



THE EFFECT OF CONTEXT IN ACHIEVEMENT VOCABULARY TESTS

Gökhan Öztürk
Afyon Kocatepe University
School of Foreign Languages
Afyonkarahisar, TURKEY
gokhanoztrk@gmail.com

Abstract

This study investigated the effect of context on the performances of students in achievement vocabulary tests. Two different tests, discrete and contextualized, having the same target vocabulary items were designed by the researcher and were administered to 123 elementary students at Afyon Kocatepe University English preparatory program. The data were analyzed through descriptive statistics. The results revealed that students performed better in the contextualized test, and there is a significant difference between the performances of students. The results also showed that there was a positive and direct correlation between the tests, and these two tests could be used instead of each other in a degree of 56%, which means context is not a must in achievement vocabulary tests.

Key Words: Vocabulary, testing vocabulary, contextualized tests.

INTRODUCTION

Background to the Study

Language tests, in general, are used for three purposes: for research, for making decisions about learners and determining the language problems. In the case of testing vocabulary, many research studies have been carried out on; how broad and deep learners' vocabulary knowledge is, how effective different testing instruments are, what the effect of other skills on testing vocabulary is, whether and how the learners infer the meanings of unknown words encountered in a text and how they deal with gaps in their vocabulary knowledge.

On the other hand, language teachers and testers who construct vocabulary tests for making decisions about learners have a different focus. They use tests for purposes such as placement, diagnosis, measuring progress or achievement, and assessing proficiency. A progress test assesses how well the students have learned the words presented in the units they have recently studied in the course book. The vocabulary section of a placement test can be designed to estimate how many high frequency words the learners already know. Apart from placing the students, the test scores would help teachers for which vocabulary items should be taught or what kind of vocabulary program should be followed during the instruction process. In the achievement test, the vocabulary section may be designed to assess how well the students have mastered a vocabulary skill that they have been taught (Read, 2000).

Statement of the Problem

Testing vocabulary and its elements have been a controversial issue in language testing for years. One of these elements is the sampling issue. Which vocabulary items the test writers should include in the tests, whether they should be chosen from active or passive vocabulary knowledge, and what the frequency of words should be are the main issues that must be considered in terms of sampling issue in vocabulary tests. On the other hand, testing the productive and the recognition vocabulary has been another issue that is open to discussion considering the criteria of knowing a word and the degree of production. Probably, the most ongoing issue among these elements is the contextualization issue. Presenting the vocabulary items in a test whether discrete or contextualized, what discrete and contextualized means, and the degree of contextualization are the main problems that have been discussed for many years. This most ongoing element of testing vocabulary, discrete or contextualized, is also the issue in this paper.

Purpose of the Study

In this study, it is aimed to find out whether presenting the items in discrete or contextualized way in an achievement vocabulary test affect the performances of students. It is also investigated whether these two types, discrete and contextualized, correlate. In order to reach these aims, two tests, discrete and contextualized, having the same target vocabulary items were constructed and administered to the same group of students as an achievement test.

Research Questions

In the light of the purposes of the study, the following research questions were studied;

1. What are the performances of students in a discrete item vocabulary test and in a contextualized vocabulary test?
2. Is there a significant difference between the performances of students in these two vocabulary tests?
3. Do these two types of tests correlate? If so, to what extent and what does it mean?

LITERATURE REVIEW

In this part of the paper, some concepts regarding vocabulary and testing it are discussed in accordance with the study.

Knowing a Word

What is “knowing a word”? In the broadest sense, knowing about a word involves knowing about its form (how it sounds, how it is spelt, the grammatical changes that can be made to it), its meaning (its conceptual content and how it relates to other concepts and words), and its use (its patterns of occurrence with other words, and in particular types of language) (Cameron, 2001).

According to Ellis and Sinclair (1989), knowing vocabulary involves understanding the word when it is spoken or written, recalling it when needed, using it with the correct meaning and in a grammatically correct way and in the right situation, pronouncing and spelling it correctly, knowing which other words can and cannot be used with it, knowing if it has positive or negative connotations and when or when not to use it.

