

Comparing the Psychometric Characteristics of Speeded and Standard C-Tests

Anita Fadaeipour¹, Zahra Zohoorian²

Received: 11 October 2016

Accepted: 02 February 2017

Abstract

The C-test as a measure of first and second language proficiency suffers from ceiling effect. That is, the C-test is very easy for native speakers and for advanced second language learners and many examinees obtain full scores. Therefore, the test cannot distinguish among advanced learners or native speakers. Speeded C-test is suggested as a remedy for this problem. Previous research has showed that if C-test is administrated under time constraint it can better differentiate among advanced learners and native speakers. In this study we aimed to compare the psychometric characteristics of speeded and standard C-tests and investigate their relationship with reading comprehension. One-hundred Iranian undergraduate English students were selected to take part in this research. A standard (power) C-test, a speeded C-test, and a reading comprehension test were given to the students. Findings revealed that both speeded C-test and standard C-test could measure reading comprehension ability but standard C-test is a better predictor of reading comprehension ability and enjoys a higher reliability. Exploratory factor analysis of the data resulted in a one-factor solution. All C-test passages (speeded and power) and reading comprehension passages loaded heavily on a single factor. Nevertheless, power C-test passages had higher loadings compared to speeded passages.

Keywords: *power C-test, speeded C-test, reading comprehension, language proficiency, validity, reliability*

1. Introduction

Reduced redundancy tests are considered as one of the most economic ways of second and foreign language proficiency assessment (Atiken, 1977; Baghaei, 2011a; Spolsky, 1968). In such tests the respondents are required to anticipate and restore elements of a mutilated text (Caulfield & Smith, 1981). In fact, reduced redundancy refers to the ability of understanding an incomplete text and make correct guesses related to the missing information. The more proficient language learners are, the more successful they will be in guessing the correct choices. In other words, the key purpose of reduced redundancy is to evaluate the learners' language level as well as assess their capacity to utilize redundancies by figuring out the

¹English Department, Islamic Azad University, Mashhad Branch, Mashhad, Iran

²English Department, Islamic Azad University, Mashhad Branch, Mashhad, Iran (corresponding author)

incomplete messages. Thus, the assumption here is that the more proficient the learners are the more deleted parts they can restore.

The most familiar reduced redundancy test is the cloze test which is used as an overall language proficiency test or a reading comprehension test (Baghaei&Ravand, 2016; Caulfield & Smith, 1981). A cloze test is a passage in which words are deleted based on a pre-specified rule. Learners need to read the passage carefully and then provide the missing words where they are blanked out. According to Oller (1973), the primary use of a cloze test is to place EFL learners at different levels of proficiency at universities. Moreover, the test is considered as an integrated test of language proficiency as it covers modules such as grammar, vocabulary, reading, etc. Bormuth (1969) also asserted that, "tests made by this procedure are simple and economical to prepare and do not confound the passage difficulty measurements with the difficulty of the language and other characteristics peculiar to the test questions, themselves." (p.189). In addition to the benefits of this test, there are however some drawbacks such as being too much clause driven. Additionally, it is considered as an inappropriate way of reading assessment (Alderson, 1979). Moreover, there are concerns related to deletion issues, scoring procedures as well as reliability and validity (Klein-Braley, 1985; Klein-Braley&Raatz, 1984).

To overcome the shortcomings of cloze, Raatz and Klein-Braley(1984) suggested the C-test. Yielding better psychometric properties (Chapelle& Abraham, 1990), the C-test as another member of the reduced redundancy tests family is employed for the assessment of general language proficiency especially in EFL learners (Atiken, 1977). In this regard, Lin, Yuan, and Feng (2008, p. 64) state that:

"The "C" in C-test was chosen as an abbreviation of the word "cloze" to emphasize the relationship between C-test and cloze test. Also a representative of the LRR family, C-test, as pointed out by Raatz and Klein-Braley, was developed not only to retain the positive aspect of cloze test (i.e., its capacity to tap an examinee's ability to process discourse and to predict from context with reduced redundancy) but also to correct the major technical defect of cloze test (i.e., the failure of its deletion technique to ensure a random sampling, which is crucial for LRR tests). Unlike cloze test in which deletion is performed at the text level, C-test was designed to achieve random sampling by performing deletion at the word level. That is, only parts of a word, rather than a whole word, are removed in C-test."

