PROMOTING ENGLISH VOCABULARY RESEARCH IN PRIMARY AND SECONDARY EDUCATION: TEST REVIEW AND TEST SELECTION CRITERIA[1]

Rosa María Jiménez Catalán & Soraya Moreno Espinosa
University of La Rioja, Spain

This article intends to fill a gap regarding critical discussions about the suitability of available tests to investigate learners' vocabulary in the context of English language teaching in Spanish Primary and Secondary schools. With this purpose in mind, we set out to compile, classify, and compare a sample of representative tests spread in English as an additional language research of the last two decades. Then we tentatively propose a practical evaluation that pays attention to different aspects such as the number and nature of dimensions of lexical competence measured by the test, the adequacy of the test for learners of particular ages, test practicability concerning its administration and its test validity and reliability. We end the article with an application of these preliminary criteria to the evaluation of the *Vocabulary Levels Test*, a well known test in English vocabulary research.

1. INTRODUCTION

From the eighties to the present, we witness a literal explosion of research in vocabulary acquisition and teaching of English as an additional language.[2] However, our knowledge of how vocabulary is acquired by primary and secondary school learners is still only in its infancy. This is due among other reasons to the fragmentation of vocabulary research, to its narrow focus on the study of

---

[2] Here, the expression 'additional language' refers to a language different from the learner's mother tongue.

vocabulary acquisition by university students, and to the lack of critical discussion regarding the suitability of available tests for investigating learners' vocabulary acquisition in primary and secondary education.

Finding out what tests could be the most adequate for assessing English vocabulary in those educational contexts is of paramount importance for teachers and researchers alike: there is a need for studying learners' progress and development, comparing learners' vocabulary sizes, selecting the lexical input to be included in language programs and course books, deciding the vocabulary level to be reached in the language by students in each stage of education, and last but not least, there is a need to come to a compromise as to how assess vocabulary development, vocabulary size, and achievement at each stage. However, although the need is evident, there is the problem of selecting the suitable test among the jungle of test formats and the jargon of vocabulary terms found in the research literature.There is only a small number of vocabulary test reviews (Read & Chapelle 2001; Melka 1982, 1997; Meara 1992; Chapelle 1994; Read 1997; Schmitt 1994, 2000) but most give partial overviews of the existing tests. The exceptions to this are Read 1997, Read 2000, and Read and Chapelle 2001, in which we find more complete reviews; nevertheless none carry out a comparative evaluation of available tests in vocabulary research nor analyse them from the perspective of the lexical competence framework or assess their adequacy for researching young learners such as those found in primary and secondary education.

In this article we will attempt to contribute to the investigation of vocabulary, first by compiling[3], classifying, and comparing tests made available for vocabulary research in English, then by proposing some preliminary criteria to evaluate the adequacy of vocabulary research tests for investigating English vocabulary learning in primary and secondary schools. Finally, we will briefly illustrate the evaluation of a vocabulary research test on the basis of the criteria discussed.

2. CLASSIFICATION OF VOCABULARY TESTS

In reviewing the literature on vocabulary research, the first thing that strikes us is the fact that tests are usually classified on the basis of dichotomies such as receptive versus productive, breadth versus depth, discrete versus embedded, context dependent versus independent, oral versus written vocabulary tests, or pen-and-paper versus computerised vocabulary tests. The second aspect that emerges

---

[3] Our compilation does not claim to be exhaustive but rather illustrative of the different types of test that are found in the vocabulary research literature.

from the literature is the fact that dichotomies usually blend giving rise to numerous types of test such as yes/no tests, word association tests, multiple choice tests, word frequency tests, translation tests, picture tests or a combination such as picture translation tests. In Figure 1 we display a classification of twelve tests which have been used in the vocabulary research of the last two decades; our classification is arranged on the basis of the receptive versus productive dichotomy; in it, different types of test format and test combinations are included.

Receptive Tests	-*Eurocentres Vocabulary Size Test* (EVST) (Meara & Buxton 1987; Meara & Jones 1987; Barrow et al 1999) -*Word Associates Test* (Read 1993)
Productive Tests	-*Lex30* (Meara & Fitzpatrick 2000) -*Lexical Frequency Profile* (Laufer & Nation 1995) -*P-Lex* (Meara & Bell 2001; Miralpeix & Celaya 2002) -*K_Lex* (Meara 2001) -*Productive Translation Test* (L1 to L2, and L2 to L1): (Arnaud 1984; Arnaud 1992; Takala 1985; Nurweni & Read 1999) -*Word Definition Tasks* (Verhallen & Schoonen 1993) -*Picture Describing Tests*:
Receptive & Productive	-*Vocabulary Levels Test* (Nation 1983; Nation 1990; Laufer & Nation 1995; Schmitt et al 2001; Schmitt & Schmitt 2001) -*Vocabulary Knowledge Scale* (Wesche & Paribakht 1996) -*A vocabular-based graded dictation test* (Fountain & Nation 2000)

