

Speeded C-Tests: Some Psychometric Analyses

Anita Fadaeipour

► **To cite this version:**

Anita Fadaeipour. Speeded C-Tests: Some Psychometric Analyses. [Research Report] Islamic Azad University. 2017. hprints-01510897

HAL Id: hprints-01510897

<https://hal-hprints.archives-ouvertes.fr/hprints-01510897>

Submitted on 20 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Speeded C-Tests: Some Psychometric Analyses

Anita Fadaei Pour

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

Winter 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. Grotjahn, Schlak, and Aguado (2010) and Grotjahn (2010) suggested Speeded C-test or Standard-C-test as a remedy for this problem. They argued 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-test and investigate the relationship between standard and speeded C-test and reading comprehension. One-hundred Iranian undergraduate EFL learners studying at Islamic Azad University and Tabaran University of Mashhad, Khorasan Razavi, Iran, 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 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 the single extracted factor. Nevertheless, power C-test passages had higher loadings compared to speeded passages.

Key terms: C-test, speeded C-test, standard C-test, reading comprehension

1. Introduction

1.1 Background and Purpose

Language testing seems to contain assessing what learners have achieved during a period of time, including both accuracy and fluency. According to Fulcher (2015, p. 1), "for some, language tests are gate keeping tools that further the agendas of the powerful. For others, they are assumed as vehicles by which society can implement equality of opportunities or learner empowerment." Learners' feelings and notions about the testing atmosphere might be quite different in each individual. In spite of the vast number of aims being claimed for language testing, Carroll (1961, p. 314) believes that "the purpose of language testing is always to render information to aid in making intelligent decisions about possible courses of actions."

Based on what Davidson and Lynch (2002, p, 76-78) believe, test purposes come from the intentions behind them. They also proposed the term mandate as authority who tends to carry out the test. These scholars also assert that mandates could be varied as internal and external, based on the situation. If the teachers, instructors, or school administrations make the test, it will be called internal. However, an external mandate is explained as an outer operator who is not much familiar with the learners and the situation. Indeed, the aims of testing regarding these different types of mandates vary completely.

There are distinct forms of language assessment which are individually appropriate for a specific purpose and setting. Brown and Hudson (1998, p. 653) assert that "language testing differs from testing in other content areas because language teachers have more choices to make". Accordingly, they believe that different types of assessments are typically categorized into three main classifications:

- Selected response assessment including true- false, matching, and multiple- choice assessments
- Constructed response assessments including fill- in the blanks, short answer, and performance assessments
- Personal response assessments including conference, portfolio, and self- or peer assessments

For each of these assessment types there could be found various examples. However, C- test as the core idea in this study could be classified in the second category. In 1981, Raatz and Klein Braley first designed C- test. They claimed that C-Test is considered as a new testing technique in light of the precepts of the cloze test with some improvements. By settling the deletion rate to 2, and expanding the quantity of sections to 4–6, they provided a promoted version of testing method. As indicated by Klein-Braley (1985, 1996), C- Tests measure test takers' language proficiency and ought to have no less than 100 blanks.

According to Khodadady and Hashemi (2011), C-Tests uncover a critical connection with built up language proficiency tests like the TOEFL. They should not be utilized alone to quantify test takers' language capability since they cannot replace other standardized tests as they are not able to prove a high correlation. Khodadady, et al, also asserts that "The C-Tests nonetheless measure a language proficiency ability which is lacking in traditional multiple choice items measuring structure, vocabulary, and reading comprehension abilities" (p. 21).

Time constraint in C-test as speeded C-Test (Grotjahn, Schlak & Agudo 2010) has been employed in the context of a research project on the effect of age on ultimate attainment in second language acquisition (Aguado, Grotjahn and Schlak, 2005, 2007), and is intended to differentiate among competent second language learners. In this study we

aim to compare the psychometric characteristics of speeded and standard C-Test¹ and investigate the relationship between standard and speeded C-Test and reading comprehension.

1.2 Statement of the Problem

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 individuals' expectations, teaching and learning techniques need to be changed and developed.

