

ENGLISH IN INDONESIAN ISLAMIC HIGHER EDUCATION

Examining The Relationship between Performance in The Yes/No Test and Reading Skills

Sahiruddin

The State Islamic University of Malang, Indonesia

Abstract: This study examines the relationship between performance in the Yes/No test of English recognition vocabulary and reading skills in Indonesian Islamic learners of English as a foreign language (EFL). Participants in the study were 83 Indonesian undergraduate students, comprising an Advanced group (n=41) and Intermediate group (n=42) of EFL learners enrolled in the English department at the State Islamic University (UIN) of Malang, Indonesia. All participants completed both tests. The results reveal that the hits accuracy performance between the Advanced EFL group and the Intermediate EFL group was statistically significant, indicating that Yes/No test performance, in context of hits accuracy, did discriminate between levels of English proficiency. However, the differences disappeared with corrected scores since both groups indicated a high false alarm rate. In addition, this study also reveals that there was no evidence of a relationship between Yes/No performance and reading scores. Several pedagogical implications for EFL language teachers are discussed.

Keyword: Islamic higher education, Yes/No Test, reading skills, English learner

Introduction

English is taught as a foreign language and is a compulsory subject in Indonesian education from elementary level to tertiary level. However, since it was first taught, there have been problems in the teaching of English as a foreign language. Changes in the curriculum

and approaches¹ along with the changes in the Ministry of Education have been common. Large class sizes and teachers with poor mastery of English are two other obvious factors that have contributed to ongoing problems in English language teaching (ELT) in Indonesia². Despite these challenges in ELT implementation, the influential role of English as international language in the world community means that English language teaching in Indonesia remains extremely important. Not only is English now taught at elementary level but it is also being taught at an even earlier stage of school, such as in playschool or kindergarten.

In the context of Islamic education, it is also evident that English is now being taught from elementary to tertiary level. What is more, it is found in all categories of education, from formal kindergarten or playgroup, to nonformal education, such as Taman Pendidikan Al-Qur'an (TPA, Islamic kindergarten for Qur'anic studies). Generally, it is expected that those who encounter English in earlier stages of their education become familiar with English and find it easier to learn it in the later periods such as tertiary level. In this way, English language teaching and learning has become vitally important in Indonesian education system.

To maximize the outcome of English learning in Indonesian education, especially at Islamic higher education level, many efforts have been made, such as modifying the curriculum or syllabus, identifying language teaching methods, and designing an appropriate evaluation format. This paper aims at investigating the validity of the Yes/No vocabulary test and examining the relationship between the performance on the Yes/No test and English L2 reading proficiency of Indonesian learners at Islamic higher education. Evidence from the study will be used to evaluate the potential of the Yes/No test as a language testing format that could be used for placement and testing purposes in Indonesian Islamic higher education and other educational institutions in Indonesia. Studies to date have shown that Yes/No test performance can be a stable measure of L2 lexical proficiency. However, these studies have only been conducted in input-rich ESL contexts. The current study will assess Yes/No in a foreign language

¹ S. Dardjowidjojo, "English Teaching in Indonesia," *EA Journal*, 18 (1) (2000): pp. 22-30.

² Ibid.

setting especially in Indonesian Islamic higher education. In addition, even though studies dealing with Yes/No task provide evidence about its validity in measuring vocabulary size of L2 learners³, little attention has been paid to how this Yes/No test performance correlates with global language skills such as reading, speaking, listening, and writing. This study tries to examine to what extent the performance on the Yes/No recognition vocabulary test correlates with reading proficiency. It is argued that reading proficiency is of importance here since this skill has been one of the main goals of English language teaching and learning in Indonesian context, especially in Islamic higher education level as the references or textbooks used during instruction are mostly written in English. The findings will be of potential relevance here.

This study addresses two questions: (a) Does accuracy in the Yes/No test improve as group proficiency and word frequency levels increase?, and (b) To what extent does Yes/No test performance correlate with reading comprehension performance? The subject of analysis in this study was a group of 83 undergraduate students enrolled in the English Department of State Islamic University (UIN) of Malang, East Java, Indonesia. The participants were from the fourth and sixth semester and were classified into two groups, Intermediate (n = 42) and Advanced (n=41) by the program staff on basis of their academic performance and overall language skills.

Theoretical Perspective of Lexical Knowledge and Reading Comprehension

Lexical knowledge is one foundation of overall language proficiency. Words are considered "the building blocks of language" and by some to be "the single most important aspect of foreign language learning"⁴. Over the past two decades the field of second language acquisition has seen considerable emphasis on the role of lexical knowledge in second language (L2) performance and vocabulary

³ A. Mochida and M. Harrington, "The Yes–No Test as a Measure of Receptive Vocabulary Knowledge," *Language Testing*, 26(1) (2006): pp. 73–98.

⁴ S. Knight, "Dictionary Use While Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities," *Modern Language Journal*, 78 (3) (1994): pp. 285-299.

acquisition research.⁵ One area of debate centers on the extent to which vocabulary knowledge can predict language performance in other domains.⁶

Word knowledge has long been recognized as an important factor in language proficiency⁷, and especially in reading comprehension⁸. Laufer⁹ asserts that vocabulary knowledge correlates with holistic assessments of writing and general proficiency, and is the best single predictor of reading comprehension. In particular, a number of studies have shown that performance on tests of breadth of vocabulary knowledge can be an excellent predictor of success in reading, writing, general proficiency and academic achievement.¹⁰

Moreover, research consistently demonstrates that vocabulary knowledge correlates more highly with reading comprehension than other factors, including morphosyntactic knowledge¹¹ and reading strategies.¹² In other words, the learners who have a larger vocabulary are assumed to have better text comprehension. Despite overwhelming evidence of this strong connection, however, there is little consensus as to the exact relationship between the two. Traditionally vocabulary has been viewed as the dominant enabling factor. A more recent view, however, suggests that both are mutually interdependent during their

⁵ I.S.P Nation, *Learning Vocabulary in Another Language* (Cambridge: Cambridge University Press, 2001).

⁶ D. D. Qian, "Investigating the Relationship between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective," *Language Learning*, 52 (3) (2002): pp. 513-536.

