

“Researching EFL Learners’ Proficiency through the Yes/No Vocabulary Test¹¹”

MONTESERÍN, ANABEL

ZINKGRÄF, MAGDALENA

Facultad de Lenguas,

Universidad Nacional del Comahue

Abstract

The Swansea Levels Test was administered to 33 university EFL students and 12 primary school learners of English. We also analysed their performance in a comprehension task and an elicited production task involving passive constructions and unaccusative verbs. Some subjects scored high in the vocabulary test and in the tasks, while others revealed an inverse relation between vocabulary knowledge and task performance. This paper presents the pros and cons of this vocabulary test to predict second language proficiency.

Introduction

In the last decades the study of vocabulary acquisition of foreign language learners has gained considerable significance. Researchers are only now beginning to acknowledge that knowing a word implies much more than just its meaning and that various types of vocabulary knowledge come into play in language production and comprehension. Vocabulary acquisition has been researched widely in connection to most of the macro-skills (Reading, Writing, Listening and Speaking) and has been found to correlate significantly especially with the receptive skills. The hypothesis has been put forward that through measuring learners’ vocabulary

¹¹ This paper presents results obtained within the Research Project (J014) *Estructuras tempranas y tardías en el desarrollo del lenguaje*, directed by Adriana Alvarez and subsidised by Universidad Nacional del Comahue.

knowledge one could quite safely predict their overall competence in the foreign language. This paper will examine this interaction.

1. Objectives

The overall aim of the research design was to ascertain EFL subjects' acquisition stage of passive constructions and unaccusative verbs¹² in English through their performance in a) a sentence- transformation activity and translation exercise in the case of passives and b) an elicited imitation task. To determine participants' overall proficiency level independently from task performance, Meara and Milton's Swansea Levels Test (2002) was administered on the same day they completed the other tasks.

2. Participants

2.1. University students

Thirty-three university teacher- and translator- trainees taking Syntax 1 (Second Year) participated in this research project after having successfully completed their first year English Language subject. Subjects' ages ranged between 18 and 25 (mean age: 20; 4). This group completed the written elicited production task.

2.2. Primary school students

Twelve 5th graders at a bilingual primary school in General Roca, Río Negro, participated in both the written and the oral elicited production tasks. Subjects' mean age was 9; 7, ranging between 9 and 12. Learners in this context are exposed to English during three 40-minute periods five times a week.

3. Methodology

3.1. The Swansea Levels Test (SLT)

This computerised version of a vocabulary recognition test (Meara and Milton, 2002) measures knowledge of vocabulary by keeping score of how many words are visually recognised by test-takers and requires learners to click on either a Yes or a No key depending on whether they know the meaning of the word shown on the screen or not. A hundred and twenty random items in isolation belonging to a) five bands of lexical items of different degrees of frequency in the foreign language (20 items each) and to b) pseudo-words collected in Meara (1992) are presented one at a time.

Results are obtained by calculating the number of hits ('Yes' responses to real words) and correct responses ('No' responses to pseudo-words), which are then

¹² Unaccusative verbs denote non-agentive events, like *exist*, or a change of state, like *fall* and *bloom*. They take only one argument, the role of which is that of "theme" or "patient". Like in passive constructions, this argument behaves as the syntactic subject of the sentence (although semantically it is not).

adjusted according to false alarms ('Yes' responses to pseudo-words) and misses ('No' responses to real words).

Each test-taker's performance is shown through a) an unadjusted raw score; b) an adjusted score and c) a profile of their answers for each of the 5 frequency bands and for the number of errors they score.

3.2. Experimental design

The experimental design included a written and an oral task that complemented each other in the type of data gathered: while the written task provided researchers with information of learners' use of passive truncated and/or full constructions and unaccusative verbs in a controlled setting, the oral task showed more spontaneous language use, even when some prompts were given.

3.2.1. Written elicited production task (WT)

Participants were required to complete 18 test items organised into three types of activities depending on the structure studied:

- 1) Transform 12 sentences in the active voice (describing an accompanying picture) into its passive counterpart starting each sentence with a given subject and using the appropriate form of the verb in brackets.
- 2) Answer the question "What's going on?" describing the situation in the adjoining picture making use of the unaccusative verb provided.
- 3) Write a possible L1 version for each of their answers (passives and unaccusatives) to ensure participants' understanding of the situation and to show their L1 knowledge of the structures.

The following verbs were selected:

Passive constructions		Unaccusative verbs		
follow	bite	arrive	appear	melt
find	push	leave	float	fall

The verbs studied in the passive constructions were presented in two conditions each: a) plural subject + singular object (*The boy and the girl pushed Tom*) and b) singular subject + plural object (*The boy found Tom and his sister*). Two of the verbs were irregular and two were regular. Subjects in the cases of the verbs *follow* and *find* varied according to the feature [+/- human] (*The cat found the parrot and the bird*).

3.2.2. The oral deferred imitation task

Participants listened to two sets of 5 and 6 situations respectively described by one researcher while looking at some pictures illustrating the actions. After each stretch, learners retold the events as they were shown the corresponding picture. Before the test proper, participants were taught how the elicited imitation task

worked with three additional verbs that were not unaccusative (*sleep, eat and kick*).

