C-test and Vocabulary Knowledge

NEDA KARIMI
Islamic Azad University, Najafabad Branch

Bio Data:
Neda Karimi received her master's degree in applied linguistics from Islamic Azad University, Najafabad Branch, Iran. She has presented papers in national and international conferences. She has also taught English in private Institutes for about five years. Her major research interests include language testing and sociolinguistics.

Abstract
This study aims to investigate the applicability of C-test in vocabulary assessment. Doing so four C-tests, two different in terms of their lexical richness (measured through P_Lex) and the other two different in terms of their lexical variation (measured through type/token ratio), were developed and the performance of 74 and 77 Iranian TEFL MA students on these tests was examined. These C-tests were identical in every term except lexical characteristics. The results revealed that the test takers performed significantly different in both situations. This led to the conclusion that C-test has the potential to tap vocabulary knowledge and with the consideration of lexical characteristics of C-test texts we can tailor C-tests that are suitable for vocabulary knowledge assessment.

Keywords: C-test, lexical richness, lexical variation, P_Lex, type/token ratio

Introduction

C-test, an integrative written test of general language proficiency, belongs to the family of reduced redundancy tests. It consists of five to six short, authentic, and complete texts. In these texts the first and the last sentences are left standing. From second sentence of the text the 'rule of two' is applied: beginning at word two in sentence two the second half of every second word is deleted. Numbers, proper names and one letter words are left undamaged, but otherwise the deletion is

entirely mechanical. The process is continued until the required number of blanks has been produced. Texts are arranged in order of difficulty with easiest texts first.

First proposed in 1981 by Klein-Braley as an alternative to Cloze test, C-test has been the subject of a massive bulk of empirical researches since its introduction to the field of language testing, some on validating C-test as a measure of general language proficiency through qualitative and quantitative methodologies (Eckes & Grotjahn, 2006; Babaii & Ansary, 2001; Dörnyei & Katona, 1992), some particularly on comparison of C-test with cloze test procedure (Jafarpur, 1995; Jafarpur, 2002), and others on studying the C-test takers' performance, the strategies used and the processes involved in C-test taking (Feldmann, & Stemmer, 1987, Babaii & Ansary, 2001; Stemmer, 1991; Kamimoto, 1992). C-tests are usually not associated with vocabulary measurement and are used to measure overall language proficiency however, second language researchers, Singleton and Little, have suggested that C-test can be used in vocabulary research (Singleton & Little, 1991; Singleton & Singleton, 1998; Singleton, 1999).

The role of vocabulary in C-test has also been challenged by Chapelle and Abraham (1990) who found that their C-test correlated highly with their multiple-choice vocabulary test (r = .862). The correlation was substantially higher than with other parts of the placement test battery, including the writing test (r = .639) and the reading test (r = .604). Looking at it the other way, the vocabulary test had a stronger association with the C-test than with any of the other three versions of the cloze procedure that Chapelle and Abraham administered. The researchers interpreted this as evidence that the C-test was particularly good as measure of what Alderson

would call 'lower-level' knowledge of lexical and grammatical elements, while at the same time it also drew on 'higher-level' contextual competence, as indicated by the substantial correlations with the writing and reading subtests.

The present study raises the question of applicability of the C-test in vocabulary assessment. Does C-test have the potential to tap vocabulary knowledge? Can it gain an insight into the vocabulary knowledge of foreign language learners? How do test takers perform on C-tests with different levels of lexical richness? These are the questions that this study attempts to seek answers to. Doing so we investigate how lexical richness of the C-test texts affects the performance of the test takers.

Propositional content and vocabulary as an important component of language has an influential effect on language use either in a natural non-testing environment or in a testing situation. The frequency of words and their degree of specialization as important characteristics of language nature have possible influence on the performance of the examinees in a testing environment.

The effect of vocabulary as an important factor on language use, mostly on the performance of test takers on text-related tests such as tests of writing, reading comprehension and reduced redundancy tests, and the fact that teachers are frequently called upon to make judgments about the types of vocabulary their students use, and to judge the adequacy of this vocabulary led to the introduction of various formal measures of lexical characteristics and richness. Lexical richness has been considered an illuminative predictor of learners' general language proficiency (Zareva, Schwanenflugel, & Nikolova, 2005).

It is important, however, to be clear what we mean by vocabulary richness and how it is measured. Clearly, it would be helpful if we could be more specific about exactly what constitutes 'wide' or 'rich' vocabulary.

The most widespread measures of lexical richness are based on the concept of vocabulary diversity or vocabulary variety. Indices of lexical variation have been widely used in studies of lexico-statistics (for example, Herdan, 1960) and L1 development (for example, Miller & Klee, 1995), and they have begun to appear with increasing frequency in work on L2 speakers (for example, Arnaud, 1984; Broeder *et al.*, 1988; Malvern & Richards, 1997). Most of this work uses measures that compare the number of Lexical Types in a text with the number of Lexical Tokens in the same text.