⁷ W. Grabe, "Current Developments in Second Language Reading Research," *TESOL Quarterly*, 25 (1991): pp. 375-406.

⁸ D.D. Qian, "Assessing the Role of Depth and Breadth of Vocabulary Knowledge in Reading Comprehension," *Canadian Modern Language Review*, 56 (1999): pp. 282-307.

⁹ B. Laufer, "The Development of Passive and Active Vocabulary in a Second Language: Same or Different," *Applied Linguistics*, 19(2) (1998): pp. 255-271.

¹⁰ P. Nation, & P. Meara, "Vocabulary," in N. Schmitt (ed.), *An Introduction to Applied Linguistics* (New York: Oxford University Press Inc., 2002), pp. 35-54.

¹¹ J. M. Ulijin and J. B. Strother, "The Effects of Syntactic Simplification on Reading EST Text as L1 and L2," *Journal of Research in Reading*, 13 (1990): pp. 38-54.

¹² M. Haynes and I. Baker, "American and Chinese Readers Learning from Lexical Familiarization in English Texts," in T. Huckin, M. Haynes and J. Coady (eds), *Second Language Reading and Vocabulary Acquisition* (Norwood, NJ: Albex, 1993), pp. 153-180.

development and must be understood as a part of complex process involving many different elements.¹³

However, it is widely agreed that successful comprehension is heavily dependent upon knowledge of individual word meanings. The often demonstrated relationship between vocabulary and reading comprehension attests to the crucial role word knowledge plays in text understanding among both L1 and L2 readers.¹⁴ Furthermore, it has been argued that a certain threshold in vocabulary size is needed for successful reading comprehension. Many studies have been conducted to investigate the quantity of vocabulary (how many words) that second language learners should have to be able to deal with English texts (written discourse). Laufer¹⁵ claimed that reading comprehension at an academic level requires 95 per cent lexical coverage, i.e. the knowledge of 95% of word tokens in a given text. Learners who knew 95 per cent of the words in the text were more likely to be successful readers and had better comprehension scores.

Nation¹⁶ argued that an ideal coverage of 98%, that is an 8,000 – 9,000 word family vocabulary, is needed for dealing with written texts. For dealing with spoken discourse the size is somewhat smaller, consisting of 6,000-7,000 word families (See also Hsueh-chao & Nation, 2000 for a similar estimate). Meanwhile, Nation and Waring¹⁷ propose a vocabulary of 15,000 to 20,000 as a prerequisite to native-like reading comprehension. Knowledge of the most frequent 10,000 words has been proposed as the minimum for handling university study requirements¹⁸. At the other end of the spectrum, they suggest that having the 3000 level mastery is needed for beginning to read

¹³ Qian, "Assessing the Role of Depth."

¹⁴ K. Koda, *Insight into Second Language Reading: A Cross-Linguistic Approach* (USA: Cambridge University Press, 2005).

¹⁵ Laufer, "The Development of Passive and Active Vocabulary."

¹⁶ I.S.P. Nation, "How Large a Vocabulary Needed for Reading and Listening?" *The Canadian Modern Language Review*, 63 (1) (2006): pp. 59-82.

¹⁷ P. Nation and R. Waring, "Vocabulary Size, Text Coverage and Word Lists," in N. Schmitt, and M. McCarthy (eds.), *Vocabulary: Description, Acquisition and Pedagogy* (Cambridge: Cambridge University Press, 1997), pp. 6-19.

¹⁸ N. Schmitt, D. Schmitt, & C. Clapham, "Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Level Test," *Language Testing*, 18 (2001): pp. 55-88.

authentic passages, and the 5000 word level makes reading authentic texts possible, allowing learners to guess the meaning of the unknown words from context.

Measuring Vocabulary Size: the Yes/No Test

From the foregoing it is evident that a basic dimension of lexical competence is size. Learners with big vocabularies are more proficient in a wide range of language skills than learners with smaller vocabularies, and there is evidence to support the view that vocabulary skills make a significant contribution to almost all aspects of L2 proficiency, including reading comprehension¹⁹. There has been considerable interest in the measurement of L2 vocabulary size and its growth. In the context of vocabulary size, there are several tests devised to measure vocabulary size. The first type of vocabulary size testing tool is Nation's Vocabulary Level Test²⁰. It is a standard test and short test which assesses a small number of words grouped by frequency, using complex multiple choice formats. This test measures passive vocabulary knowledge at 5 word-frequency levels (2000, 3000, 5000, the university word list, and 10,000 words). Each level has 6 clusters including 6 words and 3 definitions. The testees are required to match the words and the definitions.

Another type of test to provide an estimate of vocabulary size is the Yes/No test, a test developed by Meara and colleagues²¹. It is the checklist method of measuring vocabulary knowledge which simply presents the learners with a list of words and asks them to check the words they know. The test measures receptive L2 vocabulary knowledge by eliciting a simple judgment as to whether the learner knows the presented item or not. Test words are drawn from a range of frequency levels, with performance at the respective levels as the basis for inferring the size of the individual's recognition vocabulary²².

The Yes/No test was developed on the basis of a decision task in which it is a standard psycholinguistic tool for measuring word

¹⁹ P. Meara, "The Dimensions of Lexical Competence," in Brown, G., Malmkjaer, K. & Williams, J., (eds), *Performance and Competence in Second Language Acquisition* (Cambridge: Cambridge University Press, 1996), pp. 35–53.

²⁰ I.S.P. Nation, *Teaching and Learning Vocabulary* (New York: Newbury House, 1990).

²¹ P. Meara & B. Buxton, "An Alternative to Multiple Choice Vocabulary Tests," *Language Testing* 4 (1987): pp. 142–45.

²² Meara, "The Dimensions of Lexical Competence."

recognition skill²³. The decision task is often used to measure the ease with which words are activated or retrieved from lexical memory²⁴. In other words, the lexical decision task (LDT) is a forced-choice categorisation task in which the subject is presented with a set of letter strings consisting of words and pseudowords and simply asked to determine whether it is word, or known to the subject²⁵.