Consequently, C- test was introduced to the researchers and instructors community as an improved form of cloze test. A C-test battery is composed of four to six short passages where the second half of every second word is deleted. There are 20-25 gaps in each passage. Although, Baghaei (2011b, 2011c) demonstrated that C-tests with fewer gaps work efficiently, too.

As Spolsky (2001) maintains, C-tests are superior to cloze tests in measuring learners' proficiency. Klein- Braley (1996, p. 23) asserts that "tests based on the C-principle function as proficiency tests for a variety of different groups, in particular for children learning their own language (L1 learners), for children and adults learning a second language in the country in which it is spoken (L2 learners), and for foreign language learners (LF learners)". Simple

administration and easy scoring as well as correction based on the acceptable word procedure are also among the advantages. Yet, another favorable advantage of C-tests is that examinees find it less disappointing than cloze tests (Katona&Dörnyei,1992).

The other feature of C-tests include objective scoring which for the most part show high reliability (Eckes&Grotjahn, 2006). The other favorable advantage of C-test over cloze is the utilization of various passages in order to avoid content specificity and test bias (Raatz& Klein-Braley, 2002). This also allows researchers to circumvent the item local dependence problem that is common in cloze tests. Thus, by aggregating the correct replies for each passage and entering passages as super-items or test lets, the application of item response theory models and internal consistency estimates of reliability is possible (Baghaei, 2008a; Eckes&Baghaei, 2015).

Over the years C-Tests have been developed in many languages to measure first and second or foreign language proficiency. Validity evidence for C-test has been accumulated through various methods. Concurrent validity evidence (Baghaei, Monshi-Toussi, &Boori, 2009), evidence by fitting item response theory models (Baghaei, 2008a; Baghaei, 2008b), differential group studies (Baghaei, 2014), investigating mental processes (Stemmer, 1991), construct representation (Khoshdel, Baghaei, &Bemani, 2016), and other procedures have been employed to validate C-Tests.

Researchers have also shown that the construct measured by the C-test can be manipulated through deliberate manipulation of the passage content (Baghaei, et al., 2009; Baghaei, 2008b, Baghaei, 2010; Baghaei&Grotjahn, 2014a; Baghaei&Grotjahn, 2014b; Baghaei&Tabatabaee, 2015) which allows test developers to tailor the test to measure specific predefined constructs.

2. Literature Review

One drawback of C-tests is ceiling effect; that is, it is quite simple for native speakers and advanced learners. Therefore, it cannot be used as a means of distinguishing among them. To untangle this problem and to adapt the C-tests' difficulty for different levels of learners, Köberland Sigott (1994) introduced other deletion patterns. They suggested deleting more than 50% of the letters in words.

Yet, another strategy was introduced by Grotjahn, Schlak, and Aguado (2010) and Grotjahn (2010), i.e., employing speeded C-Test or S-C-test which includes imposing a relatively short time constraint on every single C-test passage. Using time limitation, the test constructors can make the C-test more difficult for advanced learners and native speakers. As mentioned above, due to various reasons, test constructors may put a time limitation on different tests to check the learners' performances in an accelerated manner. As stated by Grotjahn, et al. (2010), in the standard version of C- Test time limitation is set for the whole test rather than on individual texts due to the fact that this test type was primarily designed as a power test, not a speeded one.

In this study, the researchers impose time limitation for C-test in order to observe if the time constraint could influence learners' performance. While many researchers use C-test as a language proficiency test, finding a comparable performance on speeded C-test and standard

C-test would be advantageous as researchers can employ a new version and save the time and spend it for other services concerning language learners.

Standard C- test may be challenging to learners and teachers. As learners are free to cover the test at any rate, those who fail in time management may not be able to fulfill the tasks thoroughly. Thus, it is claimed that this could be assumed as a shortcoming, since it leads to misunderstanding and misestimating of item discrimination, text difficulty, test reliability and test validity. Hence, with the intention of compensating for the mentioned drawbacks, time limitation is considered to make the C- test a speeded one. Grotjahn, et al., (2010) argue that this way, test reliability and validity will be considerably increased. They also claimed that this variation leads to an increased test difficulty, discriminatory power, and reliability.

