

Chapter 8

Vocabulary Recognition Skill as a Screening Tool in English-as-a-Lingua-Franca University Settings

Thomas Roche, Michael Harrington, Yogesh Sinha, and Christopher Denman

Abstract Research has shown that vocabulary recognition skill is a readily measured and relatively robust predictor of second language performance in university settings in English-speaking countries. This study builds on that research by developing an understanding of the relationship between word recognition skill and Academic English performance in English-medium instruction (EMI) university programs in English-as-a-lingua-franca (ELF) contexts. The use of a Timed Yes/No (TYN) test of vocabulary recognition skill was assessed as a screening tool in two EMI university foundation programs in an Arab Gulf State: in a metropolitan state university ($N=93$) and a regional private institution ($N=71$). Pearson correlation coefficients between the TYN test and performance on university placement and final test scores ranged between 0.3 and 0.6 across the two groups and by gender within those groups. This study indicates the TYN test measures have predictive value in university ELF settings for screening purposes. The trade-off between validity, reliability, usability and the cost-effectiveness of the TYN test in academic ELF settings are discussed with consideration of test-takers' digital literacy levels.

Keywords Screening • Academic English • Placement testing • Vocabulary recognition • English as a Lingua Franca • Digital literacy

T. Roche (✉)

SCU College, Southern Cross University, Lismore, NSW, Australia
e-mail: Thomas.Roche@scu.edu.au

M. Harrington

School of Languages and Cultures, University of Queensland, Brisbane, Australia
e-mail: m.harrington@uq.edu.au

Y. Sinha

Department of English Language Teaching, Sohar University, Al Sohar, Oman
e-mail: Yogesh@soharuni.edu.om

C. Denman

Humanities Research Center, Sultan Qaboos University, Muscat, Oman
e-mail: denman@squ.edu.om

1 Introduction

With increasing numbers of English as second language users studying in English-medium instruction (EMI) university programs globally,¹ a growing number of studies have focused on the relationship between students' English language proficiency and academic performance. While Bayliss and Ingram's (2006) review reports on a number of studies that did not find a significant relationship between academic English proficiency and tertiary level academic performance (e.g. Cotton and Conrow 1998), a large number of studies have shown a significant positive relationship (albeit sometimes weak or moderate) between the two: in English-medium tertiary programs in English-speaking countries (Elder et al. 2007; Humphreys et al. 2012; Wang et al. 2008); and, in English-medium university programs in countries where English is considered a foreign language (Roche and Harrington 2013; Yushau and Omar 2007). These latter settings, where English plays an important economic, social or cultural (here educational) role, are increasingly referred to as English as a Lingua Franca (ELF) contexts (Jenkins 2007, 2012; Seidlhofer 2011), which are discussed more extensively by Read (Chap. 11, this volume). Regardless of the exact nature of the link between proficiency and performance it is widely recognised that English proficiency plays an important role in academic outcomes.² Universities in ELF settings face a fundamental challenge to ensure that incoming students have the necessary academic English skills for success in English-medium programs of study. Pre-entry foundation programs play a particularly important role in this process, especially in ELF contexts where English education in high and middle schools can be limited, and where English is not frequently used in wider society.

Placement tests are typically used in pre-entry programs to identify the level and type of English language support students need before they can undertake English-medium university study. These tests usually assess all four domains (reading, writing, listening, speaking) or subsets of these. For example, the English Placement Test (EPT) at the University of Illinois at Urbana-Champaign (UIUC) includes oral tests, as well as a reading and writing task, to provide an accurate placement (or exemption) for international students into English for Academic Purposes (EAP) writing and pronunciation classes (Kokhan 2012). Although effective, the

¹The United Nations Educational, Scientific and Cultural Organization (UNESCO) predicts that the number of internationally mobile students will increase from 3.4 million in 2009 to approximately 7 million by 2020, with a minimum of 50% of these students (some 3.5 million students) undertaking English language education (UNESCO 2012 in Chaney 2013). English is currently the most frequently used language of instruction in universities around the globe (Ammon 2006; Jenkins 2007; Tilak 2011).

²Research also points to the importance of factors such as social connections (Evans and Morrison 2011), cultural adjustment (Fiocco 1992 cited in Lee and Greene 2007) and students' understanding of and familiarity with the style of teaching (Lee and Greene 2007) as significantly contributing to students' academic success in English-medium university programs.

comprehensive nature of such tests makes them extremely time- and resource-intensive, and the feasibility of this approach for many ELF settings is questionable.

The relative speed with which vocabulary knowledge can be measured recommends it as a tool for placement decisions (Bernhardt et al. 2004). Although the focus on vocabulary alone may seem to be too narrow to adequately assess levels of language proficiency, there is evidence that this core knowledge can provide a sensitive means for discriminating between levels of learner proficiency sufficient to make reliable placement decisions (Harrington and Carey 2009; Lam 2010; Meara and Jones 1988; Wesche et al. 1996); in addition, a substantial body of research shows that vocabulary knowledge is correlated with academic English proficiency across a range of settings (see Alderson and Banerjee 2001 for a review of studies in the 1980-1990s). The emergence of the lexical approach to language learning and teaching (McCarthy 2003; Nation 1983, 1990) reflects the broader uptake of the implications of such vocabulary knowledge research in second language classrooms.

