
*Received: September 1, 2011**Accepted: October 13, 2011*

Validity and C-Tests: The Role of Text Authenticity

Ebrahim Khodadady¹, Mahmood Hashemi²

Abstract

This study explored the relationship between text authenticity and test takers' performance on C-Tests. To fulfill the objective a single C-test was developed on an authentic text (AC-test) and along with standard C-Tests (Klein-Braley 1997) and a disclosed version of the Test of English as a Foreign Language (TOEFL) was administered to one hundred and thirty five junior and senior undergraduate university students majoring in English language and literature, English translation and teaching English as a foreign language in two universities in Iran. Similar to their standard counterparts, the AC-test correlated significantly not only with the TOEFL and its structure, written expression, vocabulary and reading comprehension subtests and thus established itself as an externally valid measure of English language proficiency. Although high correlation coefficients obtained between the authentic and standard C-Tests ($r = 0.87$, $p < .01$) showed that they could be used interchangeably, the AC-Test enjoyed slightly higher internal validity and noticeably higher reliability. The implications of the findings are discussed within a foreign language testing context.

Keywords: *C-Test, cloze test, Authenticity, reduced redundancy, schema theory*

1. Introduction

C-Tests were developed by Klein-Braley and Raatz (1981) as a justified replacement for cloze test from both theoretical and psychometric perspectives (Babaii & Moghaddam, 2006). To remedy cloze test shortcomings, they modified it and produced a new deletion procedure called the C-Principle or the Rule of Two. C-Tests are developed on the basis of Reduced Redundancy Principle (RRP), postulating that native speakers are able to restore missing or distorted texts by resorting to various textual information and making use of natural redundancy in texts. The morph in the sentence, *Ali reads his Persian book every day*, for example, is redundant in that its existence is not necessary due to the presence of the name Ali and the adverb of time every day.

¹ Department of English Language and Literature, Ferdowsi University of Mashhad, Iran. E-mail: ekhodadady@ferdowsi.um.ac.ir

² Department of English Language and Literature, Ferdowsi University of Mashhad, Iran.

2. Review of Literature

Cloze tests follow RRP in that success on these measures of language achievement and proficiency depends on tester takers' ability to restore deleted parts of given texts. The deletion of the word reads in the sentence, Ali ... his Persian book every day, results in developing a typical cloze item. C-Tests follow the same principle with a significant difference. Instead of deleting the whole word, the second halves of words are deleted. That is, the C-Test item in the above example for the word 'reads' would be 're—'. Although C-Tests are a type of cloze tests, they are superior to them because they enjoy the following six advantages reproduced verbatim from Klein-Braley's (1997, p. 65) study:

1. Many more items are possible with much shorter texts. A classical cloze test using a 5th word deletion rate would have to be at least 500 words long to contain 100 items. A C-Test consisting of five texts with 20 half-deleted words would be only approximately half as long
2. C-Test scoring is exact and objective because there is almost always one possible solution. In a few cases more than one solution is possible, but almost always only two alternatives. When this happens all possible solutions are counted as correct.
3. The scoring of a C-Test is quick and easy for the native speaker or the teacher since it takes only slightly more time than is needed for simply reading the text. The original unmutated test very quickly becomes automated so that checking is unnecessary
4. C-Tests are very 'easy' for native speakers. On the other hand, someone who does not understand the language at all normally makes a score of zero or close to zero
5. Since every second word is 'damaged' the probability of obtaining a representative sample of all the word classes in the text is very much higher
6. Because the C-Test consists of a number of different texts the sampling of content classes is better. Examinees who happen to have special knowledge in certain areas no longer have substantial advantages over other examinees.

It is argued in this paper, however, the maintenance of the first half of the mutilated word solves the difficulty faced in answering cloze tests, but it renders the C-Tests themselves very easy. To solve the problem, Klein-Braley and Raatz (1985, 1990) suggested that instead of deleting certain words, every second word, starting from the second sentence be mutilated, hence the Rule of Two. In addition to applying this rule, Klein-Braley (1997) stated that between four to six short texts need to be carefully chosen and ordered intuitively in terms of their difficulty. The newly developed C-Test should be given to adult educated native speakers or English language teachers for trial. We expect native speakers to co correctly restore 90% of the mutilated words before standard C-Tests could be finalized for administration to the target population. As it can be noticed, these steps make the development of C-Tests very demanding if not too cumbersome, especially within a foreign language context where finding cooperative native speakers is too difficult, if not virtually impossible. This study is, therefore, developed to find out whether developing a C-Test on a single authentic text (henceforth AC-Test) will yield results similar to standard C-Tests.