3. Methodology

3.1 Participants and Setting

In order to collect the required data, 100 Iranian undergraduate EFL learners with different ages, genders, and various levels of proficiencies were selected randomly as the participants. It needs to be noted that participants were all non-native speakers whose first language was Persian, studying English during the first semester of 2016 at Islamic Azad University and Tabaran Institute of Higher Education, Mashhad, Khorasan Razavi, Iran. Moreover, they ranged between 20 and 35 years in age. They were also assured that their personal information would be kept confidential.

In this study, the participants' performances on standard and speeded C-tests were compared to distinguish the best method for assessing learners' general language proficiency. In this regard, the researcher provided a three-phase assessment for the group of testees. Accordingly, a speeded C-test along with a reading passage were administered to all the participants; they were required to answer completely in one session. After one week, the participants were given a standard C-test as well. It is worth noting that all the testees had the same examiner, circumstances and settings.

3.2 Instrumentation

3.2.1 Standard C-test

To measure the students' general language proficiency, a standard (power) C- test was employed. The participants were asked to read the instructions first and then fill in the 50 mutilated words provided in two short reading passages. Each passage contained 25 gaps. The texts were presented under two topics: chocolate and money. The texts were driven out of commonly used reading comprehension textbooks for EFL learners. The time allotted to the testees to complete the C-test was 10 minutes.

3.2.2 Speeded C-test

The speeded C-test like the standard C-test contained two passages, each with 25 gaps. One of the passages was about birthday and the other about Hollywood. The only difference in the administration of the C-tests was the time allotted for them. As the name justifies sufficiently, speeded C-test is administered in a shorter period of time; the testees had only five minutes to cover the whole test.

3.2.3 Reading comprehension test

The reading comprehension test consisted of four short passages including twenty questions. The questions were multiple choice items along with written ones. 25 minutes were devoted to the testees to answer this section. The test reliability was also assured having a Cronbach's alpha of 0.70. The test was the reading comprehension section of one of the past papers of Pearson Test of English General Level 3 (Pearsonpte.com).

4. Analyses and Results

As is presented in Table 1 below, the information we needed for each of the variables is summarized. Table 1, displays the descriptive statistics including the mean, standard deviation, variance, and range for the speed C-test, power C-test, and reading comprehension. The mean obtained in reading comprehension test was 6.99 out of 20. It indicates that this test is difficult for students to answer and also the mean in speeded C-test was 15.43 out of 50 and the mean in the power test was 20.14 out of 50. The means in speeded C-test and standard C-test were low. Therefore, it is observed that these two tests were also difficult for students but students were more successful in answering the power C-test.

Table 1: Descriptive statistics for the measures of the study

	Reading	Speed C-test	Power C-test
Mean	6.99	15.43	20.14
Std. Deviation	3.15	7.69	7.81
Variance	9.97	59.17	61.15
Range	16.00	39.00	40.00
Minimum	2.00	2.00	3.00
Maximum	18.00	41.00	43.00

4.1 Correlational analysis

The results obtained from the Pearson product-moment correlation coefficient presented that the relationship between speed C-test and reading comprehension [$r=.55$, $n=100$, $p<.05$], and the relationship between power C-test and reading comprehension [$r=.65$, $n=100$, $p<.05$] were statistically significant (see Table 2). The relationship between speeded C-test and power C-test [$r=.46$, $n=100$, $p<.05$] was also statistically significant. The results showed that there is a significant relationship between speeded C-test and reading comprehension and also there is a significant relationship between standard C-test and reading comprehension. The relationship between power C-test and reading comprehension is higher than the correlation between speeded C-test and reading comprehension.

Therefore, the power C -test is a better predictor of reading comprehension. The Cronbach's alpha reliability of the whole C-test battery, considering each passage as a super-item to avoid local dependence problem, with two speeded C-test passages and two standard C-test passages was 0.70. The reliability of the speeded C-test (with two super-items) was 0.41 and the standard C-test (with two super items) was 0.59; thus, the standard C-test was more reliable than the speeded C-test. Accordingly, hypothesis 4 was rejected by this result.