In this study, the researcher employs time limit for C-test in order to find out if time constraint could influence learners' performance or not. Many researchers use C-test as a language proficiency test so that if learners' performances on speeded C-test would be the same as standard C-test, the researchers can employ new version and save the time and spend it for other services regarding language learners.

Additionally, speed is one of the key factors in language learning processes such as becoming proficient in various language skills. For instance, while speaking people mostly consider fluency; this is, how quick words are set together is assumed important to everybody. The same expectations exist in reading, writing, and even listening. Thus, processing speed could be supposed as a part of proficiency. In this regard, speeded C-test could consider this factor and measure it completely.

1.3 Research Hypotheses

In this investigation following null hypotheses are formulated:

1. There is no significant relationship between speeded C-test and reading comprehension.
2. There is no significant relationship between standard C-test and reading comprehension.
3. None of the C-tests (speeded C- test or standard C- test) is a better predictor of reading comprehension.
4. There is no difference in the reliability of two types of C- tests.

1.4 Research Questions

Based on the objectives of this study, the following research questions are proposed:

1. Is there a significant relationship between speeded C- test and reading comprehension?
2. Is there a significant relationship between standard C- test and reading comprehension?
3. Which C- test (speeded C- test or standard C- test) is a better predictor of reading comprehension?
4. Is there any difference in the reliability of the standard and speeded C-tests?

¹ Standard C-test and power C-test are used interchangeably in this research

5. What is the factors structure of a composite measure of standard C- Test, speeded C- Test, and reading comprehension?

1.5 Significance of the Study

So far there have been many studies investigating different aspects of language assessment; however, there have been few inquiries about various versions of C-test as one of the popular means of assessment. The researcher believes that the findings of the current study on psychometric characteristics of speeded and standard C-test may open an unprecedented way to teaching and testing language.

The importance of this study lies in its pedagogical implications for second and foreign language teaching. This study aims at enriching the quality of teaching and testing English in educational settings.

The materials used in the past differed vastly from the ones being used in the present era. Everybody is faced with time limitation nowadays, being in a constant rush to cover the routine; thus, the teachers had better use techniques which fit this modern lifestyle.

Additionally, many studies show that C-tests normally measure general language proficiency. In other words, C- tests do not measure only vocabulary or grammar but all aspects of language achievement (Eckes & Grotjahn, 2006; Köberl & Sigott, 1995; Klein-Braley, 1985). In this study, the researcher intended to use time limit for answering C-test question. Time constraint makes the test difficult and discriminatory power increases in advanced students. The results of this research show the comparison between speeded and standard C-test and also investigate the relationship between speeded and standard C-test and reading comprehension. Besides, the outcomes also considers the matter of automaticity among learners. In the same vein, successful learners in speeded C-test are supposed to be much more automatic, since their work much faster than other students. According to Segalowitz, Trofimovich, Gatbonton and Sokolovskaya (2008), "automaticity refers to a significant change in the way processing is carried out (some form of restructuring)" (p. 403). This change could be due to fast processing in learners' minds. Many scholars including Hulstijn (2001), Robinson (2001), and Johnson (1996) have insisted on the great role of automaticity in improving SLA.

It must be noted that the results of this study are considered influential for language instructors, test- takers, school administrators, and curriculum designers who are responsible for assessing learners' general proficiency. One could even claim that students are not deprived of the benefits of this study.

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 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 testlets, 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.

The present study

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, Köberl and 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.

2. Review of the Related Literature

2.1 Conceptualization of Cloze Test

As indicated by Hinofotis (1980), "cloze" originates from the idea of conclusion utilized as a part of Gestalt psychology and alludes to the capacity to fill in the crevices in a fragmented example. According to Lin, Yuan, and Feng (2008), the LRR (Language Reduced Redundancy) standard is very much required in cloze test in the feeling that it tests natural linguistic redundancies and makes examinees use hierarchical limitations to construe the meaning and fill in the spaces.

Cloze test could be assumed as a way of language proficiency measurement. A cloze test is a collection of deletions specifically selected with various distances through a reading passage. Learners need to read the passage carefully and then find out the suitable choice for each deletion.