In accordance with the criteria of knowing a word, Hatch and Brown (1995, p.383) describe five essential steps in vocabulary learning based on research in learners’ strategies: (1) having resources for encountering new words, (2) getting a clear image, whether visual or auditory or both, for the new form of the words, (3) learning the meaning of the words, (4) making a strong memory connection between the forms and meaning of the words, (5) using the words.

Types of vocabulary: Read (2000) points out that the number of words we recognize and understand is rather larger than the number we use in our own speech and writing. Therefore, he distinguishes between receptive and productive vocabulary. (Some scholars use the terms active and passive vocabulary).

According to Lado (1964), producing a vocabulary means that a unit can be recalled almost instantaneously with its proper structural position in accordance with the context. A passive vocabulary requires only recognizing and grasping the meaning from the form in its partly redundant context.

Receptive vocabulary is described by Harmer (1991) as words which students will recognize when they meet them but they will not be able to produce. It is generally assumed that words are known receptively first and only later they become available for productive use. Melka (1997; cited in Read, 2000) regards this process as “receptive to productive continuum” representing the increasing degrees of knowledge or familiarity with a word. Thus, upon first encounter, learners have limited knowledge of the word and may not even remember it until they come across it again. It is only after they gain more knowledge of pronunciation, spelling, grammar, meaning, range of use etc. that they are able to use it themselves.



The importance of vocabulary in language learning: Vocabulary and lexical units have great importance in learning a language and communicating it. Without enough vocabulary knowledge, it is not possible to employ grammatical or other types of linguistic knowledge in communication or discourse. According to Folse (2003), without syntax, meaning is hindered; but without vocabulary meaning is impossible.

It is also emphasized that vocabulary is needed for every language skill and grammar. It can be regarded as the core of the four skills; speaking, listening, reading and writing. As Chastain (1988) states, the lack of needed vocabulary is the most common cause of students' inability to say what they want to say during communication activities. To clarify the importance of vocabulary in the development of other skills in terms of testing, Heaton (1988) points out that tests of vocabulary often provide a good guide to reading ability.

As it is seen, unless the language learner has a sufficient amount of word knowledge, s/he can neither express her/himself as intended, nor understand the speaker. Without sufficient word knowledge, a learner cannot write or read in a desired way. In such a case, when the learners are confronted with situations where they feel they lack the knowledge of vocabulary, their motivation to learn language decreases.

Testing Vocabulary

The importance of testing vocabulary in language testing: Since it is a well known fact that vocabulary has a big importance as much as the language skills in the whole language system, the importance of the assessment of vocabulary is the same as the assessment of other skills. As Schmitt (2000) states, vocabulary is an essential building block of language and, as such it makes sense to be able to measure learners' knowledge of it. With regard to second language acquisition, vocabulary is an indispensable and basic part. Both teachers and learners spend lots of time and energy on vocabulary acquisition. For this reason, due to the importance of vocabulary in the whole language system, it is quite necessary to test the vocabulary knowledge of learners.

Why do we test?: One of the most common reasons for testing vocabulary is to find out if the students have covered the words they have been taught. As Read (2000) states, in the achievement test, the vocabulary section may be designed to assess how well the students have mastered a vocabulary skill that they have been taught. Moreover, vocabulary test can be utilized as a means of motivating students to study and showing them their progress in learning new words. Also, vocabulary tests which are part of commercial tests, such as TOEFL, attempt to provide a measure of learners' vocabulary size, which is believed to give an indication of overall language proficiency (Schmitt, 1994).

What do we test?: A careful selection, or sampling, of lexical items for inclusion in a test is generally a most exciting task (Heaton, 1988). While making this selection, the test writer should consider some aspects. The first one is whether to test the students' active or passive vocabulary and whether to take the vocabulary items from spoken or written language. When the four skills are added to these dimensions, as Heaton (1988) states, selection of vocabulary can be thought of falling rough divisions according to four major language skills.

The other aspect of selecting the vocabulary items is the frequency of them. Whether to test high frequency words or more specialized technical vocabulary is the main issue of this aspect. High frequency words are so important that anything teachers and learners can do to make sure they are learned is worth doing. The 2000 most frequent words of English as high frequency words is extremely useful particularly for students who want to go on to academic study (Nation & Hwang, 1995). On the other hand, Read (2007) states that vocabulary size tests for second language learners understandably focus on a narrower range of words than those for native speakers, since low frequency words are much less likely to be known, especially by learners in a foreign language environment.

