The on-line Yes/No Test as a placement tool

Link to publication record in USC Research Bank:
http://research.usc.edu.au/vital/access/manager/Repository/4880

Document Version:
Author accepted manuscript (postprint)

Citation for published version:

Copyright Statement:
Copyright © 2009. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

This is the author’s version of a work that was accepted for publication in The on-line Yes/No Test as a placement tool. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in System, Vol. 31, No. 4, pp.614-626, DOI:10.1016/j.system.2009.09.006.

General Rights:
Copyright for the publications made accessible via the USC Research Bank is retained by the author(s) and / or the copyright owners and it is a condition of accessing these publications that users recognize and abide by the legal requirements associated with these rights.

Take down policy
The University of the Sunshine Coast has made every reasonable effort to ensure that USC Research Bank content complies with copyright legislation. If you believe that the public display of this file breaches copyright please contact research-repository@usc.edu.au providing details, and we will remove the work immediately and investigate your claim.
The on-line Yes/No Test as a placement tool

Michael Harrington  mwharr@uq.edu.au
School of Languages and Comparative Cultural Studies
Gordon Greenwood, Level 3, University of Queensland
St Lucia 4072 Qld, Australia

Michael Carey
University of the Sunshine Coast
Maroochydore DC 4558 Qld, Australia

Corresponding author.
Michael Harrington  mwharr@uq.edu.au
Tel. +61 7 3365 6719
Fax. +61 7 3365 2799

Word count: 5665 (excluding references and tables)
Abstract

This study evaluates the concurrent validity of an online Yes/No test of recognition vocabulary as a placement tool at an Australian English language school. Newly entering students (n=88) completed a Yes/No test, which measured accuracy and speed of response, and a school placement battery consisting of grammar, writing, speaking and listening measures. Yes/No test accuracy performance approximated that of the listening and grammar placement tests, discriminating between placement levels and correlating with overall placement decisions. Placement level decisions correlated with Yes/No accuracy ($r = .6$) and the discrete-point placement measures listening (.7) and grammar (.8). Yes/No test accuracy showed a consistently strong correlation (.6) across all test measures, indicating that recognition vocabulary knowledge is a fundamental element in all tests. Response time as an independent measure followed the same pattern as the accuracy results but was less sensitive to differences in placement levels. A hierarchical regression analysis showed that when combined with accuracy, response times account for an additional amount of unique variance in the placement decision. The usefulness the on-line Yes/No test format in placement testing is discussed.

Keywords

Yes/No test, recognition vocabulary, computerised testing, placement testing, recognition vocabulary, English as a second language, response time
The on-line Yes/No Test as a placement tool

1. Introduction

Placement testing is a time and resource-intensive activity for language programs. The placement of new students at the appropriate level is an important decision that affects teaching and learning outcomes and can have a pronounced effect on student satisfaction (Brown 2005). Placement testing can occupy a significant amount of teacher and program time, ranging from several hours to an entire day.

This study evaluates an on-line version of the Yes/No test (Meara and Buxton, 1987) as a tool for making placement decisions in a commercial language school. The test uses a checklist format in which word and pseudoword items are presented sequentially, with the testee indicating whether or not s/he knows the word. The format is attractive in that it is fast, easy to administer and score, and makes limited demands on the testee. The format is more a self-assessment measure than a traditional vocabulary test, and its validity as a measure of vocabulary knowledge has been the subject of research. In terms of construct validity, the test measures individual differences in the speed and accuracy of L2 word recognition, which is a fundamental aspect of performance (Koda, 1996). However, the relationship between language performance and lexical size and speed is complex, particularly as measured in the discrete, decontextualised manner of the Yes/No test.

Support for the concurrent validity of the test has come from studies showing strong correlations ($r = .6 -.8$) between the test and multiple choice measures (Meara and Buxton, 1987; Meara and Jones, 1990; Mochida and Harrington, 2006). In contrast, other research has found only weak – or no - correlations among the test and receptive and productive measures of L2 vocabulary (Cameron, 2000; Eyckman 2004; Eyckman et al., 2007).

The lack of correlation in these studies is attributable in part to the relatively high rate of “false alarms”, that is, pseudowords judged to be actual words. False alarm rates as high as 25% were reported in a study of Dutch learners of French (Eyckman et al., 2007; see also
Cameron 2000 for similar rates for ESL learners in the UK). Mochida and Harrington (2006), in contrast, reported false alarm rates of around 5% for Asian ESL learners (see also Shillaw, 1996 for a similar finding). The tendency of an individual to identify pseudowords as words, or to be overly conservative in accepting a word as known, reflects a response bias that is inherent in the test format and which must be identified and to some extent controlled if Yes/No test results are to be a valid measure of vocabulary knowledge. How to best deal with the response bias as reflected in false alarms has been examined in some depth by studies that have examined various approaches to control for guessing (Beeckman, 2001; Huibregste, Admiraal and Meara, 2002; Eyckman, 2004; Mochida and Harrington 2006; Eyckman et al., 2007).