Figure 1. Tests in vocabulary research classified according to the receptive versus productive dimensions

2.1. RECEPTIVE VOCABULARY TESTS

These tests have in common the aim of assessing learners' receptive word knowledge; however this aim can be fulfilled from either by estimating the number of words recalled by language learners or by measuring their degree of word knowledge. The former are known as 'breadth' or 'size tests', the latter as 'depth' tests. The tests included in Figure 1 share some common features but differ in

others. Among the former, we will point out the following: i) they assess 'discrete',[4] 'selective', and 'context independent' vocabulary since they measure specific isolated words; ii) the words to be tested are taken from word frequency lists; iii) they focus on the assessment of written word knowledge; and, iv) they are administered and scored by computer.

As far as the differences are concerned, we find variation in the test purpose, source for word selection, test format, and number of words to be measured. The *EVST* is a placement test that aims to place a large number of students into different levels within a short period of time; the yes/no format which comprises the test structure allows the quick and economical assessment of a great number of words. In contrast, the *Word Associate Test* is intended to be a measure of vocabulary learning for students of English for academic purposes at university level. It presents test-takers with a prompt word together with a list of eight words. Students are required to identify which words are related to the prompt and which are not. The reference list for assessing such learning is the University Word List (UWL) (Nation, 1990). Although originally devised as a vocabulary learning test, the *Word Associate Test* can be used as an association test. As Read (1993:360) remarks: "Rather than focusing on knowledge of the target word, the items could be seen as assessing the testtakers' ability to identify the lexical network formed by the stimulus plus the four associates".

2.2. PRODUCTIVE VOCABULARY TESTS

The purpose of productive vocabulary tests is to investigate learners' vocabulary production. Compared to receptive tests, more variation is found in productive vocabulary tests: *Lex 30* and *K_Lex* are breadth tests, in contrast to *Lexical Frequency Profile* and *P_Le*x which are depth tests; *Lex 30* and *Translation word lists* measure discrete vocabulary whereas *Lexical Frequency Profile*, *P_Lex*, *K_ Lex*, and *Picture Tests* assess embedded and context dependent lexis. *Word Definition Ta*sks combine selective and embedded features since assessment of specific words is aimed at in this type of test but at the same time, due to the nature of the task, the learner must provide the definition of the word in a context; *Lex 30*, *Translation Word Lists*, *Word Definition Tasks* assess selective vocabulary items whereas *Lexical Frequency Profile*, *P_Lex*, *K_ Lex* measure comprehensive vocabulary; in some cases such as *Translation Word Lists*, *Word Definition Tasks* and *Picture Tasks*, the selection of the words to be investigated is based on the assessor's own criteria; in contrast, the words assessed by means of *Lex 30*, *Lexical Frequency Profile,* and *P_Lex* are produced by the learners themselves and then

---

[4] See Read (2000) for a full review of the dimensions of vocabulary assessment.

checked and arranged according to word frequency lists; regarding *K_ Lex*, the test is based on a corpus of 100 picture descriptions produced by native speakers; four productive vocabulary tests use computer-based scoring (*Lex30*, *Lexical Frequency Profile*, *P_Lex*, and *K_Lex)*, whereas three (*Translation Word Lists*, *Word Definition Tasks*, and *Picture Tests*) have been used in the vocabulary research literature as manually scored tests. Finally, productive vocabulary tests can be classified according to the dichotomy of context independent versus context dependent. Among the former, we can place *Lex 30*, *Translation Word Lists*, and *Picture Tests*, among the latter, *Lexical Frequency Profile*, *P_Lex*, *K_Lex*, and *Word Definition Tasks*.

2.3. RECEPTIVE AND PRODUCTIVE VOCABULARY TESTS

Some tests such as *The Vocabulary Knowledge Scale*, the *Vocabulary Levels Test,* and the *Vocabulary-based graded dictation test* are receptive and productive tests. *The Vocabulary Knowledge Scale* is a depth test that comprises five graded categories aimed at eliciting learners' degree of vocabulary knowledge using on selective words chosen by the assessor. For their part, the *Vocabulary Levels Test* and the *Vocabulary-based graded dictation* measure breadth by means of assessing selective words arranged in levels. The words contained in each test are drawn from frequency lists. The three tests are manually scored, even though there is an on-line adaptation of the *Vocabulary Levels Test*.[5] The first and third tests are characterised by the former being context independent, and the latter context dependent, whereas the second one, that is, the *Vocabulary Levels Test,* can be either context independent in its receptive version, or context dependent in its productive version. Figure 2 displays the main features assigned to each of tests described in the preceding paragraphs.

