the ESP, São Paulo, vol. 21 nº 2 181-205

CONTEXTUAL VOCABULARY KNOWLEDGE: THE BEST PREDICTOR OF NATIVE AND NON-NATIVE SPEAKERS' READING COMPREHENSION ABILITY

Conhecimento Contextual de Vocabulário: a Melhor Maneira de Prever a Habilidade de Compreensão de Leitura de Nativos e Não-Nativos

Ebrahim KHODADADY (Kurdistan University)

Abstract

Two hypotheses have been generated to capture the relationship between vocabulary knowledge and reading comprehension ability: instrumental and knowledge. While the former considers contextual vocabulary knowledge, i.e., knowing the meaning of the words used in a text, as the necessary and sufficient condition for comprehending that text, the latter holds it necessary but not sufficient. This study was conducted to test the instrumental hypothesis. The disclosed vocabulary section of the Test of English as a Foreign Language (TOEFL) was adopted as a measure of global vocabulary knowledge, i.e., knowing the meaning of the words unrelated to the text of TOEFL's reading comprehension section, which was used as a measure of reading comprehension ability. For measuring contextual vocabulary knowledge, another vocabulary test was constructed on the words used in the text of the reading comprehension test. The administration of the contextual vocabulary, global vocabulary, and reading comprehension tests to 64 non-native speakers (NNSs) and 123 native speakers (NSs) showed that although NSs scored significantly higher than NNSs on all tests, the performance of both NNSs and NSs on the contextual vocabulary tests was the best predictor of their performance on the reading comprehension test.

Key words: *contextual vocabulary knowledge; global vocabulary knowledge; reading comprehension ability; English language proficiency.*

Resumo

O autor apresenta duas hipóteses para a relação entre o conhecimento de vocabulário e a habilidade de compreensão de leitura: a instrumen-

tal e a relacionada ao conhecimento. Enquanto a primeira hipótese considera o conhecimento contextual de vocabulário, i.e., saber o significado das palavras usadas em um texto, como a condição necessária e suficiente para compreender esse texto, a segunda considera-o como necessário, mas não suficiente. O presente estudo foi realizado para testar a hipótese instrumental. A seção de vocabulário do TOEFL (Test of English as a Foreign Language) foi adotada como medida do conhecimento global de vocabulário, i.e., saber o significado das palavras não relacionadas ao texto da seção de compreensão de leitura do TOEFL (tal seção foi usada como medida da habilidade de compreensão de leitura). Para medir o conhecimento contextual de vocabulário, outro teste foi elaborado a partir das palavras do texto utilizado no teste de compreensão de leitura. A aplicação dos testes de vocabulário contextual, vocabulário global e compreensão de leitura em 64 falantes não-nativos (FNNs) e 123 falantes nativos (FNs) mostrou que, embora os FNs tenham apresentado pontuação significativamente maior do que os FNNs em todos os testes, o desempenho, tanto dos FNNs como dos FNs, nos testes de vocabulário contextual foi a melhor maneira de prever seu desempenho no teste de compreensão de leitura.

Palavras-chave: *conhecimento contextual de vocabulário; conhecimento global de vocabulário; habilidade de compreensão de leitura; proficiência em língua inglesa.*

1. Introduction

Research findings indicate that vocabulary knowledge is highly related to reading comprehension ability (Anderson & Freebody, 1981; Harris & Sipay, 1990; Sternberg & Powell, 1983). The relationship between vocabulary knowledge and reading comprehension ability is expressed by Sternberg (1987: 90) in its strongest version, i.e., "one's level of vocabulary is highly predictive, if not determinative, of one's level of reading comprehension". Sternberg's assertion is based on studies showing that global vocabulary knowledge, i.e., knowledge of words not used in the text of the criterion reading comprehension tests,

is the single largest predictor of reading comprehension ability (e.g., Anderson & Freebody, 1981; Carroll, 1976; Jensen, 1980; Singer, 1964, 1965; Sternberg, Powell, & Kaye, 1983).

The present study, however, differs from the cited literature in separating global vocabulary knowledge, i.e., knowledge of words unrelated to the text of the criterion reading comprehension test, from contextual vocabulary knowledge, i.e., knowledge of words used in the text of the criterion reading comprehension test. The relationship between contextual vocabulary knowledge and reading comprehension is best captured by the instrumental hypothesis of Anderson and Freebody (1983). The instrumental hypothesis postulates that knowing the meaning of the words in a text is the necessary and sufficient condition for comprehending that text. If this hypothesis holds true, then correlations between reading comprehension tests and contextual vocabulary should surpass the correlations between reading comprehension tests and global vocabulary tests, i.e., tests constructed on words which are not related to the texts of the criterion reading comprehension tests. In addition to higher correlations, contextual vocabulary tests should be the best predictors of NNSs and NSs' performance on reading comprehension tests.

2. Method

2.1. Participants

Following an invitation sent to various departments at the University of Western Australia, a total of 187 first year undergraduate students took part in this study voluntarily. Out of these, 123 were NSs (60 female and 63 male) and 64 were NNSs (35 female and 29 male).

The participants were majoring in various fields of study and thus were not homogenous in terms of their subject knowledge. Table 1 presents the fields of study of NSs and NNSs, separately. As can be seen, the highest percentage of NSs (52.8%) was studying science, i.e., chemistry, biology and so forth, whereas the highest percentage of NNSs (46.9%) was enrolled in commerce.