The research to date indicates that Yes/No test performance can vary due to a number of factors. These include the language background of the testee, e.g. Dutch learners of French studied by Eyckman (2004) had much higher false alarm rates than the Asian background ESL learner testees in Mochida and Harrington (2006); the nature of the instructions (Eyckman, 2004); the number of items and proportion of words to pseudo word items used (Beeckmans et al., 2001) and, of particular importance here, the proficiency level of the learner (Meara, 1996). More generally, test reliability and validity increases as proficiency levels increase, meaning that the sensitivity of the measure for less proficient learners remains a concern (Mochida and Harrington, 2006). While acknowledging the role these factors can play, the Yes/No test has been recognized as a valid estimate of L2 vocabulary knowledge (Read, 2000).

The aim of this study was to assess the effectiveness of the Yes/No test as a tool for placement decisions in a commercial language school. In this setting a test that is fast, easy to administer, and of satisfactory reliability and predictive power is desirable. The study was carried out at Milton College, an established English language school located in Sydney, Australia. Milton has campuses in Australia, Taiwan and China and offers a range of English language courses, including high school preparation, English language teaching and test preparation (IELTS, TOEIC, etc). The centre in North Sydney has around 300 students from
Asia, Europe and South America. The school has rolling enrolments with 6-10 new students beginning studies every Monday. New students are placed at one of six levels in the program based on the outcome of a two-hour placement test battery that consists of assessments of Listening, Grammar, Writing and Speaking.

The concurrent validity of the Yes/No format as a placement measure will be assessed by comparing performance on the test with the Milton placement tests. The Milton Listening and Grammar tests are objective discrete point tests while Writing and Speaking measures are based on a short writing sample and an interview that are rated by the teacher using the placement levels as criteria. The writing sample and interview are rated after the Listening and Grammar tests have been marked and the student’s relative strength in these areas is known. The focus of the study will be on how the Yes-No test compares with the Listening and Grammar tests as predictors of placement levels, given that the ratings assigned to the Writing and Speaking tests are essentially the placement decision.

Unlike previous applications of the Yes/No test, the current study assesses both accuracy (number of words) and response time as measures of lexical knowledge. Response time is a stable index of lexical processing skill (Segalowitz and Segalowitz, 1998), but its application in assessing individual differences in L2 proficiency has received only limited attention (Kempe and MacWhinney, 1996; Harrington, 2006). Of specific interest here is the relationship between placement level decisions and individual differences in response time, both as an independent dimension and in combination with accuracy performance. In behavioural terms including a time constraint on performance may also lessen the use of strategic responses on the part of the testee arising from the use of reflective knowledge (Eyckman, et al., 2007).

The Yes-No test is delivered on LanguageMAP, a multimedia assessment program designed to test language proficiency on-line. LanguageMAP is a java based web application consisting of a data base management system and an html test-authoring interface for the
production and administration of multi-media enabled tests. The program was developed at the University of Queensland, Australia, and is available at www.languagemap.com. The presentation of the test on-line allows flexibility in administration, and provides an environment for test presentation that is controlled and which many students find preferable to paper-and-pencil testing formats (Jamiesen, 2005).

2. Research questions

The study has two aims.

1. To assess the Yes/No Test format as a tool for placement in the Milton College program. The effectiveness of the test in discriminating between placement levels will be compared with that of the placement test currently used at the school.

2. To assess Yes/No response time performance as a predictive measure for placement purposes in this setting. Of interest is speed of response as a means to discriminate between learners at different placement levels, both as an independent measure and in combination with accuracy.

3. The study

3.1. Participants

The participants (n=88) ranged in age from 19 to 33 (mean= 24.3, SD = 3.8) with many intending to continue on to university study in Australia and elsewhere at the end of language study. The largest number of participants was from Korea (32) and Japan (18), with the remainder from 14 different first languages. Over half of the students were from first languages that were not alphabet based or cognate with English. A possible cognate effect on pseudoword performance has been noted in previous Yes/No test research (Huibregtse, et al., 2002).
3.2. Materials

3.2.1. Milton placement tests

Milton College’s battery of multiple skill assessments consist of a listening test, a three-part grammar test, a written essay task and an oral interview with the head teacher. The tests consist of the following.

1. Listening (10 minutes). This test assesses global listening comprehension. The student listens twice to a conversation between two speakers and completes a gapped text containing 34 missing word items.

2. Grammar (45 minutes). This is a three-part test assessing knowledge of English syntax (word order). Part one consists of 8 grammatically ill-formed sentences that must be re-ordered; part two requires the identification and correction of one grammar error in each of 8 sentences, and part three requires the completion of a gapped descriptive passage in which 18 grammatical features have been removed.

3. Writing (30 min). This test assesses mastery of basic sentence structure, discourse structure and expression in expository writing. The student responds in 120 words to an essay question on the topic, “Why did you choose to come to Australia?”

4. Speaking (5-10 minutes). This test is designed to determine the student’s accuracy in spoken grammar and vocabulary, as well as fluency and pronunciation. The student is interviewed with reference to a general list of questions related to family background, interests, career goals, etc.