Based on what Oller (1973) claimed, the primary use of a cloze test goes back to EFL learners' placement at the universities. Besides, the test is considered as an integrated test of language proficiency as it covers grammar, vocabulary, reading, and etc. Bormuth (1969, p. 189) 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."

In addition to the benefits of this test, there are also some disadvantages such as being too much clause driven. Additionally, it is considered as an inappropriate way of reading assessment (Alderson, 1979).

According to Sasaki (2000), the difficulty levels of reading passages must be considered carefully. In other words, reading passages need to be related to the learners' target culture, so that they should grasp the ideas mentioned in the text and be deeply engaged in them. According to Douglas (1998), clause dependency occurs if learners are not familiar with the reading passage and it is difficult for them.

Todd (2008, p. 4) claims that, "Cloze tests are relatively easy to administer and provide valid and reliable information about a student's language proficiency by deliberate deletion of targeted language features such as verb forms, prepositions and lexical items." She also adds that the instructors could grasp some crucial points about learners' language proficiency through checking his/ her cloze- test score. The following notions could be applied in this regard:

- Control of vocabulary (word choices made correctly or incorrectly, did the student use their own word or choose from the word bank?)
- Grammatical knowledge (consistent use of tense, pronouns, correct choices based on parts of speech, observation of plural, subject verb/ agreement)
- Observance of punctuation
- Accuracy of spelling
- Ability to write neatly, form letters correctly. (p. 8)

Accordingly, based on the prevailing conditions in each test such as response time, the purpose of the test, the participants, their ages and needs, the test makers, the venue of the test, and etc the test might be either power or speed. It seems that in power tests, the response time is not considered crucially; instead, learners' ability to answer all the questions properly is assessed. Power test is also named standard test.

2.1.1 Standard Cloze Test

Among different types of cloze test, standard or settled proportion frame has been broadly researched from the methodological point of view. Standard cloze frame comprises of a content, from which a word is erased after each specific number of words as indicated by a discretionary and settled proportion technique. For instance, each seventh or tenth word is erased after maybe a couple of sentences of the unbroken content. The examinee is required to supply the missing words by deriving from the unique circumstance (Lin, Yuan, & Feng, 2008).

Oller (1973) was most acclaimed for effectively advancing it as a profoundly viable method for measuring a learner's general second/outside dialect capability. He contended that the real content and the real erasure utilized for test development are unessential in light of the fact that tests utilize diverse writings with various levels of trouble or tests and distinctive cancellation rates will be at present rank for examinees in a similar request. In other words, Oller asserted that the supposition about the test-takers' execution invariance over the standard cloze with various writings or distinctive cancellation rates is viable.

Farhady and Keramati (1996) voiced a similar supposition in their study to bolster the proposal of Weaver and Kingston's (1963) and Ohnmacht, Weaver, and Kohler's (1970) that the erasure rate of a cloze test ought to be founded on the quantity of semantic and talk structures of a content (i.e., content driven cancellation technique) and not on a discretionary number (e.g., 5, 7, or 9). The results of their study demonstrated that out of the nine cloze forms, the one in which the deletion rate depended on the quantity of existing noun phrases of the content delivered the best psychometric properties, as far as both criterion related legitimacy and unwavering quality appraisals were concerned. The second best was the one in which the erasure rate depended on the quantity of verb expressions.

2.2 Conceptualization of C- Test

Grotjahn (1986) has suggested three different approaches for the test taking process as statistical item analysis, text linguistic item analysis, and analysis of individual performance. Applying C- test, one could check all the three types of test taking process. C-test is another way of assessing general language proficiency especially in EFL learners. In this regard, Lin, Yuan, and Feng (2008, p. 64) claim 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."

Ajide and Mozaffarzadeh (2012) assert that, the basis behind the C-test is the diminished excess guideline. According to Spolsky (1969) C-tests are the best in the group of diminished repetition tests. These tests are created on the premise of Reduced Redundancy Principle (RRP). Lin, Yuan, and Feng (2008), claim that LRR goes back to the ability to understand an incomplete text and make correct guesses about the missing

information. The more proficient a language learner is, the more successful he/ she will be in guessing the correct choices. In other words, the main purpose of the LRR is to check one's language level and assess his/ her capacity to utilize redundancies by figuring out the incomplete messages.