Vocabulary knowledge is a complex trait (Wesche and Paribakht 1996). The size of an L2 English user's vocabulary, i.e. the number of word families they know, is referred to as *breadth* of vocabulary knowledge in the literature (Wesche and Paribakht 1996; Read 2004). Initial research in the field (Laufer 1992; Laufer and Nation 1999) suggested that knowledge of approximately 3000 word families provides L2 English users with 95 % coverage of academic texts, a sufficient amount for unassisted comprehension. More recent research has indicated that knowledge of as many as 8000–9000 word families, accounting for 98 % coverage of English academic texts, is required for unassisted comprehension of academic English material (Schmitt et al. 2011, 2015). Adequate vocabulary breadth is a necessary but not a sufficient precondition for comprehension.³

For L2 users to know a written word they must access orthographic, phonological, morphological and semantic knowledge. Studies of reading comprehension in both L1 (Cortese and Balota 2013; Perfetti 2007) and L2 (Schmitt et al. 2011) indicate that readers must first be able to recognize a word before they can successfully integrate its meaning into a coherent message. Word recognition performance has been shown in empirical studies to correlate with L2 EAP sub-skills: reading (Schmitt et al. 2011; Qian 2002); writing (Harrington and Roche 2014; Roche and Harrington 2013); and listening (Al-Hazemi 2001; Stæhr 2009). In addition, Loewen and Ellis (2004) found a positive relationship between L2 vocabulary knowledge and academic performance, as measured by grade point average (GPA), in English medium-university programs, with a breadth measure of vocabulary knowledge accounting for 14.8 % of the variance in GPA.

In order to engage in skilled reading and L2 communication, (i.e. process spoken language in real time) learners need not only the appropriate breadth of vocabulary, but also the capacity to access that knowledge quickly (Segalowitz and Segalowitz 1993; Shiotsu 2001). Successful text comprehension requires lower level linguistic

³Other factors have also been shown to affect reading comprehension, such as background knowledge (e.g., Pulido, 2004 in Webb and Paribakht 2015).

processes (e.g. word recognition) to be efficient, that is, fast and with a high degree of automaticity, providing information to the higher level processes (Just and Carpenter 1992). For the above reasons a number of L2 testing studies (e.g. Roche and Harrington 2013) have taken response time as an index of L2 language proficiency. In this study we focus on word recognition skill, as captured in a test of written receptive word recognition (not productive or aural knowledge) measuring the breadth of vocabulary knowledge and the ability to access that knowledge without contextual cues in a timely fashion.

2 The Study

2.1 *Motivation and Rationale*

Vocabulary recognition screening research to date has largely been undertaken in countries where English is spoken both in and outside of the university classroom, such as in the Screening phase of the Diagnostic English Language Needs Assessment (DELNA) at the University of Auckland (Elder and von Randow 2008). Of note here is that L2 English vocabulary knowledge research involving university student participants in EMI programs in China (Chui 2006), and Oman (Roche and Harrington 2013) indicates that students in such ELF university settings have comparatively lower levels of vocabulary knowledge than their L2 peers in English-medium university programs in countries traditionally considered L1 countries. The current study addresses a gap in the literature by assessing the use of a vocabulary recognition knowledge test as a screening tool in ELF contexts with low proficiency users.

The aim of the current study is to establish the sensitivity of a Timed Yes/No (TYN) test of English recognition vocabulary skill as a screening tool in two university English-medium foundation programs in the Arab Gulf State of Oman. The TYN vocabulary knowledge test assesses both breadth and the speed of access (i.e. the time test-takers needed to access that knowledge). As a follow-up question we also examine the extent to which vocabulary recognition skill predicted overall semester performance in the two English medium Foundation programs as reflected in Final Test scores. While the primary focus is on the development of a potential screening tool, the follow-up data also provides insight into the contribution of vocabulary recognition skill to academic performance.

The current study was undertaken at two university-managed and delivered foundation programs leading to English-medium undergraduate study in Oman. Both universities assess students' English language proficiency after admission to the institution but prior to program enrolment through in-house tests to determine how much English language development, if any, those applicants are likely to need (e.g. one, two or three semesters) prior to matriculation in credit courses of study (e.g. Bachelor of Engineering). At both institutions, students who can provide evidence

of English language proficiency comparable to an overall IELTS score of 5.0 can be admitted directly into an academic credit program. Typically, over 1000 students sit comprehensive English proficiency tests prior to the start of the academic year at each institution, with students then referred to one of four options: English support in a foundation program at beginner, pre-intermediate, or intermediate level; or direct entry into an award program based on their results. Performance indicators leading to placement in those respective EAP levels are determined by test designers at each institution. Outcomes for the foundation programs are specified by the national accreditation agency (Oman Academic Accreditation Authority 2008) and are assessed through exit tests at the end of each semester. The foundation programs also cover other content, such as IT and maths skills, which are not discussed here.

As noted in the Introduction, comprehensive placement tests measuring sub-skills (reading, writing, listening and speaking) which provide accurate placement (or exemption) for students into foundation-level EAP programs are time- and resource- intensive. If proven effective, the TYN test assessed here may serve as an initial tool for screening students, identifying those who have the necessary English to be admitted directly to award study, and indicating for others appropriate levels of English support necessary to reach award-level study. This screening could then be followed by more comprehensive diagnostic testing as part of on-course assessment in both foundation EAP programs, and compulsory award-level English language support units. The context- independent nature of the test also provides a readily employable method for benchmarking the proficiency of students from each institution for comparison with peers studying in English at other institutions nationally and internationally.

The study aims to:

1. Identify the relationship between recognition vocabulary skill measures (size and speed) and Placement as well as Final Test scores;
2. Assess those measures as predictors alone and in combination; and,
3. Evaluate the usability of the measures for administration and scoring in this ELF setting.

2.2 *Participants*

Participants in this study ($N=164$) were Arabic L1 using students enrolled in the English language component of general foundation programs at two institutions. They were 17–25 years old. The programs serve as pathways to English-medium undergraduate study at University A, a metropolitan national university ($N=93$), and University B, a more recently established regional private university ($N=71$) in Oman.

The primary data collection (TYN test, Screening Test) took place at the start of a 15-week semester at the beginning of the academic year. Students' consent to take

part in the voluntary study was formally obtained in accordance with the universities' ethical guidelines.