This study used TTR measure to estimate the lexical variation of the texts. The main criticism of TTR is that the measure is highly text length dependent. However, as there was control over the length of the C-test parallel texts in this study the mentioned shortcoming of this measure of lexical variation was overcame.

Lexical richness has also been studied in the context of word frequency distributions that leads to different measures of lexical sophistication. These measures would not be limited to the number of Types or Tokens appearing in an L2 text: they would supplement this information with additional information about the sorts of words being used, and the sorts of lexical choices that are being made in a particular text. These measures are based on outside-text evidences (i.e. external standards such as frequency lists).

An example of a measure of this sort is to be found in Laufer and Nation's Lexical Frequency Profile (Laufer & Nation, 1995). LFP looks at the proportion of high frequency general service and academic words in learners' writing to reflect the vocabulary size and growth in L2 learners. However a serious practical problem with LFP is that it requires relatively long texts (profiles over 200 words as claimed by Laufer and Nation) for stable measures to emerge.

Whereas the LFP requires texts over 200 tokens to obtain stable scores, P_Lex (Meara & Bell, 2001) seems to be an alternative approach to assessing the lexical complexity of short texts. It has a passing resemblance to the Lexical Frequency Profile (LFP) but Meara and Bell (2001) suggest that P_Lex works better with short texts which makes it a more practical measure for estimating the lexical sophistication of the C-test texts and therefore subject matter of this study.

P_Lex divides the text into segments of 10 words each, and then counts the number of 'difficult' words in each segment. According to Meara (2001), P_Lex considers 'difficult words': "any word which is not found in a short list of high frequency words", which in practice means any word not included in the 1,000 most frequent English content words. Thus, P_Lex looks at the distribution of 'difficult words' in texts, and provides a simple index that indicates how likely the occurrence of these words is. The higher the P_Lex score, the bigger the vocabulary size of the learner is. P_Lex v2.0 is a computer program for evaluating P_Lex.

Method

Participants

In investigating the effect of lexical richness on C-test performance the study was conducted with the participation of the graduate students majoring TEFL at the university of Najaf Abad. The C-test results of 77 students for investigating the effect of lexical variation and 74 students for investigating the effect of lexical sophistication were used for the analysis. Due to the large number of instruments and consequently large number of sessions, a few participants were absent in one session or more. Most of the participants were in their twenties. They were of both sexes and all Persian native speakers. They were attending four intact classes and their homogeneity, in terms of language proficiency, was established through the C-test validated in Dornyei and Katona (1992) as a measure of English language proficiency (Appendix 1). Therefore, the sample appears to be homogeneous with regard to age, English language proficiency, L1 background, and educational level.

Instrumentation

The following instruments were employed in this study: a) the C-test validated in Katona and Dornyei (1992) (Appendix 1), and b) four newly developed C-tests (Appendix 2, Appendix 3, Appendix 4, and Appendix 5). These instruments are presented in more detail in the next section.

Katona and Dörnyei's (1992) C-test

The C-test validated in Katona and Dornyei (1992) as a measure of language proficiency, consisting of four short texts of 24, 17, 21, and 19 mutilated words, was used to homogenize students regarding their language proficiency. Katona and Dornyei (1992) administered four different language tests (including an oral

interview) along with this C-test to subjects of two different groups of language proficiency to validate the C-test. The results of the program confirmed that the C-test is a reliable and valid measure of language proficiency.

The C-tests

Four C-tests were developed for the purpose of the study, two for examining the effect of lexical variation, and two for examining the effect of lexical sophistication. The C-test developed for examining the effect of lexical variation consisted of five short authentic texts with 20 mutilated words for each text, total of 100 mutilated words and the other two C-tests developed for examining the effect of lexical sophistication consisted of four short texts with 20 mutilated words for each text, total of 80 mutilated words. In the following sections the procedures performed for text selection and test development are explained in detail.

Text Selection

Initially an enormous number of authentic texts on 10 different topics were extracted from the internet. Because of the limitations in choosing authentic texts due to the high degree of control that was inevitable considering the research problems, the process of text selection was very time-consuming.

To develop the C-tests for the first part of the study i.e., examining the effect of lexical variation, 10 texts were chosen out of this bulk. In choosing the texts the factor of type/token ratio was considered as the criterion. That is, the parallel texts differed in their type/token ratio. However, they were two by two on the same topic and shared the same Flesch-Kincaid readability level, they were also selected with great precision with regards to their length and lexical sophistication. Therefore the

topic, readability level, length, and lexical sophistication of the texts were controlled. As a result we had five texts with lower type/token ratio and five texts with higher type/token ratio.

The study employed wordsmith 5.0 to measure the type/token ratio of the texts. Wordsmith 5.0 is a lexical analysis software published by Oxford University Press.