Khodaday and Hashemi (2011) also believe that local speakers can reestablish absent or misshaped messages by turning to different printed data and making utilization of characteristic excess in the content. Repetition is an idea created as a part of the measurable hypothesis of correspondence. As per this hypothesis, a message conveys data to the degree that it causes a lessening of instabilities in correspondence by killing certain conceivable outcomes. In regular dialect, a larger number of units are utilized than are hypothetically essential i.e. natural dialects are repetitive (Spolsky, 1968).

According to Katona and Dornyei (1993), C- test is an integrative testing technique that measures general dialect capability, particularly like cloze test. It is comprised of four to six short ideally bona fide messages in the objective dialect, to which "the rule of two" has been connected; the second half of the consistent word has been erased, starting with the second expression of the second sentence; the first and last sentences are left in place. In the event that a word has an odd number of letters, the greater part is excluded.

Katona and Dornyei (1993, p. 4) also assert that cloze test could be used for various reasons as placing learners in proper learning groups, determining learners' achievement in midterm and final exams through selected texts gathered from their own practiced materials, checking learners' grammatical knowledge and awareness through structural questions, assigning some related homework and then investigating the possible weaknesses and strengths, and finally measuring learners' specialized knowledge related to their field of study through applying some proper texts in this area.

Saeedi, Tavakoli, Rahimi Kazerooni, and Parvaresh (2011, p. 190) reported that, "there is a statistically significant correlation among participants' scores on Cloze test, C-test". The main difference between cloze test and C-test is that, in the former the whole words are deleted while in the later just some letters are omitted. In other words, cloze test could be regarded as a test of vocabulary and grammar assessment, while C-test could be assumed as a grammar, vocabulary, and dictation assessment.

While holding the idea of disguised sentence structure, the hypothetical method of reasoning behind traditional cloze methodology which is utilized as a part of all language operations, the accompanying criteria for the strategy were included to suit the important changes by Saeedi, Tavakoli, Rahimi Kazerooni, and Parvaresh (2011, p. 191):

- "Much shorter texts should be used to make up at least 100 items;
- No problems should arise in the choice of deletion
- Rate and starting point;
- The deletion should be an absolutely representative sample of the elements the texts;
- The texts should not favor the examinees;
- Only exact scoring should be used to foster score reliability;
- Native speakers are expected to achieve virtually perfect scores;
- The tests should have high reliability and validity."

Additionally, in order to adapt the tests' difficulties for different levels of learners, Koberl and Sigott (1994) have introduced three deletion patterns while Raatz and Braley have also proposed another model of deletion. As a result, C- test is introduced to the community of researchers and instructors as a developed form of cloze test. Based on what

is mentioned above and given the changes which have been implemented in this regard, it seems to be an appropriate way of assessing learners' levels of proficiency.

2.2.1 The Advantages of C-test over Cloze Test

Some scholars believe that cloze test is not fair enough. For instance, Alderson (1979) and Klein-Braley (1985) have enumerated the following problems with a cloze format:

- The test may be unfair because of text specificity
- The factors like text, deletion rate, and starting point can affect both test reliability and validity.
- Tests with high deletion rates have to be extremely long in order to ensure a sufficient number of items.
- The difficulty of the test is affected by the ration of content and structure words deleted from the text. (Cited in Grotjahn, Keith Stevenson, & Klein-Braley, 2009, p. 220)

Accordingly, as C-test is assumed as a modified version of cloze test, many scholars believe that it is preferred over Cloze Test. 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)". A quickly engaging component of C-tests is that they are exceptionally conservative estimation instruments. They are simple both to outline and score and a few unique writings can be utilized to make a total test, which are shorter what's more, contain a bigger number of cancellations than cloze tests. The second favorable position of them is that understudies think that it is less disappointing than cloze tests (Katona & Dorn1993).