The earliest versions of the Yes/No test used a paper and pencil checklist in which the testee is presented with a list containing words and pseudowords²⁶ and asked to check off which words are known. The pseudowords are possible nonwords in the language (e.g., ‘blurg’) and are included in the test items to control for guessing. More recent applications²⁷ have used a computer-driven format in which words and pseudowords are presented in a serial manner. Unlike previous studies Mochida and Harrington²⁸ collected both accuracy and response time measures of Yes/No performance. Accuracy in the lexical decision task has been used to measure L2 vocabulary size²⁹, while response time performance has been examined for insight into the development of L2 processing skills³⁰.

The accuracy response alternatives in the lexical decision task or the Yes/No test are set out in Figure 1. ‘Yes’ responses to real words (hits) reflect the individual’s vocabulary knowledge, while the rate of ‘Yes’ responses to pseudowords (false alarms) measures the individual’s tendency to guess as what they thought that the pseudowords to be the real words in English. Although both hits and

²³ M. Harrington, “The Lexical Decision Task as a Measure of L2 Lexical Proficiency,” *EUROSLA Yearbook*, 6 (2006): pp. 47–68.

²⁴ E.J. Wagenmakers et al., “A Diffusion Model Account of Criterion Shifts in the Lexical Decision Task,” *Journal of Memory and Language*, 58 (2008): pp. 140–159.

²⁵ D. A. Balota, & J. I. Chumbley, “Are Lexical Decisions a Good Measure of Lexical Access? The Role of Word Frequency in the Neglected Decision Phase,” *Journal of Experimental Psychology: Human Perception and Performance*, 10 (3) (1984): pp. 340–357.

²⁶ Meara & Buxton, “An Alternative to Multiple Choice.”

²⁷ Mochida and Harrington, “The Yes–No Test as a Measure of Receptive.”

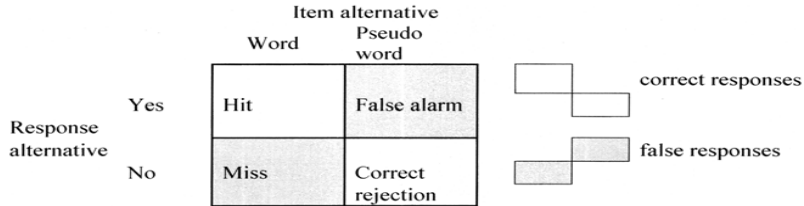
²⁸ Ibid.

²⁹ Ibid.

³⁰ N. Segalowitz and J. Hulstijn, “Automaticity in Bilingualism and Second Language Learning,” in F. F. Kroll and A.M.B. De Groot (eds), *Handbook of Bilingualism: Psycholinguistics Approaches* (Oxford: Oxford University Press, 2005), pp. 179–201.

the correct rejections of non-words are correct responses, the number of hits, adjusted by the false alarm rate, is of primary interest.

Figure 1: The item-response matrix for the Yes/No test.



The Yes/No test is unique in its use of pseudowords. These pseudowords are included to counterbalance any tendency for testees to overestimate their word knowledge by checking words that they do not actually know³¹. Claiming knowledge of pseudowords leads to a downward adjustment in the score to provide a better estimate of the knowledge of the real words. The response alternatives in the Yes/No test are the same as set out in Figure 1.

The test items are typically drawn from a range of lexical frequency levels, that is, from the most frequent words to the least frequent words (1000, 2000, 3000, 5000, and 10000 word frequency levels), with performance at the respective levels used as the basis for inferring the size of the individual’s receptive vocabulary³². Lexical frequency has been argued to be a valid means to standardize the lexical size tests³³. Lexical frequency approaches to L2 vocabulary learning are based on the assumption that the more frequently used words will be the more easily learned. Moreover, the percentage of words known at each frequency level allows an extrapolation to be made and a calculation of overall lexical knowledge in the items being tested. Thus a lexical profile can be arrived at, reflecting learners’ lexical knowledge as suggested in the Yes/No test.

³¹ L. Cameron, “Measuring Vocabulary Size in English as an Additional Language,” *Language Teaching Research*, 6 (2002): pp. 145–73.

³² Meara, “The Dimensions of Lexical Competence.”

³³ J. Milton, “Lexical Profiles, Learning Styles and the Construct Validity of Lexical Size Tests,” in Daller, J. Milton, Treffers Dallers (eds), *Modelling and Assessing Vocabulary Knowledge* (Cambridge: Cambridge University Press, 2007), pp. 47-58.

The Yes/No test also has several practical advantages. The test is very easy to construct and it requires only a few minutes for the testee to complete it. It is therefore possible to use a large number of items, which permit the construction of a reliable measurement of vocabulary size.³⁴

Studies Using the Yes/No test

Studies attempting to validate the Yes/No test against other measures of L2 vocabulary knowledge have revealed mixed results. Meara and Buxton³⁵ reported a correlation of around $r = .7$ between it and multiple-choice test measures in performance by French EFL learners. Shillaw³⁶ found significant correlations of .42 to .48 between the two measures by first year Japanese university students. However, Cameron³⁷ compared performance on Yes–No test with scores on several subsections of the Vocabulary Level Test³⁸ by secondary ESL students in the UK and found no correlation. Correlation coefficients between the Yes–No test scores and the VLT's subsections ranged from .15–.45, none of which were significant. In addition, the study also reported a large number of false alarms, suggesting that many testees tended to perform guessing in completing the test, although the actual false alarm rate was not reported. Eyckmans³⁹ also examined the concurrent validity for the Yes–No measure by having French learners of Dutch translate into their L1 Dutch words previously presented in the Yes–No format. The correlations between the translation results and various scoring methods for Yes/No performance were a modest $r = .3$ to $.5$, despite the identical content. High false alarm rates were also noted in the study, ranging from 20%–25% of the pseudowords presented.

The two studies above showed high false alarm rates. This suggests that there was significant guessing by the testees, with resulting test

³⁴ Meara, & Buxton, "An Alternative to Multiple Choice."

³⁵ Ibid.

³⁶ J. Shillaw, "The Application of Rasch Modelling to Yes/No Vocabulary Tests," Vocabulary Acquisition Research Group, University of Wales Swansea. Available at <http://www.swan.ac.uk/cals/calsres/vlibrary/js96a.htm> (September 2005).