Field of study	Percentage		Field of study	Percentage	
	NSs	NNSs		NSs	NNSs
Science	52.8	35.9	Human movement	4.1	0.0
Commerce	23.6	46.9	Agriculture	0.8	0.0
Engineering	3.3	10.9	Geography	0.8	0.0
Law	6.5	3.1	Linguistics	0.0	1.6
Arts	4.1	0.0	Medicine	0.8	0.0
Economics	3.3	1.9			

Table 1: Percentage of fields studied by NSs (n = 123) and NNSs (n = 64)

Similar to their field of study, the NNSs were not homogenous in terms of their mother language. Because of the variety of languages an analysis of their effect on the tests administered was not conducted. Out of 64 NNSs, 1.6% spoke Arabic, 1.6% Azary, 1.6% Bulgarian, 50% Chinese, 3% Czech, 1.6% Danish, 1.6% Filipino, 1.6% German, 1.6% Hindi, 1.6% Hokkein, 9.4% Indonesian, 1.6% Italian, 1.6% Japanese, 7.8% Persian, 4.7% Polish, 1.6% Russian, 1.6% Samoan, 1.6% Sinhala, 1.6% Tamil and 3.1% Thai.

2.2. Materials

In line with the study of Yano et al. (1994), who used the Structure Subtest Form A of the Comprehensive English Language Test (see Oxford, 1987 for more details) as a criterion for language proficiency, the three disclosed subtests of the TOEFL (ETS, 1991) were utilized to design the English language proficiency test (ELPT) employed in the present study. The ELPT consisted of four multiple-choice item subtests measuring structure, contextual vocabulary and global vocabulary knowledge and reading comprehension ability as follows:

2.2.a. Structure multiple choice item test

Structure multiple choice item test (MCIT) consists of 15 multiple choice items having the context or stem of an incomplete

sentence (ETS 1991:82-84). From each sentence one or more words have been deleted. The deleted word or words have some structural function in the sentences. The participants must choose one of the four options that appear below the sentence and perform the structural function under question.

The first item of the structure multiple choice item test (MCIT) has been given as Example 1 below. This item measures the test takers' knowledge of the compound conjunction *not only ... but also*. The test takers should know that the conjunction under question is always used in its fixed form in order to connect the adjective *fresh* with *canned* and *frozen*.

Example 1
Meat can be purchased not only fresh ... canned and frozen.
a. also b. not also
c. but also * d. but not also

2.2.b Contextual vocabulary multiple choice item test

For measuring the contextual vocabulary knowledge of the participants, a vocabulary MCIT was developed by the present researcher. For writing the contextual vocabulary MCIT, the reading comprehension subtest of the TOEFL 91 was administered to 22 NNSs studying at the ESL Centre of the UWA. They were asked to read the passages one by one and underline the words whose meaning they did not know and could not even guess from the context of passages (Khodadady & Herriman, 1996).

Of 60 underlined unknown words, 30 most-frequently-underlined words were selected and changed into 30 multiple choice items (the contextual vocabulary MCIT can be found in the Appendix). The stem of contextual vocabulary multiple choice items provides the participants with the definition of the words specifically used in the texts. The distracters of contextual vocabulary MCIT were selected from the list of content words prepared by Khodadady (1995), Matthiesen (1993), and Sharpe (1986). The following item utilised in the test serves as an example:

Example 2
A distinguishing mark of social disgrace
a. stigma * c. garrison b. jeopardy d. moron

The correct answer *stigma* appears in one of the sentences comprised in passage five of the reading comprehension subtest: "A fourth grader can work at a fifth-grade level in math and a third-grade level in English without the *stigma* associated with being left back or the pressure of being skipped ahead". Since *stigma* has been used in the text, Example 2 is viewed as a contextual vocabulary multiple choice item which bears directly on the context in which the word *stigma* appears.

2.2.c Global vocabulary multiple-choice item test

For measuring the global vocabulary knowledge of the participants, the disclosed vocabulary subtest of TOEFL was included in the ELPT (ETS 1991:88-91). This subtest consists of 30 multiple choice items. For developing the global vocabulary multiple choice items, compound and complex sentences are presented as the stem of the items in which one content word is underlined.

For answering the global vocabulary multiple choice items, test takers have to choose out of the four options the one which is synonymous with the underlined word. Neither the stems nor the options of the global vocabulary MCIT have any sort of relationship with the texts forming the reading comprehension subtest. As Examples 3 and 4 indicate, the underlined words encompass totally unrelated concepts as semantically different as *tranquillity* and *hailed* are. They also belong to different parts of speech. Since the vocabulary multiple choice items of TOEFL are not related to any specific texts given in the reading comprehension subtest, they are referred to as global vocabulary items.

Example 3:
The lithographs of Currier and Ives captured the <u>tranquillity</u> of nineteenth-century rural life.
a. peacefulness * b. harshness c. weariness d. happiness

Example 4:
Frederick Douglas, hailed as one of the greatest black leaders in the history of the United States, was a noted abolitionist, journalist, and politician.
a. judged b. acclaimed * d. addressed c. remembered

2.2.d Reading comprehension multiple-choice item test

For measuring the reading comprehension ability of the participants, the reading comprehension subtest of TOEFL (ETS 1991:93-100) was used in the reading section of the ELPT. The study of Duran et al. (1987) showed that the multiple choice items of the TOEFL tests address (i) the main and supporting ideas expressed in the passages, (ii) facts brought up in the passages, (iii) inferences based on the information given in the passages, (iv) the synthesis of given information with other pieces of information, and (v) the tone or attitude of the passages.