Tests 1-3 are administered on paper. Listening and Grammar scores are scored for number of items correct. The Writing and Speaking are rated by the test administrator. Each sample is rated holistically on a 6-step scale that mirrors competencies appropriate to the respective Milton College program levels.
3.2.2 The Yes/No Test

The Yes/No Test consists of two lists of target vocabulary items, a General Vocabulary List and a Program Vocabulary List. Each list contains 72 words and 28 pseudowords. The latter were included to control for guessing. The word items in the General Vocabulary List were drawn from the frequency bands used in the Vocabulary Levels Test (VLT) (Schmitt, Schmitt and Clapham, 2001). The VLT is a widely used test of vocabulary size that uses a multiple matching format to assess the learner’s ability to match target vocabulary items and their appropriate definitions (Nation, 1990). Target words were selected from the four frequency bands in the test (n=18 for each frequency band), the 2000, 3000, 5000 and 10000 most frequently occurring words. Items from the Academic Word List in the VLT were not used.

The use of the VLT target words in the Yes/No placement test yields a measure that can be related both to the Milton placement decisions and provide an estimate of general English L2 vocabulary knowledge size that can be related to language performance more broadly (Adolphs and Schmitt, 2003; Meara, 2005; Nation, 2006). However, there is the possibility that a test based on items drawn from a standard test like the VLT may not be sensitive enough for the placement decisions made in a specific setting like Milton College. As a result, a Program Vocabulary List was also used. The Program List consists of 72 vocabulary items drawn from course books and materials used at Milton College. A list of content words from elementary to advanced levels of instruction at Milton was selected from recently used texts and materials. The items were selected by the second author, an experienced ESL practitioner and former teacher at Milton. In the absence of fixed vocabulary lists corresponding to program levels, a spread of item difficulty was ensured by selecting items from a spread of frequency bands using the British National Corpus http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html. Eighteen items each were selected at the 1000, 3000, 5000 and 10,000 word frequency of occurrence levels. Both the General and Program Lists each contained 28 pseudowords.
consisting of letter strings that were orthographically and phonologically permissible in English.

Overall Yes/No test accuracy was calculated using a correction for guessing formula in which the proportion of ‘Yes’ responses to word items (hits) was adjusted by subtracting the proportion of ‘Yes’ responses to pseudoword items (false alarms). Response time was the mean reaction times for each participant, calculated on correct word responses only.

3.3. Procedure

Milton College has a rolling admissions policy with new students commencing the program every week. Placement testing is carried out every Monday year round. The data collection period spanned 15 weeks with a total of 88 students participating. The same teacher administered all 15 placements. Yes/No test data were collected using LanguageMAP. Each student completed the Yes/No task first and then completed the Milton test battery off-line. Instructions for the Yes/No Test were translated into Korean, Japanese, Chinese, Spanish, Portuguese, and Czech. The other students received the instructions in English. The students were told that they would be presented with word items on the computer screen and asked to judge “Yes” or “No” as to whether they knew the word or not. They were told that “knowing” a word means being able to understand it in a simple sentence. They were also told that some words like ‘delider’ look like English words but are not actual words, and that a ‘Yes’ response to these pseudowords would be penalised. Items were presented individually. Each presentation was preceded by a fixation point to focus attention on the screen. The item remained on the screen until the student responded. If after 5000 milliseconds (msec) there was no response, the presentation word timed out and the next word appeared. A correct response resulted if a student responded “yes” to a real word (a hit) or “no” to a pseudoword (a correct rejection). An incorrect response was registered for a “no” response to a real word (a miss) or a yes response to pseudowords (a false alarm). The score for accuracy was then
calculated as the proportion of correct hits minus the proportion of false alarms (Mochida and Harrington, 2006). The Milton tests were administered and scored by the Milton College teacher.

4. Results

4.1. Milton College test results and placement outcomes

On the basis of the placement tests, students entering Milton College are placed at one of six levels depending on level of proficiency: Beginner, Elementary, Lower Intermediate, Upper Intermediate, Advanced and English for Academic Purposes. During the data collection only three students were placed in the EAP level. Thus, for analysis purposes these students were combined with the Advanced group resulting in five placement levels examined. The Beginner (n =10) and Elementary (n = 12) levels were less than half the size of the three higher levels, which ranged from 20-26 students each. The Milton College placement test results and the placement decisions will first be described, followed by a presentation of the Yes/No test results. The Milton placement findings and the Yes/No test results will then be compared as predictors of placement decisions. Finally, the interaction of accuracy and response time performance will be examined.

The results for the Milton College tests for the five placement levels are presented in Table 1. The Listening and Grammar tests are measures of listening comprehension and grammar knowledge performance on prepared materials. The Writing and Speaking tests are tasks administered and subjectively scored by the teacher in reference to the placement levels. The ratings and the placement decision were both done by the same teacher, with the ratings assigned according to the perceived capacity to function at the corresponding program level. Furthermore, the teacher assigns the ratings after scoring the Listening and Grammar tests and these results provide general guidance as to where the student should be placed. This relationship is evident in the strength of relationship between the individual measures and the
placement decisions. Table 2 presents correlations for the Spearman’s rho test which was used to assess the relationship between the interval and rating scores. The Writing and Speaking ratings both correlated with Placement at around .9, while the Grammar and Listening test scores were at .8 and .7 respectively.