3. Methodology

3.1 Participants

One hundred thirty five, 34 male (25.2%) and 101 female (74.8%), undergraduate students whose age ranged between 21 and 35 (mean = 22.50, SD = 1.71) participated voluntarily in the study. They were junior (n = 88, 65.2%) and senior (n = 47, 34.8%) undergraduate students majoring in English Language and Literature (n = 46, 34.1%), English Translation (n = 73, 54.1%), and Teaching English as a Foreign Language (n = 16, 11.9%) at Islamic Azad University of Quchan (n = 93, 68.9%) and Teacher Training University of Tehran (n = 42, 31.1%). They spoke Persian (n = 87, 64.4%), Turkish (n = 30, 22.2%) and other unspecified languages (n = 18, 13.3%) as their mother tongues. The scores obtained on the tests were reported to the participants because they intended to sit for the MA Entrance Examination part of which requires taking a language proficiency test such as the TOEFL.

3.2 Instruments

Three instruments were employed in the study, i.e., TOEFL, a standard C-Test and a C-Test developed on a single authentic and unmodified text.

3.2.1 Test of English as a Foreign Language

Following Khodadady (2007) and Khodadady and Herriman (2000), the structure, written expression, vocabulary and reading comprehension sections of some disclosed TOEFL tests (Educational Testing Service, 1991) were employed to explore the empirical validity of the AC-Test. The total TOEF comprised 115 items, including 30 structure items, 25 written expression items, 30 reading items and 30 vocabulary items.

3.2.2. Standard C-Test

Four C-Tests developed by Klein-Braley (1997, pp. 79-80) were used in this study. They consist of 99 items developed on four short texts. With the exception of C-Test 2, which had 24 items, the other three C-Tests had 25 items each. The validity and reliability of the C-Test battery was demonstrated by Klien-Braley (1997) and Khodadady (2004).

3.2.3 Authentic C-Test

Gholami (2006) used the *New Scientist* article “*why don't we just kiss and make up*” (Dugatkin, 2005) to explore the effect of semantic schema types, i.e., adjectives, adverbs, nouns and verbs, on EFL learners' performance on tests. The selection of this particular journal was based on Clapham's (1996) observation that its articles are “more academic than ... articles in quality newspapers” (p. 145) and they provide standard scientific texts for public readership. The readability Ease score of Flesch (57.4) indicated that the text used in developing the AC-Test of this study was fairly difficult for high school students at grades 10-12 (Flesch 1948, 1949).

Based on the microstructural approach of schema theory, Khodadady, Hosseini and Etminan (2012) treated the words comprising Dugatkin's (2005) article as schemata whose processing and understanding depends not only on the readers' background knowledge of what they stood for but also on their syntactic, semantic and discoursal relationships they hold with

each other. Following Khodadady and Elahi (2012), Khodadady, Hosseini and Etminan divided the schemata of the text into three main domains, i.e., syntactic, semantic and parasyntactic. Syntactic schemata are few in type but many in frequency whereas semantic schemata are many in type but few in frequency. The parasyntactic domain includes the schemata which might be either few in types as syntactic schemata or many as semantic schemata are but always behave like syntactic schemata as shown in Table 1.

