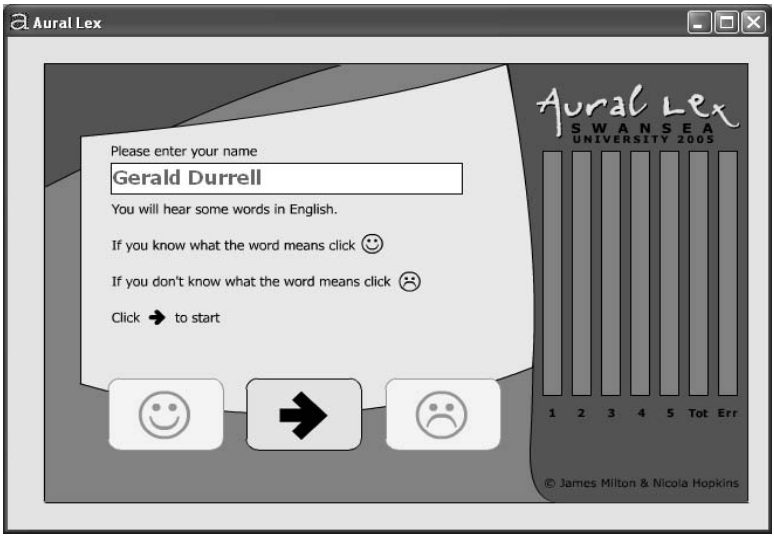
The participants further included 38 Arabic native speakers ranging in age from seven to adult, and in ability from beginner to very advanced users of English with TOEFL scores in advance of 600 and preparing for postgraduate study in Britain and the United States. Twenty-seven were students at a language school in Amman, Jordan, and 11 were Saudis recently arrived and studying in Swansea, United Kingdom. Arabic native speakers were chosen, obviously, to allow comparison with the Greek group in their performance on the vocabulary-size tests.

Data were collected such that the two groups were as similar in overall ability and in distribution of age and ability as could reasonably be attained.

Test administration

Both Aural Lex and X-Lex were delivered in computer labs and, in the case of Aural Lex, using headphones. Word selection was different in each test, and the order of the tests was randomly assigned, with half the learners in each group taking Aural Lex first and the other half X-Lex first.

FIGURE 2
Aural Lex test format



Expectations and interpretation

The first objective in the study is to design a test of vocabulary size based on the phonological rather than the written form of words, which gives credible scores. X-Lex, the orthographic vocabulary-size test, has proved its reliability, which gives reason to think that Aural Lex, which has a similar format but differs in one key aspect, will perform reliably too. It will be regarded as working satisfactorily if it can be demonstrated to be reliable. The test, in addition to producing an overall vocabulary-size estimate produces sub-scores for each of the five 1,000 word-frequency bands. These sub-scores have been used as items to calculate reliability using Cronbach's alpha for both Aural Lex and X-Lex. For comparison, the same calculation will be made for X-Lex. It will also be regarded as working satisfactorily if scores correlate with X-Lex scores significantly. A very strong correlation might be a surprise, especially given the problems we know that Arabic speakers labour under when taking written tests, but it would be very surprising if phonological knowledge of words did not increase at all in relation to increases in overall ability and orthographic vocabulary size.

The second objective is to compare orthographic and phonological vocabulary size in learners. It might be broadly expected that the two

will be comparable. The mean scores on the two tests will be compared in order to see if any significant difference between the two tests emerges.

The final objective is to test whether the Arabic learners show a pattern of scores on the two tests that is different from the Greek learners'. The Arabic learners, it has been argued, are more likely to experience problems with the orthography of English. They seem likely, therefore, to do comparatively less well on the orthographic form of the vocabulary test than on the phonological test. The group scores on the two tests will be compared to see if significant differences emerge.

Results

Aural Lex performance

Reliability statistics produced applying Cronbach's alpha to the test's sub-scores produced the results shown in Table 4. It can be concluded that the test appears to be performing reliably.

The correlation between scores on Aural Lex and X-Lex for the participants as a single group is shown in Table 5. The correlation between the two tests shows a moderate relationship. While this finding is hard to interpret in the absence of further evidence, some of which is presented below, it might suggest that Aural Lex appears to be performing satisfactorily and is providing a reasonable estimate of learners' phonological vocabulary size.