Table 2: Correlations between the tests in the study

	1	2	3
Speed C-test	1	.47	.55
Power C-test		1	.65
Reading Comprehension			1

Note: All correlations are significant $p<0.01$ (2-tailed)

4.2 Regression analysis

Multiple regressions was run to evaluate the explanatory power of the four independent variables (two speed C-test passages and two standard C-test passages), in explaining L2 reading comprehension. Preliminary analyses were conducted to ensure there was no violation of the assumptions of normality, linearity, multicollinearity, and homoscedasticity.

The results revealed that the model explains a significant portion of the variance in the reading comprehension scores ($F(4,95) = 28.96$, $p<.000$, $R^2 = .54$, $R^2_{Adjusted} = .53$). The analysis showed that speeded C-test passage 1 (Beta = .38, $p = 0.00$), standard C-test passage 1 (Beta = 0.32, $p = 0.00$), and standard C-Test passage 2 (Beta = .27, $p = 0.001$) could significantly predict reading comprehension, only speeded C-Test passage 2 did not significantly predict reading comprehension (Beta = .061, $p = .41$). Therefore, standard C-test is a better predictor for reading comprehension. Moreover, the adjusted R^2 was 0.53 which means

that these four independent variables explained 53% of the variance in the reading comprehension scores.

Table 3: Multiple regression analysis for predicting reading comprehension using speeded and standard C-test passages

Independent variable	Beta	T	P	Part correlation
		-.04	.96	
S.Test1	.38	4.4	.00	.30
S.Test2	.06	.82	.41	.05
Standard1	.32	3.96	.00	.27
Standard2	.27	3.39	.00	.23

4.3 Factor analysis

The four C-test passages and the four reading comprehension passages were subjected to principal component analysis (PCA) using SPSS version 21. Prior to performing PCA, the suitability of data for factor analysis was assessed. The Kaiser-Meyer-Okin value was 0.83, exceeding the recommended value of 0.60 and Bartlett’s test of Sphericity reached statistical significance, ($p=0.00$) supporting the factorability of the correlation matrix. Principal components analysis illustrated the presence of one component with eigenvalue exceeding 1, explaining 41% of the variance. Inspection of the scree plot revealed one clear break after component one.

Figure 1: Scree plot showing the number of factors to extract from the data



Table 4 depicts the factor loadings of each variable on the single extracted factor. As the table shows all the variables have high loadings on the factor. All the loadings are above 0.40 and we can name the factor as reading comprehension factor as all reading passages load on this factor. The C-test passages also load on this factor. The power C-test passages, however, have higher loadings on this factor. This can be interpreted as the superiority of power C-test in measuring reading comprehension compared to speeded C-test. This is in line with the outcome of the regression analysis.

Table 4: Factor loadings for the variables on the single extracted factor

Variable	Component
	1
Speed-C 1	.684
Speed-C 2	.496
Standard-C 1	.743
Standard-C 2	.705
Passage 1	.405
Passage 2	.590
Passage 3	.678
Passage 4	.762

5. Discussion and conclusion

The aim of this study was to compare the psychometric characteristics of speeded and standard C-tests among Iranian EFL learners. One hundred undergraduates of English were selected using convenience sampling. Findings indicated that the correlation between speeded C-test and reading comprehension was significant and also the correlation between standard C-test and reading comprehension was highly significant. The correlation between the speeded and power C-test was smaller than the correlation between these two test types and reading comprehension. This finding suggests that power and speed C-tests may not share much variance and doing C-tests under time constraints probably alters the construct of C-test to a great extent. Further research is needed to corroborate this finding. Findings also indicated that standard C-tests had a higher reliability.

Multiple regression analysis of the data indicated that the betas in speeded C-test passage 1, standard C-test passage 1, and standard C-test passage 2 were highly significant which confirms that these three variables were good predictors of reading comprehension. The researchers performed factor analysis on data which revealed that all the variables loaded on one component. This means that the speeded C-test, the standard C-test and the four reading comprehension passages probably measured reading comprehension ability. This can be considered as validity evidence for the entire test as a measure of reading comprehension (see Baghaei&Tabatabaee, 2016).