The length of the passages used in this study varies from 179 to 298 words. According to Flesch's readability formula, passage four has the readability grade of graduate books (17+) though Fry's graph assigned it to senior undergraduate level (16). Passages one and three were suitable for undergraduate students, whereas passages two and five had the characteristics of senior high school textbooks. The readability indices thus showed that only two texts were suitable for the sample, i.e., first-year undergraduate students. Of the three remaining texts, one text was difficult for participants whereas the other two were easy.

2.3. Procedure

The researcher met with the participants in the lecture theatre of the Department of Accounting and Finance (DAF) of the UWA and explained the purpose of the study. Following this, the ELPT consisting of structure MCIT, contextual vocabulary MCIT, global vocabulary MCIT and reading comprehension MCIT was administered under standard conditions. The participants were not allowed to have anything other than

the tests on their desks. No answer sheets were used and the participants were required to mark their answers on their tests. When everyone had finished, the researcher collected the tests. The responses were manually scored at a later date according to the keys provided by ETS (1991).

2.4. Data analysis

The fundamental assumption underlying this study is the difference between native speakers (NSs) and non-native speakers (NNSs) in terms of their structure, contextual and global vocabulary knowledge on the one hand and their reading comprehension ability on the other. On the basis of this assumption the following two hypotheses are generated in this study:

1) The scores of NSs on MCITs measuring structure, contextual vocabulary and global vocabulary knowledge and reading comprehension ability will be significantly higher than the scores of NNSs.

2) Contextual vocabulary knowledge predicts the reading comprehension ability of NNSs and NSs' better than global vocabulary knowledge.

For determining the psychometrics of multiple choice items comprising the four subtests of the ELPT, *p*-values and point biserial correlation coefficients (r_{pbi}) were employed. *P*-value was calculated as the proportion of correct responses given to an item. Items having *p*-values less than 0.30 or greater than 0.70 were considered unacceptable or misfitting (Reynolds, Perkins & Brutten, 1994). The r_{pbi} coefficients were estimated by correlating the responses given to each individual item with the total test score, and these coefficients thus provided an index of item discrimination. According to numerous texts, an absolute value around 0.30 is usually sufficient to retain an item in a text (e.g., Kelin-Braley, 1981).

The scores of NNSs and NSs on the reading comprehension MCIT were also correlated with their scores on the structure MCIT,

contextual vocabulary MCIT and global vocabulary MCIT to determine the relationship between reading comprehension ability and structure knowledge, contextual vocabulary knowledge and global vocabulary knowledge. The correlation coefficients (r_{xy}) were based on the Pearson product-moment formula.

For determining the best predictor of variance in reading comprehension MCIT, the performance of NNSs and NSs were examined by using stepwise regression analysis. The participants' scores on the structure, contextual vocabulary, global vocabulary, and reading comprehension subtests were also subjected to a multivariate analysis of variance to find out whether NNSs and NSs performed differently on these subtests. The internal consistency of the four subtests was also assessed based on Cronbach's alpha. All of the statistical analyses were carried out using SPSS Release 6.1, standard version.

3. Results and Discussion

3.1. Psychometrics of structure multiple choice items

The psychometrics of the structure multiple choice items are provided in Table 2. As can be seen, out of 15 items, only one item (14) functioned well for both NNSs and NSs. These results indicate that since most of structure multiple choice items were very easy, i.e., their *p*-values fell between 1.00 and 0.80, they functioned very poorly for both NNSs and NSs.

Item	NNSs		NSs		Item	NNSs		NSs	
	p-value	r_{pbi}	*p*-value	r_{pbi}		*p*-value	r_{pbi}	*p*-value	r_{pbi}
1	1.00	.00	1.00	.00	9	.98	.21	.99	.02
2	1.00	.00	1.00	.00	10	.80	.48	.93	.48
3	.98	.33	1.00	.00	11	.91	.49	.99	.17
4	.95	.33	.98	.35	12	.97	.05	.98	.02
5	.98	.21	1.00	.00	13	.92	.38	.99	.17
6	.94	.49	.99	.33	14	.47	.38	.30	.58
7	.98	.09	.99	.17	15	.98	-.02	.97	.43
8	.92	.59	1.00	.00					

Table 2: Item psychometrics of the structure multiple-choice item test

The *p*-value of item 14 indicates that the structure under question was more difficult for NSs (0.30) than for NNSs (0.47). The stem of structure item 14, *Certain ancient societies used wood ashes and water for washing ... the resulting skin irritation with oil*, provides the necessary context to measure the test takers' structure knowledge dealing with the usage of the verb *use to* which requires an infinitive without to if it is followed by another verb. Test takers who sit for English language proficiency tests must have learned that someone *uses* something to do something. This very structural knowledge of test takers limits the range of attractive options and thus leads to the most critical deficiency of traditional MCITs, i.e., what attractive and plausible distracters should be developed to distract the less knowledge test takers (Popham, 1990).