For second part P_Lex was employed to measure the lexical sophistication of the texts. P_Lex is a measure of lexical richness and P_Lex v2.0 is a program for evaluating the vocabulary used in short texts. P_Lex is based on the idea that 'difficult' words occur only infrequently in texts. P_Lex looks at the distribution of difficult words in a text, and returns a simple index that tells us how likely the occurrence of these words is (Meara & Bell, 2001). This index is called lambda. Lambda values typically range from 0 to about 4.5, with higher figures corresponding to a higher proportion of infrequent words. The P_Lex methodology gives lambda scores that are reasonably stable with very short texts, therefore suitable to be utilized to evaluate the degree of lexical richness in the short texts of the C-tests. The current version of P_Lex is based on Nation's (1984) word lists.

Like the previous part, 10 texts were chosen and this time the lambda value was considered as the criterion. The corresponding texts were chosen in a way that they were different in their lambda value but shared almost the same type/token ratio. In addition, the parallel texts were on the same topic and shared the same readability level and length like the first part of the study. Again, we had five texts with lower lambda value and five texts with higher lambda value.

Development of the C-tests

Based on guidelines given by Raatz and Klein-Braley (1982) the texts were used for the development of the C-tests. The texts were arranged from easy to difficult (based on Flesch-Kincaid readability level). The newly developed C-tests were examined with a group of 10 EFL teachers. Their scores on all the five C-tests were over 90% correct on average. Furthermore, the C-tests were piloted with a group of subjects similar to the target group. An original version of TOEFL was also utilized to serve as the validation criterion for the C-tests. Due to some practical limitations, however, the 'listening comprehension' and the 'writing' sections were excluded. That is, TOEFL comprised 100 multiple-choice items on the 'grammar and written expressions' (40 items), and 'vocabulary and reading comprehension' (60 items) to be completed in 70 minutes.

Reliability of the C-tests was measured by K-R21. The information about the psychometric characteristics of the C-tests employed in the study is provided in table 1, below.

Table 1

Psychometric Characteristics of the C-tests

	K	*KR-21*	*Correlation Coefficient*
C-test with texts of higher TTR	100	0.81	0.80
C-test with texts of lower TTR	100	0.80	0.76
C-test with texts of higher lambda value	100	0.81	0.79

C-test with texts of lower lambda value	100	0.74	0.81

Finally, based on the results of item analysis, both text number four of the C-tests developed to answer the second question of the study i.e. C-tests examining the effect of lexical sophistication were removed from the tests due to the relatively high number of malfunctioning items in text four with higher lambda value. The facility index for this text was 0.786 which corresponds to "too easy."

As shown in table 1, all four C-tests were reliable and appropriate for the research purpose. In fact, the newly developed C-tests as reliable and valid measuring instruments could provide us with interpretable data to probe the questions set forth in the study.

Procedure

The participants attended 4 intact classes and the data collection was done in the first 30 minutes of each class in five sessions.

At first the participants were fully informed about the C-test and how they should function on the test; a sample C-test was practiced by the participants and the researcher to familiarize them with the test.

In the first session the C-test comprising texts with lower lambda value was administered to the participants. The rationale behind administering the C-tests comprising texts of lower lexical sophistication in the first session was the assumption that these texts are easier. In other words, administering more

sophisticated C-tests to a group of subjects with no familiarity with the test format could influence their performance.

Since the parallel texts in the C-tests were on the same topic, counter balancing design was used. This could minimize the effect of topic familiarity on the performance of the participants. The second session was allotted to the administration of the C-test comprising the texts with lower TTR. The C-test comprising the texts with higher Lambda value and the C-test comprising the texts with higher TTR were administered in the third and fourth session respectively. Finally in the last and fifth session all participants took the C-test taken from Katona and Dornyei (1992).

Meanwhile due to a number of absentees, there were two additional sessions for the absent participants.

Results

Results of Lexical Variation and C-test Performance

The effect of lexical variation on the C-test performance is tested. That is, the performance of the participants on the C-test with texts of higher type/token ratio and the C-test with texts of lower type/token ratio is compared by running a paired *t*-test. Table 2 displays the results of the descriptive statistics and 3 reports the results of paired *t*-test for lexical variation and C-test performance.

Table 2

Results of Descriptive Statistics for Lexical Variation and C-test Performance

	Mean	N	Std. Deviation	Std. Error Mean
LTTR*	77.73	77	9.052	1.032
HTTR*	71.60	77	10.117	1.153

*LTTR= Lower Type/Token Ratio

*HTTR= Higher Type/Token Ratio

Table 3

Results of Paired t-test for Lexical Variation and C-test Performance

	Paired Differences					T	Df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
LTTR & HTTR	6.130	7.419	.845	4.446	7.814	7.251	76	.000

As can be observed in table 3, given a critical level for rejection at .05 the null

hypothesis that there is no difference between the mean scores of the participants on

the C-test with text of higher TTR and the C-test with texts of lower TTR is not accepted (p=.000 < .05).

The results of the descriptive statistics displayed in table 2 indicates that the mean score of the C-test with texts of higher TTR is lower than the mean score on the other C- test i.e. C-test with lower TTR texts (77.73>71.60). Consequently, the performance of the participants on the C-test with texts of lower lexical variation is better than their performance on the parallel C-test i.e. C-test with texts of higher lexical variation.