Test	Distinguishing features
Eurocentres Vocabulary Size Test (Meara & Jones 1987)	Receptive. Breadth. Discrete. Selective. Frequency word list. Computer-based scored. Context independent. Yes/no Format.
Word Associate Test (Read 1993)	Receptive. Breadth. Discrete. Selective. University frequency word list. Computer-based scored. Context independent. Word association Format.

---

[5] Version developed by Tom Cobb. See http://www.er.uqam.ca/nobel/r21270/levels/

Lex 30 (Meara & Fitzpatrick 2000)	Productive. Breadth. Discrete. Comprehensive. Word Frequency. Computer-based scored. Context independent. Word Association Format.
Lexical Frequency Profile (Laufer & Nation 1995)	Productive. Depth. Discrete. Comprehensive. Word Frequency. Computer-based scored. Context dependent.
P_Lex (Meara & Bell 2001)	Productive. Depth. Embedded. Comprehensive. Word Frequency. Computer-based scored. Context dependent.
K_Lex (Meara 2001)	Productive. Breadth. Embedded. Comprehensive. Picture description Corpus based. Computer-based scored. Context dependent.
Productive Translation Test (Arnaud 1984; Arnaud 1992	Productive. Discrete. Selective. Word selection done by the researcher. Manually scored. Context independent.
Word Definition Tasks (Verhallen & Schoonen 1993)	Productive. Depth. Embedded. Selective. Word selection done by the researcher. Manually scored. Context dependent.
Picture Describing Tests	Productive test. Discrete. Embbeded. Word selection done by the researcher. Manually scored. Context dependent.
Vocabulary Levels Test (Nation 1983; Nation 1990; Schmitt et al 2001; Schmitt & Schmitt 2001)	Receptive & Productive (Laufer & Nation 1995). Breadth. Discrete. Selective. Word Frequency List. Manually Scored (optional online computerised version). Context dependent in the productive version and independent in the receptive one.
Vocabulary Knowledge Scale (Wesche & Paribakht 1996)	Receptive & Productive. Depth. Discrete. Selective. Word selection done by the researcher. Manually scored. Context independent.
Vocabulary-graded dictation Test (Fountain & Nation 2000)	Receptive & Productive (Receptive spoken form, Productive written form). Embedded. Selective. Word Frequency List. Manually Scored. Context dependent.

Figure 2. Main features of our exemplary vocabulary tests

3. TEST SELECTION CRITERIA

In the previous section we discussed the main tests used in vocabulary research and highlighted their main distinguishing features. Here, we will attempt to outline the minimum criteria that should be taken into account when selecting a test for

researching young learners' English vocabulary when studying this language in Primary and Secondary Education. What we propose is a practical evaluation of the tests that pays attention to different aspects such as the number and nature of dimensions of lexical competence measured by the test, the adequacy of the test for the age of the learners, test practicability concerning its administration and scoring in the context of Primary and Secondary Schools, and test validity and reliability. Let us look at each of these issues in turn.

3.1. DIMENSIONS OF LEXICAL COMPETENCE

The issue of the lexical dimensions that are assessed by a given test is naturally linked to the content of the test as well as its purpose; to discover its nature and purpose we find it useful to pose the following questions: 'What does the test intend to measure?' 'What lexical dimensions are addressed by the test?' 'How many lexical dimensions are covered?' 'What is the test used for?'. However, in order to carry out a systematic evaluation we propose to adopt the concept of lexical competence as a framework of reference. Traditionally, this concept has been treated in English vocabulary research from the angle of what it means to know a word. Since Richards' 1976 seminal article entitled "The Role of Vocabulary Teaching" —in which he specified seven aspects of what it means to know a word— we have been aware of the fact that vocabulary knowledge is many-faceted. Following Richards, quite a number of vocabulary researchers have developed this idea and added new aspects to the list of dimensions covered by the concept of lexical competence.[6] These dimensions involve knowledge of at least the following aspects: 'receptive and productive knowledge of the word', 'the word grammar, pronunciation, and spelling', 'word morphology', 'word collocation', 'syntactic restrictions on the word', 'word frequency', 'word context' 'semantic and syntactic relationships of the word with other words', 'conceptual meaning of the word'. In addition to these dimensions, the concept of lexical competence also involves being capable of recalling the word in both its oral and written mode, as well as retrieving the word when is needed both receptively and productively.