³⁷ Cameron, "Measuring Vocabulary Size."

³⁸ Nation, *Teaching and Learning Vocabulary*.

³⁹ J. Eyckmans, *Measuring Receptive Vocabulary Size* (Utrecht: LOT, 2004).

scores of low reliability. However, a recent study by Mochida and Harrington⁴⁰ provides stronger support for the concurrent validity of the Yes/No format. This study examined accuracy in the Yes/No test format as a predictor of VLT performance by Advanced ESL learners studying at an Australian university. A strong correlation (.8) between accuracy on the two formats was evident at all levels. The study also reported a lower overall false alarm rate (5%) than the subjects in either Cameron or Eyckmans above, possibly due to the Asian L1 backgrounds of the subjects in Mochida and Harrington.

The Yes/No studies mentioned above have only been conducted in input-rich ESL contexts. The current study will assess Yes/No in a foreign language setting. It is assumed here that learning context may have an effect on the test's performance, in this instance, the lack of English exposure, the amount of time spent in language learning, motivation for learning English or doing the test, the proportion of vocabulary learning in the classroom, and other contextual issues.

Results *The Yes/No Test Performance*

The first question concerned whether or not accuracy in the Yes/No test performance improves as group proficiency and word frequency level increase.

Reliability measures (Cronbach's alpha) for the Yes/No test were calculated for both word and pseudoword items. Overall reliability for the real words was .72 and .74 for the pseudowords. Reliability coefficients of .8 - .9 are considered strong, so the test here has only moderate reliability⁴¹ (Brown, 1996). The reliability of pseudowords was similar to that of words, indicating that there seems to be systematicity to how the pseudowords function in the test. The reliability score in this study was slightly lower than the .72 (the Yes/No test for Computer Application test B) to .79 (the Yes/No test for two Computer Application test A) for the real words and the reliability of the .70 (the Yes/No test for Computer Application test B) and the .78 (the Yes/No test for Computer Application test A) for the

⁴⁰ Mochida and Harrington, "The Yes-No Test as a Measure of Receptive."

⁴¹ J. D. Brown, *Testing in Language Programs* (Upper Saddle River, NJ: Prentice Hall, 1996).

pseudowords reported in Eyckmans et. al.⁴². The value for the reliability test for this study was lower than many previous studies. Some of the plausible reasons could be the number of test items and different test items being used, as well as the sample size.

Table 1. presents the group differences for the hits ('yes' responses to word items) at each word frequency level. The bottom part of the table also contains the overall false alarm rate ('yes' responses to pseudowords) and corrected scores (hits – false alarms) that are given by group.

Table 1 : Differences between Intermediate and Advanced Groups by Frequency Levels for Hits, False Alarms and Corrected Scores.

Frequency Level	Item (no.)	Intermediate N = 42		Advanced n= 41		t(81)
		Mean (%)	SD (%)	Mean (%)	SD (%)	
500	20	100	0	100	0	-
1000	20	99	1.4	100	0	2.05*
2000	20	97	4.2	99	1.9	3.43**
3000	20	85	9.2	89	7.2	2.20*
5000	15	50	10.6	57	7.9	3.38**
10000	15	29	10.9	32	9.3	1.25
Overall Hits	110	77	6.1	80	4.3	See F
False alarm (%)	40	16	9.5	18	11.4	0.77
Corrected score		61	8.3	62	11.1	0.58

*p < .05,** p < .01

The t values for the planned pairwise comparisons carried out as part of the analysis of variance test are also reported for the respective levels⁴³. These indicate whether group performance at that level was statistically different at either the .05 or .01 levels of significance. The group difference between the overall mean hits was tested in the two-way ANOVA F-test below.

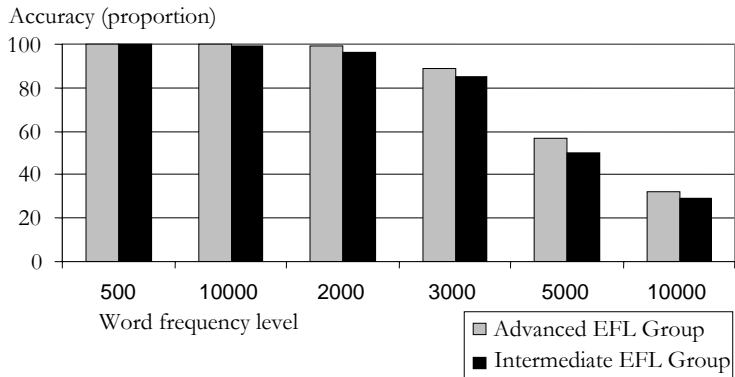
⁴² J. Eyckmans et al., "Learners' Response Behavior in Yes/No Vocabulary Test," in Daller, J. Milton, Treffers Dallers (eds), *Modelling and Assessing Vocabulary Knowledge* (Cambridge: Cambridge University Press, 2007), pp. 59-76.

⁴³ D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistics* (Bacon Raton, FL: Chapman and Hall, 2007).

There was a ceiling effect in performance by both Groups for the first three frequency levels, 500, 1000 and 2000, while the Advanced group outperformed the Intermediate group for the last three levels, 3000, 5000 and 10,000. The mean differences are presented in Figure 2.

Figure 2 : Mean accuracy scores by proficiency and frequency level

The performance of the Yes/No test by group and frequency level



At the 3K word frequency level, there was a slight decrease in scores for both groups, though performances were still high. The mean accuracy for the advanced group was 89% compared to the Intermediate group that achieved 85%.

This overall score at this level was similar to what was found in Harrington⁴⁴ with a value of 90 for the Advanced ESL group and 76 for the Intermediate ESL group. At the 5000 word level the Advanced group had a mean of 57% compared to 50% for the Intermediate group. At the 10000 word level performance fell sharply to under 50% for both groups, with the Advanced group scoring 32% and the Intermediate group 29%. This pattern was similar to the one reported in Harrington’s study for both Advanced and Intermediate ESL group whose scores were under 50% overall.

A two-way repeated measures ANOVA was done to assess the mean differences observed. The between-group factor was Group (Intermediate vs Advanced) and the within-group (repeated) factor was Frequency level (500- 10000). Overall there was a significant effect for Group, $F(1,81) = 10.48$, $p = .002$, $p = .001$, partial eta squared

⁴⁴ Harrington, “The Lexical Decision Task.”