How do orthographic and phonological size compare?

Aural Lex has been constructed to estimate phonological vocabulary size in EFL learners in a manner that is as similar as possible to X-Lex, an

TABLE 4
Cronbach's alpha statistics

	Cases	Cronbach's alpha	Number of items
X-Lex	126	0.95	5
Aural Lex	126	0.96	5

TABLE 5
Correlation between Aural Lex and X-Lex

N	Correlation	Sig.
126	.68	< .01

orthographic test. It uses the same principles of word selection, the same numbers of words, and the same computer-delivery format. It should be possible therefore to compare scores from the two tests meaningfully. Results from the two tests are shown in Table 6.

It appears, at least in the learners taking part in this study, that their knowledge of the written form of words exceeds their knowledge of the aural form of words, despite the presence among the subjects of a number of learners – approximately a third – who might have been expected to have problems with orthographic form and to favour the phonological side of their lexicon. The difference between the means is significant: t (125) = 5.12, $p < .01$.

> *Do Arabic and Greek speakers differ in their phonological and orthographic word knowledge?*

The test scores have been subdivided according to the learners' native language to allow the Arabic-speaking learners' scores to be compared with the Greek-speaking learners. The expectation is that the Arabic-speaking learners might know more vocabulary phonologically than they do orthographically when compared with the Greek learners. The means scores are shown in Table 7.

The results show that the two groups appear very similar in their performance on X-Lex, but that they differ quite noticeably on Aural Lex. The similarity of the X-Lex scores is reassuring, since the two groups were constructed to be as similar to each other as possible in overall ability and the distribution of age and of language ability. The close means and standard deviations suggest that this aim was achieved quite well. The mean scores on X-Lex cannot be demonstrated to differ statistically: t (124) = 0.76, p = .48. Group scores on Aural Lex differ

TABLE 6
Mean scores for Aural Lex and X-Lex

Test	*M*	*SD*
Aural Lex	2,260.32	1,090.29
X-Lex	2,655.56	1,059.38

TABLE 7
Aural Lex and X-Lex means by language groups

Learners	*n*	Aural Lex *M*	Aural Lex *SD*	X-Lex *M*	X-Lex *SD*
Arabic group	38	2,823.68	1,242.64	2,553.94	984.43
Greek group	88	2,017.05	923.45	2,699.43	1,092.65

much more noticeably, and the Arabic group's score is significantly higher than the Greek group's score, t (124) = 4.04, $p < .01$.

The Greek group have scored higher on the written form of the test than on the aural form, and the difference in means is significant: t (87) = 10.02, $p < .01$. The result suggests they recognize more words orthographically than they do phonologically. This is not the case for the Arabic group, where the mean score on the aural test is higher than on the written test. This difference, however, is not significant: t (37) = 1.73, $p = .09$. The Arabic native speakers as a group, therefore, do not appear to have the imbalance in favour of the orthographic side of their lexicon that the Greek learners display.

Discussion

We began this paper by pointing out that existing tests of vocabulary breadth, which are widely used, are all orthographic, and that in our testing and consideration of learners' developing vocabulary we have failed to consider another important aspect of vocabulary breadth, that of phonological vocabulary knowledge. As a consequence, we can argue that we can form a good idea about how many words learners can recognize in foreign language writing, but not what they might or might not recognize in sound, when they hear the language spoken. We can only guess how large a learner's phonological vocabulary size is and how a learner's orthographic and phonological lexicons compare. The difficulties that native Arabic speakers experience in reading English suggests that, in some learners, the phonological and orthographic aspects of the lexicon may be quite different.

The creation of an aural version of a vocabulary-size tests allows this kind of calculation and comparison to be made. The results from the learners in this study suggest that the phonological and orthographic aspects of the lexicon are, overall, quite closely connected. If one aspect of the lexicon increases, the other is likely to increase too. The correlation between orthographic and phonological scores is surprisingly robust. However, the two aspects need not grow equally, and this study suggests that learners tend to grow larger orthographic lexicons than phonological.