3.2. ADEQUACY WITH RESPECT TO THE AGE OF THE LEARNERS

The adequacy of vocabulary research tests with respect to the learners' age is related to what Read and Chapelle (2001:18) call 'impacts of the test'; as they

---

[6] See Nation 1999 for a review and Jiménez Catalán 2002 for a compilation and systematisation of the different dimensions proposed by scholars.

remark ' …it is necessary to consider how and to whom the test is to be presented". Our intended audience here is Spanish school learners whose ages run from 6 to 12 in primary education and from 12 to 16 in compulsory education. Increasing age usually goes hand in hand with cognitive, psychological, and social changes. One well known theory of cognitive development is that of Piaget,[7] which involves a gradual transition through four stages: the sensorimotor stage —from birth to two—, the preoperational stage —from two to seven—, the concrete operational stage —from seven to eleven—, and formal thinking —from eleven onwards—. On the basis of this theory, primary school learners experience more intellectual changes than secondary school learners: the fact that they start at six and end around twelve means that they go through three cognitive stages: The pre-operational stage, which is characterised by the children's development of schemata and discovery of rules that help them make sense of the world; the concrete operational stage, which has to do with the development of abstract concepts and principles although the child is still tied to the concrete world; and the formal operational stage, which brings a peak in cognitive development and, as a result, the child's ability to deal with abstract concepts. In contrast, secondary school learners experience greater psychological and social changes that have to do with physical bodily changes, with the search for identity and autonomy and the development of moral and social consciousness.

It is not within the scope of this paper to discuss the implications of primary and secondary school learners' cognitive, psychological and social traits for English vocabulary learning, but when analysing the vocabulary assessed by a given test it seems obvious that we should have in mind learners' ages and their corresponding cognitive stages. In this regard, it is quite probable that if abstract words are contained in the test, if the test aims to elicit such words, most 6 to 10 year olds and some 11 to 12 year olds will fail the test. Likewise, it is important to be aware of the fact that primary school pupils have a very short memory span diminishes their capacity for long tests. For both primary and secondary school learners, motivational aspects are important: words that have nothing to do with their needs and interests are likely to cause problems in their responses to the test.

3.3. TEST PRACTICABILITY

Test practicability has to do with its administration and scoring. When dealing with vocabulary testing in primary and secondary schools we need to bear in mind

---

[7] For a description of this theory see Piaget and Inhelder 1969. A description of the psychological and social changes that occur in children and adolescents can be found Hayes 1994. Finally, for a discussion of the implications of these changes for English Language Teaching in primary education see Jiménez Catalán 1996.

the existence of fixed characteristics that may impose limitations on the administration of the test. English, like certain other subjects, is compulsory in the school curricula, and the syllabus has fixed objectives, it is taught in large classes of no less than twenty-five students, and the timetime is fixed. We also need to take into account the differences among individual learners, as well as the fact that classrooms may not be equipped with computers and other technologies. If we pay attention to these factors, a number of things seem clear: both the test purpose and the test requirements should be clear; the test should be within the reach of this intended audience, and adequate for the contexts of primary and secondary education; furthermore, it should not be long, yet it should be capable of eliciting as much as information as possible, and ideally it should have been employed with students of similar ages and educational levels.

3.4. TEST VALIDITY AND TEST RELIABILITY

Many types of validity have been put forward in the language testing literature.[8] Those most closely related to our purpose here are content validity, predictive validity, construct validity, and concurrent validity. Content validity is defined as the extent to which the test measures what it intends to measure. For instance, a test of vocabulary should measure only lexical items not grammar or phonology. Predictive validity occurs when the results of the test can predict learners' future performance. Construct validity refers to the harmony between the items in a test and the theoretical constructs underlying the test. For instance, if a test claims to be communicative there must be a symbiotic relationship between the theory of communicative competence and the items in the test. Finally, concurrent validity refers to "the extent to which a test correlates with some other test that is aimed at measuring the same skill, or with some other comparable measure of the skill being tested" (Richards and Schmitt: 2002:105). As can be observed, validity can be considered from an internal and external perspective; the former has to do with content and construct validity, the latter is related to concurrent validity, in which similar results are found in different situations.