Learners were interviewed one at a time and their responses were recorded. Each interview took approximately 10 minutes. Because they need to finish looking at one whole set of pictures before turning to their own production, the time elapsed between the stimulus and their response is long enough for them not to be able to “imitate” the stimulus unless equipped with the linguistic resources to do so. Learners’ production was elicited through the question “What’s going on here/in this picture?” Below is a table with the verbs selected for this task (those in bold-face type were also used in the written task).

Unaccusative verbs					
melt	float	arrive	Stand	break (2)	slip
Fall		appear	leave/go home	fly	live

4. Results

Results are presented in terms of groups of participants.

4.1. University group’s performance

Taking into account the results in the written task for both structures analysed and the scores obtained in SLT, we have divided participants into four groups ranging from higher to lower achievers. For reasons of space, only four out of the 33 will be analysed to illustrate the characteristic performance of these subgroups in Table 1. For each of them the number of participants within the group has been included together with the mean adjusted score performance and the mean number of errors. The final exam column indicates the number of participants in the group who have passed the final exam for their English I course. Within SLT, a) represents the raw score; b) the adjusted score and the following figures (B1 to Error) characterise each learner’s profile. In the last two columns appear the rates obtained for both structures under study, 36 and 18 being their maximum scores. Each answer for these structures was rated according to the following scale: correct answers got 3 points and 1 point was deducted for each incorrect aspect in the structures analysed (word order, agreement, auxiliary verb used and/or form of the main verb)¹³.

Part per group	Mean adj score	Mean error	Part. N ^o	Final exam	Swansea Levels Test						Passives (n=12)	Unacc. verbs (n=6)	
					a) Error	b)	B1	B2	B3	B4			B5
n= 6	4457,698	0	32	5/6	4800:	4550:	20	20	20	19	17:	36	18

¹³ Only correct instances of the structures studied were considered, while typical mistakes and other aspects of their production were excluded.

n= 20	3848,06 3	1	31	7/20	4750: 4250: 20 20 20 18 17: 2	35	16
n= 6	3092,07 9	3	20	2/6	3050: 3050: 20 14 11 12 4: 0	33	18
n= 1			18	0	4900: 1900: 20 20 19 19 20: 12	35	18

Table 1: Subcategories according to scores in SLT

Table 1 shows that more than two-thirds fall into the Intermediate and Advanced levels in terms of their SLT scores and seven and five participants in these levels have passed the English I final exam¹⁴. Accordingly, the number of errors increases as the mean adjusted score decreases. Participants' profiles in SLT present what Meara and Milton (2002:5) characterise as "a normal cline" from higher to lower scores as the frequency levels for each band fall. In the first three sub-groups the number of pseudo-words which has been accepted as real words is considerably small (3 or less), which is directly related to the expected scores for the last 2 bands. For participants like 18 who claim to know 19 and 20 words out of the 20 items presented for B4 and B5 and register a high number of errors, the results may not be reliable because s/he may be "guessing far too much to allow his or her level to be confidently assessed" (Meara and Milton, 2002:6). In such cases, it is advisable to administer the test again.

Results means for passives and unaccusative verbs are 34, 44 and 17, 50 respectively, which clearly show that, overall, participants' performance in the tasks is almost error-free. These university students' awareness of syntactic forms and constructions might account for these scores in such a controlled-production activity. Yet, the WT results correlate significantly with the SLT scores for all the subgroups, proving its usefulness in placing these EFL learners in corresponding proficiency levels.

Test-taker 20 shows an expected normal cline which renders his SLT performance reliable though at a lower level as that of the rest of his/her subgroup.

4.2. 5th Form participants' performance

Responses were rated according to accurate forms of the analysed structures, included in Table 2¹⁵. There is a clear correspondence between years of exposure, correct answers in tasks and scores for the yes/no vocabulary test, especially in tandem with their profiles.

Table 2 shows students arranged in three different sub-groups from higher to lower SLT scores. As in the case of university students, the number of errors suddenly soars with decreasing levels of proficiency. The abrupt increase occurs between participants B and J, who scored 1 and 7 errors respectively.

¹⁴ This factor was considered an independent indicator of learners' recognised proficiency level at the institution. Lack of a passed final exam does not necessarily mean that the student has not reached the desired standard, but rather that s/he has chosen to postpone sitting for it.

¹⁵ The data were rich and enlightening in terms of the different types of errors organised mainly around a) correct or incorrect use of the auxiliary verb 'to be'; and b) different forms of the main verb (*The ice-cream is melt; *The moon appear, *The parrot and the bird are foundid from a cat) but for reasons of space, they are not presented here.

For participants K, L, F and C, the low scores obtained for SLT and those for correct instances of the structures (see footnote 4) studied do not seem to correlate. We can interpret this mismatch in two ways. Participants may not have been aware of the existence of pseudo-words (in spite of the fact that this was part of the instructions given before the test was administered) and the test-taking procedure failed to capture relevant data as regards vocabulary size. In these cases, the test responses provide no basis for making a meaningful estimate of her/his vocabulary size and general proficiency level.