2.3 *Materials*

Recognition vocabulary skill was measured using an on-line TYN screening test. Two versions of the test were used, each consisting of 62 test items. Items are words drawn from the 1,001st-4,000th most frequently occurring word families in the British National Corpus (BNC) in Part A; and from the 1 K, 2 K, 3 K, and 5 K frequency bands in Test B. Test A therefore consists of more commonly occurring words, while Test B includes lower frequency 5 K items thought to be facilitative in authentic reading (Nation 2006), and an essential part of academic study (Roche and Harrington 2013; Webb and Paribakht 2015). The composition of the TYN tests used here differed from earlier TYN testing research by the authors (Roche and Harrington 2013; Harrington and Roche 2014), which incorporated a set of items drawn from even lower frequency bands, i.e., the 10 K band. Previous research (Harrington 2006) showed that less proficient learners found lower frequency words difficult, with recognition performance near zero for some individuals. In order to make the test more accessible to the target group, the lowest frequency band (i.e. 10 k) was excluded. During the test, items are presented individually on a computer screen. The learners indicate via keyboard whether they know each test item: clicking the right arrow key for 'yes', or the left arrow key for 'no'. In order to control for guessing, the test items consist of not only 48 real word prompts (12 each from four BNC frequency levels) but also 12 pseudoword prompts presented individually (Meara and Buxton 1987). The latter are phonologically permissible strings in English (e.g. *stoffels*). The TYN test can be administered and scored quickly, and provides an immediately generated, objectively scored measure of proficiency that can be used for placement and screening purposes.

Item accuracy and response time data were collected. Accuracy (a reflection of size) was measured by the number of word items correctly identified, minus the number of pseudowords the participants claimed to know (Harrington 2006). Since this corrected for guessing score can result in negative values, 55 points were added to the total accuracy score (referred to as the *vocabulary score* in this paper). Participants were given a short 12-item practice test before doing the actual test. Speed of response (referred to as *vocabulary speed* here) for individual items was measured from the time the item appeared on the screen until the student initiated the key press. Each item remained on the screen for 3 s (3000 ms), after which it was timed out if there was no response. A failure to respond was treated as an incorrect response. Students were instructed to work as quickly and as accurately as possible. Instructions were provided in a video/audio presentation recorded by a local native speaker using Modern Standard Arabic. The test was administered using *LanguageMap*, a web-based testing tool developed at the University of Queensland, Australia.

Placement Tests. Both of the in-house placement tests included sections on reading, writing and listening, using content that was locally relevant and familiar to the students. A variety of question types, including multiple-choice, short answer, as well as true-and-false items, were used.

Final Test. Overall performance was measured by results on end of semester exams. These determine whether students need to take another semester of EAP, or if they are ready to enter the undergraduate university programs. At both institutions the Final Test mirrors the format of the Placement Tests. All test items are based on expected outcomes for the foundation program as specified by the national accreditation agency (Oman Academic Accreditation Authority 2008).

2.4 Procedure

Placement tests were administered in the orientation week prior to the start of the 15-week semester by university staff. The TYN tests were given in a computer lab. Students were informed that the test results would be added to their record along with their Placement Test results, but that the TYN was not part of the Placement Test. The test was introduced by the first author in English and then video instructions were given in Modern Standard Arabic. Three local research assistants were also present to explain the testing format and guide the students through a set of practice items.

The Final Tests were administered under exam conditions at both institutions at the end of the semester. Results for the Placement and Final Tests were provided by the Heads of the Foundation programs at the respective universities with the participants' formal consent. Only overall scores were provided.

3 Results

3.1 Preliminary Analyses

A total of 171 students were in the original sample, with seven removed for not having a complete set of scores for all the measures. Students at both institutions completed the same TYN vocabulary test but sat different Placement and Final Tests depending on which institution they attended. As such, the results are presented by institution. The TYN vocabulary tests were administered in two parts, I and II, to lessen the task demands placed on the test-takers. Each part had the same format but contained different items (see above). Reliability for the vocabulary test was measured using Cronbach's alpha, calculated separately for the two parts. Analyses were done by word (N=48), and pseudoword (N=14) items for the vocabulary score and vocabulary speed measures (see Harrington 2006). Each response type is

assumed to represent a different performance dimension. For vocabulary score performance on words, the reliability coefficients were: Part A = 0.75 and Part B = 0.77; for pseudowords, Part A = 0.74 and Part B = 0.73. Vocabulary speed response reliability for the words was Part A = 0.97 and Part B = 0.97, and for the pseudowords, Part A = 0.92 and Part B = 0.91. This indicates that test-taker response times were acceptably consistent on items within each test version. The reliability coefficients are satisfactory for both the vocabulary score and vocabulary speed measures.

Prior to the main analysis, performance on the respective parts was compared for a possible order effect. This was done first for the vocabulary size scores and then the vocabulary speed measures. For University A, the mean vocabulary scores (and standard deviations) were: Part I = 99.80 (18.93) and Part II = 98.87 (16.19). A paired t-test indicated that the difference in the means was not significant $t(92) = 0.338$, $p = 0.736$ (two-tailed). For University B, the statistics were: Part I = 81.20 (33.26) and Part II = 78.03 (26.97). Again the difference was not significant, $t(70) = 0.81$, $p = 0.417$. Given the lack of difference between the mean scores on the two parts, they were averaged into a single score for the respective groups.

The vocabulary speed data is reported in milliseconds by mean (and standard deviation). For University A, the figures for Part I were 1393 (259) and for Part II, 1234 (222). The 160 ms. difference between the two parts was significant, $t(92) = 8.45$ $p < 0.001$ (two-tailed), $d = 0.86$; with the latter statistic indicating a large effect for order. A similar pattern emerged for University B: Part I = 1543 (396) and Part II = 1415 (356). The mean difference was significant $t(70) = 4.18$, $p < 0.001$, $d = 0.34$. The effect size here is in the small to medium range. Response time performance was significantly faster in Part II for both groups. However, to facilitate the presentation of the placement finding, an average of the two vocabulary speed measures will be used in parallel with the vocabulary breadth measures. The vocabulary speed differences will be addressed in the Discussion.