Table 1. Schema domains, genera, tokens and types comprising the text of AC-Test

Schema Domain	Schema Genus	Tokens	Percent	Types	Percent
Semantic	Adjectives	141	8.2	97	13.8
	Adverbs	43	2.5	34	4.8
	Nouns	417	24.3	209	29.8
	Verbs	265	15.5	158	22.5
	Total	866	50.5	498	70.9
Syntactic	Conjunctions	116	6.8	13	1.9
	Determiners	153	8.9	24	3.4
	Prepositions	203	11.8	31	4.4
	Pronouns	113	6.6	32	4.6
	Syntactic verbs	53	3.1	21	3.0
	Total	638	37.2	121	17.2
Parasyntactic	Abbreviations	8	.5	5	0.7
	Names	41	2.4	28	4.0
	Numerals	4	.2	4	0.6
	Para-adverbs	117	6.8	45	6.4
	Particles	41	2.4	1	0.1
	Total	211	12.3	83	11.8
Total Schemata		1715	100.0	702	100.0

The article “*why don’t we just kiss and make up*” consists of three sections, i.e., introduction, stress-busting, and feuding families. As can be seen in Table 1, 1715 schemata constituted the entire text and since all of them could not be included in the AC-Test, the introduction section consisting of seven paragraphs and 615 schemata were arbitrarily chosen and the other two sections were deleted. Following Klein-Braley (1997) the first and last sentences of the seven paragraphs were kept intact and every second schema of the second sentence was mutilated. This procedure resulted in mutilating 180 schemata as the final items of AC-Test. (The AC-Test is given in the Appendix.)

3.3 Procedure

After talking to participants in person and ensuring them of the reliability and validity of tests employed in the study, they agreed to take the three tests, i.e., the TOEFL, the standard C-Tests and the AC-Test in two sessions, on the condition that they would be provided with their scores on the tests upon being corrected. After making the most suitable arrangement in terms of their availability, the TOEFL was administered in the first session and after two weeks they took the standard C-Tests and AC-Test in a counterbalanced manner. They were divided into two groups and while the first group did the standard C-Tests first, the second group took the AC-Test. Then they took the standard C-Tests and the AC-Test in the same session in reversed order. All the tests were administered under standard conditions.

3.4 Data Analysis

For estimating the internal validity of the tests, Baker (1989) was followed and p -values falling within the range of 0.25 to 0.75 were considered acceptable. Similar to Khodadady (2004, 2007), each gap was considered as an item and Cronbach Alpha was used as an index of reliability not only for the standard C-Tests and AC-Test but also for the TOEFL. For determining the discrimination power of items point biserial correlations (r_{pbi}) between the total test score and individual items were calculated and coefficient of 0.25 and higher were used along with acceptable p -values as indices of well functioning items. For establishing the external validity of the AC-Test, the test takers' scores on this test were correlated with the TOEFL and standard C-Tests. All statistical analyses were conducted via SPSS version 19.0 to test the following three hypotheses:

H1. The AC-Test will be as reliable as the standard C-Test.

H2. The AC-Test will be internally as valid as the standard C-Test.

H3. The AC-Test will correlate significantly with the standard C-Test and TOEFL.

4. Results and Discussion

Table 2 presents the descriptive statistics of the tests administered in the study. As can be seen, among the three tests of language proficiency, AC-Test has the highest standard deviation, i.e., 21.6, followed by the TOEFL, i.e., 12.5, indicating that the AC-Test distinguishes among the test takers better than both the TOEFL and the standard C-Tests. Similarly, the reliability coefficient of the AC-Test, i.e., .92, is noticeably higher than the standard C-Tests, i.e., .82. These results not only *confirm* the first hypothesis that *the AC-Test will be as reliable as the standard C-Test* but also indicate that the AC-Tests are superior to their standard counterparts in terms of their reliability.

Table 2. Descriptive Statistics of the standard C-Tests, AC-Test and TOEFL and its subtests

Tests	No of items	Minimum	Maximum	Mean	Std. Deviation	Alpha
C-Tests	99	29	78	53.44	11.358	.821

C-Test1	25	8	20	14.59	2.800	.252
C-Test2	24	3	19	12.49	3.204	.443
C-Test3	25	7	21	13.18	3.190	.415
C-Test4	25	4	21	13.19	3.497	.519
AC-Test	180	40	137	96.43	21.589	.921
TOEFL	115	26	85	61.92	12.515	.833
Structure	30	6	30	17.16	4.019	.597
Written expressions	25	4	19	13.16	3.276	.463
Reading	30	4	30	14.27	3.490	.411
Vocabulary	30	6	24	17.33	4.013	.573

While it is argued that higher standard deviation (SD) and reliability coefficient of the AC-Test is due to its length, it does not necessitate standardizing the SDs as suggested to compare them with each other for two reasons. First, SDs are standardized by their very nature and secondly there is no theoretically sound basis to establish a cut off number for the items comprising the C-Tests, i.e., 100, as Klein-Braley (1997) did. It is, in fact, argued in this paper that the inclusion of more items in the AC-Tests provides a more reliable and valid measure of test takers' ability as the results presented in Table 2 show.