The psychometrics of structure item 14 are given in Table 3. As can be seen, 11 NNSs (34%) 28 NSs (46%) in upper group (the top half) chose the incorrect distracter *a*. These results suggest that either the structure multiple choice item was ill constructed or native test takers were not competent in the structure under question. Given the educational level of native test takers, i.e., first-year undergraduate students, the former argument holds more plausible than the latter. Although the item has acceptable psychometrics, one of its distracters attracted a large percentage of high ability NSs and NNSs because there was no sound item writing theory which could be used by the item writer (Khodadady 1999).

Alternatives	NNSs (n = 64)		NSs (n = 122)		Item statistics	
	UG	LG	UG	LG	NNSs	NSs
a. and relieved	11	20	28	54	*p*-value: 0.47	*p*-value: 0.30
b. and the relieving	0	2	1	1	r_{pbi} : 0.38	r_{pbi} : 0.58
c. to relieve *	21	9	32	5		
d. to have relieved	0	1	0	1		

Table 3: Statistics of structure multiple choice item 11

3.2. Psychometrics of contextual vocabulary multiple choice items

Table 4 provides the psychometrics of the contextual vocabulary multiple choice items. As can be seen, while 8 items (3, 4, 6, 7, 10, 11, 19, and 25) functioned well for NNSs, only 3 items (5, 20, and 26) had acceptable item statistics for NSs, i.e., r_{pbi} greater than 0.30 and *p*-values between 0.30 and 0.70. These results indicate that since most of the contextual vocabulary multiple-choice items were very easy for both NNSs and NSs, i.e., their *p*-values fell between 1.00 and 0.80, they functioned poorly for both groups.

Item	NNSs		NSs		Item	NNSs		NSs	
	p-value	r_{pbi}	*p*-value	r_{pbi}		*p*-value	r_{pbi}	*p*-value	r_{pbi}
1	.92	.31	1.00	.00	16	.98	.34	1.00	.00
2	.80	.32	.98	.22	17	.83	.52	1.00	.00
3	.59	.43	.76	.44	18	.77	.50	.98	.24
4	.69	.54	.89	.38	19	.64	.50	.95	.16
5	.27	.30	.54	.44	20	.53	.21	.51	.46
6	.59	.40	.79	.47	21	.95	.08	.97	.07
7	.58	.46	.83	.22	22	.94	.28	1.00	.00
8	.97	.40	1.00	.00	23	.73	.58	.97	.15
9	.83	.51	.99	-.08	24	.17	.32	.28	.40
10	.52	.55	.78	.44	25	.58	.70	.77	.47
11	.59	.58	.85	.18	26	.25	.31	.45	.44
12	.72	.46	.86	.25	27	.81	.34	.95	.04
13	.83	.17	.91	.32	28	.86	.36	.95	.47
14	.83	.53	.99	.06	29	.78	.15	.84	.24
15	.86	.42	.96	.30	30	.89	.54	.98	.19

Table 4: Psychometrics of contextual vocabulary multiple choice items

As shown in Table 4, with the exception of contextual vocabulary multiple choice item 20, having the *p*-values of 0.53 and 0.51 for NNSs and for NSs, respectively, the *p*-values of all items are lower for NNSs than for NSs, indicating that the contextual vocabulary MCIT was easier for NSs. The psychometrics of item 20 are shown in Table 5. As can be seen, item 20 was more difficult for NSs because its distracters

functioned better for NSs than for NNSs. Distracters *a*, *b* and *d* attracted 25% of NNSs from the upper group whereas only distracter *b* appealed to 31% of NSs in the upper group. Since distracter *b*, i.e., morsel, has a semantic interrelationship with the keyed response *host*, it attracted more NSs from the upper group and could not discriminate among NNSs (r_{pbi} = 20).

Alternatives	NNSs (n = 64)		NSs (n = 122)		Item statistics	
	UG	LG	UG	LG	NNSs	NSs
a. glossary	1	2	0	11	*p*-value: 0.53	*p*-value: 0.51
b. morsel	6	15	19	28	r_{pbi} : 0.20	r_{pbi} : 0.46
c. host *	23	11	42	21		
d. sentry	2	1	0	1		
Omit	0	3	0	0		

Table 5: Statistics of contextual vocabulary multiple-choice item 20

None of the remaining seven items which had acceptable psychometrics for NNSs functioned well for NSs, i.e., their *p*-values were higher for NSs. Similarly, the two items (5 and 26) which had acceptable psychometrics for NSs did not function well for NNSs, i.e., their *p*-values were lower for NNSs, indicating that there is a relationship between the difficulty of items and their being answered by native and non-native test takers. This relationship can best be explained in the psychometrics of item 11which are provided in Table 6. The stem of item 11 *a membrane that lines body cavities* provides the definition of the keyed response *mucous*. While 31% of NNSs in the lower group chose the keyed response, 79% of NSs in the lower group (the bottom half) did so, indicating that contextual multiple-choice item 11 was very easy for NSs but of middle difficulty for NNSs.