The high correlation between the placement decisions and the Writing and Speaking ratings reflect the dependent relationship between the measures and placement decisions. The main focus will thus be on how Yes/No test performance compares with the Listening and Grammar results as the basis for placement decisions. These comparisons will be made after the descriptive statistics for the Yes/No test are presented.

4.2 Yes/No Test Results

The reliability of the two Yes/No lists was calculated using Cronbach’s alpha, a measure of internal consistency. Separate analyses were carried out for performance on the words and pseudowords for each list (Beeckmans et al., 2001). Both tests had a high degree of reliability. For the word results the coefficient estimate was .94 for the General list and .95 for the Program list, while pseudoword performance was at .85 and .86, respectively. Table 3 presents the scores, response times, and error rates (false alarms) for the Yes/No tests by Placement level. The three measures represent performance averaged across all frequency levels. Performance is reported by the respective General and Program Lists and by a Combined List, which is an average of the two. Error represents the false alarm rate, that is, the proportion of ‘Yes’ responses to pseudoword items. The score is the percentage of ‘Yes’ responses to word items minus the error rate.

As was the case with the Milton placement scores, the Combined Yes/No test accuracy means showed a systematic increase across the five placement levels, with the lowest for Beginners (mean = 23) and highest for Advanced (mean = 67). The response times were more varied. Although the Advanced learners responded approximately 600 msec faster than the
Beginner, the decrease as a function of higher placement level was not as systematic as the accuracy responses, with the Lower Intermediate Group showing a faster mean response time than the Higher Intermediate Group for the Combined list. The mean differences are tested for statistical significance in the next section.

Overall Mean performance on the Program list (60% corrected score) was superior to that on the General list (44%). A paired t-test showed the difference to be significant $t(85) = 10.55, p = .000$. Note two sets of results for the General List were missing for technical reasons. The analysis below will discuss both the respective list results as well as the Combined scores that pools both results.

The error rate ranged from 25%, for the Beginners in the Program list to 10% for the Advanced Level for all tests, with a mean overall rate of 17%. This was higher than the 5% reported in Mochida and Harrington, 2006, but somewhat less than the 25% reported in other studies (Eyckman, 2004; Cameron, 2002). For the other level and test combinations the error rate varied from 16% to 20% in a non-systematic way, suggesting that the false alarm rate is only correlated in an approximate way.

The error rate standard deviations were also very high, due to some learners having a 50% error rate, which is chance performance. High error rates pose a substantial threat to the reliability and interpretability of Yes/No test results. This will be discussed below.

4.3. Comparing performance on the Milton Placement Battery and the Yes/No Test as placement predictors.

Accuracy results. Figure 1 compares mean accuracy performance for the Yes/No Combined, Milton Listening and Grammar tests. The comparative change in the scores across placement levels is the focus here, and not the relative magnitude of the scores within a placement level (i.e., it is coincidental that the Listening and Vocabulary scores are both around 23% for the Beginner level).
The mean differences varied systematically across the Milton Placement and Yes-No test measures. The mean differences were tested using Kruskal-Wallis one-way analysis of variance, the nonparametric counterpart to the standard (parametric) one-way analysis of variance. A distribution-free statistic was used given the differences in n sizes across the placement levels and the small numbers in Beginner (n=10) and Elementary (n=12) groups. Reported in Table 4 are the results of the individual analyses for the accuracy scores for the Yes/No Tests (General, Program and Combined) and the Listening and Grammar results from the Milton placement tests. The response time results are examined in a separate analysis below. Overall mean differences for all five tests were statistically significant at $p < .000$. Mann-Whitney $U$ tests were also performed to compare pairwise differences. A Bonferroni adjustment for multiple comparisons set the significance level at $p < .008$ (.05/6 comparisons = .0083) for all comparisons.

The sensitivity to differences in adjacent placement levels was highest for the Milton Grammar test and lowest for the Yes/No General list. Examining the results by level, the difference between the Beginner and Elementary levels were significant for the Yes/No Program list and the Milton Grammar and Listening tests. The Elementary-Lower Intermediate differences were not significant for any of the tests and the Lower Intermediate-Upper Intermediate difference only reached significance for the Milton Grammar test. The difference between the two highest levels, Upper Intermediate and Advanced, was significant for the two Milton tests and the Yes/No Combined results.

A three-way contrast was also done comparing Beginners (Level 1) - Lower Intermediate (Level 3), and Lower Intermediate – Advanced (Level 5). For the lower level only the Yes/No General list was not significant and for the higher level all the tests yielded statistically significant differences at the .001 level.