Table 3 presents the ordered psychometrics of standard C-Test items. As can be seen, out of 99 items, 46 have functioned well because their p-values fall between .25 and .75 and their r_{pbi} is .25 or higher, indicating that the standard C-Tests enjoy acceptable internal validity because 46 percent of their items discriminate well among high ability and low ability test takers. They do not, however, support Jafarpur's (1999) observation that "a sizeable number of them are either extremely easy or extremely difficult to fill in" (p. 79) because p-values of the standard C-Test range between .37 and .73 (Mean = .54, SD = .08).

Table 3. P-value (PV) and r_{pbi} of 99 items comprising the standard C-Test

Item	PV	r_{pbi}	Item	PV	r_{pbi}	Item	PV	r_{pbi}	Item	PV	r_{pbi}	Item	PV	r_{pbi}
I2	.67	.39	I18	.68	.30	I65	.44	.27	I85	.53	.22	I90	.37	.16
I71	.57	.39	I23	.56	.30	I36	.59	.26	I50	.53	.21	I1	.73	.15
I93	.47	.39	I91	.52	.30	I57	.55	.26	I53	.59	.21	I39	.66	.15
I44	.59	.38	I15	.45	.29	I24	.59	.25	I58	.60	.21	I92	.42	.15
I95	.50	.38	I25	.48	.29	I35	.39	.25	I61	.39	.21	I20	.59	.14
I70	.47	.37	I27	.43	.29	I96	.56	.25	I62	.46	.21	I21	.56	.14
I29	.61	.36	I43	.48	.29	I9	.48	.24	I6	.58	.20	I4	.55	.13
I84	.45	.36	I46	.41	.29	I16	.64	.24	I14	.56	.20	I22	.46	.13
I88	.47	.36	I63	.50	.29	I38	.54	.24	I72	.66	.20	I77	.58	.13

I33	.56	.33	I83	.65	.29	I47	.47	.24	I7	.73	.19	I12	.56	.11
I48	.57	.33	I99	.57	.29	I49	.58	.24	I82	.52	.19	I42	.61	.11
I59	.42	.33	I8	.66	.28	I68	.41	.24	I40	.45	.18	I73	.45	.11
I66	.46	.33	I34	.48	.28	I69	.59	.24	I51	.49	.18	I94	.62	.11
I76	.62	.33	I45	.57	.28	I75	.47	.24	I5	.55	.17	I41	.50	.09
I97	.63	.33	I54	.56	.28	I98	.58	.24	I17	.50	.17	I64	.67	.05
I28	.53	.32	I81	.53	.28	I79	.42	.23	I32	.50	.17	I10	.67	.01
I37	.43	.32	I87	.56	.28	I11	.71	.22	I52	.61	.17	I86	.65	.01
I89	.50	.32	I26	.56	.27	I19	.47	.22	I74	.61	.17	I13	.59	-0
I67	.50	.31	I30	.45	.27	I60	.59	.22	I78	.51	.17	I31	.52	-0
I3	.56	.30	I55	.47	.27	I80	.44	.22	I56	.60	.16			

Table 4 presents the ordered psychometrics of the items comprising the AC-Test. As can be seen, out of 180 items, 97 have functioned well in terms of item their difficulty level and discrimination power, i.e., 54 percent. Comparing the percentage of well functioning items on the standard C-Test, i.e., 46%, with that of the AC-Test, i.e., 54%, not only confirms the second hypothesis that the AC-Test will be internally as valid as the standard C-Test but also indicate that the AC-Tests are superior to their standard counterparts in terms of their internal validity. This supervisory, however, comes at a price because the item psychometrics support Jafarpur's (1999) observation regarding five (3%) and 17 (9%) items being extremely difficult and easy, respectively. The p-values of the AC-Test range between .16 and .91 (Mean = .54, SD = .17).