The third aspect of what to test is the breadth or depth. Vocabulary breadth is defined as the quantity of words for which students may have some level of knowledge (Anderson & Freebody, 1981, cited in Stahl & Bravo, 2010). On the other hand, vocabulary depth refers to how much students know about a word and the dimensions of word learning addressed previously. So, according to the purpose of test writer, the dimension of breadth or depth affects the selection of items in a vocabulary test.



How do we test?: Deciding how to test vocabulary is still a controversial issue in the world of language testing. The dominance of communicative approach to language teaching in the past thirty years has thrown up various challenges to the validity of the conventional vocabulary test and this has prompted some re-thinking of the nature of lexical ability as well as how it can best be assessed (Read, 2007).

Depending on this re-thinking, some dimensions on how to test vocabulary have been improved. One of these dimensions is the discrete-embedded issue. This distinction addresses whether vocabulary is regarded as a separate construct with its own separate set of test items and its own score report, which is the discrete end of the continuum, or whether vocabulary is an embedded construct that contributes to, but is not regarded as separate from, the larger construct of text comprehension (Pearson & Hiebert & Kamil, 2007).

The issue of contextualized-decontextualized is the second one in the dimension of how to test vocabulary. The issue involves the presentation of the vocabulary in a context or in an isolated way. According to Pearson & Hiebert & Kamil (2007), this continuum refers to the degree that textual context is required to determine the meaning of a word. Any word can readily and easily be assessed in a decontextualized format. But simply assessing a word in a contextualized format does not necessarily mean that context is required to determine its meaning. In order to meet the standard of assessing students' ability to use context to identify word meaning, context must actually be used in completing the item.

The Importance of Context in Testing Vocabulary

An important dimension of vocabulary assessment is concerned with the role of context. This dimension considers the extent to which the test-taker is required to engage with and utilize the context in which the vocabulary items appear. The decontextualized formats present students with words in isolation and require them to select meanings for the words without reference to any linguistic context (Read & Chapelle, 2001). However, in context-dependent vocabulary tests students need to make use of contextual clues. Many scholars believe that vocabulary testing in context can offer a real situation for test takers and, the usage of words is tested in such real situations.

Related Studies

The issue of testing vocabulary has been one of the interests of the researchers in the field of language testing. Many different studies have been carried out to test vocabulary knowledge of learners. Read (1997) ascertains that if vocabulary knowledge is accepted as a fundamental component of second language proficiency, it is natural to accept that one of the primary goals of language testing will be to assess whether learners know the meanings of the words they need to communicate in a successful way in L2. Here are some of the studies related to the assessment and testing of vocabulary knowledge;

Qian (2008) conducted a research on the predictive power of discrete and contextualized vocabulary items on assessing the reading performance. According to results of this research, in assessing reading performance, discrete-point vocabulary items and fully contextualized vocabulary items provide a similar amount of prediction. However, in the context of considering educational impact, the article argues in favor of the continued adoption of the fully contextualized vocabulary item format because it will more likely induce beneficial backwash effects than the discrete-point vocabulary item format. It is also stated that the contextualized format also has the advantage of bringing vocabulary testing closer to real-life communicative application of the English language and therefore has more positive implications for the language classroom.

A very recent study was carried out by Fitzpatrick and Clenton (2010) on the performance of vocabulary test, Lex30, designed to measure second language productive vocabulary knowledge. In this article, they presented evidence which makes a significant contribution to the inquiry into the validity of the Lex30 test of vocabulary knowledge. Their research encourages further investigation of its validity in an instructional context.

In their research study, Laufer and Goldstein (2004), they test the size and strength of vocabulary with the trial of bilingual computerized test. They consider the size as the number of words the learners know, and the



strength as a combination of four aspects of knowledge of meaning that are assumed to constitute a hierarchy of difficulty: passive recognition, active recognition, passive recall and active recall. They investigated whether this hierarchy was valid and which modality correlated best with classroom language performance. The participants were 435 learners of English as a second language. Results showed that the hierarchy was present at all word frequency levels and passive recall was the best predictor of classroom language performance.