Therefore, the results suggest a significant difference between the performance of test takers on C-tests with different degrees of lexical variation.

Results of Lexical Sophistication and C-test Performance

This time the effect of lexical sophistication on C-test performance is being tested and once more in order to determine whether there is a statistically significant difference between the performance of the participants on the C-tests with texts of higher lexical sophistication and their performance on the parallel C-test, a paired *t*-test is carried out. Table 5, below displays the results from the descriptive statistics and table 6 shows the results of the paired *t*-test for lexical sophistication and C-test performance.

Table 5

Results of Descriptive Statistics for Lexical Sophistication and C-test Performance

	Mean	N	Std. Deviation	Std. Error Mean
HLS*	53.1622	74	8.79606	1.02252

LLS*	57.0000	74	7.01075	81498.

Table 6

Results of paired t-test for lexical sophistication and C-test performance

	Paired Differences					T	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
HLS* & LLS*	-3.83784	7.97088	.92660	-5.68454	-1.99113	-4.142	73	000.

*HLS= Higher Lexical Sophistication

*LLS= Lower Lexical Sophistication

Looking at the results of table 6, the p value (p=0.000) is smaller than the alpha value (α= 0.05). Therefore, the mean scores of the C-test with texts of higher lexical sophistication and the parallel C-test, i.e. the C-test with lower lexical sophistication are significantly different. This conclusion is confirmed again once we look at the descriptive statistics of participants' scores shown in table 5.

The mean of participants' scores on the C-test with texts of lower lexical sophistication is significantly higher than the mean of participants' scores on the C-test with texts of higher lexical sophistication i.e. the texts of lexically more sophisticated words. As a result, lexical sophistication of the C-test texts significantly influences C-test performance.

Results of Lexical Variation and Lexical Sophistication

In addition, supplementary analyses were performed to examine first, the relationship between the performance of the participants on the C-test with texts of higher lexical variation and their performance on the C-test with texts of higher lexical sophistication, and second, the relationship between the participants performance on the C-test with texts of lower lexical variation and their performance on the C-test with texts of lower lexical sophistication.

To examine the relationship between the performance of the participants on the C-test with texts of higher TTR and their performance on the C-test with texts of higher lexical sophistication, a paired sample *t*-test is run to study the participants' mean score difference on the mentioned C-tests. However, to be able to run the *t*-test the two variables must be of the same scale. Consequently, the scores on the C-test with texts of higher lexical sophistication need to be rescaled. These scores were from a total score of 80 and to rescale the variable, they are multiplied by 100/80.

Table 7 shows the results of the descriptive statistics and table 8 illustrates the results of the paired sample *t*-test.

Table 7

Results of Descriptive Statistics for Higher Lexical Variation Performance and Higher Lexical Sophistication Performance

	Mean	N	Std. Deviation	Std. Error Mean
HTTR	71.80	66	9.911	1.220
HLS	66.7614	66	11.31493	1.39277

Table 8

Results of Paired t-test for Higher Lexical Variation Performance and Higher Lexical

Sophistication Performance

	Paired Differences							
			Std. Error	95% Confidence Interval of the Difference				Sig. (2-tailed)
	Mean	Std. Deviation	Mean	Lower	Upper	T	Df	
HTTR HLS	5.04167	7.27794	.89585	3.25253	6.83081	5.628	65	.000

As table 8 suggests there is a significant difference in the performance of the participants on the C-test with texts of higher TTR and their performance on the C-test with texts of higher lexical sophistication (p=0.000 < 0.05). A look at the results of the descriptive statistics of the two variables in table 7 reveals that the performance of the participants was better on the C-test with texts of higher lexical variation.

This time to study the relationship between the performance of the participants on the C-test with texts of lower TTR and their performance on the C-test with texts of lower lexical sophistication, a paired *t*-test will be run to study the participants' mean score difference on the mentioned tests. Again to rescale the scores on the C-test with texts of lower TTR the scores of this group are being multiplied by 100/80.

Table 9 reveals the results of descriptive statistics and table 10 shows the results of the paired *t*-test.

Table 9

Results of Descriptive Statistics for Lower Lexical Variation Performance and Lower Lexical Sophistication Performance

	Mean	N	Std. Deviation	Std. Error Mean
LTTR	78.11	66	8.608	1.060
LLS	71.5341	66	9.05590	1.11470

Table 10

Results of Paired t-test for Lower Lexical Variation Performance and Lower Lexical Sophistication Performance

	Paired Differences							Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		T	Df	
				Lower	Upper			
LTTR & LLS	6.57197	8.76913	1.07941	4.41625	8.72769	6.089	65	.000

Table 10 reveals the results of paired sample *t*-test. As the results show the performance of the participants on the mentioned C-tests are significantly different (p=.000 < .05). Moreover, the results of the descriptive statistics of the C-tests illustrated in table 9 demonstrates that the participants performed better on the C-test with texts of lower TTR than the C-test with texts of lower lexical sophistication.