This overall observation may be the product of the sample of learners used in this study. Two thirds are Greek and one third Arabic, and it was anticipated that there would be some differences between the groups on their scores. The two have to be proved different. It was observed from the data that the learners with a Greek first language demonstrate a significant difference between their scores on the two

tests, with orthographic scores tending to exceed phonological, while no such difference could be demonstrated among the learners with Arabic as a first language. This finding was not exactly what was expected. It seems that the native Arabic speakers do not have large phonological vocabularies that the orthographic tests are unable to tap. The fear, suggested by Ryan (1997) and Milton (2005), that vocabulary-size tests delivered in written form would underestimate native Arabic speakers' knowledge has not been borne out. Rather, if anything, the usual written tests we use will overestimate the phonological knowledge of the Greek native speakers. The relationship between these two aspects of the lexicon need not be fixed and universal, it would seem, but is related in some way to learners' background and first language.

Differences between the two language groups may also extend to the degree with which the X-Lex scores predict the Aural Lex scores. When the correlation between the two is calculated for each group separately, it appears the two are more closely connected in the Greek group, $r = .81$, $p < .01$, than for the Arabic group, $r = .65$, $p < .01$. This result may be only a product of the destabilizing effect of the false words in the checklist test format. A comparison of these is shown in Table 8.

In both groups the misidentification of false words as real is greater in the aural test than the written one. But, as noted by Al-Hazemi (1993), false alarms appear consistently high in the Arabic-speaking group, and more frequently than among the Greek speakers. There is no clear dividing line between acceptable and unacceptable levels of guesswork in these tests, but an average of more than six false alarms on the aural form of the test suggests that over 30% of answers may have an element of guesswork in them. It is nowhere near the rates reported among Dutch learners, which can be more than double this figure (Beeckmans, Eyckmans, Janssens, Dufranne, & Van de Velde, 2001), but intuitively the Arabic speakers' false alarm rates feel uncomfortably large. It suggests that native Arabic-speaking learners have a problem identifying English words as words, whatever form they encounter them in. Whatever problems they may have, they are not merely those of orthographic decoding. It might be suspected that learners from some

TABLE 8
Mean false alarm rates by group and test

	Number of pseudo-words	X-Lex	Aural Lex
Greek	20	1.91	4.63
Arabic	20	4.70	6.14

Note: A false alarm is a pseudo-word identified as real.

language backgrounds have a more flexible approach to word creativity than we would normally allow in English. Most native speakers of English have fairly fixed ideas about what words exist and are 'correct' in English. Neologisms, even if they are transparent in meaning and follow all the rules for correct word formation, are frequently resisted with accusations that language standards are declining. Other language users may be much more flexible and open to accepting words they may never have encountered, provided that the meaning can be understood. This would make a word recognition exercise, whether written or spoken, a much trickier exercise than we have previously considered. The structure and selection of pseudo-words in particular needs much more consideration and investigation if their use is to be continued.

The analysis thus far has concentrated on the characteristics of the groups in this study. However, the presentation of mean scores, and the fact that the two language groups have conveniently divided in performance as expected, rather disguises the degree of individual variation and the degree of overlap that the results display. To give an impression of the way individuals within these groups have performed, the scores on Aural-Lex and X-Lex have been plotted against each other. Figure 3 is a scattergram of the Greek group's scores and Figure 4 is a scattergram of the Arabic group's scores.