METHODOLOGY

Participants

The participants of this study include 123 students from Afyon Kocatepe University English preparatory program. Their level of English is elementary as all the students in this preparatory program. Since most of the students come from state schools, their level of proficiency is low. They have 25 hours of English every week. English is taught integratively in English preparatory program at Kocatepe University. Students have 15 hours main course for the course book and 10 hours of writing and reading skills with a different instructor.

The participants were chosen from five different classes which are close to each other in terms of proficiency level. Their proficiency levels are indicated with their official mid-term grades in the appendix.

Instruments

Two multiple choice vocabulary tests were designed by the researcher himself, one of which consisted of discrete vocabulary items (called d- test in the paper), and the other consisted of contextualized items (called c- test in the paper). The researcher benefited from the guidelines for writing test items described by Heaton (1988). Both tests had 25 items having the same target vocabulary items to be tested. The target vocabulary items were chosen from the vocabulary items that had been taught to the students in their lessons. In other words, the tests were administered as an achievement test which is designed to assess how well the students have mastered a vocabulary skill that they have been taught (Read, 2000).

The d-test consisted of 25 target vocabulary items. The items in this test were presented as discrete-point items. In discrete-point items, the target vocabulary is presented in a sentence context or an isolated way. Read (2000) states that in discrete items, the context most commonly consists of a sentence in which the target word occurs. Both ways of presenting the target word were designed in such a way that makes the test a totally discrete-point test and distinguishes it from the c-test in terms of the issue of contextualization.

The c-test, on the other hand, was designed in a different way. Although it had the same target vocabulary items, the presentation of them was totally different from the d- test. It is a known fact that the context for a target vocabulary item should provide a meaningful way for students in making their responses to the items. Read (2000) states that the test writers should broaden the notion of context, to include whole texts and more generally discourse. Depending on this clear definition, the researcher tried to create discourses for the target words in each item to provide contextual clues for the students.

By designing two different tests and presenting the same target words to the test takers in two different ways, the researcher tried to reach his aim which is to learn whether the students perform differently in these two different tests having the same target vocabulary items. In other words, the researcher tried to find out whether the presentation of words to the students in a discrete way and contextualized way affects the students' performances.

Procedure

Before the study, the researcher made an evaluation on the classes that would form the participants of the study by examining their official mid-term grades. After deciding on the participants, the instruments of the study (d- test, c- test) were examined and evaluated by the colleagues of the researcher. One of them had had her MA degree on teaching vocabulary and contributed a lot to examining the instruments.

After this process, the tests were administered to the participants. First, the d-test was administered because the researcher believed that if the c- test had been administered first, the students would have had the meanings from the contexts and would have done the d- test easily. For this reason, in order to prevent this, it was more appropriate to administer the d- test first.

The next day, the c-test was administered. Although the same target vocabulary items were included in the c- test, the order of the items and the order of distractors were changed in order to prevent memorization. The reason for administering the c-test on the next day is that if it had been administered on the same day with d- test, the researcher thought that the students would have given the same answers without reading the texts and dialogues. The administration of c-test on the next day made students perform a new and different test. The researcher believed that this application affected the reliability of the study in a positive way.

Data analysis

The data collected from the participants was analyzed using the Statistical Package for Social Sciences (SPSS version 17.0). The mean scores for both tests were calculated and analyzed by means of T-test to see whether there was a significant difference between the performances of participants in two tests. To see the correlation of the tests, the Pearson Correlation was used.

RESULTS

Reliability of the tests

Before the data analysis in the light of the three research questions, the reliability and scores of the tests administered for the study were computed using split-half method.

Table 1: Reliability Statistics - d test (discrete – point test) and c test (contextualized test)

	Guttman Split-Half Coefficient
c-test (contextualized test)	0,784
d-test (discrete-point test)	0,706

As it is shown in Table 1, the reliability score (Guttman Split-Half Coefficient) for the d-test is 0,706 and for the c-test, it is 0,784. In terms of reliability, since it is known that the closer the score to 1, the higher is the degree of reliability, it can be said that the reliability scores of these two tests are satisfactory. These scores contribute a lot to the reliability of the study.