The descriptive statistics for the test scores and vocabulary measures are presented in Table 8.1. Vocabulary measures consist of accuracy, response time and false alarms, the latter being the percentage of incorrect ('Yes') responses to pseudowords. The individual's false alarm rate is used to adjust the overall accuracy score by subtracting the proportion of 'Yes' responses to pseudowords from the overall proportion of 'Yes' responses to words (Harrington 2006). The false alarm rate alone also indicates the extent to which the test-takers were prone to guessing. An initial analysis showed a high false alarm rate for the University B students, which in turn differed by gender. As a result, the descriptive statistics are presented both by university and gender.

Table 8.1 presents the descriptive statistics for the TYN vocabulary, placement test and false alarm scores by university and gender. At both universities the male respondents typically underperformed when compared with their female peers. At University B, the regional private university, female participants had slightly better scores on all measures when compared with their male peers. For University A, the metropolitan state university, participants performed better on the vocabulary test measures (vocabulary score, vocabulary speed, false alarms) than University B participants.

Table 8.1 Descriptive statistics for TYN vocabulary, vocabulary speed, placement test and false alarm scores by university and gender

			Range		Score		Score confidence intervals (95%)	
			Low	High	<i>M</i>	<i>SD</i>	Lower	Upper
Vocabulary test (Total points = 155)	Uni A	Female n=63	72	128	98.97	11.56	96.05	101.88
		Male n=30	76	125	101.16	18.12	95.75	104.47
		Total N=93			99.34	11.55	96.96	101.71
	Uni B	Female n=37	27	122	88.75	21.52	81.57	95.92
		Male n=34	14	119	69.68	25.99	60.61	78.75
		Total N=71			79.62	25.47	73.59	85.64
Response time	Uni A	Female	967	1910	1335	213	1281	1388
		Male	739	1648	1270	214	1179	1360
		Total			1314	223	1268	1360
	Uni B	Female	953	2336	1390	294	1292	1488
		Male	836	2185	1575	391	1439	1712
		Total			1479	354	1395	1563
False alarms	Uni A	Female	4	68	28.69	14.24	25.10	32.27
		Male	0	64	23.69	13.96	18.48	28.90
		Total			27.07	14.27	24.13	30.01
	Uni B	Female	00	82	32.24	20.75	25.32	39.16
		Male	11	86	49.69	19.47	42.98	56.48
		Total			40.59	21.84	35.42	45.77
Placement score % (Uni specific)	Uni A	Female	41	65	54.93	5.64	53.26	56.60
		Male	39	65	54.10	7.07	51.46	56.74
		Total			54.66	6.75	53.27	56.05
	Uni B	Female	3	90	46.65	27.54	37.47	55.83
		Male	0	78	24.50	26.53	15.24	33.76
		Total			36.04	29.08	29.16	42.93
End of semester test % (Uni specific)	Uni A	Female	60	86	75.80	5.35	74.44	78.00
		Male	62	84	75.15	5.64	73.18	77.12
		Total			75.59	5.43	74.48	77.71
	Uni B	Female	35	86	73.37	10.99	70.50	76.24
		Male	21	85	61.18	12.86	58.18	64.17
		Total			67.53	13.33	64.37	70.68

Uni A University A, a national public university; *Uni B* University B, a private regional university

3.2 *False Alarm Rates*

The false alarm rates for both University A males and females (24% and 29%, respectively), as well as the University B females (32%) were comparable to the mean false alarm rates of Arab university B students (enrolled in first and fourth year of Bachelor degrees) reported in previous research (Roche and Harrington 2013). They were also comparable to, though higher than, the 25% for beginners and 10% for advanced learners evident in the results of pre-university English-language pathway students in Australia (Harrington and Carey 2009). As with other TYN studies with Arabic L1 users in academic credit courses in ELF university contexts (Roche and Harrington 2013; Harrington and Roche 2014), there were students with extremely high false alarm rates that might be treated as outliers. The mean false alarm rate of nearly 50% by the University B males here goes well beyond this. The unusually high false alarm rate for this group, and the implications it has for the use of the TYN test for similar populations, will be taken up in the Discussion.

The extremely high false alarm level for the University B males and some individuals in the other groups means the implications for Placement and Final Test performance must be interpreted with caution. As a result, two sets of tests evaluating the vocabulary measures as predictors of test performance were performed. The tests were first run on the entire sample and then on a trimmed sample in which individuals with false alarm rates that exceeded 40% were removed. The latter itself is a very liberal cut-off level, since other studies have removed any participants who incorrectly identified pseudowords at a rate as low as 10% (Schmitt et al. 2011). The trimming process reduced the University A sample size by 16%, representing 19% fewer females and 10% fewer males. The University B sample was reduced by over half, reflecting a reduction of 17% in the total for the females and a very large 68% reduction for the males. It is clear that the males in University B handled the pseudowords in a very different manner than either the University B females or both the genders in University A, despite the use of standardised Arabic-language instructions at both sites.

The University A students outperformed the University B students on the vocabulary size and vocabulary speed measures. Assumptions of equality of variance were not met, so an Independent-Samples Mann-Whitney U test was used to test the mean differences for the settings. Both vocabulary score and mean vocabulary speed scores were significantly different: $U=4840$, $p=0.001$, $U=6863$, $p=0.001$, respectively. All significance values are asymptotic, in that the sample size was assumed to be sufficient statistically to validly approximate the qualities of the larger population. For gender there was an overall difference on test score $U=4675$, $p=0.022$, but not on vocabulary speed, $U=7952$ ($p=0.315$). The gender effect for test score reflects the effect of low performance by the University B males. A comparison of the University A males and females showed no significant difference between the two, while the University B female vocabulary scores ($U=948$, $p=0.001$) and vocabulary speed ($U=1155$) ($p=0.042$) were both significantly higher than those for their male peers.