To summarize the accuracy findings, the Milton tests were better overall than the Yes/No tests in discriminating between placement levels. The Grammar test differentiated
between three of the four adjacent level comparisons, the Listening test between two levels, Levels 1 and 2 and Levels 4 and 5; and Yes/No Combined list only between the latter, Levels 4 and 5. When the five levels were conflated to three, all the tests except the Yes/No General tests reliably discriminated between placement levels, with the latter able to only differentiate between the middle (Lower Intermediate) and the high (Advanced). As the Milton tests have been developed and regularly modified for placement purposes it is expected that they would be relatively more sensitive than the Yes-No results. However, what is striking is the very modest size of the difference in that sensitivity between the two types of tests. The Yes/No scores provide a very reasonable approximation to placement levels, and does so in a test that is significantly faster and easier to administer.

**Response time.** The response time means are based on correct responses to words. As word items were timed out after 5000 msec, individual response times were not screened for outliers. A log transformation was performed on the raw response times prior to statistical analysis. The interpretability of the response time performance depends on minimizing tradeoffs in speed and accuracy by individuals. A positive correlation between speed and accuracy, in which higher accuracy scores are accompanied by higher (=slower) response times would indicate that relative response speed is less a measure of proficiency than a reflection of individual differences in response style, i.e. slower and more accurate versus faster and less accurate. Table 5 presents bivariate correlations for the accuracy and response times. There is no evidence for a systematic trade-off in accuracy and speed.

Overall, the mean response times for the three Yes/No tests decreased as a function of placement level, with the Advanced level 400-600 msec faster than the Beginner Level (see table 3). However, the response time means were more variable, with mean response time differences not always reflecting a systematic decrease. As with the accuracy results, the mean response time differences were tested using a Kruskal-Wallis one way analysis of variance with placement as the grouping variable. All three tests were significant ($df = 4$ for all tests).
General List, chi-square = 10.19, \( p = .037 \); Program List, chi-square = 16.04, \( p = .003 \); and Combined List, chi-square = 13.75, \( p = .008 \). Mann-Whitney U tests of the level differences were also carried out but none of the comparisons carried out for the accuracy results were significant. A post hoc test comparing means for Level 1 and Level 5 was significant for the Program List and Combined List. General List, \( Z = -2.55, p = .010 \); Program List, \( Z = -2.88, p = .005 \); Combined List, \( Z = -2.88, p = .006 \).

The first aim of the study was to compare the effectiveness of the Yes/No test measures with the placement test battery currently used at Milton. The results above showed that the accuracy results approximated the Milton Listening and Grammar results as discriminators of placement levels. A table of bivariate correlations showing the strength of association among the various measures is presented in Table 6.\(^1\) The Milton tests have a stronger correlation with the placement level decisions \( .72 – .80 \) than the Yes/No accuracy scores \( .54 – .64 \), but given the former tests have been developed and modified specifically for Milton placement purposes, the strength of the Yes/No test correlations is notable. The better the student scored on the placement test, the faster they responded on the Yes/No test, as evident in the inverse correlation between these measures. There was also a consistent strength of association between the General, Program and Combined scores and the Milton Listening, Grammar, Writing and Speaking tests. This reflects the fact that recognition vocabulary knowledge is a basic foundation for performance across all domains (Koda, 1996), and a dimension that can be effectively measured for placement purposes.

The second aim of the study was to examine response time differences as a discriminator of placement levels, both as an independent measure and in combination with

\(^1\) The high level of intercorrelations between the test scores does not make it possible to examine all the results in a single regression model. The standard caution about interpreting bivariate correlations applies.
accuracy. As an independent measure, response time proved to be much less sensitive to differences between levels and had a weaker overall association with the various placement scores. A possible effect for response time in combination with accuracy was assessed by hierarchical regression analyses of the three Yes/No tests in which placement level was the criterion variable and the accuracy score was entered as the first predictor variable followed by response time. Of interest was the degree to which response time accounted for additional unique variance in the placement variable. A summary of the results are presented in Table 7.

Response time accounted for an additional unique amount of variance of around 5% in all three vocabulary measures, indicating that speed of response does improve the predictive validity of the Yes/No test.

5. Discussion
5.1 The Yes/No test as a placement tool.

This study evaluated a computerised version of the Yes/No test as a tool for language program placement decisions. The concurrent validity of Yes/No accuracy and response time measures was established by comparing test performance with that of a battery of placement tests used to make weekly placement decisions at a language school. Correlations with the individual test results and the resulting placement decisions showed that Yes/No test performance serves as a sensitive measure of differences in learner proficiency. There was a highly consistent correlation between the Yes/No vocabulary scores and the Listening, Grammar and Writing tests, indicating that recognition vocabulary knowledge is a dimension fundamental to performance on these tests, and L2 proficiency in general (Koda, 1996). The correlation was not as strong for the Speaking ratings but was also substantial. Note that this stable pattern of differences across levels was evident despite the small numbers in the Beginning and Elementary placement groups.
Despite the patterns that emerged there was also significant variability within the levels for both accuracy and response time performance. On the whole this variability decreased as proficiency levels increased, but it is evident that individual differences in task performance means the relationship between Yes/No test performance and more global language proficiency, reflected here in program levels is not uniform across learners.