Table 4. P-value (PV) and r_{pbi} of 180 items comprising the AC-Test

Item	PV	r_{pbi}	Item	PV	r_{pbi}	Item	PV	r_{pbi}	Item	PV	r_{pbi}	Item	PV	r_{pbi}
I151	.67	.58	I153	.47	.35	I21	.64	.28	I172	.44	.23	I176	.84	.17
I51	.49	.52	I180	.16	.35	I57	.49	.28	I23	.51	.22	I1	.46	.16
I89	.47	.48	I24	.40	.34	I68	.32	.28	I27	.3	.22	I2	.72	.16
I160	.65	.46	I31	.62	.34	I86	.43	.28	I32	.39	.22	I4	.65	.16
I100	.53	.45	I35	.50	.34	I142	.70	.28	I96	.27	.22	I12	.28	.16
I137	.39	.45	I45	.66	.34	I161	.30	.28	I144	.51	.22	I74	.69	.16
I141	.38	.44	I60	.70	.34	I17	.42	.27	I162	.64	.22	I109	.30	.16
I40	.57	.42	I73	.26	.34	I108	.70	.27	I67	.67	.21	I6	.86	.15
I138	.39	.42	I152	.27	.34	I124	.63	.27	I78	.29	.21	I38	.24	.15
I59	.46	.41	I29	.62	.33	I130	.42	.27	I116	.38	.21	I80	.78	.15
I113	.44	.41	I46	.45	.33	I170	.36	.27	I128	.59	.21	I83	.72	.15

I56	.76	.40	I104	.76	.33	I173	.46	.27	I131	.50	.21	I16	.76	.14
I105	.56	.40	I106	.50	.33	I3	.39	.26	I8	.61	.2	I39	.86	.14
I26	.44	.39	I115	.47	.33	I62	.48	.26	I15	.39	.2	I171	.54	.14
I146	.26	.39	I139	.74	.33	I125	.61	.26	I43	.67	.2	I178	.76	.14
I166	.49	.39	I157	.57	.33	I5	.57	.25	I94	.44	.2	I20	.40	.13
I66	.50	.38	I81	.56	.32	I11	.84	.25	I99	.75	.2	I64	.74	.13
I90	.25	.38	I126	.43	.32	I18	.71	.25	I102	.48	.2	I88	.73	.12
I95	.44	.38	I143	.77	.32	I69	.51	.25	I127	.39	.2	I114	.79	.12
I101	.33	.38	I19	.78	.31	I77	.51	.25	I167	.50	.2	I168	.76	.12
I133	.36	.38	I42	.19	.31	I82	.34	.25	I33	.53	.19	I175	.53	.12
I140	.82	.38	I92	.39	.31	I132	.41	.25	I72	.56	.19	I7	.91	.11
I155	.27	.38	I107	.45	.31	I145	.54	.25	I84	.67	.19	I58	.76	.11
I163	.38	.38	I120	.40	.31	I158	.53	.25	I22	.68	.18	I117	.27	.11
I14	.62	.37	I47	.35	.30	I165	.52	.25	I25	.56	.18	I156	.29	.11
I65	.57	.37	I63	.76	.30	I53	.28	.24	I36	.57	.18	I136	.82	.10
I112	.33	.37	I135	.63	.30	I91	.69	.24	I71	.61	.18	I164	.68	.10
I147	.44	.37	I174	.77	.30	I93	.33	.24	I79	.70	.18	I41	.64	.09
I150	.64	.37	I28	.67	.29	I103	.64	.24	I85	.26	.18	I49	.56	.08
I55	.75	.36	I30	.39	.29	I121	.48	.24	I159	.77	.18	I34	.71	.07
I70	.48	.36	I37	.77	.29	I123	.56	.24	I13	.30	.17	I154	.53	.07
I111	.39	.36	I50	.39	.29	I54	.73	.23	I44	.73	.17	I9	.76	.05
I122	.52	.36	I52	.38	.29	I87	.53	.23	I48	.76	.17	I61	.23	.04
I10	.61	.35	I110	.30	.29	I97	.42	.23	I76	.55	.17	I169	.50	.02
I118	.64	.35	I148	.87	.29	I129	.33	.23	I98	.52	.17	I75	.21	0
I149	.45	.35	I179	.82	.29	I134	.81	.23	I119	.47	.17	I177	.81	0