Table 8.2 Bivariate correlations between vocabulary and test measures for University A and University B, complete data set

University A n=93	University	Vocab speed	Placement test	Final test
University B n=71				
Vocabulary size	A	-0.12	0.34**	0.10
	B	-0.59**	0.65**	0.50**
Vocabulary speed	A	-	-0.11	0.04
	B	-	-0.48**	-0.33**
Placement test	A	-	-	0.08
	B	-	-	0.53**

** $p < 0.001$ All tests two-tailed

Table 8.3 Bivariate correlations between vocabulary and test measures, data set trimmed for high false alarm values (false alarm rates >40 % removed)

University A n=78	University	Vocab speed	Placement test	Final test
University B n=38				
Vocabulary size	A	-0.13	0.37**	0.10
	B	-0.37*	0.34*	0.24
Vocabulary speed	A	-	-0.34*	0.03
	B	-	-0.31	-0.19
Placement test	A	-	-	0.12
	B	-	-	0.27

* $p < 0.05$, ** $p < 0.001$, All tests two-tailed

3.3 Vocabulary Measures as Predictors of Placement and Final Test Performance

The sensitivity of the vocabulary measures as predictors of placement and final test performance was evaluated first by examining the bivariate correlations among the measures and then performing a hierarchical regression to assess how the measures interacted. The latter permits the respective contributions of size and speed to be assessed individually and in combination. Table 8.2 presents the results for the entire data set. It shows the vocabulary measures were better predictors for University B. These results indicate that more accurate word recognition skill had a stronger correlation with Placement Test Scores and Final Test scores for University B than for University A.

When the data are trimmed for high false-alarm rates, the difference between the two universities largely disappears (see Table 8.3). The resulting correlation of approximately 0.35 for both universities shows a moderate relationship between vocabulary recognition skill test scores and Placement Test performance. In the trimmed data there is no relationship between TYN vocabulary recognition test scores and Final Test scores.

Regression models were run to assess how much overall variance the measures together accounted for, and the relative contribution of each measure to this amount.

Table 8.4 Hierarchical regression analyses of placement test scores with vocab score and speed as predictors for complete and trimmed data sets

Predictor	R^2	Adjusted R^2	R^2 change	B	SEB	β
Complete data set						
University A						
Vocab score	0.117	0.107	0.117*****	0.200	0.058	0.334**
Vocab speed	0.121	0.102	0.005	-6.051	8.813	-0.068
University B						
Vocab score	0.419	0.411	0.419*****	0.739	0.130	0.567**
Vocab speed	0.432	0.415	0.012	0.642	32.14	-0.136
Trimmed data set						
University A						
Vocab score	0.134	0.123	0.134*****	0.212	0.064	0.359**
Vocab speed	0.139	0.116	0.005	-5.889	9.276	-0.068
University B						
Vocab score	0.111	0.087	0.111***	0.390	0.261	0.251
Vocab speed	0.152	0.104	0.031	-6.227	50.048	-0.219

t significant at * $p < 0.05$, ** $p < 0.001$. F significant at *** $p < 0.05$ and **** $p < 0.001$

Table 8.4 reports on the contribution of vocabulary speed to predicting the Placement Test score criterion after the vocabulary scores were entered. Separate models were calculated for University A and University B.

As expected from the bivariate correlations, the regression analysis shows that the word measures and reaction time predicted significant variance in Placement Test scores. The University A model accounted for nearly 12% of the Placement Test score variance while the University B one accounted for over 40%.

Table 8.5 shows the ability of the vocabulary score (and speed) to predict Final Test performance. The vocabulary measures served as predictors of the Final Test for University B, where there was a moderate correlation (0.5) between the vocabulary scores and overall English proficiency scores at the end of the semester. There was no significant correlation with TYN Test scores and Final Test scores at University A, but it is of note that the Final Test scores for this group had a truncated range, which may reflect the higher academic entry requirements and concomitant English proficiency levels of students at the national metropolitan university as discussed below in 4.1. The results not only indicate that the TYN word recognition test is a fair predictor of performance, but also reinforce the importance of the vocabulary knowledge. The results of the study at University B are a confirmation of the significant role the vocabulary knowledge of L2 English users plays in achieving success in higher education contexts, given what is already known about the importance of other factors such as social connections (Evans and Morrison 2011), cultural adjustment (Fiocco 1992 cited in Lee and Greene 2007) and students' understanding of and familiarity with the style of teaching (Lee and Greene 2007).

The regression analyses for Final Test scores also reflect the results of the bivariate correlations: the word measures and vocabulary speed predicted significant vari-

Table 8.5 Hierarchical regression analyses of final test scores with vocab score and speed as predictors for complete and trimmed data sets

Predictor	R^2	Adjusted R^2	R^2 change	B	SEB	β
Complete data set						
University A						
Vocab score	0.117	0.107		0.200	0.058	0.334**
Vocab speed	0.121	0.102	0.005	-6.05	8.81	-0.068
University B						
Vocab score	0.251	0.240		0.248	0.068	0.567**
Vocab speed	0.252	0.230	0.001	-5.881	16.891	-0.136
Trimmed data set						
University A						
Vocab score	0.010	-0.004	0.010	0.052	0.058	0.102
Vocab speed	0.011	-0.015	0.005	3.048	8.504	0.041
University B						
Vocab score	0.059	0.033	0.111	0.096	0.058	0.203
Vocab speed	0.068	0.015	-10.754	18.374	8.81	-0.103

t significant at * $p < 0.05$, ** $p < 0.001$

ance in the Final Test results. The model based on the Vocabulary test measure accounted for nearly 22 % of the Final Test variance (total adjusted $R^2 = 0.219$) while the vocabulary speed word model accounted for only 9 % (0.094).