Performances of the participants

In the light of the first research question, performances of the participants and related statistics were shown in Table 2.

Table 2 : Performances of Students

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	d_total	12,4472	123	4,34044	,39136
	c_total	13,8943	123	4,77496	,43054

Although there is not a big difference between them, the mean scores in Table 2 shows that the participants performed better in the contextualized test. While the participants had a score of 12.4 out of 25 in the test having discrete point items (d-test), they had 13.8 out of 25 in the test having contextualized items (c-test), which shows a 1.4 difference in their performances.

T-test results

To answer the second research question, the difference between the tests was computed by means of t-test.

Table 3: Differences Between Performances

	Mean	N	Std. Deviation	Sig. (2-tailed)
d_total	12,4472	123	4,34044	,000
c_total	13,8943	123	4,77496	

As it is understood from Table 3, the sig. value for the performances in two tests is 0.00 ($p < 0.05$), which means there is a statistically significant difference between the participants' performances in two tests. This significant difference can be interpreted as students perform better in a vocabulary test having contextualized items than a vocabulary test having discrete point items.

Correlation results

In the light of the third research question, the correlation between the results of two tests was computed by using Pearson Correlation.

Table 4: Correlations

		d_total	c_total
d_total	Pearson Correlation	1	,754**
	Sig. (2-tailed)		,000
	N	123	123
c_total	Pearson Correlation	,754**	1
	Sig. (2-tailed)	,000	
	N	123	123

** . Correlation is significant at the 0.01 level (2-tailed).

As it is shown in Table 4, the correlation coefficient (r) between the results of these two tests is 0.754, which means that the results of these two tests have a high and positive correlation. Depending on this finding, the square of correlation coefficient is calculated as 0,56 (56%) which explains the correlation of these two tests as a percentage.

DISCUSSION

This study aimed at finding answers to three research questions: a. What are the performances of students in a discrete item vocabulary test and in a contextualized vocabulary test? b. Is there a significant difference between the performances of students in these two vocabulary tests? c. Do these two types of tests correlate? If so, to what extent and what does it mean?

For the first question, the mean scores of two tests were computed. The participants had a score of 12.44 over 25 in the test having discrete-point items, and 13.8 in the test having contextualized items. It is seen that the students did better in the c-test than the d-test although they included the same target vocabulary items.

For the second questions, the performances of participants were compared by using t-test in order to see whether there was a significant difference between their performances. The findings showed that there was a significant difference between the performances of the participants. This significant difference reveals that participants performed better in a vocabulary test in which the target vocabulary items were presented in a contextualized way. This contextualization helped students grasp the meaning of the target word and created a meaningful way for the students in making responses to the items. On the other hand, presenting the target items in a discrete way provided no clues to the students for grasping the meaning and, they performed poorer.



For the third question, Pearson correlation coefficient was used to reveal the relation between these two tests. The finding showed that there was a high and positive correlation ($r : 0.754$) between these two different tests. This statistical finding reveals that there is a direct and positive relationship between these tests. That means, if the performances of students in the d-test increases or decreases, their performances in the c-test will also increase or decrease in the same way. Furthermore, the square of the correlation coefficient which was calculated as “R: 0.56” which means that the results of the d-test can be explained by the results of the c-test or vice versa in a degree of 56%. In other words, these two tests can be administered as an achievement test instead of each other in a degree of 56%, which is a very important statistical finding of this study.

One of the most important points that cannot be ignored in this study is that the tests were administered as an achievement test. Since the participants had already been taught the target vocabulary items, it was observed that they responded correctly in both tests if they remembered the target word, and they couldn't respond correctly in d-test if they did not remember the target words. Just in this case, the contexts helped them to respond correctly. For this reason, there is not a big difference between the mean scores of these tests. However, the researcher believes that if the tests had been designed as proficiency test the vocabulary items of which had not already been taught, the difference between the mean scores would have been broader and the contexts would have played a more important role in the responses of the participants.

CONCLUSION

Testing vocabulary has been a very hot debate in the world of testing for many years. Having too many dimensions such as what to test, how to test, criteria of knowing a word etc., the issue becomes much more challenging both for the instructors and test writers.