FIGURE 3
Scattergram of greek native-speaker scores

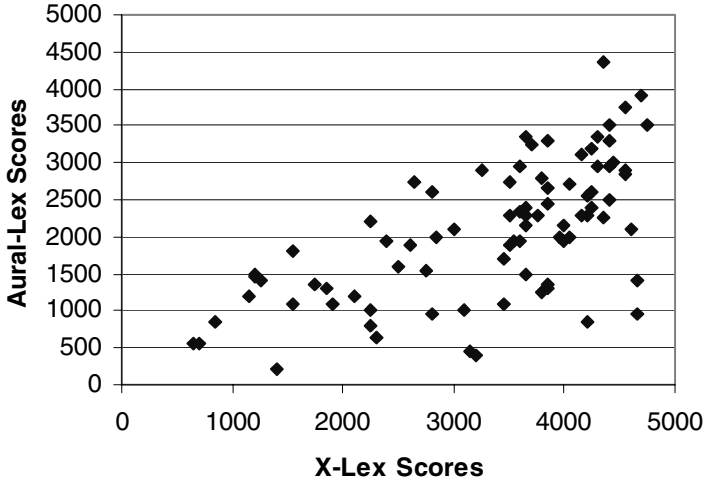
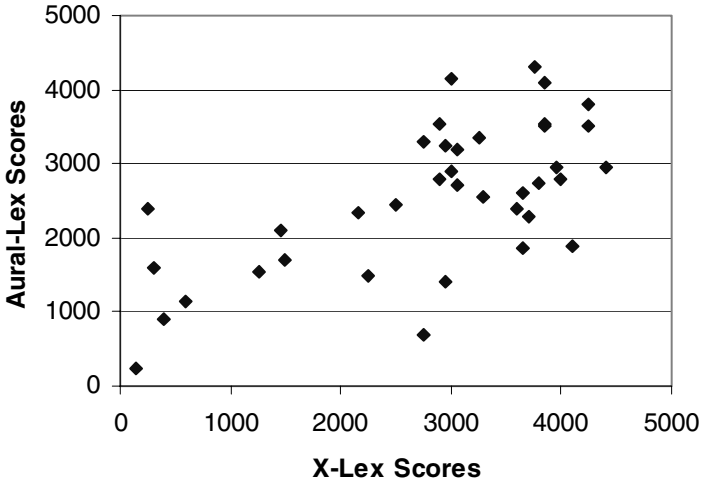


FIGURE 4
Scattergram of Arabic native-speaker scores



Presenting the results this way reveals some astonishing contrasts. In Figure 4 one Arabic learner appears to recognize about 2,500 words aurally but almost none in writing, while another appears to recognize nearly 3,000 in writing but only 700–800 in speech. These figures ought to suggest that the first learner could be highly competent in speech, where with approximately 2,000 to 3,000 words in English you might hope to understand almost everything in general conversation (Adolphs & Schmitt, 2003; Schonell, Middleton, & Shaw, 1956). But the same learner might well be a comparatively poor performer in reading and writing, since he or she would seem to have very few words to work with. With the second learner, the reverse would be true. The scattergrams also suggest that whatever the group differences, many learners across the two groups are performing comparably. Most of the Arabic learners' scores would be lost within the Greek data if the two scattergrams were overlapped.

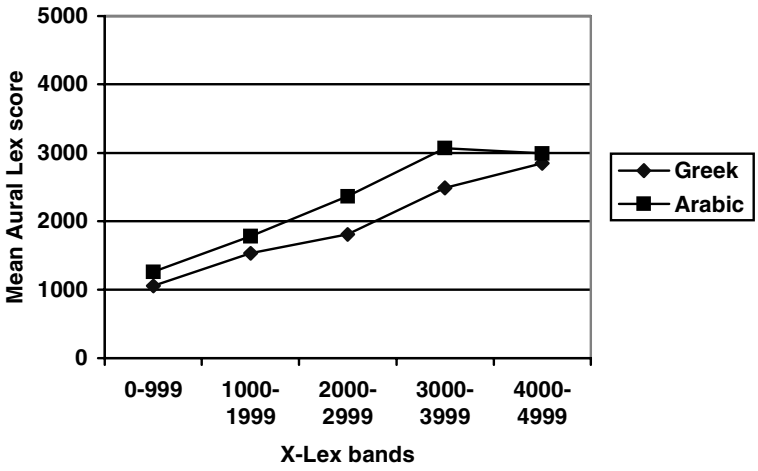
Nonetheless, the scattergrams also hint at a more subtle relationship in the data, one that the analysis has not been able to show as yet. In both language groups, low scores tend to be associated with a greater tendency for phonological vocabulary knowledge to exceed orthographic vocabulary knowledge. The tendency for scores to lie above the diagonal line showing parity of scores, especially in the low-scoring Arabic speakers' data, makes this very clear. For high scorers, the reverse is true, and this point is particularly clear in the highest scores

in the Greek data, which fall consistently below the diagonal. Among Greek learners in this study, high ability is always associated, it would seem, with knowing many more words by sight than by hearing. A reworking of the data, Figure 5, makes this relationship clearer.