(measure of effect size) = .115 .For the Frequency Levels, $F(5,405)=1955$, $p = .000$, partial eta squared = .96. There was also a significant Group x Level interaction $F(5,405) = 3.68$, $p = .003$, partial eta squared = .043. Planned comparisons were also carried out to test the differences between the levels, summing across the groups. The small mean difference between the 500 and 1000 level was significant at $p < .05$, while the remaining level differences were all statistically significant at $p < .001$. Group performance at each frequency level was compared with independent group t-test. The differences between the Advanced EFL group and the Intermediate EFL at the 1000, 2000 3000 and 5000 levels were all statistically significant. See Table 1.

The performance discussed thus far is on the raw hits, that is, the percentage of 'yes' responses to words. However, as discussed in the literature review, the number of hits alone does not take into account possible guessing by the testee. In order to get a better picture of the individual's vocabulary knowledge, it is also necessary to correct for guessing, which is reflected in false alarms.

The false alarm rate for the Advanced group was around 18%, while the Intermediate group was around 16% but the difference was not statistically significant. The false alarm rate in this study was lower than the study by Eyckmans⁴⁵ which reported false alarm rates ranging from 20% to 25%. A more recent study by Eyckmans et. al.⁴⁶ also reveals a false alarm rate which was higher than this study ranging from 23% to 24%. However, the false alarm rate noted in this study was higher than that was found by Harrington⁴⁷ and Mochida and Harrington⁴⁸, who reported less than 5% overall.

The corrected scores were calculated for the results across all the frequency levels and they showed very little difference across the two groups (Intermediate = 41, Advanced = 42). The difference was not statistically significant. The corrected scores of this study were lower than the ones reported in Eyckmans (2007) ranging from 81%-84% and in Mochida and Harrington (2006) with 81% overall. The lower score was evident despite the fact that the present test included higher frequency items (1-500 and 501-1000 levels).

⁴⁵ Eyckmans, *Measuring Receptive Vocabulary Size*.

⁴⁶ Eyckmans et al., "Learners' Response Behavior."

⁴⁷ Harrington, "The Lexical Decision Task."

⁴⁸ Mochida and Harrington, "The Yes-No Test as a Measure of Receptive."

Thus, the first finding reveals that the Yes/No test did discriminate the two group performances in terms of hits accuracy but the difference disappeared with corrected scores. The higher proficiency group has more accurate responses in every word frequency level except the very highest frequency level (500). These results support for the claim that the Yes/No test is a stable measure of L2 receptive lexical proficiency.⁴⁹

The participants (Advanced EFL group and Intermediate EFL group) in the study reached ceiling performance at the 500, 1000 and 2000 word levels. However, the Advanced group outperformed the Intermediate, suggesting the robust pattern of the results obtained for the lower frequency levels here (3K, 5K and 10K). The same pattern was also observed in Harrington⁵⁰, in which the group performances indicated significant differences at these three low frequency levels. This suggests that the Yes/No test can serve as a useful tool for lexical assessment. Further work is still needed to see if these patterns could be observed in learners at lower levels of proficiency, especially in an EFL context. To what degree the patterns evident for the low frequency items in this study hold for performance by lower proficiency learners on higher frequency items is an open question.

Although the Advanced group outperformed the Intermediate level for the Hits, this was not the case for corrected scores where there was no difference between the two groups. The lack of difference between the two groups is partly due to the false alarm rate, which was higher for the Advanced group, but not in a statistically significant way. Taking into account the high false alarm rate, this evidence suggests that the actual vocabulary knowledge of the two groups was reasonably similar.

The false alarm rate in this study ranges from 18 % (the Advanced group) to 16 % (the Intermediate group) between the two groups or 17 % overall. This false alarm rate is larger than that reported in Mochida and Harrington⁵¹ who reported less than 5% overall for two groups. On the other hand, this false alarm rate was smaller than that was reported by Eyckmans⁵² which ranged from 20%-25% and Eyckmans

⁴⁹ Harrington, "The Lexical Decision Task."

⁵⁰ Ibid.

⁵¹ Mochida and Harrington, "The Yes–No Test as a Measure of Receptive."

⁵² Eyckmans, *Measuring Receptive Vocabulary Size*.

et.al⁵³ which ranged from 23% to 24%. However, the false alarm rate in this study did not affect the basic finding that the performance of the two groups on the Yes/No test was statistically significant.

In the present study, there are several possible explanations for the guessing behavior (false alarm rate) in this EFL context. First, as the Yes/No test was designed on the basis of decision task format, it is important to note here that basically there are two basic stages in a lexical decision, a perceptual recognition or discrimination stage and a decision stage, with performance affected by both.⁵⁴ A variant of the two stage model developed by Atkinson and Juola⁵⁵ for the memory search task could be used to account for the decision process. The basic notion is that words and pseudo words differ on familiarity/meaningfulness (FM) dimension. The first stage of the decision process involves a global computation of the FM value of the letter string. That is, the participant makes a quick check to determine if the stimulus is producing any meaning or is very familiar (frequently found). The second stage of the decision process occurs when the participants have low familiarity with the items, so that the participant needs more information before a decision can be made. In this context, guessing responses by participants in this study occur in the global analysis and analytic stage when words or pseudo words have an extremely low FM value. Since the display duration for tested items was set by the researcher in this study, the participant was pushed to make a decision within a certain amount of time.

Also in the analytic stage, the error could occur when the participant has a lack of knowledge about the appropriate spelling of the word⁵⁶. In other words, the participants appeared to misread or misjudge the items, especially those with orthographically or phonologically similar items. Kanwisher and Potter⁵⁷ propose that the participants of the lexical decision task as also reflected in the Yes/No

⁵³ Eyckmans et. al., "Learners' Response Behavior."

⁵⁴ A. M. Jacobs & J. Grainger, "Models of Visual Word Recognition — Sampling the State Knowledge," *Language Testing*, 26(1) (1994): pp. 73–98.