It is definite that the perspective on testing vocabulary will differ depending on the purpose of the test writers. The issues that must be taken into consideration in terms of the dimensions of testing vocabulary mentioned above will change depending on the testing objective. In other words, for example, the issues considered in an achievement vocabulary test will be different from the issues considered in the vocabulary part of a proficiency or certification test. No matter what the objectives or purposes are, since the importance of testing vocabulary has been revealed in the literature, it is crucial that this important language development skill is assessed validly and reliably. To accomplish this, several important issues need to be taken into account and decisions need to be made about how best to assess vocabulary.

This study focused on one of the issues in the dimension of how to test vocabulary, which is “contextualization”. The researcher provided a little contribution to the importance of contextualization in vocabulary testing by revealing the difference of a contextualized test and a discrete point test which were administered as achievement tests. Although the issue covered in this paper constitutes a small aspect of testing vocabulary, it provides a good start for the researcher for future studies on testing vocabulary.

WJEIS's Note: This article was presented at World Conference on Educational and Instructional Studies - WCEIS, 07- 09 November, 2012, Antalya-Turkey and was selected for publication for Volume 2 Number 4 of WJEIS 2012 by WJEIS Scientific Committee.

REFERENCES

- Cameron, L. (2001) *Teaching Languages to Young Learners*. UK: Cambridge University Press.
- Chastain, K. (1988). *Developing Second Language Skills: theory and practice*. (3rd Ed.). New York: Harcourt Brace Javanovich, Inc.
- Ellis, R & Redman, S. (1989) *A way with words*. UK: Cambridge University Press.



- Fitzpatrick, T. & Clenton, J. (2007) The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*. 27(4) 537–554.
- Folse, K. & Bologna, D. (2003). *Targeting Listening and Speaking: Strategies and Activities for ESL/EFL Students*. Ann Arbor: Univ. of Michigan Press.
- Harmer, J. (1991). *The Practice of English Language Teaching*. London: Longman.
- Hatch, E. & Brown, C. (1995) *Vocabulary, Semantic and Language Education*. New York: Cambridge University Press.
- Heaton, J.B. (1988) *Writing English Language Tests*. London. Longman.
- Lado, R. (1964) *Language testing; the construction and use of foreign language tests: a teacher's book*. New York, McGraw-Hill.
- Laufer, B. & Goldstein, Z. (2004) Testing Vocabulary Knowledge: Size, Strength and Computer Adaptiveness. *Language Learning*. 53 (4), 399-436.
- Nation, I. S. P. and Hwang K. (1995) Where would general service vocabulary stop and special purposes vocabulary begin? *System* 23, 35-41.
- Pearson, P. D. & Hiebert, E. H. & Kamil, M. L. (2007) Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*. 42 (2), 282-296.
- Qian, David D. (2008) *From Single Words to Passages: Contextual Effects on Predictive Power of Vocabulary Measures for Assessing Reading Performance*. *Language Assessment Quarterly*, v5 n1 p1-19 2008. 19 pp.
- Read, J. (1997a) Assessing vocabulary in a second language. In Clapham, C. & Corson, D. (eds.) *Encyclopedia of Language and Education. Language Testing and Assessment*. 7, 99-107.
- Read, J. (1997b) Vocabulary and testing. In Schmitt, N. & McCarthy, M. (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. (303-320) Cambridge: Cambridge University Press.
- Read, J. (2000) *Assessing Vocabulary*. UK: Cambridge University Press.
- Read, J. and Chapelle, C. (2001) A framework for second language vocabulary assessment. *Language Testing* 18 (1), 1-32.
- Read, J. (2007) Second Language Vocabulary Assessment: Current Practices and New Directions. *International Journal of English Studies*. 7 (2), 105-125.
- Schmitt, N. (1994) Vocabulary Testing: Questions for Test Development with Six Examples of Tests of Vocabulary Size and Depth. *Thai TESOL Bulletin*. 6 (2), 9-16.
- Schmitt, N. (2000) *Vocabulary in Language Teaching*. New York: Cambridge University Press.
- Stahl, K. A. D. & Bravo, M. A. (2010) Contemporary Classroom Vocabulary Assessment for Content Areas. *The Reading Teacher*. 63(7), 566–578.