The Yes/No test format has potential application in placement testing programs like the one run at Milton College. Given the speed of the format and the advantages of computer-driven testing, the test warrants further attention as an alternative or complement to existing measures. The results do not suggest that the Yes/No test can replace the global assessment of the student’s proficiency available through spoken and written production, though it can possibly reduce the need for extensive testing in other areas. The test format also provides a reliable means to assess L2 lexical knowledge, which is a fundamental dimension of language proficiency. The format may be usefully applied in other types of “low-stakes” testing (Read and Chappelle, 2001), where test outcomes do not have the longer term consequences of gatekeeper tests like IELTS, or where the testing is done for diagnostic purposes (Hsueh-chao and Nation, 2000).

Superior performance on items in the Program list demonstrate the usefulness of using items adapted from program materials over those from a standard test. This most likely reflects the difference between items drawn from a pool of ‘textbook’ English, as in the Program list, versus those sampled more broadly, as in the case of the VLT. Note that the Academic Word List items (Coxhead, 2000) in the VLT were not included in the test. The Program word list also contained more higher frequency words. The inclusion of frequency bands in test content provides the means to estimate the student’s vocabulary size in addition to placement decisions. These estimates can be related to language performance in other areas (Adolphs and Schmitt 2003; Meara, 2005; Nation, 2006) in a way that students and teachers alike find useful. The sensitivity of approaches like the lexical frequency profile (Laufer and Nation, 1995;
Milton, 2007) in making the relatively fine grain distinctions needed for placement purposes has been challenged (Meara, 2005; for a reply see Laufer, 2005), but the findings here indicate that continued work in the area is warranted.

Response time measures of proficiency also deserve further attention. Increasing lexical retrieval speed subserves the development of L2 fluency in quantitative and possibly even qualitative terms (Segalowitz, Segalowitz and Wood 1998; Snellings, et al., 2002; Harrington, 2006). The implications of this research for the type of testing examined here needs further exploration. The increasing use of computer-controlled tests makes the control and manipulation of these variables increasingly possible, even in everyday classroom contexts.

A key question concerns what response time values tell us about learner proficiency. Higher accuracy generally correlates with faster responses, though the findings here showed that response time was not a very sensitive discriminator of placement levels. However, the combination of word accuracy and response time did improve the predictive power of the Yes/No measure, suggesting a role for response time measures in placement testing. It is possible that stable individual differences in response time will only be evident in more advanced learners, thus limiting potential applications in placement settings that include lower proficiency learners (Harrington, 2006). More generally, the results draw attention to the need for a better understanding of how accuracy and response time interact in the development of L2 proficiency (Segalowitz, 2003). Proficiency in the current study was operationalized as language skills (grammar, writing, listening and speaking) at respective placement levels in a language program.

5.2 Methodological issues

The use of pseudowords sets the Yes/No test apart from other vocabulary tests used in L2 testing and instruction. Performance on pseudowords, as reflected in the false alarm rate, is used to correct for guessing. The false alarm rate for a given testee is assumed to reflect, in
part, the decision criterion adopted by the individual as to when a word is accepted as known. This criterion is affected by motivational, attitudinal and situational factors arising from the testing purposes and the test session. A very conservative strategy results in a ‘yes’ response being given only when the testee is very confident that the presented item is a known word. This results in a low false alarm rate but also in a probable underestimation of the testee’s knowledge of actual words (Mochida and Harrington, 2006). Alternatively, a liberal response strategy would result in more ‘yes’ responses to words and pseudowords. The various methods used to correct for guessing attempt to control for the latter, but provide no insight into potential underestimation of performance (Huibregste et al., 2002).

The effectiveness of the Yes/No format as a measure of L2 vocabulary depends on being able to adequately control for response bias. This ensures the resulting score is reliable and related to L2 performance in a meaningful way. Research to date, including the present study, has shown that the degree to which Yes/No test performance meets these two aims varies by setting and purpose. Eyckman (2004) has proposed an alternative to the Yes/No format that avoids the potential response bias inherent in Yes/No responses to single items, while retaining the other advantage of the checklist approach. In the Recognition-based Vocabulary Test word-pseudoword pairs are presented and the testee chooses which one is a word. While avoiding the response bias problem, the various factors arising from the use of pseudowords remain (van Ee, 2007). It is possible that pseudowords are not necessary at all, and that the appropriate instruction set will prove adequate to minimise guessing, especially for particular learner backgrounds and settings (Shillaw, 1996).

The use of response time as an independent measure differentiates the current study from other applications of the Yes/No test. As noted, response time alone was not a very sensitive measure for the participants in this study, but did provide additional information when combined with the accuracy results. The lack of sensitivity may be due in part to using only the mean response times as a measure of processing performance. Baseline information
processing rates vary across individuals, and an inclusion of a baseline processing measure, in
the L1 or L2, may provide a way to better assess the degree to which response time differences
are affected by task-specific factors (Faust, Balota, Spieler and Ferraro, 1999).

In this study time was both a behavioural measure and a task condition. The testees
were instructed to respond as accurately and quickly as possible. The time condition was
included to examine response latency as a dependent measure and to minimise the use of
strategic responses arising from the use of reflective knowledge. These are related but
independent aims and it may be that the presence of a time constraint alone, whether controlled
by the testee or the test itself, results in a more sensitive measure of the recognition vocabulary
skills needed in settings like the one examined here.