Part	Years of exposure	Swansea Levels Test					Passives	Unacc. verbs			
		a) Error	b)	B1	B2	B3		B4	B5	WT OT	(n=12) (n=11)
H	6	3650: 1	3400:	18	18	12	12	13:	2	1	8*
E	9	3100: 1	2850:	17	10	14	12	9:	4	1	--
B	4	2700: 1	2450:	15	13	9	12	5:	0	6	7
J	7	3600: 7	1850:	20	13	12	14	13:	12	6	9
A	8	4150: 10	1650:	19	16	17	15	16:	3	0	3
G	6	3250: 7	1500:	18	15	13	10	9:	4	1	6*
I	6	2950: 6	1450:	16	11	12	12	8:	7	0	7*
C	8	3700: 11	950:	17	15	15	14	13:	3	2	9
D	2	3250: 10	750:	16	13	16	10	10:	0	0	5
K	7	3250: 11	550:	16	15	10	13	11:	4	1	5*
L	9	3950: 14	500:	19	19	14	14	13:	9	3	7*
F	7	4700: 18	200:	19	19	18	19	19:	12	4	4

Table 2: 5th form participants' performance per task/* Subjects who used external negation

Conversely, SLT adjusted scores can be seen as reflecting these participants' level of foreign language performance in spite of the high proportion of correct answers for the structures analysed. An analysis of the less controlled production in the oral task reveals a quite early stage in their language development, which is

reflected in their inaccurate use of negation (**Franco no have one*) and bare forms (**The balloon go to sky*). However, in the light of the variable "years of exposure", this interpretation seems to lack credibility and the first explanation seems to fully account for these participants' performance.

D's results for passive and unaccusative constructions and his SLT scores indicate that this participant "lacks a good basis for distinguishing words from non-words and tends to respond in inconsistent and unpredictable ways" (Read, 2000:131), proved by the fact that he has only been learning English for two years. The SLT score truly depicts his beginner level of performance, also reflected in the lack of accurate responses in the written task.

5. Discussion and conclusion

Criticism has been raised against this test as to its capacity to measure vocabulary knowledge (Eyckmans, 2004) since there is more to knowing a word than what can be assessed by recognising it (Nation, 2001). The whole issue of word recognition as an indicator of knowledge is also questioned on the grounds that there might be different degrees to which a word can be said to be known. Moreover, its design rests upon test-takers' awareness of their knowledge, which should not be taken as a given (Eyckmans, 2004).

Although the data in our study against which the SLT results are compared are compartmentalised and cannot lend support or disclaim the validity of the test, the results obtained in this study seem to indicate it is a trustworthy tool for our purposes. There are few "awkward" profiles but these fall within the expected response patterns according to Meara (1996: 43-44) as cited in Read (2000: 131). Our purpose was not to research vocabulary acquisition and knowledge in depth, but rather to broadly characterise participants' overall performance independently of the data obtained. Had this been our main aim, different instruments would have been necessary.

In sum, as a widely recognized predictor of general language knowledge (Meara and Milton, 2002), the test offers a simple, time-saving computerised means to determine vocabulary size and learners' knowledge of samples of words corresponding to frequency bands, thus indicating level of proficiency. The short time that its completion requires leaves no time for participants to ponder long on each test item so there is little chance of external or individual factors affecting word recognition. For research purposes, the software also provides immediate, easily accessible feedback on test-takers' scores which are compatible with standardised proficiency EFL levels.

While in some cases subjects' performance in the tasks obtained similar results to those predicted by SLT, there were other cases in which there was a significant mismatch between learners' scores in SLT and those in the activities related to the structures under study. 'High error' cases might point to the need of a) administering the test more than once as in the case of some 5th Form participants, or b) combining independent data on the learners' overall linguistic performance, in the case of our university learners, where their meta-knowledge of syntax might have biased the responses for the written task.

References

- Eyckmans, J. (2004) *Measuring Receptive Vocabulary Size: Reliability and Validity of the Yes/No Vocabulary Test for French-speaking Learners of Dutch*. Dissertation, Utrecht University, last retrieved at <http://www.lotpublications.nl/http://www.lotpublications.nl/publish/articles/000794/bookpart.pdf>
- Meara, P. and J. Milton (2002) *The Swansea Levels Test*. (CD-Rom). Express Publishing.
- Meara, P. (1992) *EFL Vocabulary Tests*. Swansea: Centre for Applied Language Studies, University of Wales.
- Meara, P. (1996) The dimensions of lexical competence. In G. Brown, K. Malmkjaer and J. Williams (Eds.) *Performance and Competence in Second language Acquisition* Cambridge: Cambridge University Press, pp 35-53.
- Nation, I.S.P. (2001) *Learning Vocabulary in another Language*. Cambridge: Cambridge University Press.
- Read, J. (2000) *Assessing Vocabulary*. Cambridge: Cambridge University Press.