

Fairclough, M. (2011). Testing the lexical recognition task with Spanish/English bilinguals in the United States. *Language Testing*, 28, 273-297.

1. Introduction

- The increase in the number of Hispanics in the U.S universities needs to create more reliable and valid placement test for them into right Spanish programs and levels.
 - There are two kinds of learners of Spanish: second language (Spanish) learners (SLL) and heritage language learners (HLL).
 - Heritage language learner is a student (a) who is raised in a country within a non-English language, (b) who speaks or merely understands the heritage language, and (c) who is to some degree bilingual in English and the heritage language (Valdés, 2000).
- Practicality is one of the key issues in creating a placement exam (Bachman & Palmer, 1996). A test ideally requires a short time to complete, computerized one, automaticity, and instant scoring.
- From the view point of practicality, there are some arguments for using vocabulary testing (Meara's Yes/No Vocabulary Tests, lexical recognition test: Meara, 1996) as a basis for language placement.
 - Several studies found a strong correlation between
 - (a) Yes/No lexical recognition test and other vocabulary tests (Mochida & Harrington, 2006)
 - (b) lexical knowledge and reading comprehension (Koda, 1989; Laufer, 1992)
 - (c) lexical knowledge and listening comprehension (Kelly, 1991; Mecarty, 2000)
 - Vocabulary size is a useful predictor of abilities in English (Schmitt et al., 2000).
- However, there is little research that examined the use of lexical recognition as a measure of Spanish language proficiency of students at the university level.
 - This study addressed the following research question: Is a lexical recognition test using the Yes/No format a reliable and valid measure of general Spanish proficiency?
 - More specifically, this research focused on: (a) test quality, (b) differences in performance between SLL and HLL, (c) differences of performance among the levels of each group, and (d) correlation with more general proficiency measures.

2. Methodology

2.1 Participants

T1 Distribution of the Participants by Group and Level

Level	HLL	SLL	Total
First year	28 (15.3%)	95 (64.6%)	123 (37.3%)
Second year	89 (48.6%)	22 (15.0%)	111 (33.6%)
Third year	38 (20.8%)	17 (11.6%)	55 (16.7%)
Fourth year	28 (15.3%)	13 (8.8%)	41 (12.4%)
Total	183 (55.5%)	147 (44.5%)	330 (100%)

→ A control group (16 graduate students of Spanish/English bilingual) was set.

2.2 Tasks

a. *Lexical recognition test*: 120 words were randomly selected from a Spanish dictionary on the basis of frequency (1,000- to 5,000-word levels); each level had 24 words (5 levels × 24 words = 120). Moreover, 80 pseudowords were added.

b. *Measure of general language proficiency*: About half of the students completed a cloze test, and the others took the multiple-task test.

(1) The cloze test consisted of a paragraph in which every fifth word was omitted (23 items).

(2) The multiple-task test had four tasks:

- Partial translation (10 items)
- Dictation (20 items)
- Fill-in-the-blank task (30 items)
- Multiple-choice task to measure grammatical knowledge (25 items)

2.3 Scoring

- In the cloze test and multiple-task test, all acceptable answers were counted as 1 for correct.
- The lexical recognition test was scored as Figure shows:

		Response pattern	
		Yes	No
Item	Target word	Hit	Miss
alternative	Pseudoword	False alarm	Correct rejection

Figure 1. The item-response matrix of the test. Hit and Correct rejection were regarded as correct (1 point) and the others were incorrect (0 point).

- To adjust the results taking into account a random guess, this study used Index of Signal Detection (I_{SDT}).

$$I_{SDT} = 1 - \{4h(1-f) - 2(h-f)(1+h-f) / 4h(1-f) - (h-f)(1+h-f)\}$$

h = hit rate, f = false alarm rate.

3. Results

T2 Mean Proportions and Standard Deviations (SD) for Target Words Based on Frequency Level

Frequency level	HLL (n = 183)	SLL (n = 147)	Total (n = 330)	Control (n = 16)
1,000	.96 (.05)	.73 (.23)	.86 (.13)	1.00 (.01)
2,000	.92 (.09)	.56 (.30)	.76 (.18)	1.00 (.01)
3,000	.86 (.10)	.39 (.20)	.65 (.13)	.99 (.02)
4,000	.80 (.22)	.34 (.26)	.60 (.22)	.99 (.03)
5,000	.78 (.15)	.31 (.21)	.57 (.16)	.97 (.07)

- The internal consistency of the test was very high (Cronbach's $\alpha = .972$).
- This Table shows that the higher the frequency level, the higher the mean for all groups; however, there were significant differences between HLL and SLL.

T3 Results of the Lexical Recognition Test: Means of the Two Groups by Level (I_{SDT} , MPS = 1.00)

Group	n	Mean	SD	Min	Max
HLL					
First year	28	.44	.14	.22	.81
Second year	89	.68	.12	.37	.90
Third year	38	.70	.15	.36	1.00
Fourth year	28	.75	.16	.37	.99
Total	183	.66	.17	.22	1.00
SLL					
First year	95	.32	.10	.07	.72
Second year	22	.39	.09	.19	.57
Third year	17	.48	.11	.35	.68
Fourth year	13	.59	.09	.43	.75
Total	147	.37	.13	.07	.75

- The mean scores increase with proficiency level.
- However, one-way ANOVA demonstrated that there were no significant differences among three proficiency levels in HLL group (i.e., First < Second = Third = Fourth). This is considered as a ceiling effect.

[Correlation between the lexical recognition test and general language proficiency]

- Cloze test × Lexical recognition test
 - The Pearson coefficient indicates strong correlations ($r = .872$)
 - Correlations for each group were still high ($r = .786$ for HLL, and $r = .674$ for SLL).
- Multiple-task test × Lexical recognition test
 - The Pearson coefficient indicates slightly lower correlations than those of Cloze test ($r = .792$).
 - Correlations for each group were moderate ($r = .584$ for HLL, and $r = .444$ for SLL).

4. Discussion and Conclusions

4.1 Test quality

- The quality of the test was confirmed by the results; the higher the frequency level was, the higher the means for all groups were. Moreover, the alpha value was very high.

4.2 Difference in performance between the two groups

- This study showed SLL got lower mean scores than HLL, which indicates that HLL had been exposed to Spanish much more than SLL, increasing the number of target words they recognized.

4.3 Difference in performance among the levels of each group

- The results demonstrated that a steady increase in the means from one proficiency level to the next (see T3) and the test was able to discriminate between HLL and SLL.
- These findings provide supporting evidence for the validity of the test. However, a 5000-word lexicon was not sufficiently large to clearly distinguish between the more advanced levels of HLL.
- Therefore, a wider range of words is necessary to avoid the ceiling effect.

4.4 Correlation with more general proficiency measures

- The findings from this study suggest a relation between vocabulary size and the results of the other types of tests (cloze test and multiple-task test).
- Considering the strong points of the lexical recognition test (e.g., this test is easier and faster to conduct, computer scores are instantly available, scoring is completely objective), these can make the test more practical.

<Comments>

The strong relationship between the Yes/No vocabulary test and the general L2 proficiency tests was found in this study. This indicates that the Vocabulary test can be used as a placement test instead of

other language proficiency tests, theoretically. However, we consider that this kind of tests will have negative wash-back effects; particularly, L2 learners focus on only mechanical vocabulary learning (e.g., they learn only word form from a word list/card). Therefore, whether Vocabulary tests as a placement test should be used or not is carefully argued.