⁵⁵ R.C. Atkinson & J.F. Juola, "Factors Influencing Speed and Accuracy of Word Recognition," in S. Kornblum (ed.), *Attention and Performance IV* (New York: Academic Press, 1973).

⁵⁶ Balota & Chumbley, "Are Lexical Decisions a Good Measure."

⁵⁷ N. Kanwisher and M. Potter, "Repetition Blindness: Levels of Processing," *Journal of Experimental Psychology: Human Perception and Performance*, 16 (1990): pp. 30–47.

test were simply poorer at reading rapidly presented words or pseudo words in the neighbor condition (phonologically or orthographically).

Also, it is worth noting here that the serial presentation for the Yes/No test in this study used a PowerPoint presentation in which the reaction time for the test was controlled at 2000 milliseconds. The instrument for presenting tested items through the PowerPoint presentation in the Yes/No test here was different from other previous studies. For instance, Mochida and Harrington⁵⁸ and Eyckmans et al.⁵⁹ incorporated a computer interface in the designing the Yes/No test. The use of visual time serial presentation using PowerPoint slides during the test in this study perhaps caused participants to misread presented items. Technically, those participants who sat next to the presentation screen saw the items more clearly compared to those who sat at the back of the class during the test. In the computer-driven Yes/No test, however, the testee could see clearly the items from the computer screen and also control the speed of response or reaction time in which this condition has been argued to reduce misreading of the items.

Another factor influencing guessing behavior in the Yes/No test is the response latency or display duration of the items during the tests. It has been suggested⁶⁰ that the use of a longer latency for lexical decision for pseudo words appeared to have been effective in increasing participants' accuracy when faced with pseudo words. It may be that a longer time duration for item display would improve pseudo word accuracy for the groups in this study. In line with this, Harrington (2006) suggests that the higher proficiency group responded faster and more accurate than the lower proficiency group, indicating that the proficiency level contributes to reaction time (RT) in the Yes/No test.

Test instructions may also have played a role in the false alarm rate. As this study employed minimum-instruction conditions merely the original instructions from the Yes/No test, the test takers were only asked to judge the words they knew without warning them that they might be tested again later. These instructions might have led to higher

⁵⁸ Harrington, "The Lexical Decision Task."

⁵⁹ Eyckmans et al., "Learners' Response Behavior."

⁶⁰ M. L. Still and A.L. Morris, "Now You See It, Now You Don't: Repetition Blindness for Nonwords," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1 (2008): pp. 146–166.

false alarm rate behavior, as suggested by Eyckmans⁶¹ because test takers tend to overestimate their word knowledge during the test. Also with these instructions, the test itself could not measure the participant's level of knowing. It is evident in this study that the participants demonstrate a higher false alarm rate compared to the previous study by Mochida and Harrington⁶² which showed a lower false alarm rate using another type of Yes/No test instructions, that is, the maximum-instruction method. In maximum-instruction mode, the participants were asked to consider carefully to what extent they know the items as they are warned that they might be tested again later. Accordingly, further studies should always try to provide clearer and better instructions to avoid a high false alarm rate or overestimation response style.

The last possible factor is the participants' attitude toward taking the test. In the context of Harrington's study, the participants participated in a partial fulfillment of course requirements, by which the participants were valued for their participation on the test. Their participation on the test would affect their grade during the course. In contrast, the participants in this study were volunteers and were not give any credit toward their course for participation. In this case, as the researcher himself experienced, most of the participants were likely not very serious in doing the test, indicating that they just did not really put much effort into the test. As Nation⁶³ suggests, the most important factor that influences a test's result is the participant's attitude to taking it. In this instance, they did not take it seriously and probably did it quickly without giving it much attention. The participants' attitude is hard to control in test administration, especially in EFL contexts, but it is very important.

Overall, the study shows that the participants have limited vocabulary especially at the level of 5K and 10K word frequency levels in the test. This suggests that language teachers need to give more attention to developing the needed vocabulary knowledge, either in

⁶¹ Eyckmans, *Measuring Receptive Vocabulary Size*.

⁶² Mochida and Harrington, "The Yes–No Test as a Measure of Receptive."

⁶³ I.S.P. Nation, "Fundamental Issues in Modelling and Assessing Vocabulary Knowledge," in Daller, Milton, J., Treffers Dallers (eds), *Modelling and Assessing Vocabulary Knowledge* (Cambridge University Press, 2007), pp. 35-43.

receptive or productive knowledge of vocabulary. As Meara⁶⁴ suggested, the greater their knowledge of vocabulary, the more learners can handle the language, and perform language skills such as reading, writing, listening, and speaking.

This study examined the Yes/No format using serial presentation. The time of exposure was limited in order to minimize the testees' use of strategic processes and reflective knowledge. Future applications could incorporate this feature into a computer-driven test.

In a global context of performance, the Yes/No must be interpreted within a larger theoretical framework that specifies the role of word recognition in L2 processing and development (Koda, 1996). Word recognition plays an important role especially in the level of L2 sentence processing (Fender, 2002; Harrington, 2001). It is also assumed here that this word recognition or lexical access performance also relates to further global comprehension processes, especially in reading comprehension. In more practical terms, we still need to address the issue of how response accuracy could be the predictive power of global reading comprehension.

Above all, when considering the use of the Yes/No test format, there are several things to consider in this test format. First, clearer or better instructions and directions should be provided to reduce the false alarm rate and thereby obtain a valid test result. Test results that include high guessing behavior (false alarm rate) may reflect learner attitude rather than language skills, because the learner uses test-taking strategies such as guessing. These need to be minimized. Second, one of the ways to avoid or reduce response bias may be through the use of the computer interface as suggested by Eyckmans et. al.⁶⁵ If students receive regular feedback concerning pseudo word performance this might limit the number of false alarms produced. Third, the use of the test should be incorporated into the learning context so that students see it as a natural and important part of the language learning process.

⁶⁴ Meara, "The Dimensions of Lexical Competence."

⁶⁵ Eyckmans et al., "Learners' Response Behavior."