6. Conclusion

Placement testing is an important but time consuming activity in language programs
like Milton College. The purpose of this study was to assess the Yes/No test as a tool in
placement testing. Of interest was both the effectiveness of the test format and the vocabulary
recognition measure in providing a level of discrimination between learners that was fine
enough to inform placement decisions at the school. The Yes/No test yielded results similar to
the Milton placement battery, especially for the Listening and Grammar tests. The use of
program based vocabulary items, in particular, approached the sensitivity of the placement
tests, indicating that it can play a role in placement testing. When combined with an entry
interview, where the students’ learning goals are elicited and an impression of the students’
spoken fluency is gained, an on-line Yes/No test like LanguageMAP offers an effective
alternative and complements conventional multi-skills placement test. The test may also have a
role in estimating the proficiency levels of prospective students located offshore, for use in
both the enrolment and placement process.
Acknowledgements

This research was funded by a research contract between Milton College and Uniquest Pty, Ltd, University of Queensland.
References


Table 1
Milton College placement test battery scores by placement level

<table>
<thead>
<tr>
<th>Placement Level</th>
<th>Placement Tests</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Listening 34 items % correct</td>
<td>Grammar 34 items % correct</td>
<td>Writing 120 word essay Rating 1-6</td>
<td>Speaking 5-10 dialogue Rating 1-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Beginner (n = 10)</td>
<td>22.30</td>
<td>15.71</td>
<td>20</td>
<td>10.12</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Elementary (n = 12)</td>
<td>45.00</td>
<td>12.00</td>
<td>34.17</td>
<td>7.79</td>
<td>2</td>
<td>0.47</td>
</tr>
<tr>
<td>Lower Intermediate (n = 20)</td>
<td>55.15</td>
<td>13.05</td>
<td>40.40</td>
<td>7.78</td>
<td>2.9</td>
<td>0.45</td>
</tr>
<tr>
<td>Upper Intermediate (n = 26)</td>
<td>61.15</td>
<td>14.87</td>
<td>51.15</td>
<td>8.47</td>
<td>3.88</td>
<td>0.52</td>
</tr>
<tr>
<td>Advanced (n = 20)</td>
<td>76.55</td>
<td>10.67</td>
<td>62.80</td>
<td>15.91</td>
<td>4.95</td>
<td>1.03</td>
</tr>
<tr>
<td>Overall (n = 88)</td>
<td>56.67</td>
<td>20.77</td>
<td>45.59</td>
<td>16.72</td>
<td>3.3</td>
<td>1.41</td>
</tr>
</tbody>
</table>
Table 2

Spearman’s rho correlations for Milton test scores and placement

<table>
<thead>
<tr>
<th></th>
<th>Grammar</th>
<th>Listening</th>
<th>Writing</th>
<th>Speaking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placement</td>
<td>.795</td>
<td>.718</td>
<td>.906</td>
<td>.858</td>
</tr>
<tr>
<td>Grammar</td>
<td></td>
<td>.656</td>
<td>.842</td>
<td>.704</td>
</tr>
<tr>
<td>Listening</td>
<td></td>
<td></td>
<td>.781</td>
<td>.610</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td></td>
<td></td>
<td>.809</td>
</tr>
</tbody>
</table>

All correlations are significant at the 0.01 (2-tailed).
Table 3
Yes/No Test accuracy by placement level. Percentage means, (standard deviations) for scores, response times, and error rates for the three vocabulary measures by placement level

<table>
<thead>
<tr>
<th>Level</th>
<th>General List</th>
<th>Program List</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score %</td>
<td>Time msec</td>
<td>Error %</td>
</tr>
<tr>
<td>1. Beginner (n = 10)</td>
<td>22 (21)</td>
<td>1981 (603)</td>
<td>17 (23)</td>
</tr>
<tr>
<td>2. Elementary (n = 12)</td>
<td>39 (12)</td>
<td>1752 (647)</td>
<td>18 (17)</td>
</tr>
<tr>
<td>3. Lower Intermediate (n = 20)</td>
<td>42 (14)</td>
<td>1507 (324)</td>
<td>20 (14)</td>
</tr>
<tr>
<td>4. Upper Intermediate (n = 26)</td>
<td>46 (15)</td>
<td>1606 (399)</td>
<td>17 (18)</td>
</tr>
<tr>
<td>5. Advanced (n = 20)</td>
<td>57 (15)</td>
<td>1375 (222)</td>
<td>10 (14)</td>
</tr>
<tr>
<td>Overall (n = 88)</td>
<td>44 (18)</td>
<td>1593 (451)</td>
<td>16 (17)</td>
</tr>
</tbody>
</table>

Error rate represents the false alarm rate, proportion of ‘Yes’ responses to pseudowords items. 28 pseudoword items each for General and Program; 56 items for Combined.