4 Discussion

The findings are consistent with previous work that indicates vocabulary recognition skill is a stable predictor of academic English language proficiency, whether in academic performance in English-medium undergraduate university programs in ELF settings (Harrington and Roche 2014; Roche and Harrington 2013), or as a tool for placement decisions in English-speaking countries (Harrington and Carey 2009). The current study extends this research, showing the predictive power of a vocabulary recognition skill test as a screening tool for English-language university foundation programs in an ELF context. It also identifies several limitations to the approach in this context.

4.1 The TYN Test as a Placement Test for Low Proficiency Learners

Vocabulary recognition skill is a good indicator of student Placement Test performance in the target ELF context. The mid 0.3 correlations between vocabulary and Placement Tests observed for both universities in the trimmed data were at the lower

end of previous findings correlating TYN test performance with academic writing skill, where correlations ranged from 0.3 to 0.5 (Harrington and Roche 2014; Roche and Harrington 2013). Higher correlations were observed at University B when the students with very high false alarm rates were included. Although the inclusion of the high false alarm data improves predictive validity in this instance, it also raises more fundamental questions about this group's performance on the computerised test and therefore the reliability of the format for this group. This is discussed below in 4.2. The vocabulary speed means for the present research were comparable to, if not slightly faster than, those obtained in previous studies (Harrington and Roche 2014; Roche and Harrington 2013). Mean vocabulary speed was, however, found to be a less sensitive measure of performance on Placement Tests, in contrast to previous studies where vocabulary speed was found to account for a unique amount of variance in the criterion variables (Harrington and Carey 2009; Harrington and Roche 2014; Roche and Harrington 2013).

Other TYN studies in university contexts (Harrington and Carey 2009; Roche and Harrington 2013) included low frequency band items from the BNC (i.e. 10 k band), whereas the current test versions did not. Given research highlighting the instrumental role the 8000–9000 most commonly occurring words in the BNC play in reading academic English texts (Schmitt et al. 2015, 2011), it is possible that including such items would improve the test's sensitivity in distinguishing English proficiency levels between students. This remains a question for further placement research with low proficiency English L2 students.

Results show a marked difference in performance between University A and University B on all dimensions. This may reflect differences in academic standing between the groups that are due to different admission standards at the two institutions. University A is a prestigious state institution, which only accepts students who score in the top 5% of the annual graduating high school cohort, representing a score of approximately 95% on their graduating certificate. In contrast, University B is a private institution, with lower entry requirements, typically attracting students who score approximately 70% and higher on their graduating high school certificate. The differences between the two groups indicate their relative levels of English proficiency. It may also be the case that the difference between groups may be due to the digital divide between metropolitan and regional areas in Oman, with better connected students from the capital at University A more digitally experienced and therefore performing better on the online TYN test. This issue is explored further in 4.2.

4.2 *The TYN Test Format*

Participants in the study had much higher mean false-alarm rates and lower group means (as a measure of vocabulary size) than pre-tertiary students in English-speaking countries. A number of authors have suggested that Arabic L1 users are likely to have greater difficulties with discrete-item English vocabulary tests than

users from other language backgrounds due to differences between the Arabic and English orthographies and associated cognitive processes required to read those systems (Abu Rabia and Seigel 1995; Fender 2003; Milton 2009; Saigh and Schmitt 2012). However, as research has shown that word recognition tests do serve as effective indicators of Arabic L1 users' EAP proficiency (Al-Hazemi 2001; Harrington and Roche 2014; Roche and Harrington 2013), this is unlikely to be the reason for these higher rates. As indicated in 2.3 the test format, in particular the difference between words and pseudowords, was explained in instructions given in Modern Standard Arabic. It is possible that some students did not fully understand these instructions.

The comparatively high false-alarm rates at the regional institution may also reflect relatively low levels of digital literacy among participants outside the capital. The TYN test is an on-line instrument that requires the user to first supply biodata, navigate through a series of computer screens of instructions and examples, and then supply test responses. The latter involves using the left and right arrow keys to indicate whether the presented item is a word or a pseudoword. It was noted during the administration of the test at University B that male students (the group with the highest false-alarm rates) required additional support from research assistants to turn on their computers and log-in, as well as to start their internet browsers and enter bio-data into the test interface. As recently as 2009, when the participants in this study were studying at high school, only 26.8% of the nation's population had internet access, in comparison to 83.7% in Korea, 78% in Japan, and 69% in Singapore (World Bank 2013); and, by the time the participants in the present study reached their senior year of high school in 2011, there were only 1.8 fixed (wired) broadband subscriptions per 100 inhabitants in Oman, compared to a broadband penetration of 36.9/100 in Korea, 27.4/100 in Japan, and 25.5/100 in Singapore (Broad Band Commission 2012). Test-takers in previous TYN test studies (Harrington and Carey 2009) had predominantly come from these three highly connected countries and were likely to have brought with them higher levels of computer skills and digital literacy. It is also of note that broadband penetration is uneven across Oman, with regional areas much more poorly served by the telecommunications network than the capital, where this study's national university is located, and this may in part account for the false-alarm score difference between the two universities. Previous studies in Arab Gulf States were with students who had already completed foundation studies and undertaken between 1 and 4 years of undergraduate study; they were therefore more experienced with computers and online testing (Harrington and Roche 2014; Roche and Harrington 2013) and did not exhibit such high false-alarm scores. The current study assumed test-takers were familiar with computing and the Internet, which may have not been the case. With the spread of English language teaching and testing into regional areas of developing nations, it is necessary to be context sensitive in the application of online tests. Improved results may be obtained with an accompanying test-taking video demonstration and increased TYN item practice prior to the actual testing.