Furthermore, the results revealed that both the speeded C-test and the standard C-test could measure reading comprehension ability; however, the standard C-test could be regarded as a better measure for this goal. Based on the findings of the present study teachers are recommended to employ standard C-test if there is no time pressure. However, the findings of the study contradict those of Grotjahn et al. (2010) who demonstrated that time limitation improves C-test's discriminatory power and reliability.

When time constraints exist employing speeded C-test is highly recommended by this study. The findings also showed that the test composed of speeded C-tests, standard c- tests, and multiple choice reading comprehension items yields a uni-factorial solution. This indicates that all subtests measure one specific construct which is most probably the reading comprehension ability.

Nowadays, there is a lack of time in the lives of all human beings. All areas related to human lives including teaching and learning need to be covered rapidly. In addition, due to the progress of societies, drastic changes of human needs, and the great increase in every individual's expectations, teaching and learning techniques need to be modified and developed. Speeded C-test is a step toward the need for this alteration.

Additionally, speed is one of the key factors in language learning process when the purpose of testing is measuring proficiency in various language skills. For instance, while speaking people mostly consider fluency; i.e., how quick words are set together is assumed important. The same expectations exist for other language skills of reading, writing, and listening. Thus, processing speed could be considered as a part of proficiency. In this regard, speeded C-tests could incorporate this factor by measuring it. Future research should examine the utility of speeded C-test in other population of students at different proficiency levels and age groups. In this study we only investigated the relationship of these two types of C-Test with reading comprehension. Examining and comparing the association between these two test types and other language skills and compost scores of various language tests is also needed.

References

- Aitken, K.G. (1977). Using Cloze Procedure as an Overall Language Proficiency Test. *TESOL Quarterly*, 11, 59-67.
- Alderson, J. C. (1979). "The cloze procedure and proficiency in English as a foreign language." *TESOL Quarterly*, 13, 219-227.
- Baghaei, P. (2008a). The effects of the rhetorical organization of texts on the C-Test construct: A Raschmodelling study. *Melbourne Papers in Language Testing*, 13, 32-51.
- Baghaei, P. (2008b). An attempt to fit the Rasch model to a C-Test. *Iranian EFL Journal*, 2, 6-15.
- Baghaei, P., Monshi-Toussi, M. T., &Boori, A. A. (2009).An Investigation into the validity of conversational C-Test as a measure of oral abilities.*Iranian EFL Journal*, 4, 94-109.
- Baghaei, P. (2010). An investigation of the invariance of Rasch item and person measures in a C-Test. In R. Grotjahn (Ed.).*Der C-Test: Beiträge aus der aktuellenForschung/ The C-Test: Contributions from Current Research*(pp.100-112). Frankfurt/M.: Lang.
- Baghaei, P. (2011a). *C-Test construct validation: A Rasch modeling approach*. Saarbrücken: VDM VerlagDr Müller.