The other feature of C-tests is that they permit profoundly objective organization and scoring, and for the most part show high unwavering quality (Eckes and Rudiger, 2006). The other favorable position of C-test over cloze is the utilization of various entries in order to dispense with content specificity and test inclination (Baghaei, 2008).

Braley and Raatz (1984) also assert that C-tests usually appreciate a few favorable cons over cloze tests as the likelihood of setting up a firmer stay point in local speaker utterances, high dependability, simple and target stamping, and the way that the C-tests are irregular and more illustrative examples of writings and language than Cloze tests. Moreover, Katona and Dornyei (1993, p. 2) believe that "C- test appears to have some advantages over the cloze test:

- As students are confronted with a variety of short passages, a better sampling of content areas is possible. Also, a person with special knowledge in a certain field does not have an unfair advantage.
- By damaging every second word, we can obtain a more representative sample of all the different language elements in the text than in the cloze, where normally only every fifth or sixth word is left out.
- Many more items can be included in much shorter texts, making the test less time-consuming for the students than cloze.
- Unlike the cloze test, scoring is easy and objective, as there is only one acceptable solution in most cases.
- As a rule, students actually like doing C-tests, whereas the cloze test is one of the most frustrating test types for learners.

As mentioned above, some of the problems with cloze tests were presented, and some characteristics germane to C-test were introduced and explained. Comparisons between cloze tests and C-tests were mentioned by scholars in the field. At the end, some advantages of c-test were enumerated.

2.2.2 Disadvantages of C- test

In spite of the great cons found in C- test, alert should be practiced when utilizing the C-test as a part of testing local language abilities, as it focuses on perusing and composing aptitudes, the testing of talking and listening abilities, and local fitness is said much to be craved. As such, the topic of what the C-test truly measures has been highlighted. The positive relationships between the C-test scores and instructors, judgment of language skill is suspected as educators may have excessively limited an idea of language capability (Carroll 1987, p. 105).

Another problem with C- test could be claimed as test takers with high perusing understanding capacity may score low on the C-test in view of an absence of gainful abilities in the dialect. Another essential issue is the topic of what C-test really measures which has not yet been settled (Ajideh & Mozaffarzadeh, 2012).

2.2.3 The Rule of Two

According to Klein-Braley & Raatz (1984), in this rule the second half of the consistent word in the content is erased until the required number of mutilations is achieved, leaving the first and last sentence of entry in place to give enough settings. This rule is also called C- principle by Khodadady and Hashemi (2011). As cited in Coleman, Grotjahn, and Raatz (2002, p. 182), "Metalinguistic processes are less effective in children than in adults, and that language proficiency depends on age. So children need additional proficiencies and strategies in C-test taking".

Accordingly, learners and instructors had better be aware of the applied techniques in C- test and also rational behind each one. This way, they could employ it with a higher awareness and appropriate purposes.

2.2.4 Desired Number of Gaps in a C- test Passage

Both cloze test and C- test consist of differing number of gaps to assess learners' awareness and knowledge regarding various points such as parts of speech and structures. Baghaei (2011) deliberately checked and analyzed the psychometric properties of C-Tests with various number of holes. He looked at eight C-Tests which contained 5, 10, 15, 20, 25, 30, 35, and 40 holes in every entry and reasoned that as the quantity of holes in every entry expanded, the interior consistency, unwavering quality, and factorial legitimacy expanded in the same manner.

He also concluded that, as the quantity of gaps in C-Test entries expands the dependability, factorial legitimacy goes up to a specific point in the same manner. However, while the quantity of gaps expanded from 35 to 40 no adjustment in dependability was watched. Besides, the change in unwavering quality in tests with 20 to 35 gaps was little. A similar example was likewise watched for factorial legitimacy. There is an awesome jump in the change of factorial legitimacy from five to 20 crevices yet from 20 gaps ahead the change is practically nothing.

This demonstrates that as the number of crevices in every entry expands we get more data about examinees. Nonetheless, after a specific point the data that additional crevices include get to be distinctly dull and we do not take in more about the respondents. Grotjahn

(1987) also states that a few number of gaps in a C-test reading passage will not lead to a proper text- level macro skills' measurement. He also claims that the range of 25- 30 gaps are required in order to achieve the goal.