Alternatives	NNSs (n = 64)		NSs (n = 122)		Item statistics	
	UG	LG	UG	LG	NNSs	NSs
a. segment	0	12	0	4	*p*-value: 0.59	*p*-value: 0.85
b. tyro	3	3	4	1	r_{pbi} : 0.57	r_{pbi} : 0.18
c. mucous *	28	10	55	49		
d. intestine	1	6	2	7		
Omit	0	1	0	0		

Table 6: Statistics of contextual vocabulary multiple-choice item 11

3.3. Psychometrics of global vocabulary multiple choice items

The psychometrics of global vocabulary multiple-choice items are presented in Table 7. As can be seen, out of 30 items, only 4 items (24, 25, 27 and 28) functioned well for NNSs. None of the global vocabulary multiple choice items had acceptable psychometrics for NSs. These results indicate that since most of global vocabulary multiple choice items were very easy, their *p*-values fell between 1.00 and 0.80, they functioned very poorly for both NNSs and NSs.

Item	NNSs		NSs		Item	NNSs		NSs	
	p-value	r_{pbi}	*p*-value	r_{pbi}		*p*-value	r_{pbi}	*p*-value	r_{pbi}
1	.92	.10	.98	.17	16	.91	.47	.97	.24
2	1.00	.00	1.00	.00	17	.94	.39	.99	-.03
3	.95	.11	.98	.17	18	.83	.50	.99	.55
4	.97	.38	.98	.63	19	.81	.39	.98	.53
5	.98	.23	.99	.34	20	.81	.63	.98	.49
6	.97	.48	.98	.45	21	.94	.41	.97	.40
7	1.00	.00	.98	.30	22	.89	.42	.98	.17
8	.92	.08	.84	.44	23	.81	.47	.98	.49
9	.98	.08	.99	.34	24	.64	.39	.84	.44
10	.91	.45	.98	.53	25	.70	.43	.90	.35
11	.95	.36	1.00	.00	26	.78	.51	.99	.13
12	.89	.36	.96	.08	27	.66	.47	.76	.52
13	.78	.23	.77	.18	28	.52	.44	.82	.32
14	.98	.30	1.00	.00	29	.80	.47	.88	.42
15	.81	.35	.94	.09	30	.98	.26	.99	.03

Table 7: Psychometrics of global vocabulary multiple choice items

As shown in Table 7, the *p*-values of global vocabulary multiple-choice items for NNSs are lower than those of NSs, indicating that all of the items were easy for NSs. As an example, the psychometrics of item 28 are given in Table 8. The stem of item 28, *A goal worked toward halfheartedly is seldom achieved,* provided test takers with the context of the underlined word which could best be substituted with the keyed response *apathetically.* Distracters *a, b* and *d* appealed to 22% and 75% of NNSs in the upper and lower groups, respectively, and thus rendered item 28 of middle difficulty for NNSs. In contrast, only distracter *b* attracted 5% of NSs in upper group whereas distracters *b* and *d* were chosen by 32% of NSs in the lower group, indicating that item 28 was easy for native test takers.

Alternatives	NNSs (n = 64)		NSs (n = 122)		Item statistics	
	UG	LG	UG	LG	NNSs	NSs
a. customarily	2	1	0	0	*p*-value: 0.52	*p*-value: 0.82
b. bipartisanly	4	11	3	19	r_{pbi} : 0.44	r_{pbi} : 0.32
c. apathetically *	25	8	58	41		
d. unconventionally	1	12	0	1		

Table 8: Statistics of global vocabulary multiple choice item 28

3.4. Psychometrics of reading comprehension multiple choice items

Table 9 provides the psychometrics of the reading comprehension multiple choice items. As can be seen, while two items (2 and 13) functioned well for NNSs, none of the 30 items had acceptable statistics for NSs. These results indicate most of the reading multiple choice items functioned very poorly for both NNSs and NSs because they were very easy, i.e., their *p*-values fell between 1.00 and 0.70.

Item	NNSs		NSs		Item	NNSs		NSs	
	p-value	r_{pbi}	*p*-value	r_{pbi}		*p*-value	r_{pbi}	*p*-value	r_{pbi}
1	.77	.49	.89	.27	16	.91	.49	.98	.48
2	.70	.43	.81	.26	17	.88	.50	.86	.56
3	.98	.09	.99	-.03	18	.94	.26	.99	.57
4	.89	.28	.94	.53	19	.75	.30	.77	.57
5	.95	.21	.99	.00	20	.77	.26	.92	.44
6	.88	.15	.96	.36	21	.80	.27	.93	.34
7	.97	.19	.97	.36	22	.88	.45	.99	.45
8	.95	.27	.97	.16	23	.83	.28	.92	.51
9	.84	.65	.98	.04	24	.89	.56	.95	.30
10	.88	.48	.88	.51	25	.84	.51	.95	.53
11	.86	.43	.96	.44	26	.83	.48	.85	.50
12	.89	.52	.97	.64	27	.83	.63	.93	.63
13	.63	.37	.89	.39	28	.94	.33	.94	.44
14	.77	.36	.88	.33	29	.86	.51	.95	.44
15	.95	.32	.98	.59	30	.83	.24	.85	.56

Table 9: Psychometrics of reading comprehension multiple choice items

As shown in Table 9, the *p*-values of reading comprehension multiple choice items were lower for NNSs than for NSs, indicating that they were easier for NSs. For example, item 2 was of low difficulty for NNSs (*p*-value = 0.70) whereas it was very easy for NSs (*p*-value = 0.81). The stem of item 2 poses the question: *According to the passage, Julia Morgan most likely had her office in which of the following locations?* The keyed response *San Francisco* and the distracters of item 2 appear in the following context.