- Baghaei, P. (2011b). Do C-Tests with different number of gaps measure the same construct? *Theory and Practice in Language Studies, 1*, 688-693.
- Baghaei, P. (2011c). Optimal number of gaps in C-Test passages. *International Education Studies, 4*, 166-171.
- Baghaei, P. (2014). Construction and validation of a C-Test in Persian. In R. Grotjahn (Ed.), *Der C-Test: Aktuelle Tendenzen/The C-Test: Current Trends*, 301-314. Frankfurt am Main: Lang.
- Baghaei, P., & Grotjahn, R. (2014a). The validity of C-Tests as measures of academic and everyday language proficiency: A multidimensional item response modeling study. In R. Grotjahn (Ed.). *Der C-Test: Aktuelle Tendenzen/The C-Test: Current trends* (pp. 163-171.). Frankfurt/M.: Lang.
- Baghaei, P., & Grotjahn, R. (2014b). Establishing the construct validity of conversational C-Tests using a multidimensional Item Response Model. *Psychological Test and Assessment Modeling, 56*, 60-82.
- Baghaei, P., & Ravand, H. (2016). Modeling local item dependence in cloze and reading comprehension test items using testlet response theory. *Psicológica, 37*, 85-104.
- Baghaei, P., & Tabatabaee, M. (2015). The C-Test: An integrative measure of crystallized intelligence. *Journal of Intelligence, 3*, 46-58. Available: <http://www.mdpi.com/2079-3200/3/2/46>
- Baghaei, P., & Tabatabaee-Yazdi, M. (2016). The logic of latent variable analysis as validity evidence in psychological measurement. *The Open Psychology Journal, 9*, 168-175.
- Bormuth, J. R. (1969). Factor Validity of Cloze Tests as Measures of Reading Comprehension Ability. *Reading Research Quarterly, 4*, 358-365.
- Byram, M., & Hu, A. (2013). *Routledge Encyclopedia of Language Teaching and Learning*. Canada: Routledge.
- Caulfield, J. and Smith, W.C. (1981). The reduced redundancy test and the cloze procedure as measures of global language proficiency. *Modern Language Journal, 65*, 54-58.
- Chapelle, C., & Abraham, R., (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.
- Eckes, T., & Baghaei, P. (2015). Using testlet response theory to examine local dependency in C-Tests. *Applied Measurement in Education, 28*, 85-98.
- Grotjahn, R. (2010). Gesamtdarbietung, Einzeltextdarbietung, Zeitbegrenzung und Zeitdruck: Auswirkungen auf Item- und Testkennwerte und C-Test-Konstrukt. In Rüdiger Grotjahn (Ed.), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from current research* (pp. 265-296). Frankfurt am Main: Lang.
- Grotjahn, R., Schlak, T., & Aguado, K. (2010). S-C-Tests: Messung automatisierter sprachlicher Kompetenzen anhand von C-Tests mit massiver textspezifischer Zeitlimitierung. In

- Rüdiger Grotjahn (Ed.), *Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from current research* (pp. 297-319). Frankfurt am Main: Lang.
- Katona, L., & Dörnyei, Z. (1992). The C-test: A teacher-friendly way to test language proficiency. *English Teaching Forum*, 31, 34-35.
- Khoshdel, F. Baghaei, P., & Bemani, M. (2016). Investigating Factors of Difficulty in C-Tests: A Construct Identification Approach. *International Journal of Language Testing*, 6(2), 113-122.
- Klein-Braley, C. (1996). Towards a theory of C-Test processing. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 3, pp. 23-94). Bochum: Brockmeyer.
- Köberl, J. & Sigott, G. (1994). Adjusting C-test difficulty in German. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 2, pp. 179-192). Bochum: Brockmeyer.
- Klein-Braley, C., (1985). A cloze-up on the C-Test: A study in the construct validation of authentic tests. *Language Testing*, 2, 76-104.
- Klein-Braley, C., & Raatz, U., (1984). A survey of research on the C-test. *Language Testing*, 1, 134-146.
- Lin, W., Yuan, H., & Feng, H. (2008). Language reduced redundancy tests: A reexamination of cloze test and C-test. *Journal of PanPacific Association of Applied Linguistics*, 12, 61-79.
- Oller, J. W. jr. (1973). Cloze tests of second language proficiency and what they measure. *Language Learning*, 23, 105-118.
- Raatz, U., & Klein-Braley, C. (2002). Introduction to language testing and to C-Tests. In J. A. Coleman, R. Grotjahn & U. Raatz (Eds.), *University language testing and the C-test* (pp. 75-91). Bochum: AKS-Verlag.
- Spolsky, B. (2001). Closing the cloze. In H. Pürschel & U. Raatz (Eds.), *Tests and Translation: Papers in Memory of Christine Klein-Braley* (pp. 1-20). Bochum: AKS-Verlag.
- Spolsky, B. (1968). What does it mean to know a language? Or how do you get someone to perform his competence? Paper presented at the Second Conference on Problems in Foreign Language Testing, University of Southern California.
- Stemmer, B. (1991). *What's on a C-test taker's mind: Mental processes in C-test taking*. Bochum: Brockmeyer.