In Figure 5 the data have been grouped by X-Lex scores. Learners with X-Lex scores between 0 and 999, and who would be beginners, form the first group, learners whose X-Lex scores fall between 1,000 and 1,999 form the second, and so on, up to learners with scores between 4,000 and 5,000, who form the top group and will be highly proficient in English. A division of the data by level in this way coincidentally produces a cline of general language level. The mean Aural Lex score for each of these groups has been calculated, revealing the relationship between language level and how much the lexicon is dominated by orthographic word knowledge. At the outset of learning, phonological word knowledge tends to exceed orthographic word knowledge. As learners become linguistically better, the orthographic vocabulary knowledge becomes more prominent, and it grows disproportionately in relation to phonological word knowledge. By the time learners become advanced-level users of English, and a score over 4,000 on X-Lex indicates a high level of proficiency, as much as 40% of the words they know are recognized by sight only.

In a study that has concentrated on the likely differences between the vocabularies of Arabic-speaking and other learners, the most surprising

FIGURE 5
X-Lex bands and Aural Lex mean scores



and noticeable feature of this diagram is just how similar the two groups of learners are. The native Arabic speakers display only a slight preference for phonological vocabulary compared to their Greek equivalents over the entire scale on knowledge. In both groups, the most able learners have much larger orthographic vocabularies than phonological. Apparently, if you want to become a really good EFL user, at least as measured by formal exams, then you must develop this large orthographic knowledge. It may be that the inability of some native Arabic learners to develop this vocabulary holds them back from making measurable progress beyond intermediate level on most formal language tests, even when their vocabulary allows them to be orally proficient.

Why should this be so? Part of the answer may lie in with the way in which we teach language skills, especially at the most able end of the learning scale, and the way we test overall ability. We tend to be rather academic and book orientated. Growing a large lexicon in any language and becoming academically able is often associated with extensive reading. Nation (2001) reports repeated studies that emphasize the importance of extensive reading in developing this aspect of knowledge, rather than extensive listening. It seems likely that many of the words that are acquired by this process are only ever encountered in writing, and learners may never actually hear them.

This possibility would tie in too with the effects of word frequency in learning. There is a strong relationship between the frequency with which a word occurs and the likelihood it will be learned, and there are differences between the word-frequency distributions in written and spoken corpora. Spoken text, it seems, makes particularly heavy use of the most frequently occurring words in English, and in a least one study the most frequently occurring 2,000 words gave 95% coverage of a spoken corpus (Schonell, Middleton and Shaw, 1956). By contrast, at least double this figure would be required for similar coverage of a written English corpus (Nation, 2001, p. 147). This ought to mean that it is easier to learn comparatively infrequent words from written sources than from oral sources where these words are much rarer.

Another factor that may be at play in the trends noted in these data has already been alluded to. In the sample of learners used in this study, low language level is also associated with young age. Children as young as seven are only just coming to grips with orthography in any language, so it is not surprising, perhaps, if low language level associates with a preference for phonological word knowledge. This possibility would seem likely to change as age and orthographic ability both increase. It is not clear from these data whether the effect of language

level on the relationship between orthographic and phonological vocabulary size will stay the same once the age factor is discounted.

A final factor that may contribute to this trend is the effect of teaching and learning preferences. The Greek learners in this study undergo a course teaching that has a particular focus on introducing reading in English at even the earliest stages. The parents want this emphasis, the pupils expect it, and society values this skill quite highly. The Arabic-speaking groups may not have had all these advantages. Among the Saudi learners, teacher preference appeared to be to introduce the taxing elements of English reading and writing rather later in the syllabus, and this approach may reflect the current state of pedagogical practice in Saudi generally (Al-Akloby, 2001). The trend noted in this study may simply be a reflection of what the two groups have been taught, rather than an inevitable result of the nature of a developing lexicon.