As many scholars remark (Heaton 1987; Bachman 1990; Alderson *et al*., 1995) test reliability is an essential condition for a test to be good. It is defined in terms of obtaining the same results every time that test is given either to the same informants or to a sample of a population of the same characteristics. Reliability is closely connected to replicability which, in turn, is what allows test scores to be

---

[8] Richards and Schmidt 2002 list twelve types of validity: concurrent validity, consequential validity, construct validity, content validity, convergent validity, criterion measure, criterionrelated validity, discriminant validity, divergent validity, empirical validity, face validity, and predictive validity.

generalised. There are three ways of measuring test reliability: test-retest reliability, split-half reliability, and equivalent form reliability. In the first case, the test is given twice to the same group of students in order to correlate the scores. In the second case, the items of the test are distributed into two equal halves and both parts given to the same students on two different occasions and the scores obtained correlated. Finally, in equivalent form reliability two different versions of the same test are given to the students in order to see if the same scores are obtained on both versions. A high positive correlation means a high degree of reliability.

4. APPLICATION

Due to space limitations a detailed evaluation of each of the twelve tests included in Figure 1 and 2 is beyond the scope of this paper. Instead, we will focus on applying the above criteria to a test that has been widely used in vocabulary research: *The Vocabulary Levels Test.*

In agreement with what was put forward in the previous section, the criteria for the evaluation will be: a) the number of dimensions of lexical competence assessed by this test; b) the adequacy of the test for English learners in primary and secondary education; c) test practicability; d) validity and reliability.

VOCABULARY LEVELS TEST

DIMENSIONS OF LEXICAL COMPETENCE ASSESSED BY THIS TEST

As was noted in section 2, this test has two versions: receptive and productive. Here, we will analyse only the first of these. As many scholars have claimed since Richards 1976, knowing a word means knowing it receptively, as well as knowing the word frequency. The receptive version of the *Vocabulary Levels Test* (VLT) covers both dimensions of lexical competence.

Adequacy for English learners in primary and secondary education

The test was designed by Nation (1983, 1990) as a practical tool to assess breadth of vocabulary knowledge in university students, who are learners of English as a second language. It can, however, also be used to assess primary and

secondary students learners of English as a foreign language, as the test has five frequency bands that range from 2**,**000 up to 10,000 words. In addition to this, Nation (1993) designed a 1,000 word test that seems to be suitable for young learners with low levels.

Students are presented with six words[9] and three definitions, and they have to choose which of the six words match the three meanings. Thus, the test format seems to be appropriate to primary and secondary school learners, since they are used to do matching exercises. The test assesses words appropriate to the cognitive stage of primary and secondary school learners and, furthermore, EFL teachers can decide how many levels learners are to be tested on, an important issue in order to deal with the short memory span of certain types of learner, such as those in primary school.

Moreover, the procedures followed in the test are clearly stated by the author, and permit further adaptations if needed.

TEST PRACTICABILITY

The *Vocabulary Levels Test* assesses a large number of words in a short time and is easy to mark. As Nation (1990) states, the information retrieved from the test can be used: (a) to find learners' vocabulary size; (b) to compare vocabulary knowledge before and after the course; (c) to keep a continuing check on progress; (d) to encourage learning by setting short-term goals; (e) to see the effectiveness of one's teaching; and (f) to investigate learning, amongst others.

We agree with Read & Chapelle (2000), when they say that we can interpret test results at different levels, since we can draw conclusions not only at whole test level, but also at sub-test level. Thus, can we get not only an estimate of testees' receptive vocabulary size, from the proportion of correct responses to the whole battery of test items, but also an estimate of learners' vocabulary size at each of the different frequency levels covered by the test. In our view, approaching the outcome of the test from a multiple component profile will provide a wider range of information, since teachers and researchers will be able to know approximately what stage of vocabulary development testees are at.

By taking into account, Read & Chapelle (2000)'s practical outcomes of test results. *The Vocabulary Levels Te*st can be used not only for instruction –for diagnostic and placement purposes– but also for research uses, as it has already been used in primary (Jiménez Catalán & Moreno Espinosa 2005), secondary

---

[9] Thirty words are matched at the 1,000 word level, and eighteen words are matched at each level from 2,000 up to 10,000.

(Cameron 2002, Laufer 1998, Laufer & Paribakht 1998) and tertiary education (Laufer & Paribakht 1998, Fan 2000, Waring 1997).

TEST VALIDITY AND RELIABILITY

According to Schmitt *et al* (2001), despite *VLT* being a widely used test, it has not been properly validated. We can observe a first validation attempt in the work of Read (1988) who claims that the *VLT* is a useful tool for diagnostic purposes, even though he pinpoints three possible shortcomings: (a) it tests a small sample of words at each level; (b) the definitions provided are dictionarytype, since they are written within a controlled vocabulary,[10] and sometimes they are awkwardly expressed, which may cause learners to experience problems in making sense of them; (c) the influence of the test format on the testee performance remains to be studied.