Table 5 presents the correlation coefficients obtained among the three language proficiency tests and their subscales. As can be seen, the correlation coefficients (CCs) obtained between the AC-Test and the TOEFL ($r = .83, p < .01$) is almost the same as that of the standard

C-Test ($r = .84, p < .01$). The CCs between the standard C-Tests and AC-Test ($r = .88, p < .01$) is even higher than those obtained among the standard C-Tests, AC-Test and TOEFL. These results confirm the third hypothesis that *the AC-Test will correlate significantly with the standard C-Test and TOEFL*.

Table 5. Correlation coefficients obtained among the tests and their subscales

Tests	C-Test	CT1	CT2	CT3	CT4	ACT	TOEFL	Str.	Wr(Exp.)	Read	Voc
C-Test	1										
C-Test1 (CT1)	.871**	1									
C-Test2 (CT2)	.909**	.730**	1								
C-Test3 (CT3)	.911**	.755**	.778**	1							
C-Test4 (CT4)	.887**	.672**	.741**	.728**	1						
AC-Test (ACT)	.878**	.781**	.755**	.826**	.780**	1					
TOEFL	.840**	.734**	.762**	.767**	.742**	.834**	1				
Structure (Str.)	.730**	.636**	.661**	.673**	.641**	.695**	.886**	1			
Written Expression	.764**	.663**	.653**	.709**	.704**	.768**	.854**	.677**	1		
Reading	.628**	.526**	.661**	.563**	.499**	.602**	.800**	.666**	.567**	1	
(Voc)abulary	.718**	.652**	.605**	.650**	.662**	.754**	.838**	.631**	.676**	.494**	1

* Correlations are significant at .01 (Two tailed)

5 Conclusions

This study distinguished between short passages chosen by experts in C-Tests and the authentic passages which are written by speakers of a given language such as English in order to provide their literate public with as much information as possible as regards a topic of mutual interest. When the scores of 135 undergraduate students on the standard C-Tests developed on short passages and the one constructed on a single authentic passage, i.e., AC-Test, were compared with each other it was found that while both tests enjoy almost the same degree of external validity, i.e., they both correlate significantly and highly with the TOEFL, the AC-Tests proved to be superior to their standard counterparts in terms of reliability and internal validity.

The results of the present study show that choosing *only* a number of short passages in order to have a more representative sample of target texts does not necessarily result in developing empirically valid measures of language proficiency. This argument holds equally true for the AC-Tests developed on short authentic texts chosen by some scholars such as Babaii and Ansary (2001). This argument holds equally true for the C-Tests developed on short AC-Tests (see Babaii & Ansary, 2001). Nor is presenting C-Test items to native speakers or teachers of languages necessary to ensure the successful functioning of the whole test and its constituting items. Developing C-Tests on even a section of an authentic text can provide testing experts and language teachers alike with reliable and valid measures of their learner's proficiency. However, future research must show whether replicating the study with a sample of both undergraduate and graduate students and employing only the well functioning items of all the tests administered will confirm the findings of this study.