> The first woman to graduate in architecture from the Ecole des Beaux– Arts in Paris, she was one of *San Francisco's* most renowned and prolific architects. Starting in 1902, she designed close to 800 buildings throughout California and the West, most of them as understated and beautifully proportioned as San Simeon is extravagant and jumbled. Several, like the Berkeley City Club and the YWCA of Oahu, have landmark status; two – San Simeon and the Asilomar Conference Center on the Monterey Peninsula – are California state monuments.

The distracters of reading comprehension multiple choice item 2 are specified in Table 10. As can be seen, none of NNSs and NSs in the lower groups chose distracter *c*, indicating that it was not attractive at all. *San Simeon* as distracter *b*, however, attracted some NNSs and NSs not only from lower groups but also from upper groups because it shares the initial *San* with *San Francisco*. Furthermore, the stem of item 2 deliberately misled the test takers by using the term *location* to make them believe that *San Simeon* is a city like *San Francisco* where one can have one's own office. These results suggest that in addition to ambiguous stems, some of the distracters designed by experts do not function as they should.

Alternatives	NNSs (n = 64)		NSs (n = 122)		Item statistics	
	UG	LG	UG	LG	NNSs	NSs
a. Oahu	0	2	0	3	*p*-value: 0.70	*p*-value: 0.81
b. San Simeon	6	11	3	17	r_{pbi} : 0.43	r_{pbi} : 0.26
c. Berkeley	0	0	0	0		
d. San Francisco *	26	19	58	41		

Table 10: Statistics of reading comprehension multiple-choice item 2

3.5. Test psychometrics

Table 11 presents the descriptive statistics for the scores of the participants on the four subtests of the ELPT. Among the four subtests, the structure MCIT was the least reliable test for NNSs (0.39) and NSs (0.36). The reading comprehension MCIT proved to be the most reliable subtest for NNSs (0.82) and NSs (0.80). The contextual vocabulary MCIT was as reliable as the reading comprehension test (0.82), whereas the reliability coefficient of the global vocabulary MCIT was lower than the contextual vocabulary MCIT (0.75) for NNSs. Both contextual and global vocabulary MCITs, however, proved to be of low reliability for NSs (0.58).

Test Takers	Tests	Mean	Standard Deviation	Skew	Cronbach's ∝
NNSs	Structure	13.8†	1.2	-1.66	.39
	Contextual vocabulary	21.3††	4.9	-.31	.82
	Global vocabulary	26.1††	3.3	-.93	.75
	Reading comprehension	25.7††	3.9	-1.14	.82
NSs	Structure	14.5†	.78	-2.00	.36
	Contextual vocabulary	25.7††	2.5	-.63	.58
	Global vocabulary	28.5††	1.7	-.2.68	.58
	Reading comprehension	27.8††	3.0	-.3.38	.80

Table 11: Descriptive statistics for the scores of NNSs (n = 64) and NSs (n = 123) on the four subtests of the ELPT

Note: † Mean score reported out of 15
†† Mean score reported out of 30

3.6. Effect of language

A multivariate analysis of variance was undertaken to determine whether the mother language of test takers brings about significantly different performance on the structure, contextual vocabulary, global vocabulary and reading comprehension MCITs. Wilks Lambda test indicated that the means of NNSs and NSs on the four subtests were significantly different ($F = 182$, $df = 4$, $p < 0.0001$). The follow up Univariate F-Tests shown in Table 12 revealed that the means of NNSs and NSs are substantially different on the structure MCIT, contextual vocabulary MCIT, global vocabulary MCIT and reading comprehension MCIT ($p < 0.0001$). These results *support* the first hypothesis which stated that the scores of NSs on MCITs measuring structure, contextual vocabulary and global vocabulary knowledge and reading comprehension ability will be significantly higher than the scores of NNSs.

Tests	Hypoth SS	Error SS	Hypoth MS	Error MS	F
Structure	18.774	155.857	18.773	.842	22.28*
Contextual vocabulary	819.328	2263.784	819.328	12.237	66.96*
Global vocabulary	249.534	1048.088	249.538	5.665	44.05*
Reading comprehension	194.572	2096.450	194.572	11.332	17.17*

*$p < .0001$

Table 12: Multivatiate analysis of variance for the scores of NNSs and NSs on the four subtests of the ELPT

3.7. Validity

The correlations between the four subtests of the ELPT are given in Table 13. As can be seen, the reading comprehension MCIT correlated significantly with the structure MCIT (0.33), contextual vocabulary MCIT (0.45) and global vocabulary MCIT (0.44) for NNSs, indicating that there is a positive relationship between the reading comprehension ability of NNSs and their structure, contextual and global vocabulary knowledge. The performance of NSs, however, correlated only with their performance on the contextual and global vocabulary MCITs, indicating that there is no relationship between their structural knowledge and reading comprehension ability.

Participants	Tests	Contextual vocabulary	Global vocabulary	Reading comprehension
NNSs	Structure	.50***	.45***	.33*
	Contextual vocabulary		.82***	.45***
	Global vocabulary			.44***
NSs	Structure	.17	.12	.12
	Contextual vocabulary		.35***	.48***
	Global vocabulary			.29**

* $p < .01$, ** $p < .001$, *** $p < .0001$

Table 13: Correlation coefficients between the structure, contextual vocabulary, global vocabulary and reading comprehension subtests of the ELPT

As the results presented in Table 13 indicate, while the performance of NNSs on contextual vocabulary MCIT correlated slightly higher than global vocabulary MCIT with the reading comprehension MCIT (0.45 and 0.44, respectively) the correlation coefficients between the contextual vocabulary MCIT and reading comprehension MCIT, i.e., 0.48, surpassed that of global vocabulary MCIT, i.e., 0.29, for NSs, indicating that the relationship between reading comprehension ability and contextual vocabulary knowledge is stronger than the relationship between reading comprehension ability and global vocabulary knowledge, particularly for NSs.