Discussion and Conclusion

This study examined the performance of test-takers on C-tests with different levels of lexical richness and the findings suggest a significant difference in their performance. In fact, the lower the lexical richness of the texts, the better the test takers perform.

Getting back to the questions raised in the beginning of the paper, since the performance of the subjects was different on two C-tests of the same topic, length, and difficulty level but different lexical characteristics, the findings suggest that the C-test has the potential to assess vocabulary knowledge. In fact, we can conclude that with the introduction of lexical characteristic measures in the construction of C-test, C-test can be used in vocabulary assessment. The results of cross analyses can provide more information on which lexical measure is more suitable to be applied in the development of C-test.

The cross analyses of the test takers' performance on lower TTR and lower lexical sophistication tests, and higher TTR and higher lexical sophistication tests shows that the scores were significantly different. In fact, test takers performed significantly better on lower TTR and higher TTR C-tests compared with lower lexical sophistication and higher lexical sophistication C-tests respectively. This suggests that lexical sophistication measure (P_Lex) seems to tap test takers' vocabulary knowledge more successfully, an argument also suggested by Meara and Bell (2001). In fact, Meara and Bell (2001) argues that there is no clear agreement about exactly what features of texts type/token ratio measure describes.

On the other hand, P_Lex is a measure that works well with short texts and is based on Nation's (1984) word lists. Therefore, by considering it as a criterion in the selection of C-test texts we would be able to contextualize the standard context-independent traditional vocabulary tests. The target instrument is therefore not restricted to the measurement involving knowledge of individual content words. This is in line with what Chapelle (1994) calls systematic item selection when he suggests that interpretation of the C-test results in terms of test-takers' vocabulary size can be done only if items are selected more systematically.

Limitations and Directions for Future Research

Future research into the employment of such tests in vocabulary assessment would be needed. In particular, a follow-up on the comparison of a context-independent vocabulary test and its contextualized C-test alternative seems necessary. However, one problem with these objective methodologies is that they observe only the end product whereas to understand how C-test can be used in vocabulary assessment there is a need for scrutinizing the processes involved in C-test taking. One possible approach would be to use think aloud protocols with participants taking a C-test developed based on the study's suggestions. This will enable us to observe the processes involved in taking a C-test especially designed for vocabulary assessment which makes more significant contribution to our understanding about the applicability and use of C-test in vocabulary knowledge assessment.

References

Arnaud, P. J. L. (1984). The lexical richness of L2 written productions and the validity of vocabulary tests. In Culhane, T., Klein Bradley, C., & Stevenson, D. K. (Eds.), *Practice and problems in language testing: Papers from the International Symposium on Language Testing* (pp. 14-28). Colchester: University of Essex.

Babaii, E., & Ansary, H. (2001). The C-test a valid operationalization of reduced redundancy principle? *System, 29*, 209-219.

Broeder, P., Extra G., Van Hout, R., Strömqvist, S., & Voionmaa, K. (1988). *Processes in the developing lexicon.* Strasbourg: ESF (Final report III of the European Science Foundation Project, "Second language acquisition by adult immigrants").

Chapelle, C. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research, 10(2),* 157-187.

Chapelle, C. A., & Abraham, R. A. (1990). Cloze method: What difference does it make? *Language Testing, 7,* 121-146.

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing, 23,* 290-325.

Feldmann, U., & Stemmer, B. (1987). Thin___ aloud a___ retrospective da___ in C-te___ taking: diffe___ languages – diff___ learners – sa___ approaches? In Claus Færch, & Gabriele Kasper (Eds.), *Introspection in second language research* (pp. 251-267). Clevedon: Multilingual Matters.

Herdan, G. (1960). *Type token mathematics*. The Hague: Mouton.

Jafarpur, A. (1995). Is C-testing superior to cloze? *Language Testing, 12,* 194-216.

Jafarpur, A. (2002). A comparative study of a C-Test and a cloze test. In Grotjahn, R.

(Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen [The C-

Test: theoretical foundations and practical applications]* (pp. 31-51). Bochum: AKS-

Verlag.

Kamimoto, T. (1992). An inquiry into what a C-test measures. *Fukuoka Women's

Junior College Studies, 44*, 67-79.

Katona, L., & Dornyei, Z. (1992). Validation of the C-test amongst Hungarian EFL

learners. *Language Testing, 9*, 187-206.

Laufer B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2

written production. *Applied Linguistics, 16*(3), 307-322.

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A

Ryan and A Wray (Eds.), *Evolving models of language*. Clevedon: Multilingual

Matters.

Meara, P., & Bell, H. (2001). P_Lex A simple and effective way of describing the

lexical characteristics of short L2 texts. *Prospect, 16*(1), 5–24.

Meara, P. (2001). *PLEX v1.1.* University of Wales Swansea. Retrieved from

http://www.swan.ac.uk/cals/calsres

Miller, J., & Klee, T. (1995). Quantifying language disorders in children. In P.