References

- Babaii, E., & Ansary, H. (2001). The C-test: a valid operationalization of reduced redundancy principle. *System*, 29, 209–219.
- Babaii, E., & Moghaddam, M. J. (2006). On the interplay between test task difficulty and macro-level processing in the C-test. *System*, 34, 586–600.
- Baker, D. (1989). *Language Testing: A critical survey and practical guide*. London: Edward Arnold.
- Clapham, C. (1996). *The development of IELTS: A study of the effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.
- Dugatkin, L. (May 07, 2005). Why don't we just kiss and make up? *New Scientist* 2498, p. 35. Retrieved May 23, 2005, from <http://www.newscientist.com/channel/life/mg18624981.300>.
- Educational Testing Service. (1991). *Reading for TOEFL*. Princeton, NJ: ETS.
- Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221–33.
- Flesch, R. F. (1949). *The art of readable writing*. New York: Harper & Row.
- Gholami, M. (2006). *The effect of content schema type on Iranian test takers' performance*. Unpublished MA thesis, Ferdowsi University of Mashhad, Iran.
- Jafarpur, A. (1999). Can the C-test be improved with classical item analysis? *System*, 27, 79–89.
- Khodadady, E. (2004). Schema-based cloze multiple choice item tests: Measures of reduced redundancy and language proficiency. *ESpecialist*, 25(2), 221–243.
- Khodadady, E. (2007). C-Tests method specific measures of language proficiency. *Iranian Journal of Applied Linguistics (IJAL)*, 10(2), 1–26.
- Khodadady, E., & Elahi, M. (2012). The Effect of Schema-Vs-Translation-Based Instruction on Persian Medical Students' Learning of General English. *English Language Teaching*, 5(1), xx–xx.
- Khodadady, E., & Herriman, M. (2000). Schemata theory and selected response item tests: from theory to practice. In A. J. Kunnan (Ed.), *Fairness and validation on language assessment* (pp. 201–222). Cambridge: CUP.
- Khodadady, E., Hosseini, A. F., & Etminan, S. (2012). Cognitive Styles and Performance on Schema-Based Cloze Multiple Choice Item Tests: A Fairness Issue. *Journal of Language Teaching and Research*, 3(4), xx–xx.
- Klein-Braley, C. (1981). *Empirical investigations of cloze tests*. Unpublished PhD dissertation, University of Duisburg.
- Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: an appraisal. *Language Testing*, 14/1, 47–84.
- Klein-Braley, C., & Raatz, U. (1990). Die objektive Erfassung des Sprachstands im mutter- und fremdsprachlichen Unterricht durch C-Tests. In A. Wolff, & H. Rössler (Eds.). *Deutsch als Fremdsprache in Europa*. Regensburg: Arbeitskreis Deutsch als Fremdsprache, 239–50.
- Klein-Braley, C., & Raatz, U. (Eds.) (1985). *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis*. Bochum: AKS.

Appendix

Authentic C-Test

Directions: The second half of every second word in the passage below has been removed. If there is an odd number of letters in the word, then the larger “half” is removed. Read the texts and do as follows:

- You have to restore the second deleted **half**.
- Your restored word should be only **one word**.
- Pay attention to the tense and subject-verb agreement.
- Your spelling should be **correct**.
- Use a pencil so that you can correct your mistakes.

Why don't we just kiss and make up?

LOOK at the world's worst trouble spots and you can't fail to notice they have one thing in common: tit-for-tat attacks between warring parties. Escalation o_____ (1) violence i_____ (2) incredibly destr_____ (3), yet w_____ (4) humans fi_____ (5) it ve_____ (6) difficult t_____ (7) break t_____ (8) vicious cy_____ (9). It se_____ (10) we a_____ (11) not go_____ (12) at conf_____ (13) resolution. Perhaps we could learn a lesson or two from the spotted hyena.

Spotted hyenas are highly sociable. Like ot_____ (14) animals th_____ (15) live i_____ (16) close-knit gro_____ (17), they d_____ (18) not alw_____ (19) get al_____ (20). But spo_____ (21) hyenas do n_____ (22) hold a gru_____ (23). Within ab_____ (24) 5 min_____ (25) of a fi_____ (26), the erst_____ (27) combatants c_____ (28) often b_____ (29) seen pla_____ (30), licking o_____ (31) rubbing o_____ (32) another, o_____ (33) engaging i_____ (34) other frie_____ (35) acts t_____ (36) dissipate t_____ (37) tension. A_____ (38) they a_____ (39) not t_____ (40) only ani_____ (41) with a penc_____ (42) for kis_____ (43) and mak_____ (44) up. I_____ (45) their bo_____ (46) *Natural Conflict Resolution*, Filippo Aureli from Liverpool John Moores University, UK, a_____ (47) Frans de Waal from t_____ (48) Yerkes Primate Center at Emory University, Atlanta, Georgia, docu_____ (49) reconciliation i_____ (50) no le_____ (51) than 27 spe_____ (52) of prim_____ (53). Bottlenose dolp_____ (54) also d_____ (55) it. Ev_____ (56) goats. So why can't we be more forgiving?