A step wise regression analysis on the scores of NNSs on the reading comprehension MCIT as a dependent variable and structure, contextual vocabulary and global vocabulary MCITs as three independent variables indicated that the contextual vocabulary MCIT is the best predictor of their reading comprehension ability. This accounts for 20% of variance in the reading comprehension MCIT for NNSs ($F = 15.98$, $df = 1$, $p < .0001$). Due to the high correlations of the contextual vocabulary MCIT with the other independent variables, the structure and global vocabulary MCITs made no further contribution to the reading comprehension MCIT.

Similarly, when a step wise regression analysis was run on the performance of NSs on the reading comprehension MCIT as a dependent variable and structure, contextual vocabulary and global vocabulary MCITs as three independent variables, the contextual vocabulary MCIT emerged as the best predictor, accounting for 23% of variance in the reading comprehension MCIT ($F = 36.04$, $df = 1$, $p < .0001$). Due to the significant correlations between the contextual and global vocabulary MCITs, the global vocabulary MCIT made no further contribution to the reading comprehension MCIT. These results *support* the second hypothesis that contextual vocabulary knowledge is the best predictor of reading comprehension ability for both NNSs and NSs.

4. Summary and conclusion

Factors such as the authenticity of texts, expertise of test writers, participants, gender, age, heterogeneity of non-native test takers in terms of their mother language, format of the test and the difficulty level of texts might have contributed to the results obtained in this study. For example, Khodadady (1997) developed a reading comprehension test consisting of 25 multiple choice items on four authentic and unmodified newspaper and magazine articles and administered it to 44 year 12 high school students (22 NSs and 22 NNSs). The administration of the test did not show any significant difference in the reading comprehension ability of NSs and NNSs. The results of the present study, however, demonstrate that the performance of NNSs and NSs on structure, contextual vocabulary, global vocabulary and reading comprehension multiple choice item tests designed by experts is significantly different.

The disagreement between the findings of the present study and those of Khodadady (1997) calls for further research. It is suggested that the study be replicated with speakers of English as a foreign language, i.e., in countries where English is not the language of communication, and their academic fields as well as mother language be controlled. Furthermore, it would be interesting to find out whether the performance of NNSs and NSs is significantly different on contextual vocabulary MCITs developed on reading comprehension tests other than the TOEFL.

The results obtained in this study also indicate that although NNSs and NSs differ significantly in terms of their structure knowledge, global vocabulary knowledge and contextual vocabulary knowledge, the contextual vocabulary knowledge of both NNSs and NSs is the best predictor of their reading comprehension ability. This finding has an important implication for the field of language testing. Since contextual vocabulary knowledge is the best predictor of reading comprehension ability, the structure and global vocabulary sections of language proficiency tests should be replaced with contextual vocabulary MCITs.

Recebido em: 10/97. Aceito em: 03/99.

References

ANDERSON, R. C. & P. FREEBODY 1981 Vocabulary knowledge. IN: J. GUTHRIE (ed.) *Comprehension and teaching: research reviews.* International Reading Association.

ANDERSON, R. C. & P. FREEBODY 1983 Reading comprehension and the assessment and acquisition of word knowledge. IN: B. HUTSON (ed.) *Advances in reading/language research: a research annual*. JAI Press.

CARROLL, J. B. 1976 Psychometric tests as cognitive tasks: a new "structure of intellect." IN: L. B. RESNICK (ed.) *The nature of intelligence.* Lawrence Erlbaum Associates.

DURAN, R. P., M.CANALE, J.PENFIELD, C.W. STANFIELD & J.E.LINSKIN-CASPARRO 1987 TOEFL from a communicative viewpoint on language proficiency: a working paper. IN: R. O. FREEDLE & R. P. DURAN (eds.) *Cognitive and linguistic analyses of test performance*.

EDUCATIONAL TESTING SERVICE. 1991 *Reading for TOEFL*. ETS.

HARRIS, A. J. & E. R. SIPAY 1990 *How to increase reading ability: a guide to developmental & remedial methods* (9th ed.) Longman.

JENSEN, A. R. 1980 *Bias in mental testing*. Free Press.

KELIN-BRALEY, C. 1981 *Empirical investigations of cloze tests: an examination of the validity of cloze tests as tests of general language proficiency in English for German university students*. Unpublished Ph.D. dissertation, Liverpool, England.

KHODADADY, E. 1995 *Textual analysis of testing material: validity readdressed.* Paper presented at the 8th educational conference of the ELICOS Association of Australia, Fremantle, Western Australia, September 1995.

_____ 1997 *Schemata theory and multiple choice item tests measuring reading comprehension*. Unpublished Ph.D. dissertation, the University of Western Australia, Perth.