Beglar and Hunt (1999) revised and validated the 2,000 word level and UWL.[11] The follow up to these initial validation studies can be found in the article of Schmitt *et al* (2001), in which evidence for its validity is presented and the equivalence of the two new versions of the *VLT* explored. From the results obtained Schmitt et al (2001) believe that there is at least initial evidence of the test providing accurate estimates of the vocabulary size of testees at the targeted frequency levels. Furthermore, the reliability indices are high, and they seem to be in line with the figures reported by Read (1988). Therefore, the test is efficient in terms of the reasonable amount of time required.

Schmitt et al point out that the rubric of the test discourages testees from guessing blindly, since that could pose a problem, especially with proficient examinees who seem to be quite successful when guessing. This issue seems, however, to be a less serious problem for low proficiency testees because their guesses are usually unsuccessful. In order to check the degree of guessing, results were correlated with an oral interview undertaken with a small sample of

---

[10] This is not seen as a problem by Schmitt *et al* (2001), who point out that amongst other things, the definitions are kept short in order to provide a minimum of reading. Furthermore, the higher frequency of words in the definitions, in comparison with the target words, ensures that the ability of the testees to demonstrate knowledge of the target words is not jeopardized by a lack of knowledge of the defining words.

[11] Schmitt *et al* (2001) criticise this validation, since they claim that a complete section of the test, corresponding to a single level, should be carried out, rather than individual items being subjected to factor analysis -which is the approach taken by Beglar and Hunt (1999) -. Schmitt *et al* believe that the separate constructs of section and item, can be confounded when undertaking factor analysis of individual items.

informants. The interviews suggested that many of the mismatches –i.e. when the informants selected the distractors instead of the right elicited words-were the result of partial knowledge rather than of guessing.

5. CONCLUSION

In this article, we have attempted a classification and a comparison of twelve tests found in the English vocabulary research literature of the last two decades. In our review, we observed that vocabulary tests are usually classified on the basis of manifold dichotomies which usually blend and result in countless types of test; all this may prevent researchers from having a comprehensive view of tests as a whole, let alone establishing a general classification that could embrace all the tests that have been used in vocabulary research. To this we should add the difficulty of classifying a given test under a single category as different traits and dimensions constantly intermingle in them. In this situation, we opted for what we believe is the less problematic solution: adopting the receptive versus productive dichotomy as the basis for classifying other taxonomies as well as for agglutinating the different features assigned to each test.

Once our samples of tests had been classified and contrasted, we went on to draw up some tentative criteria that might serve to evaluate vocabulary tests from the standpoint of investigating English vocabulary learning in primary and secondary education, contexts which have so far been neglected in the vocabulary research agenda. Our proposal was a practical evaluation that focuses on the number and nature of dimensions of lexical competence measured by the test, the adequacy of the test for learners of particular ages, test practicability, and test validity and reliability.

We concluded with an example of application of the criteria to the evaluation of *The Vocabulary Levels Test*, in its receptive version. Our brief analysis revealed that by means of this test it is possible to measure two dimensions of lexical competence: receptive word knowledge and knowledge word frequency. On the whole, the test is adequate for primary and secondary school learners thanks to the six word frequency bands (from 1,000 to 10,000 words) that make up the structure of the test and correspond to the beginners/low intermediate/intermediate levels that are to be expected in primary and secondary education. Furthermore, the range of frequency bands permits the use of the test from 1st form in primary education (six year olds) to the 4th form of secondary education (sixteen year olds), and even further than this, in higher compulsory education. The test can be shown to have validity regarding: i) content, as it focuses on words and measures the number of words from each band known by the learner and checks whether learners' word

knowledge corresponds to the most frequently words used by native speakers; ii) power to discriminate among students, as test scores reveal learners' word knowledge on each of the bands which can be a basis for comparing the vocabulary sizes of learners of different educational levels both, synchronically and over time; iii) the existence of different versions of the test which allow test rest of the same students. Finally, test reliability has been demonstrated in Schmitt et al (2001), even though some previous attempts had already been carried out by Read (1988) and Beglar and Hunt (1999).

However, the *Vocabulary Levels Te*st has received some criticism, mainly on the basis of the number of words tested and the sources from which words are taken to make up the frequency bands. Regarding the number of words, Read (1988) and Meara & Fitzpatick (2000) have pointed out that the test does not provide a good estimate of learner's vocabulary size because few words out of the corresponding bands are tested. However, we believe this is a problem for most vocabulary tests because of the lack of agreement concerning estimates of native vocabulary size. There is also the difficulty of evaluating learners' whole vocabularies by means of a single test, let alone at one session: sitting testees to recall 1,000 words one by one would be impracticable, if not impossible in normal class conditions. As to the other main criticism, several scholars (Engels 1968, Richards 1974) have claimed that Michael West 1953's frequency list (which is the main source of this test) is too old to reflect frequency in today's language, since some words contained in the test might be oldfashioned, infrequent, or simply no longer used nowadays. However, Nation & Waring (1997:13) claim that: "In spite of its age, some errors, and its solely written base, it still remains the best of the available lists because of its information about the frequency of each word's various meanings, and West's careful application of criteria other than frequency and range".