L2 Reading Performance and the Relationship between the Yes/No Test Performance and L2 Reading Performance

Table 2 reports the mean proportions of reading performance between the Advanced EFL group and the Intermediate EFL group. The reliability measures (Cronbach's alpha) were calculated for both groups; Advanced and Intermediate. The reliability scores ranged from .092 to .152. This indicates that this reading test had little reliability in this context. Besides, when looking at the item discrimination indices in the reading test, it was found that only 1 out of the total 20 reading items (no. 10) had a good item value ($D=.30$), while the rest were poor ($D=.01-.19$). Interestingly several items (2, 6, 7, 11, 16, and 18) even indicated a negative figure of discrimination index ranging from $D= -.01$ to $-.11$. The low reliability of the test makes interpretation of the results difficult.

The study reveals that there was no significant differences at value $p=.516$ between the two group performances in reading comprehension test. The mean accuracy for the advanced group was 46%, and that of the Intermediate group 44%. Both groups scored poorly, indicating that the test was too difficult, although the source of the difficulty is self-evident. It is important to note that, even though the reading performances of the two groups were not statistically different, there were still slight mean differences. The low reliability value for this test was assumed to be one major factor affecting the low performance for both groups in this study.

Table 2 : Reading scores by groups

Group	Mean	Std. Deviation
Advanced	45.61	11.093
Intermediate	43.89	8.807

The result indicates that the advanced group had higher Yes/No test scores and also did slightly better on the reading test, though the latter turned out to be an unreliable measure of reading skill for the learners in the study.

A potential association between Yes/No test results and reading scores was tested by performing a Pearson product moment correlation. The results are reported in Table 3 and show no correlation between reading scores and either Hits or the Corrected scores.

The results showed that the advanced group did better on the Yes/No Test and the reading test, but that the difference on the reading test was not significant. There are several possible reasons for this. The first is that the reading test used in this study had a low reliability value, indicating that it was not appropriate for the group tested. Most of the items in the reading test could not discriminate the test taker's performance.

Table 3 : Correlation between vocabulary performance and reading performance

The Yes/No test		Reading test
Hits	Pearson Correlation	.057
	Sig. (2-tailed)	.663
Corrected scores	Pearson Correlation	.040
	Sig. (2-tailed)	.763
None of the correlations are significant at $p < .05$		

Second is that both groups have limited knowledge of vocabulary, especially at the 5000 and 10000 word levels. The learners in this study knew as little as half of the vocabulary at the 5000 word level and 30% of the vocabulary at the 10,000 word level. This may have limited their ability to do the reading test. Specifically, a vocabulary of 8,000-9,000 word families is required for 98% text coverage, which in turn is needed for adequate text comprehension⁶⁶. Since the learners in this study possessed limited vocabulary, they were considered inadequate readers. In a larger perspective, having small vocabulary size has been argued to be a stumbling block in the path to academic success⁶⁷.

Finally, even though vocabulary plays a central role in reading comprehension⁶⁸, there are many other factors influencing the global reading comprehension processes, such as morpho-syntactic knowledge⁶⁹ and reading strategies⁷⁰ and these may also have contributed to the relatively poor reading performance observed here.

⁶⁶ Nation, "How Large a Vocabulary Needed."

⁶⁷ A. Biemiller, "Teaching Vocabulary: Easy, Direct and Sequential," *American Educator*, 25 (1) (2001): pp. 24-28.

⁶⁸ Qian, "Investigating the Relationship."

⁶⁹ Ulijin and Strother, "The Effects of Syntactic Simplification."

⁷⁰ Haynes and Baker, "American and Chinese Readers Learning."

Conclusion

The results of Yes/No task test reveal that the hits accuracy performance between the Advanced EFL group and the Intermediate EFL group was statistically significant, indicating that Yes/No test performance in context of hits accuracy did discriminate between levels of English proficiency. However, the differences disappeared with corrected scores since both groups indicated high false alarm rates. Future studies should put an effort to avoid such high false alarm rates through providing better instruction.

On the other hand, the overall reading performance between the Advanced EFL group and the Intermediate EFL group was not statistically significant, although the differences were observed between the two groups in their mean scores. One plausible reason for this absence of statistical difference between the two levels is the presence of a floor effect, suggesting that both groups had low performance on reading task.

This study reveals that there is no statistical correlation between the Yes/No test performance and reading performance because the two group performances were very low in the reading test and the reading test itself had a low reliability value. Further studies using other reading tests with high validity and reliability values are needed to examine the relationship between the Yes/No test performance and the reading proficiency.

Accordingly, there are several aspects of the above-mentioned discussion which have a direct implication for language teaching in an Indonesian Islamic educational context or Indonesian context in general. First, the Yes/No test may be used by language teachers as a measure for testing and placement purposes. This may especially be the case if the items in the test are taken from Indonesian EFL textbooks and examination materials. In broader context, this test could be used to measure the development of L2 lexical processing skill which includes both accuracy and speed of processing.

Second, there is a need for language teachers to place more emphasis on the development of their students' vocabulary knowledge, especially at higher levels - between 5000 and 10,000 word frequency levels. The main reason is that the mastery of vocabulary at such levels is required for understanding written academic discourse. Teaching vocabulary also promotes reading comprehension and language proficiency in general.

In an Indonesian context, English language teaching in Islamic higher education undoubtedly plays a significant role, both in developing professional academic life and also preparing learners, when they graduate, to become well-prepared and trained English language teachers, whether at elementary level, secondary level or even for tertiary level. As is generally known, one of the programs provided in Islamic higher education is the English education program, the final outcome of which is to prepare learners to be English teachers at elementary and secondary level. Thus, the results of this study should be taken into account as one entry point to develop the quality of English language teaching in Islamic higher education in Indonesia.

Finally, it has been argued that English language proficiency in this era is of importance in developing academic life professionally and socio-cultural life in general. English proficiency is now believed to be a significant means to develop academic knowledge or world knowledge in general since the majority of information nowadays is in English written discourse, such as worldwide magazines or newspapers, textbooks, and electronic information, as well as in English spoken discourse broadcast by international television programs. In this context, Indonesian Islamic higher education learners in particular, as with Indonesian learners in general, are expected to be literate in English so that they can play an active role in the international, academic community. Therefore, Islamic higher education in Indonesia should continuously strive to develop its quality and produce the best possible graduates - ready to face an increasingly challenging world. []

Bibliography

- Atkinson, R.C., and Juola, J.F. "Factors Influencing Speed and Accuracy of Word Recognition." in S. Kornblum (ed.). *Attention and Performance IV*. New York: Academic Press, 1973.
- Balota, D. A. and Chumbley, J. I. "Are Lexical Decisions a Good Measure of Lexical Access? The Role of Word Frequency in the Neglected Decision Phase." *Journal of Experimental Psychology: Human Perception and Performance*, 10, 3 (1984): pp. 340–357.
- Biemiller, A. "Teaching Vocabulary: Easy, Direct, and Sequential." *American Educator*, 25, 1 (2001): pp. 24-28.