_____ 1999 *Multiple choice items in testing: practice and theory*. Rahnama Publications.

KHODADADY, E. & M. HERRIMAN 1996 *Contextual lexical knowledge and reading comprehension: relationship and assessment.* Paper presented at the 9th Educational conference of the ELICOS Association of Australia, Sydney, New South Wales, September 1996.

MATTHIESEN, S. J. 1993 *Essential words of the TOEFL*. Baron's Educational Series.

OXFORD, R. 1987 Review of the Comprehensive Language Test. IN: J. C. ALDERSON, K. J. KRANHKE, & C. W. STANSFIELD (eds.) *Reviews of English Language Proficiency Tests*. Waschington, DC: TESOL.

POPHAM, W. J. 1990 *Modern educational measurement: a practitioner's perspective*. Prentice Hall.

REYNOLDS, T., K. PERKINS & S. BRUTTEN 1994 A comparative item analysis study of a language testing instrument. *Language Testing*, **11.1:** 1-13.

SHARPE, P. J. (1986). *Baron's how to prepare for the TOEFL test of English as a foreign language* (5th ed.) Baroon's Educational Series.

SINGER, H. 1964 Substrata factor pattern accompanying development in power of reading, elementary through college level. IN: E. THURSTON & L. HAFNER (eds.) *Philosophical and sociological basis of reading. Fourteenth Yearbook of the National Reading Conference*, **14:** 41-56.

SINGER, H. 1965 A developmental model of speed of reading in grades 3 through 6. *Reading Research Quarterly*, **1:** 29-49.

STERNBERG, R. J. 1987 Most vocabulary is learned from context. IN: M. G. MCKEOWN & M. E. CURTIS (eds.) *The nature of vocabulary acquisition*. Lawrence Erlbaum Associates.

STERNBERG, R. J. & J. S. POWELL 1983 Comprehending verbal comprehension. *American Psychologist*, **38:** 878-893.

STERNBERG, R., J. S. POWELL & D. B. KAYE 1983 The nature of verbal comprehension. IN: A. C. WILKINSON (ed.) *Communicating with computers in classroom: prospects for applied cognitive science*. Academic Press.

YANO, Y., M.H. LONG & S. ROSS 1994 The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, **44.2**:bb 189-219.

APPENDIX: Contextual vocabulary test

Directions: There are 30 definitions or words in this test. Beneath each definition or word you will see four words, marked a, b, c, and d. Choose the one word that fits or provides a synonym for the given definition or word. Mark the corresponding box on your answer sheet.

1. existing in plentiful supply
   a. pensive b. abundant*
   c. incessant d. ominous

2. to put in peril
   a. placate b. prevail
   c. reverse d. endanger*

3. any of various burrowing bivalve molluscs
   a. rat b. crayfish
   c. clam* d. frog

4. incapable of producing offspring, seed or fruit
   a. barren* b. negligent
   c. prudent d. tentative

5. a great number
   a. glossary b. morsel
   c. host* d. sentry

6. pretentious
   a. vulnerable b. ostentatious*
   c. affluent d. treacherous

7. in effect though not in fact
   a. incisively b. virtually*
   c. obscurely d. deliberately

8. to become smaller in size
   a. shrink* b. cruise
   c. dissect d. grope

9. a small pool of rain
   a. hoax b. puddle*
   c. oasis d. ordeal

10. a distinguishing mark of social disgrace
   a. stigma* b. jeopardy
   c. garrison d. moron

11. a membrane that lines body cavities
   a. segment b. tyro
   c. mucous* d. intestine

12. to adapt so as to make suitable
   a. pilfer b. scoop
   c. rebut d. tailor*

13. to be omitted as in passing
   a. roam b. skip*
   c. deride d. vanish

14. a protective layer of dead cells on woody plants
   a. pillar b. timber
   c. bark* d. trunk

15. rich or brilliant in colour, design
a. amicable b. surly
c. flamboyant* d. vanish

16. performing without delay
a. promptly* b. hilariously
c. frigidly d. permanently

17. excessively fat
a. obese* b. agile
c. limpid d. profound

18. to conjecture without knowing the complete facts
a. speculate* b. flicker
c. traverse d. ponder

19. to say something in restrained terms
a. understate* b. ramble
c. swerve d. tamper

20. producing constant or successful results
a. audacious b. prevalent
c. prolific* d. uncouth

21. the relative position of a person or thing
a. avarice b. precept
c. status* d. troupe

22. to use up
a. scrape b. consume*
c. terminate d. reiterate

23. to feel strong disgust for
a. loathe* b. laud
c. peek d. abject

24. a feeling of grudging
a. wrath b. rancour*
c. envy d. vigilance

25. keep away from
a. menace b. shun*
c. reproach d. quell

26. to feel confusion
a. molest b. embarrass*
c. concoct d. dilate

27. not explicit
a. vague* b. sporadic
c. tepid d. callous

28. the residence of a religious community living in seclusion
a. cathedral b. revenue
c. monastery* d. shutter

29. to hold in check
a. suppress* b. abate
c. ferry d. litter

30. a person who has given long service
a. accomplice b. veteran*
c. loafer d. invalid

Acknowledgements

I am grateful for the comments provided by two anonymous *the ESPecialist* referees.

Ebrahim Khodadady is an assistant professor of Applied Linguistics at Kurdistan University, Sanandaj, Iran. He received his Ph.D. from the University of Western Australia in 1998 and has been the director general of research at Kurdistan University since then. His main interests are language testing, research and teaching methodology.