In our view, one of limitations of *The Vocabulary Levels Test,* at least from the perspective of using this test to investigate primary and secondary school learners' vocabulary, is the reduced number of lexical dimensions that the test aims to measure. Receptive word knowledge and word frequency within learners' vocabulary size are two important dimensions but by no means all the possible dimensions that make up lexical competence. But again, this might be an unfair criticism as the author only claims that the test measures vocabulary size based on word frequency.

The many dimensions embraced in the concept of lexical competence as well as the different domains involved in these dimensions (linguistic, sociolinguistic, psycholinguistic, and pedagogic) make it almost impossible to investigate learners' receptive and productive vocabulary by means of a single test. In our view, this is the reason for the existence of multiple dichotomies and test typologies: instead of opting for improving an already existing test, vocabulary researchers tend to opt for designing different tests that focus on reduced dimensions of lexical competence.

REFERENCES

Alderson, J. Charles; Clapham, Caroline and Dianne Wall. *Language test construction and evaluation.* Cambridge: Cambridge University Press, 1995.

Arnaud, Pierre. "The lexical richness of L2 written productions and the validity of vocabulary tests." *Practice and Problems in Language Testing*. Eds. Terry Culhane, Christine. Klein-Braley and Douglas K. Stevenson . Colchester: University of Essex. 1984, 14-28.

Arnaud, Pierre. "Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests." *Vocabulary and Applied Linguistics*. Eds. Pierre Arnaud and Béjoint, H. (eds.) London: Macmillan, 1992, 83-97.

Bachman, L.F. *Fundamental Considerations in LanguageTesting*. Oxford: Oxford University Press, 1990.

Barrow, J, Y Nakanishi and Ishino, H. "Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey." *System* 27 (1999): 223-247.

Beglar, David and Alan Hunt. "Revising and Validating the 2000 Word Level and University Word Level Vocabulary Tests." *Language Testing* 16 (1999): 131-162.

Cameron, Lynne. "Measuring Vocabulary Size in English as an Additional Language." *Language Teaching Research* 6, 2 (2002): 145-173.

Chapelle, Caroline A. "Are C-tests valid measures for L2 vocabulary research ?" *Second Language Research* 10 (1994): 157-187.

Engels, Leopold K. "The fallacy of word counts." *IRAL* 6 (1968): 213-231.

Fan, May. "How Big is the Gap and How to Narrow it? An Investigation into the Active and Passive Vocabulary Knowledge of Low Learners." *RELC Journal* 31 (2000): 105-119.

Fountain, Ron L. and Nation, Paul. "A Vocabulary-Based Graded Dictation Test. "*RELC Journa*l 31 (2000): 29-44.

Hayes, Nicky. *Foundations of Psychology* : an introductory text. London: Routledge, 1994.

Heaton, J.B. *Writing English Language Tests*. New York: Longman, 1987.

Jiménez Catalán, Rosa María. "Psicología Evolutiva y aprendizaje del inglés en Primaria." Revista de *Psicodidáctica*, 2 (1996): 65-81.

Jiménez Catalán, Rosa María and Moreno Espinosa, Soraya. "Using *Lex30* to Measure the L2 Productive Vocabulary of Spanish Primary Learners of EFL." (2005) forthcoming.

Jiménez Catalán, Rosa María. "El concepto de competencia léxica en los estudios de aprendizaje y enseñanza de segundas lenguas." *ATLANTIS* 24 (2002): 149-162.

Laufer, Batia and Nation, Paul. "Vocabulary size and use: Lexical richness in L2 written production." *Applied Linguistics* 16 (1995): 307-22.

Laufer, Batia and Nation, Paul. "A vocabulary size test of controlled productive ability. "*Language Testing 16* (1999): 33-51.

Laufer, Batia and T. Sima Paribakht. "The Relationship between Passive and Active Vocabularies: Effects of Language Learning Context." *Language Learning* 48, 3 (1998): 365-391.

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

Meara, Paul. *K-Lex*, University of Wales Swansea (2001) Downloaded from
http://www.swan.ac.uk/cals/calsres/lognostics.htm

Meara, Paul and Jones, G. 1990. *Eurocentres vocabulary size tests 10KA*. Zurich: Eurocentres Learning Service.