Fletcher, & B. MacWhinney (Eds.), <u>Handbook of child language (pp. 545-572)</u>.

Basal Blackwell.

Nation, I. S. P. (1984). V*ocabulary lists: Words, affixes and stems.* Victoria University of

Wellington: English Language Institute. Occasional Publication No 12.

Raatz, U., & Klein-Braley, C. (1981). The C-test—a modification of the cloze

procedure. In Culhane, T., Klein-Braley, C., & Stevenson, D. K. (Eds.). *Practice*

and problems in language testing, University of Essex Occasional Papers (pp. 113-

148). Colchester: University of Essex.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge:

Cambridge University Press.

Singleton, D., & Little, D. (1991). The second language lexicon: some evidence from

university-level learners of French and German. *Second Language Research, 7*,

62-81.

Singleton, D., & Singleton, E. (1998). The C-test and L2 acquisition/processing

research. In J. A. Coleman (Ed.), *University language testing and the C-test* (pp.

150-178). Portsmouth: University of Portsmouth.

Stemmer, B. (1991). *What's on a C-test taker's mind: Mental processes in C-test taking*.

Bochum: Brockmeyer.

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical

competence and language proficiency. *Variable sensitivity, 27*(4), 567-595.

APPENDICES

Appendix 1

Dörnyei and Katona's (1992) C-test

Extract 1

One cool autumn evening, Bob L., a young professional, returned home from a trip to the supermarket to find his computer gone. Gone! All so--- of cr--- thoughts ra--- through h-- mind: H-- it be—stolen? H-- it be-- kidnapped? H- searched h-- house f-- a cl-- until h- noticed a sm--- piece o- printout pa--- stuck un--- a mag--- on h-- refrigerator do--. His he--- sank a- he re-- this sim--- message: CAN'T CONTINUE, FILE CLOSED, BYE.

Extract 2

There is a third factor besides farming and herding in the spread of man-made deserts: deforestation. The progr------ destruction o- the Th--- World's st--- of tr--- is dama---- not on-- in d-- regions: every----- it occ--- it c-- accelerate t-- decay o- the so-- and red--- its capa---- to fe-- people. It can reduce rainfall and lead to drought.

Extract 3

There are certain things which no student can do without and others which may not be as necessary as you thought. It m-- be worth considering so-- small hi---. You m-- find your---- in ne-- of elect----- appliances su-- as li--- bulbs, adap---- or pl---. These c-- be obta---- from ma-- places. GILL i- a go-- hardware sh-- and try--- to fi-- it i- a chal-----. It is hidden in a little alley leading off High Street called Wheatsheaf Yard.

Extract 4

The private conscience of the leader, rather than his public responsibilities, becomes the focal point of politics. Internal crit---- (possession o-, devotion t-, and stan---- up f-- private princ-----) become t-- standards o- political judg-----. Constituents disa-----, and w- are le-- with a poli----- leader deter------ policy o- the ba--- of compat------ with h-- private princ-----. From this perspective we can better understand why Goldwater voted against the Civil Rights Act of 1964.

Appendix 2

Lower lexical variation C-test

Directions: In this test parts of some of the words have been damaged. Please replace the missing parts. Each '-' represents one letter.

Example:

Millions of children in the United States go to summer camp. Some g- to pl-- outdoors a- traditional ca--- in t-- woods, i- the moun----- or o- a la--. But fami---- now ha-- many cho---- of spec----- camps. Th--- can b- in t-- middle o- nature o- a b-- city. Spec----- camps of--- young peo--- the cha--- to le--- about diff----- subjects. Anything from space exploration to business to medicine.

Your job is to complete the text as:

Millions of children in the United States go to summer camp. Some go to play outdoors at traditional camps in the woods, in the mountains or on a lake. But families now have many choices of specialty camps. These can be in the middle of nature or a big city. Specialty camps offer young people the chance to learn about different subjects. Anything from space exploration to business to medicine.

Text 1

Slave traders are trafficking boys ranging from the age of 12 to 16 from their home countries and are selling them to cocoa farmers in Cote d'Ivoire. They wo-- on sm--- farms acr--- the cou----, harvesting t-- cocoa be--- day a-- night, un--- inhumane condi-----. Most o- the bo-- come fr-- neighboring Mali, wh--- agents ha-- around b-- stations loo---- for chil---- that a-- alone o- are beg---- for food. They lure the kids to travel to Cote d'Ivoire with them, and then the traffickers sell the children to farmers in need of cheap labor.

Text 2

Seasonal affective disorder (SAD) is a type of depression that is triggered by the seasons. The mo-- common ty-- of S-- is cal--- winter on--- depression. Symp---- usually be--- in la-- fall o- early win--- and g- away b- summer. A mu- - less com--- type o- SAD, kn--- as sum--- onset depre-----, usually beg--- in t-- late spring or early summer and goes away by winter. Seasonal affective disorder may be related to changes in the amount of daylight during different times of the year.