- Brown, J. D. *Testing in Language Programs*. Upper Saddle River, NJ: Prentice Hall, 1996.
- Cameron, L. "Measuring Vocabulary Size in English as an Additional Language." *Language Teaching Research*, 6 (2002): pp. 145–73.
- Dardjowidjojo, S. "English Teaching in Indonesia." *EA Journal*, 18, 1 (2000): pp. 22-30.
- Eyckmans, J. et. al. "Learners' Response Behavior in Yes/No Vocabulary Test." in Daller, Milton, J., Treffers Dallers (eds.). *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press, (2007): pp. 59-76.
- Eyckmans, J. *Measuring Receptive Vocabulary Size*. Utrecht: LOT, 2004.
- Grabe, W. "Current Developments in Second Language Reading Research." *TESOL Quarterly*, 25 (1991): pp. 375-406.
- Harrington, M. "The Lexical Decision Task as a Measure of L2 Lexical Proficiency." *EUROSLA Yearbook*, 6 (2006): pp. 47–68.
- Haynes, M. and Baker, I. "American and Chinese Readers Learning from Lexical Familiarization in English Texts." in T. Huckin, M. Haynes and J. Coady (eds.). *Second Language Reading and Vocabulary Acquisition*. Norwood, NJ: Albex, (1993): pp. 153-180.
- Jacobs, A. M. & Grainger, J. "Models of Visual Word Recognition — Sampling the State Knowledge." *Language Testing*, 26, 1 (1994): pp. 73–98.
- Kanwisher, N., and Potter, M. "Repetition Blindness: Levels of Processing." *Journal of Experimental Psychology: Human Perception and Performance*, 16 (1990): pp. 30–47.
- Knight, S. "Dictionary Use While Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities." *Modern Language Journal*, 78, 3 (1994): pp. 285-299.
- Koda, K. *Insight into Second Language Reading: A Cross-Linguistic Approach*. USA: Cambridge University Press, 2005.
- Laufer, B. "The Development of Passive and Active Vocabulary in a Second Language: Same or Different." *Applied Linguistics*, 19, 2 (1998): pp. 255-271.

- Meara, P. & Buxton, B. "An Alternative to Multiple Choice Vocabulary Tests." *Language Testing*, 4 (1987): pp. 142–45.
- Meara, P. "The Dimensions of Lexical Competence." in Brown, G., Malmkjaer, K. & Williams, J., (eds.). *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge University Press, (1996): pp. 35–53.
- Milton, J. "Lexical Profiles, Learning Styles and the Construct Validity of Lexical Size Tests." in Daller, Milton, J., Treffers Dallers (eds.). *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press, (2007): pp. 47-58.
- Mochida, A. and Harrington, M. "The Yes–No Test as a Measure of Receptive Vocabulary Knowledge." *Language Testing*, 26, 1 (2006): pp. 73–98.
- Nation, I.S.P. "Fundamental Issues in Modelling and Assessing Vocabulary Knowledge." in Daller, Milton, J., Treffers Dallers (eds.). *Modelling and Assessing Vocabulary Knowledge*. (Cambridge University Press, (2007). pp. 35-43.
- Nation, I.S.P. "How Large a Vocabulary Needed for Reading and Listening?" *The Canadian Modern Language Review*, 63, 1 (2006): pp. 59-82.
- Nation, I.S.P. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press, 2001.
- Nation, I.S.P. *Teaching and Learning Vocabulary*. New York: Newbury House, 1990.
- Nation, P., & Meara, P. "Vocabulary." in N. Schmitt (ed.). *An Introduction to Applied Linguistics*. New York: Oxford University Press Inc., 2002: pp. 35-54.
- Nation, P., and Waring, R. "Vocabulary Size, Text Coverage and Word Lists." in N. Schmitt, and M. McCarthy (eds.). *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, (1997): pp. 6-19.
- Qian, D. D. "Investigating the Relationship between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective." *Language Learning*, 52, 3 (2002): pp. 513-536.

- Qian, D.D. "Assessing the Role of Depth and Breadth of Vocabulary Knowledge in Reading Comprehension." *Canadian Modern Language Review*, 56 (1999): pp. 282-307.
- Schmitt, N., Schmitt, D., & Clapham, C. "Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Level Test." *Language Testing*, 18 (2001): pp. 55-88.
- Segalowitz, N. and Hulstijn, J. "Automaticity in Bilingualism and Second Language Learning." in F. F. Kroll and A.M.B. De Groot (eds), *Handbook of Bilingualism: Psycholinguistics Approaches*. Oxford: Oxford University Press, (2005): pp. 179–201.
- Sheskin, D. J. *Handbook of Parametric and Nonparametric Statistics*. Bacon Raton, FL: Chapman and Hall, 2007.
- Shillaw, J. "The Application of Rasch Modelling to Yes/No Vocabulary Tests." *Vocabulary Acquisition Research Group*, University of Wales Swansea. Available at <http://www.swan.ac.uk/cals/calsres/vlibrary/js96a.htm> (September 2005).
- Still, M. L. and Morris, A.L. "Now You See It, Now You Don't: Repetition Blindness for Nonwords." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1 (2008): pp. 146–166.
- Ulijin, J. M. and Strother, J.B. "The Effects of Syntactic Simplification on Reading EST Text as L1 and L2." *Journal of Research in Reading*, 13 (1990): pp. 38-54.
- Wagenmakers, E.J., et al. "A Diffusion Model Account of Criterion Shifts in the Lexical Decision Task." *Journal of Memory and Language*, 58 (2008): pp. 140–159.