Meara, Paul and Buxton , B. "An alternative to múltiple choice vocabulary tests. *Language Testing*." 4 (1987): 142-154.

Meara, Paul and Bell, Huw "P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. " *Prospect* 16 (2001): 323-337.

Meara, Paul and Fitzpatrick, Tess. "Lex30: an improved method of assessing productive vocabulary in an L2. " *System* 28 (2000): 19-30.

Meara, Paul and Jones, G. "Tests of vocabulary size in English as a foreign language." *Polyglot* 8, 1 (1987): 1-40.

Meara, Paul. *EFL vocabulary tests*. University of Wales, Swansea: Centre for Applied Language Studies, 1992.

Meara, Paul. *K-Lex*: http://www.swan.ac.uk/cals/calsres/lognostics.htm (2001).

Melka, Francine. "Receptive versus productive vocabulary: a survey. *Interlanguage Studies Bulletin* ." 6 (1982): 5-33.

Melka, Francine. 1997. "Receptive vs. productive aspects of vocabulary. " *Vocabulary: Description, acquisition and pedagogy*. Eds. Norbert, Schmitt and McCarthy, Michael. Cambridge: Cambridge University Press, 1997: 84-102.

Nation, Paul and Waring, Robert. "Vocabulary Size, Text Coverage and Word Lists." *Vocabulary: Description, Acquisition and Pedagogy*. Eds. Norbert Schmitt and Michael McCarthy(eds): Cambridge: Cambridge University Press, 1997. 6-19.

Nation, Paul. "Measuring readiness for simplified material: a test of the first 1,000 words of English." *Simplification:* Theory and Application . Ed. Makhan Lal .Tickoo. RELC Anthology Series, 31 (1993): 193-203.

Nation, Paul. *Teaching and Learning Vocabulary*. New York: Newbury House, 1990.

Nation, Paul. *Learning Vocabulary in Another Language*. Victoria University of Wellington: English Language Institute Occasional Publication, 19 New Zealand, 1999.

Nation, Paul. *Teaching and Learning Vocabulary.* Victoria University of Wellington: English Language Institute Occasional Publication , 19 (1993).

Nurweni, Ari and Read, John. "The English vocabulary knowledge of Indonesian University students. "*English for Specific Purposes* 18 (1999): 161-175.

Piaget, Jean and Inhelder, Bärber. *The psychology of the child*. New York: Basic Books,1969.

Read, John and Chapelle, Caroline A. "A framework for second language vocabulary assessment". *Language Test*ing 18 (2001): 1-32.

Read, John. "The development of a new measure of L2 vocabulary knowledge. " *Language Testing* 10 (1993): 355-371.

Read, John. *Assessing vocabulary*. Cambridge: Cambridge University Press, 2000.

Read, John. "Measuring the Vocabulary Knowledge of Second Language Learners". *RELC Journal* 19, 2 (1988): 12-25.

Read, John. "Vocabulary and testing. "*Vocabulary: Description, acquisition, and pedagogy*. Eds. Schmitt, Norbert and McCarthy, Michael. Cambridge: Cambridge University Press, 1997: 303320.

Richards, Jack and Schmitt, Norbert. *Longman Dictionary of Language Teaching & Applied Linguistics*. London: Pearson Education Limited, 2002.

Richards, Jack. "The role of vocabulary teaching." *TESOL Quarterly* 10 (1976): 77-89.

Richards, Jack. "Word Lists: Problems and Prospects." *RELC Journal* 5 (1974): 69-84.

Schmitt, Norbert; Schmitt, Dianne, and Caroline Clapham. " Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18,1 (2001): 55-88.

Schmitt, Norbert. "Assessing vocabulary knowledge" *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press, 1997: 163-180.

Schmitt, Norbert. "Vocabulary testing: Questions for test development with six examples of tests of vocabulary size and depth." *Thai TESOL Bulletin* 6 (1994): 9-16.

Takala, Samuli. *Evaluation of students' knowledge of English vocabulary in Finnish comprehensive school*. Jyväskylä: Institute of Educational Research, 1985.

Verhallen, Marianne and Schoonen, Rob. "Lexical knowledge of monolingual and bilingual children." *Applied Linguistics* 14 (1993): 344-356.

Waring, Robert. "A Comparison of the Receptive and Productive Vocabulary Sizes of Some Second Language Learners." *Immaculata*, 1 (1997): 53-68.

Wesche, Marjorie and T. Sima Paribakht. "Assessing second language vocabulary knowledge: Depth versus breadth." *The Canadian Modern Language Review* 53, 1 (1996): 13-40.