Score = Percentage of ‘Yes’ responses to word items (72 word items each for General and Program Lists; 144 items for Overall) – Error rate.
Table 4
Kruskal-Wallis one way analysis of variance for Yes/No test accuracy scores (General, Program and Combined) and the Milton Placement tests (Listening and Grammar) and Mann-Whitney U tests of placement level differences for the tests.

Kruskal-Wallis one way analysis of variance. df = 4 for all tests

<table>
<thead>
<tr>
<th>Yes/No Test</th>
<th>Milton Placement Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>Program</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>Sig</td>
</tr>
<tr>
<td>26.04</td>
<td>.000</td>
</tr>
<tr>
<td>55.68</td>
<td>.000</td>
</tr>
</tbody>
</table>

Mann-Whitney U statistic.

<table>
<thead>
<tr>
<th>Level comparisons</th>
<th>Yes/No Test</th>
<th>Milton Placement Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General</td>
<td>Program</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>Sig</td>
</tr>
<tr>
<td>1. Beginner – 2. Elementary</td>
<td>-1.85</td>
<td>.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Elementary – 3. Lower Intermediate</td>
<td>-0.05</td>
<td>.567</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Lower Intermediate – 4. Upper Intermediate</td>
<td>-1.16</td>
<td>.245</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Lower Intermediate – 5. Advanced</td>
<td>-3.08</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Upper Intermediate – 5. Advanced</td>
<td>-2.44</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All significance levels are for asymptotic significance (2-tailed)
Table 5

Spearman’s rho correlations for Yes/No test accuracy scores and response times.

<table>
<thead>
<tr>
<th></th>
<th>Program Accuracy</th>
<th>Combined Accuracy</th>
<th>General Time</th>
<th>Program Time</th>
<th>Combined Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>General List Accuracy</td>
<td>0.82**</td>
<td>0.95**</td>
<td>-0.12</td>
<td>-0.34**</td>
<td>-0.25*</td>
</tr>
<tr>
<td>Program List Accuracy</td>
<td>--</td>
<td>0.95**</td>
<td>-0.22*</td>
<td>-0.69**</td>
<td>-0.42**</td>
</tr>
<tr>
<td>Combined List Accuracy</td>
<td>--</td>
<td>--</td>
<td>-0.19</td>
<td>-0.41**</td>
<td>-0.33**</td>
</tr>
<tr>
<td>General List Response</td>
<td>--</td>
<td>--</td>
<td>0.69**</td>
<td>0.91**</td>
<td></td>
</tr>
<tr>
<td>Program List Response</td>
<td>--</td>
<td>--</td>
<td></td>
<td>0.91**</td>
<td></td>
</tr>
</tbody>
</table>

**Significant at the 0.01 level (2-tailed); * Significant at the 0.05 level (2-tailed).
Program List and Combined List, n = 88; General List, n = 86.
Table 6
Spearman’s rho correlations for Yes/No test accuracy and response time scores, Milton test scores and placement.

<table>
<thead>
<tr>
<th></th>
<th>Milton Entrance Test Battery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Placement</td>
</tr>
<tr>
<td>Placement</td>
<td>--</td>
</tr>
<tr>
<td>General List Accuracy</td>
<td>.54**</td>
</tr>
<tr>
<td>Program List Accuracy</td>
<td>.64**</td>
</tr>
<tr>
<td>Combined List Accuracy</td>
<td>.63**</td>
</tr>
<tr>
<td>General List Response Time</td>
<td>-.42**</td>
</tr>
<tr>
<td>Program List Response Time</td>
<td>-.43**</td>
</tr>
<tr>
<td>Combined List Response Time</td>
<td>-.38**</td>
</tr>
</tbody>
</table>

**Significant at the 0.01 level (2-tailed).
* Significant at the 0.05 level (2-tailed).
Table 7
Summary of hierarchical regression analyses of Yes/No test results with placement as criterion variable and accuracy and response time as ordered predictor variables.

<table>
<thead>
<tr>
<th>Test</th>
<th>n</th>
<th>order of entry</th>
<th>R²</th>
<th>Standard error of estimate</th>
<th>R² change</th>
<th>F change</th>
<th>Sig F change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>88</td>
<td>a</td>
<td>.386</td>
<td>1.02</td>
<td>.386</td>
<td>54.03</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>.430</td>
<td>.968</td>
<td>.044</td>
<td>6.53</td>
<td>.012</td>
</tr>
<tr>
<td>Program</td>
<td>88</td>
<td>a</td>
<td>.376</td>
<td>1.02</td>
<td>.376</td>
<td>51.87</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>.427</td>
<td>.988</td>
<td>.051</td>
<td>7.55</td>
<td>.007</td>
</tr>
<tr>
<td>General</td>
<td>86</td>
<td>a</td>
<td>.288</td>
<td>1.09</td>
<td>.288</td>
<td>34.00</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>.342</td>
<td>1.06</td>
<td>.054</td>
<td>6.79</td>
<td>.011</td>
</tr>
</tbody>
</table>

a = accuracy as 1st predictor variable
b = response time as 2nd predictor variable
Figure 1. A comparison of mean accuracy on the Yes/No Combined and Milton Listening and Grammar tests by placement level.