Admittedly, human interactions are far more complex. But per_____ (57) we c_____ (58) draw so_____ (59) lessons fr_____ (60) the st_____ (61) of conf_____ (62) resolution i_____ (63) nature. N_____ (64) only ha_____ (65) ethologists disco_____ (66) that i_____ (67) is a l_____ (68) more com_____ (69) than y_____ (70) might exp_____

(71), they a_____ (72) also wor_____ (73) out t_____ (74) costs a_____ (75) benefits o_____ (76) conflict resol_____ (77). Their ideas about when, where and how reconciliation works in nature, could help us to improve the chances of settling our own disputes.

When it comes to making up, our primate cousins get top marks. De Waal desc_____ (78) a typ_____ (79) incident i_____ (80) which Hennie, a yo_____ (81) female ch_____ (82), has be_____ (83) slapped dur_____ (84) a pas_____ (85) charge b_____ (86) Nikkie, t_____ (87) leader o_____ (88) the gr_____ (89). Hennie ret_____ (90) from t_____ (91) fray, a_____ (92) first care_____ (93) the sp_____ (94) where s_____ (95) was h_____ (96) and th_____ (97) just ly_____ (98) in t_____ (99) grass a_____ (100) staring in_____, (101) the dist_____ (102). "More th_____ (103) 15 min_____ (104) later Hennie slo_____ (105) gets u_____ (106) and wa_____ (107) straight t_____ (108) a gr_____ (109) that incl_____ (110) Nikkie," de Waal wri_____ (111). "Hennie

appro_____ (112) Nikkie, gree_____ (113) him wi_____ (114) a ser_____ (115) of so_____ (116) pant gru_____ (117). Then s_____ (118) stretches o_____ (119) her a_____ (120) to of_____ (121) Nikkie t_____ (122) back o_____ (123) her ha_____ (124) for a ki_____ (125). Nikkie's ha_____ (126) kiss cons_____ (127) of tak_____ (128) Hennie's wh_____ (129) hand rat_____ (130) unceremoniously in_____ (131) his mo_____ (132). This contact is followed by a mouth-to-mouth kiss."

Note, it is Hennie, the chimp who came off worst in the argument, who instigates the reconciliation. In fa_____ (133), this i_____ (134) a gen_____ (135) pattern f_____ (136) most inst_____ (137) of conf_____ (138) resolution. Gabriele Schino fr_____ (139) the National Research Council i_____ (140) Rome, Italy, even fo_____ (141) it i_____ (142) goats. Af_____ (143) she h_____ (144) induced conf_____ (145) over fo_____ (146), she fo_____ (147) that 16 per_____ (148) of a _____ (149) interactions bet _____ (150) goats we_____ (151) reconciliatory, consi_____ (152) of frie _____ (153) acts su_____ (154) as groo_____ (155) and muz_____ (156) rubbing bet_____ (157) animals th_____ (158) had be_____ (159) fighting previ_____ (160). As in primates, this was most often initiated by the loser of the fight.

Like Hyenas and most primates, goats are sociable animals. And th_____ (161) seems t_____ (162) be o_____ (163) of t_____ (164) key attri _____ (165) of spe _____ (166) that g_____ (167) out o_____ (168) their w_____ (169) to res_____ (170) their conf_____ (171). Hardly surpr _____ (172) then, th_____ (173) such beha _____ (174) is al _____ (175) found i _____ (176) dolphins. Alth _____ (177) they se _____ (178) to ha _____ (179) happy gr _____ (180) perpetually plastered on their faces, dolphins are surprisingly aggressive. And, sure enough, they are big on conflict resolution, as Amy Samuels from Woods Hole Oceanographic Institute, Massachusetts, and Cindy Flaherty from Brookfield Zoo in Chicago, Illinois, have observed. While studying a small group of bottlenose dolphins at the zoo, Samuels and Flaherty noticed that after a fight opponent often engaged in "gentle rubbing" or "contact swims", in which one dolphin towed another through the water.

This is the end of the test.