The poor performance by some of the participants may also be due to the fact that the test was low-stakes for them (Read and Chapelle 2001). The TYN test was

not officially part of the existing placement test suite at either institution and was administered after the decisive Placement Tests had been taken; high-false alarm rates may be due to some participants giving acquiescent responses (i.e. random clicks that brought the test to an un-taxing end rather than reflecting their knowledge or lack of knowledge of the items presented, see Dodorico-McDonald 2008; Nation 2007). It is therefore important that test-takers understand not only how to take the test but also see a reason for taking it. For example, we would expect fewer false alarms if the test acted as a primary gateway to further study, rather than one of many tests administered as part of a Placement Test suite, or if it was integrated into courses and contributed towards students' marks.

4.3 Gender

An unexpected finding to emerge in the study was the difference in performance due to gender. The role of gender in second language learning remains very much an open question, with support both for and against gender-based differences. Studies investigating performance on discrete item vocabulary tests such as *Lex30* (Espinosa 2010) and the *Vocabulary Levels Test* (Agustín Llach and Terrazas Gallego 2012; Mehrpour et al. 2011) found no significant difference in test performance between genders. Results reported by Jiménez Catalán (2010) showed no significant difference on the receptive tests between the male and female participants, though it was noted that girls outperformed boys on productive tests. The stark gender-based differences observed in the present study are not readily explained. Possible reasons include the low-stakes nature of the test (Nation 2007; Read and Chapelle 2001), other personality or affective variables or, as noted, comparatively lower levels of digital literacy among at least some of the male students.

5 Conclusion

Testing is resource-demanding activity involving a trade-off between the time and money available and the reliability and sensitivity of the testing instruments used. Many in-house university placement tests provide detailed placement (and potentially diagnostic) information about test-takers' English language ability across a range of sub-skills (e.g., reading, speaking, listening and writing). Such tests however, have significant resource implications for institutions, taking a great deal of time to develop, administer and mark. For ELF institutions such as the ones studied here, administering upwards of 1000 placement tests at the start of each academic year, the resource implications are considerable. The TYN test is an attractive screening tool given the limited resources needed for its administration and generation of results, thereby enabling higher-education providers in ELF contexts to use their limited resources more efficiently.

Results here show that the TYN test is a fair measure of English proficiency in tertiary ELF settings, though with some qualification. It may serve an initial screening function, identifying which EAP levels (beginner, pre-intermediate and intermediate) students are best placed in, prior to more comprehensive in-class diagnostic testing, but further research is needed to identify the TYN scores which best identify placement levels. The tests' predictive ability could be improved, potentially through adding lower-frequency test items or adding another component, such as grammar or reading items, to replace the less effective higher-frequency version I of the two vocabulary tests trialled in this study.

The use of the TYN test with lower proficiency learners in a context like the one studied here requires careful consideration. Experience with implementing the test points to the importance of comprehensible instructions and the test-taker's sense of investment in the results. The findings also underscore the context-sensitive nature of testing and highlight the need to consider test-takers' digital literacy skills when using computerised tools like the TYN test. As English continues to spread as the language of tertiary instruction in developing nations, issues of general digital literacy and internet penetration become educational issues with implications for testing and assessment.

Finally, the findings here contribute to a growing body of literature emphasising the fundamental importance of vocabulary knowledge for students studying in ELF settings. In particular, it shows the weaker a student's vocabulary knowledge, the poorer they are likely to perform on measures of their academic English proficiency and subsequently the greater difficulties they are likely to face achieving their goal of completing English preparation courses on their pathway to English-medium higher education study.

Acknowledgements This research was supported by a grant from the Omani Research Council [Grant number ORG SU HER 12 004]. The authors would like to thank the coordinators and Heads of the Foundation Programs at both institutions for their support.

References

- Abu Rabia, S., & Seigel, L. S. (1995). Different orthographies, different context effects: The effects of Arabic sentence context in skilled and poor readers. *Reading Psychology, 16*, 1–19.
- Agustín Llach, M. P., & Terrazas Gallego, M. (2012). Vocabulary knowledge development and gender differences in a second language. *Estudios de Lingüística Inglesa Aplicada, 12*, 45–75.
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (part 1). State-of-the-art review. *Language Testing, 18*, 213–236.
- Al-Hazemi, H. (2001). Listening to the Y/N vocabulary test and its impact on the recognition of words as real or non-real. A case study of Arab learners of English. *IRAL, 38*(2), 81–94.
- Ammon, U. (2006). The language of tertiary education. In E. K. Brown (Ed.), *Encyclopaedia of languages & linguistics* (pp. 556–559). Amsterdam: Elsevier.
- Bayliss, A., & Ingram D. E. (2006). *IELTS as a predictor of academic language performance*. Paper presented at the Australian International Education Conference. Retrieved from <http://www.aiec.idp.com>.