Based on studying different papers in this field the research came up with the results that the desired and optimal number of gaps depends on many elements like: interior consistency, unwavering quality, and factorial legitimacy. Moreover, it is of high importance to consider enough items in C-tests because since a few items do not measure the actual learners' macro-skill appropriately.

2.3 Speeded Tests

There are various limitations teachers and learners may face with during the learning, teaching, and testing processes. Time constraint could be named as one of these obstacles. Teaching and learning processes might sometimes be difficult to be covered in a shorter amount of time; however, testing is one of the procedures which could be done in a limited period of time. Meanwhile, power test and speed test are introduced as two opposing definitions for any type of situation.

- Power test: A kind of test which is applied to check the learners' ability and how well they can achieve.
- Speed test: Another version of tests which is employed to determine how fast learners are able to fulfill the tasks.

Generally, in a speed test the whole means and methods are clear. In other words, the questions seem to be almost straightforward. It must be noted that speed test is normally concerned with the number of questions one is able to answer correctly in the considered time. In the opposite manner, a power test does not assume time as a key factor. It rather provides some complex sets of questions and challenge learners' understanding. The methods one should apply in answering these questions are clearly set. Speed tests usually include more items than power tests while they mostly have the same amount of time limitation. Speed tests are almost used in selection at the administrative and clerical level. However, power tests tend to be employed more at the graduate, professional or managerial level. Although, the situation may vary, the results of a speed test could be the mirror of a power test's outcome. In other words, if one does well in speed tests then he/ she will do well in power tests as well.

2.3.1 Speeded C- Test vs. Classical C- Test

As generally 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. According to Grotjahn, Schlak, and Aguado (2010), in the standard version of C- Test time limitation is set for the whole items rather than some specific texts. That is because this test was designed as a power test, not a speeded one.

However, this type of C- test may provide some troubles for learners and teachers. As learners are free to cover the test at any rate of time, some weak ones in time management may not be able to fulfill the tasks thoroughly. They claim 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, in order to compensate for the mentioned defects, a time limitation is considered to make the C- test a speeded one. This way, test reliability and validity will be considerably increased (Grotjahn, Schlak, and

Aguado, 2010). This type of C- test is Called S-C Test (Speeded C- Test). They also claimed that this change led to difficulty, discriminatory power, and reliability increase.

2.4. Empirical Studies on C-test

Developing valid and reliable tests for measuring learners' proficiency in second language is an arduous and costly labor which needs time, money and hard endeavor. Sun Young Lee and Ellis (2009) developed a 30-minute C-test for Korean learners. This Korean C-Test was developed with the specifics of Korean language structure in mind, and Interagency Language Roundtable (ILR) skill-level descriptions were utilized in passage selection in order to test a wide range of participant proficiency levels.

The final outcome was a test and a self-assessment questionnaire (Kondo-Brown, 2005) which were administered to 37 learners of Korean. In order to examine the reliability and concurrent validity of the measure Rasch analysis (Bond & Fox, 2007) was used, and Rasch measurement statistics such as separation reliability, difficulty measures, and model fit statistics were run to suggest further improvement to the C-Test. The reliability and validity indices of the developed C-test were proved to be very excellent, and the results revealed the potential of the C-Test as a quick proficiency indicator.

Moreover, Klein-Braley and Raatz (1984) developed a C-test which functioned as a proficiency test for the following groups: 1- for children who learn their own language (L1), 2- for children and adults who learn an L2 in a country where that language was spoken, and 3- for EFL learners; those who learn a language as the language which is not spoken in their country. In their paper, they pointed to some strategies and techniques for constructing successful C-tests.

In another study conducted by Klein-Braley (2002), it is concluded that language operations must take place on several different levels in order to restore the mutilated text to its original form. He came up with the results that less proficient learners who worked on lower level texts had a better performance on higher level texts in the future.