Text 3

Every day, the average person inhales about 20,000 liters of air. Every ti-- we bre----, we ri-- inhaling dang----- chemicals th-- have fo--- their w-- into t-- air. A-- pollution incl---- all contam------ found i- the atmos-----. These dang----- substances c-- be eit--- in t-- form o- gases o- particles. A-- pollution can be found both outdoors and

indoors. Pollutants can be trapped inside buildings, causing indoor pollution that lasts for a long time.

Text 4

A child is secure if the parents are mature, loving and considerate toward each other. The ch--- is inse---- if t-- parents a-- immature, spit----, selfish a-- argumentative tow--- each ot---. The ch--- is suff----- traumatic str--- if t-- parents a-- violent o- threatening tow--- each ot---. Even a- that po---, if t-- parents re--- resolution and grow toward mature, loving behavior, the child will overcome their insecurity. If the parents divorce, it teaches the child that there is no stable, secure, loving place.

Text 5

Acid rain is produced when chemical pollutants interact with sunlight to produce sulfuric and nitric acids. These ac--- combine wi-- moisture i- the a-- (groundwater sou---- that ha-- evaporated a-- condensed) t- produce ac-- rain, sn--, etc. Si--- there i- usually mo-- sunlight i- the sum--- months, ac-- rain te--- to b- more aci--- in t-- summer. Acid rain is the result of industrial pollution, which causes rainwater to carry small quantities of acidic compounds such as sulphuric and nitric acid.

Appendix 3

Higher lexical variation C-test

Directions: In this test parts of some of the words have been damaged. Please replace the missing parts. Each '-' represents one letter.

Example:

Millions of children in the United States go to summer camp. Some g- to pl-- outdoors a- traditional ca--- in t-- woods, i- the moun----- or o- a la--. But fami---- now ha-- many cho----

of spec----- camps. Th--- can b- in t-- middle o- nature o- a b-- city. Spec----- camps of--- young peo--- the cha--- to le--- about diff----- subjects. Anything from space exploration to business to medicine.

Your job is to complete the text as:

Millions of children in the United States go to summer camp. Some go to play outdoors at traditional camps in the woods, in the mountains or on a lake. But families now have many choices of specialty camps. These can be in the middle of nature or a big city. Specialty camps offer young people the chance to learn about different subjects. Anything from space exploration to business to medicine.

Text 1

Every time you pick up a chocolate bar, you're telling that chocolate company that there's nothing wrong with the way they do business. In Afr---, farmers oper----- cocoa fa--- are pa-- so lit--- that th-- turn t- slave la--- , usually th-- of chil----. This i- not ch--- labor – th-- are n-- paid, th-- are for--- to st--, locked u- at ni---, beaten a-- killed for doing an unsatisfactory job or trying to escape. This is modern day slavery in no uncertain terms, and though it is well known, nothing is being done.

Text 2

Hormone levels and hence our moods may be affected by the weather. Gloomy wea---- can ca--- depression, b-- sunshine app---- to ra--- the spi----. In Britain, f-- example, t-- dull wea---- of win--- drastically cu-- down t-- amount o- sunlight th-- is exper------ - which stro---- affects so-- people. Th-- become s- depressed a-- lacking in energy that their work and social life are affected. This condition has been given the name SAD (Seasonal Affective Disorder).

Text 3

Many people spend large portion of time indoors - as much as 80-90% of their lives. We wo--, study, e--, drink a-- sleep i- enclosed enviro------ where a-- circulation m-- be restr-----. For th--- reasons, so-- experts fe-- that mo-- people suf--- from t-- effects o-indoor a-- pollution th-- outdoor poll-----. There a-- many sou---- of indoor air pollution. Tobacco smoke, cooking and heating appliances, and vapors from building materials, paints, furniture, etc. cause pollution inside buildings.

Text 4

Acid rain describes any form of precipitation with high levels of nitric and sulfuric acids. It c-- also oc--- in t-- form o- snow, f--, and ti-- bits o- dry mate---- that set--- to Ea---. Rotting veget----- and erup---- volcanoes rel---- some chem----- that c-- cause ac-- rain, b-- most ac-- rain fa--- because o- human activities. The biggest culprit is the burning of fossil fuels by coal-burning power plants, factories, and automobiles.

Text 5

Researchers have filled volumes studying and analyzing the effect of divorce on children. Studies exam----- the chil---- of div---- have fo--- that mo-- suffer a se--- of lo-- that c-- manifest i- many diff----- ways, depe----- on t-- children's ag-- and uni--- personalities, a- well a- on h-- parents han--- the div---- themselves. You---- children may regress in areas such as sleeping and toilet training, or throw more tantrums. School-age and teenage children may show symptoms of depression, rebel against discipline, or change their eating and sleeping habits.

Appendix 4

Lower lexical sophistication C-test

Directions: In this test parts of some of the words have been damaged. Please replace the missing parts. Each '-' represents one letter.