Conclusion

The concerns described at the outset of this paper – that vocabulary-size tests delivered in written form only may underestimate the vocabulary knowledge of Arabic native-speaking learners of EFL – appear to have been allayed. The results of this study suggest that while the Arabic-speaking learners do have a proportionately greater knowledge of aural word knowledge than their Greek counterparts, it does not exceed their knowledge of the written form of words. A written form of a vocabulary test is unlikely systematically to under-report their knowledge or give a misleading impression of overall language ability. This means we may have to seek a new explanation for why native Arabic speakers can be so good in speech with comparatively few lexical resources.

The study has also suggested that there may be an interesting relationship between level of language ability and the proportion of words in the foreign language learners' lexicon that are known in their written form only. Both Arabic and Greek native-speaking learners appeared to favour phonological knowledge of words at the outset of learning and to develop an ability to recognize the written form of words disproportionately thereafter. Very good learners of English are particularly characterized by having a phonological recognition vocabulary much smaller than their written vocabulary recognition. Perhaps this is a necessary condition of developing a large vocabulary. This effect may be the result of age factors that were not controlled for in this study or may be the result of the language exposure the learners have received.

James Milton worked as head of English in both Tudun Wada Teachers College, Nigeria, and the University of Sebha, Libya, and has been head of Applied Linguistics at the University of Wales Swansea since 1986. In addition to research in vocabulary learning and testing, he also publishes ELT learning materials with Express Publishing.

Nicola Hopkins has worked as a tutor in the Centre for Applied Language Studies in Swansea University since 2002 and previously worked in journalism and teaching in France. She currently mixes teaching English in Swansea with research in the areas of vocabulary assessment and the semiotics of cinema.

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425–438.

Al-Akloby, S.A.A. (2001). *Teaching and learning English vocabulary in Saudi Arabian public schools.* Unpublished doctoral dissertation, University of Michigan, Ann Arbour.

Al-Hazemi, H. (1993). *Low-level EFL vocabulary tests for Arabic speakers.* Unpublished doctoral dissertation, University of Wales, Swansea.

Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & Van de Velde, H. (2001). Examining the Yes/No vocabulary test: Some methodological issues in theory and practice. *Language Testing, 18*, 235–274.

Council of Europe. (2001) *Common European framework of reference for languages.* Cambridge: Cambridge University Press.

Fountain, R.L., & Nation, I.S.P. (2000). A vocabulary-based graded dictation test. *RELC Journal: Guidelines, 1,* 76–80.

Haynes, M. (1984). Patterns and perils of guessing in second language reading. In J. Handscombe, R.A. Orem, and B.P. Taylor (Eds.), *On TESOL '83* (pp. 163–176). Washington, DC: TESOL.

Hindmarsh, R. (1980). *Cambridge English lexicon.* Cambridge: Cambridge University Press.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4,* 142–151.

Meara, P., & Milton, J. (2003). *The Swansea levels test.* Newbury: Express.

Milton, J. (2005). Vocabulary size testing in the Arabic-speaking world. In P. Davison, P.C. Coombe, & W. Jones (Eds.), *Assessment in the Arab World* (pp. 337–353). Dubai, UAE: TESOL Arabia.

Milton, J. (2006). Language lite: Learning French vocabulary in school. *Journal of French Language Studies, 16*(2), 1–19.

Milton, J., & Hopkins, N. (2005). *Aural lex.* Swansea: University of Wales Swansea.

Nation, I.S.P. (1983). Testing and teaching vocabulary. *Guidelines, 5,* 12–25.

Nation, I.S.P. (Ed.). (1984). *Vocabulary lists: Words, affixes and stems.* English University of Wellington, New Zealand: English Language Institute.

Nation, I.S.P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Randall, M. (1990). *Recognising words in English and Arabic.* Unpublished doctoral dissertation, Birkbeck College, University of London.

Ryan, A.M.G. (1997). Learning the orthographical form of L2 vocabulary: A receptive and productive process. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 181–198). Cambridge: Cambridge University Press.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55–89.

Schonell, F.J., Middleton, I.G., & Shaw, B.A. (1956). *A study of the oral vocabulary of adults.* Brisbane: University of Queensland Press.