In a recent study, Baghaei and Tabatabaee (2015) argued that the basic construct for the C-test, as a general proficiency test, is in harmony with the abilities underlying the language component of crystallized intelligence, as defined in the well-established theories of intelligence. They also recommended that by carefully selecting texts from pertinent knowledge domains, the factual knowledge component of crystallized intelligence could also be measured by the C-test.

In a different study, Lin, Yuan and Feng (2008) compared the psychometric properties of cloze test and C-test, both of which belong to the family of language reduced redundancy test. Based on two authentic texts with different rhetoric modes, three cloze forms with text-driven deletion method, along with one standard cloze and one form of C-test, were constructed and randomly administered to 237 student subjects at one private university in northern Taiwan. They concluded that neither the three cloze forms nor the C-test was substantially superior to the standard cloze in terms of reliability and validity.

The next research question they investigated to answer was to test whether the assumption of test-takers' performance invariance across different texts hold for both cloze test and C-test. They came up with the results that no proof was detected to support the claim that both cloze test and C-test meet the assumption of test-takers' performance invariance across different texts.

Moreover, Baghaei (2011) investigated the optimal number of gaps in C-test passages. He used four passages out of a C-test English battery each of which had 40 blanks. The participants of his study were 104 undergraduate English students. The results of his study

revealed that the more gaps in each passage, the better item discrimination, the higher reliability, and more factorial validity are achieved.

Furthermore, Ajideh and Mozaffarzadeh (2012) compared the results of multiple-choice cloze test with those of C-test as measures of reading comprehension. They selected 27 female English learners in advanced level and gave them one traditional C-test and one fixed ratio (n=7) multiple-choice cloze test. The results of the study showed that multiple-choice cloze is a better measure of reading comprehension in comparison to C-test.

In a recent study, sixty Iranian female junior university students were selected by Azimi (2016) to investigate the possible relationship between the kinds of test given to the students. She gave three types of tests to her participants: an anxiety test, a C-test and a Cloze-test. She gathered and analyzed data by running correlations between cloze test and anxiety and C-test and anxiety. Finally, she came up with the results that participants felt more anxious when they took C-test in comparison with taking cloze test.

Moreover, Karimi (2011) gave 151 EFL learners four types of C-tests, two of which were 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 the participants in two groups of (74 and 71 members) on these tests was examined. The only difference between the C-tests was in their lexical characteristics. The findings of her study revealed that participants had significantly different performances in the two circumstances. Thus, she concluded 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.

Also, Saeedi, Tavakoli, Rahimi Kazerooni, and Parvaresh (2011) conducted a study to measure the validity of C-test and Cloze test. They concentrated on criterion-related validity based on the results of the integrative language tests. They selected 90 EFL participants who were senior students in universities. Their findings revealed that there is a statistically significant correlation among participants' scores on cloze test, C-test, and the TOEFL test.

Salehi and Bagheri Sanjareh (2013) explored and compared the extent to which the C-test and cloze test, as measures of the reduced redundancy principle, tap macro-level and micro-level strategies. The results of the analyses showed that C-test and cloze test trigger both types of macro-level and micro-level strategies although the former strategies were less elicited compared to the latter strategies.

Many attempts were made to construct reliable and valid tests for measuring learners' proficiency, among them C-test was of high significance. Some scholars devised different types of C-test and piloted them, some of them compared and contrasted C-test with other types of tests like Cloze and enumerated some advantages and disadvantages for each of these types. The main purpose of this chapter was to provide a comprehensive background knowledge in the field of language tests in general and C-test in particular.

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 sample members. It must be noted that participants were all non-native speakers of English whose first language was Persian, studying English during the autumn semester the Islamic Azad University and Tabaran University of Mashhad, Khorasan Razavi, Iran. Moreover, they ranged between 20 and 35 in age. They were also assured that their personal information would be confidential. The researcher appreciated them for their cooperation.

In this study, the participants' performances on Standard and speeded C-tests were compared to check 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 asked to answer completely in one session. After one week, participants were given a standard C-test as well. It should be justified that the time gap is chosen due to the time limitation as all experimental processes have to be done in one semester.