Example:

Millions of children in the United States go to summer camp. Some g- to pl-- outdoors a- traditional ca--- in t-- woods, i- the moun----- or o- a la--. But fami---- now ha-- many cho---- of spec----- camps. Th--- can b- in t-- middle o- nature o- a b-- city. Spec----- camps of--- young peo--- the cha--- to le--- about diff----- subjects. Anything from space exploration to business to medicine.

Your job is to complete the text as:

Millions of children in the United States go to summer camp. Some go to play outdoors at traditional camps in the woods, in the mountains or on a lake. But families now have many choices of specialty camps. These can be in the middle of nature or a big city. Specialty camps offer young people the chance to learn about different subjects. Anything from space exploration to business to medicine.

Text 1

When parents decide to divorce or separate, their child or children are faced with multiple stressors. Just kno---- that thi--- are go--- to b- different af--- a div----, but n-- knowing exa---- how c-- be fright------ for mo-- children. Chil---- from divo---- families ha-- to le--- to co-- with t-- many cha---- in th--- family. T-- amount of contact with one parent, often their father, will be reduced. Children may have to move from their family home or change schools.

Text 2

Studies have shown that the air inside buildings, including our homes, can often be many more times polluted than the air outside. This is--- surprising wh-- you cons---- that t-- air ins--- is a combi------ of t-- air out---- with a-- the poll----- generated ind----. Indoor a-- pollutants oc--- within buil----- or ot--- enclosed spa--- such a- cars. Pollu----- occur bec---- of t-- things we do inside like cleaning and pest control as well as what personal care products we might use. These chemical-containing products release their ingredients into the air and create indoor pollution.

Text 3

Anger is a completely natural and instinctive emotion. In fa--, it c-- be a ca--- for con---- if a per--- does n-- experience a-- tinge o- anger, espec----- when th-- person i- subject t- great dur--- or provo------. Angry feel---- may b- triggered b- a per---, object, inci----, event or even memory. Reactions may vary from mild annoyance to aggression.

Text 4

Probably the biggest cause of obesity is lack of or very little participation in physical activity. The n-- cyber a-- has defin----- added t- the incre----- problem o- obesity. Ev--- the num--- of ch--- obesity ca--- has ri--- to alar---- levels. Ki-- today a-- spending mo-- time i- front o- the T-, the comp---- or pla---- video games, and less time playing outside. In fact child obesity has reached epidemic proportions in the US today.

Appendix 5

Higher lexical sophistication C-test

Directions: In this test parts of some of the words have been damaged. Please replace the missing parts. Each '-' represents one letter.

Example:

Millions of children in the United States go to summer camp. Some g- to pl-- outdoors a- traditional ca--- in t-- woods, i- the moun----- or o- a la--. But fami---- now ha-- many cho---- of spec----- camps. Th--- can b- in t-- middle o- nature o- a b-- city. Spec----- camps of--- young peo--- the cha--- to le--- about diff----- subjects. Anything from space exploration to business to medicine.

Your job is to complete the text as:

Millions of children in the United States go to summer camp. Some go to play outdoors at traditional camps in the woods, in the mountains or on a lake. But families now have many choices of specialty camps. These can be in the middle of nature or a big city. Specialty camps offer young people the chance to learn about different subjects. Anything from space exploration to business to medicine.

Text 1

Children sometimes get the brunt of the whole divorce, having feelings of sadness, guilt and anger. They fe-- helpless th-- they can--- solve th--- parents' prob---- or ev-- may fe-- that h- or s-- is t-- source o- the prob----. It i- a fa-- that i- a chi---- feelings a-- ignored a-- not de--- with prop----, the child will grow up having a hard time developing intimate relationships and will have severe self-image problems. Children of divorced parents are more likely to go through a divorce themselves later in life.

Text 2

Most people spend more than half of their lives indoors. The signif------ of ind---- air-quality h-- become mo-- important i- recent ye--- as a res--- of eff---- to ma-- our ho--- more ene--- efficient. A- we tig---- up o-- homes t- prevent he-- exchange, w- also pre---- air exch---- and pollu----- released into the home environment are trapped for long periods of time. Second-hand cigarette smoke, gas stoves, and wood-burning stoves and fireplaces are major sources of indoor air pollution.

Text 3

Overweight and obese conditions can be prevented or treated with exercise along with a healthy diet. Activity he--- to red--- body f-- and incr---- muscle ma--, thus impr----- your bod--- ability t- burn calo----. The combi------ of red---- calories a-- daily exer---- is t-- ticket t- weight lo--. And contr------ obesity i- critical, a- it i- a major risk factor for many diseases. Lowering your body mass index (BMI) is a sure way to reduce your risk of dying early and to live a healthier life.

Text 4

There are health benefits in acquiring constructive ways to manage anger. As an--- causes t-- heart be-- to ra-- and bl--- pressure t- rise, a prol----- angry st--- has undes------ consequences. A pers---- character c-- be deve----- by posse----- effective an--- management ski---. He o- she te--- to b- less emot----- during dur--- and are more focused in problem-solving. Consequently, the person enjoys better quality of social relationships.