- Bernhardt, E., Rivera, R. J., & Kamil, M. L. (2004). The practicality and efficiency of web-based placement testing for college-level language programs. *Foreign Language Annals*, 37(3), 356–366.
- Broad Band Commission. (2012). *The state of broadband 2012: Achieving digital inclusion for all*. Retrieved from www.broadbandcommission.org.
- Chaney, M. (2013). *Australia: Educating globally*. Retrieved from <https://aei.gov.au/IEAC2/theCouncilsReport>.
- Chui, A. S. Y. (2006). A study of the English vocabulary knowledge of university students in Hong Kong. *Asian Journal of English Language Teaching*, 16, 1–23.
- Cortese, M. J., & Balota, D. A. (2013). Visual word recognition in skilled adult readers. In M. J. Spivey, K. McRae, & M. F. Joannis (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 159–185). New York: Cambridge University Press.
- Cotton, F., & Conrow, F. (1998). *An investigation of the predictive validity of IELTS amongst a sample of international students studying at the University of Tasmania* (IELTS Research Reports, Vol. 1, pp. 72–115). Canberra: IELTS Australia.
- Dodorico-McDonald, J. (2008). Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, 1(1). Retrieved from www.nottingham.ac.uk/shared/McDonald.pdf.
- Elder, C., & von Randow, J. (2008). Exploring the utility of a web-based English language screening tool. *Language Assessment Quarterly*, 5(3), 173–194.
- Elder, C., Bright, C., & Bennett, S. (2007). The role of language proficiency in academic success: Perspectives from a New Zealand university. *Melbourne Papers in Language Testing*, 12(1), 24–28.
- Espinosa, S. M. (2010). Boys' and girls' L2 word associations. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 139–163). Chippenham: Macmillan.
- Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, 30(3), 198–208.
- Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics*, 24(3), 289–316.
- Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. *EUROSLA Yearbook*, 6, 147–168.
- Harrington, M., & Carey, M. (2009). The on-line yes/no test as a placement tool. *System*, 37, 614–626.
- Harrington, M., & Roche, T. (2014). Post-enrolment language assessment for identifying at-risk students in English-as-a-Lingua-Franca university settings. *Journal of English for Academic Purposes*, 15, 37–47.
- Humphreys, P., Haugh, M., Fenton-Smith, B., Lobo, A., Michael, R., & Walkenshaw, I. (2012). Tracking international students' English proficiency over the first semester of undergraduate study. *IELTS Research Reports Online Series*, 1. Retrieved from <http://www.ielts.org>.
- Jenkins, J. (2007). *English as a Lingua Franca: Attitudes and identity*. Oxford: Oxford University Press.
- Jenkins, J. (2012). English as a Lingua Franca from the classroom to the classroom. *ELT Journal*, 66(4), 486–494. doi:10.1093/elt/ccs040.
- Jiménez Catalán, R. M. (2010). Gender tendencies in EFL across vocabulary tests. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 117–138). Chippenham: Palgrave Macmillan.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 98(1), 122–149.
- Kokhan, K. (2012). Investigating the possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing*, 29(2), 291–308.

- Lam, Y. (2010). Yes/no tests for foreign language placement at the post-secondary level. *Canadian Journal of Applied Linguistics*, 13(2), 54–72.
- Laufer, B. (1992). Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading*, 15(2), 95–103.
- Laufer, B., & Nation, I. S. P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, 16(1), 33–51.
- Lee, Y., & Greene, J. (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Methods Research*, 1, 366–389.
- Loewen, S., & Ellis, R. (2004). The relationship between English vocabulary skill and the academic success of second language university students. *New Zealand Studies in Applied Linguistics*, 10, 1–29.
- McCarthy, M. (2003). *Vocabulary*. Oxford: Oxford University Press.
- Meara, P., & Buxton, B. (1987). An alternative multiple-choice vocabulary test. *Language Testing*, 4, 142–145.
- Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 80–87). London: Centre for Information on Language Teaching and Research.
- Mehrpour, S., Razmjoo, S. A., & Kian, P. (2011). The relationship between depth and breadth of vocabulary knowledge and reading comprehension among Iranian EFL learners. *Journal of English Language Teaching and Learning*, 53, 97–127.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Nation, I. S. P. (1983). Learning vocabulary. *New Zealand Language Teacher*, 9(1), 10–11.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes*, 63(1), 59–81.
- Nation, I. S. P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 35–43). Cambridge: Cambridge University Press.
- Oman Academic Accreditation Authority. (2008). *The Oman academic standards for general foundation programs*. Retrieved from <http://www.oac.gov.om/>.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513–536.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 209–227). Amsterdam: John Benjamins.
- Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1), 1–32.
- Roche, T., & Harrington, M. (2013). Recognition vocabulary knowledge as a predictor of academic performance in an English-as-a-foreign language setting. *Language Testing in Asia*, 3(12), 133–144.
- Saigh, K., & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System*, 40, 24–36. doi:10.1016/j.system.2012.01.005.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95, 26–43. doi:10.1111/j.1540-4781.2011.01146.x.
- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2015). How much vocabulary is needed to use English? Replication of Van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 1–15.
- Segalowitz, N., & Segalowitz, S. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word. *Applied Psycholinguistics*, 14, 369–385.

- Seidlhofer, B. (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.
- Shiotsu, T. (2001). Individual differences in L2 word recognition speed: a measurement perspective. *Bulletin of the Institute of Foreign Language Education Kurume University*, 8, 63–77. Retrieved from <http://www.lognostics.co.uk>
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31, 577–607.
- Tilak, J. B. G. (2011). *Trade in higher education: The role of the General Agreement on Trade in Services (GATS)* (Fundamentals of Educational Planning, Vol. 95, p. 154). Paris: UNESCO.
- Wang, L., Eignor, D., & Enright, M. K. (2008). A final analysis. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the test of English as a foreign language* (pp. 259–318). New York: Routledge.
- Webb, S., & Paribakht, T. S. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes*, 38, 34–43.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53, 13–40.
- Wesche, M., Paribakht, T. S., & Ready, D. (1996). A comparative study of four ESL placement instruments. In M. Milanovic & N. Saville (Eds.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium (LTRC), Cambridge and Arnhem* (pp. 199–209). Cambridge: Cambridge University Press.
- World Bank. (2013). *Internet users (per 100 people)*. Retrieved <http://data.worldbank.org/indicator/IT.NET.USER.P2>.
- Yushau, B., & Omar, M. H. (2007). Preparatory year program courses as predictors of first calculus course grade. *Mathematics and Computer Education*, 41(2), 92–108.