3.2 Instrumentation

3.2.1 Standard C-test

In order to check the students' general language proficiency, a normal 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 questions and were presented under two topic: chocolate and money. Besides, the texts were driven out of Reading skillfully1, written by Akbar Mirhassani and Navid Rahmani. The testees were assumed to answer the questions all in 10 minutes.

3.2.2 Speeded C-test

Speeded C-test has the same nature as standard C-test. The only difference is the time allowed for each one. As the name justifies, speeded C-test is administered in a shorter period of time; so that learners had only five minutes to cover the whole test. This test consisted of two short reading passages with 25 gaps in each. The two passages were titled as Birthday and Hollywood.

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 learners in order to answer this section .Besides, the reliability of these tests were calculated by Cronbach's alpha (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).

3.3 Procedure

In this study, 100 Iranian undergraduate EFL learners studying at Islamic Azad University and Tabaran University of Mashhad, Khorasan Razavi, Iran, were selected randomly to take part in this research. The learners were both males and females. As mentioned before, the three tests were given to students in two distinct processes. In order to start, a speeded C-test along with a reading comprehension test were administered. Five minutes were devoted to the first test and 25 minutes to the second one.

In the second phase, after two weeks the same participants took a standard C-test. The data were compared with the ones obtained from the first phase. Finally, the data were analyzed by using SPSS software 22.0 to investigate the outcome of comparing the psychometric characteristics of speeded and standard C-test.

When the required data was obtained from both phases, the data were analyzed to find out if there is any significant relationship between speeded C- Test and Standard C- test with reading comprehension.

3.4 Research Design and analysis

In order to conduct this research, the researcher employed a correlational design with a quantitative approach. Two forms of C-tests were considered as the independent variables while reading comprehension was assumed as the dependent one in this study. Intact classes were assigned randomly to collect the required number of participants.

29

4. Analyses and results

4.1 Comparing psychometric characteristics of speeded and standard C-test

To test the research hypotheses of this study concerning the psychometric characteristics of speeded and standard C-tests and their prediction of reading comprehension the researcher opted for Pearson correlation coefficient, multiple regression and factor analysis and determine which C-test is a better predictor of reading comprehension.

4.2 Restatement of hypotheses

The researcher proposed the following null hypotheses:

1. There is no significant relationship between speeded C-test and reading comprehension.
2. There is no significant relationship between standard C-test and reading comprehension.
3. None of the C-tests (speeded C- test or standard C- test) is a better predictor of reading comprehension.
4. There is no difference in the reliability of two types of C- test.

4.3 Interpretation of the results

In Table 4.1 presented below, the information we need for each of the variables is summarized. Table 4.1, 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 is 6.99 out of 20. It showed that this test is difficult for students to answer and also the mean in speeded C-test is 15.43 out of 50 and the mean in power test is 20.14 out of 50. The means in speeded C-test and standard C-test were low. Therefore, it showed that these two tests were also difficult for students to answer but students were more successful in answering the power C-test.

Table 4.1: Descriptive Statistics

	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

4.4 Correlation analysis

The results obtained from Pearson Product-moment Correlation Coefficient showed 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 4.2). The relationship between speed C-test and power C-test [$r=.46$, $n=100$, $p<.05$] was also statistically significant. The results show 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 so we can reject Hypotheses 1 and 2. The relationship between power C-test and reading comprehension is higher than the correlation between speed C-test and reading comprehension.

Therefore, the power C -test is a better predictor of reading comprehension. The reliability of speeded C-test and standard C-test were calculated by Cronbach's alpha (0.70). The reliability of speeded C-test was .41 and standard C-test was .59 so standard C-test was more reliable than speeded C-test and hypothesis 4 was rejected by this result.

Table 4.2: *Correlations between the tests*

	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)

4.5 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 no violation of the assumptions of normality, linearity, multicollinearity, and homoscedasticity.

Results showed 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 so the hypothesis 3 was rejected.

The adjusted R^2 was .53 which means that these four independent variables explained 53% percent variance in the reading comprehension scores.

Table 4.3: *Multiple regression analysis for predicting reading comprehension using speeded and standard C-test passages*

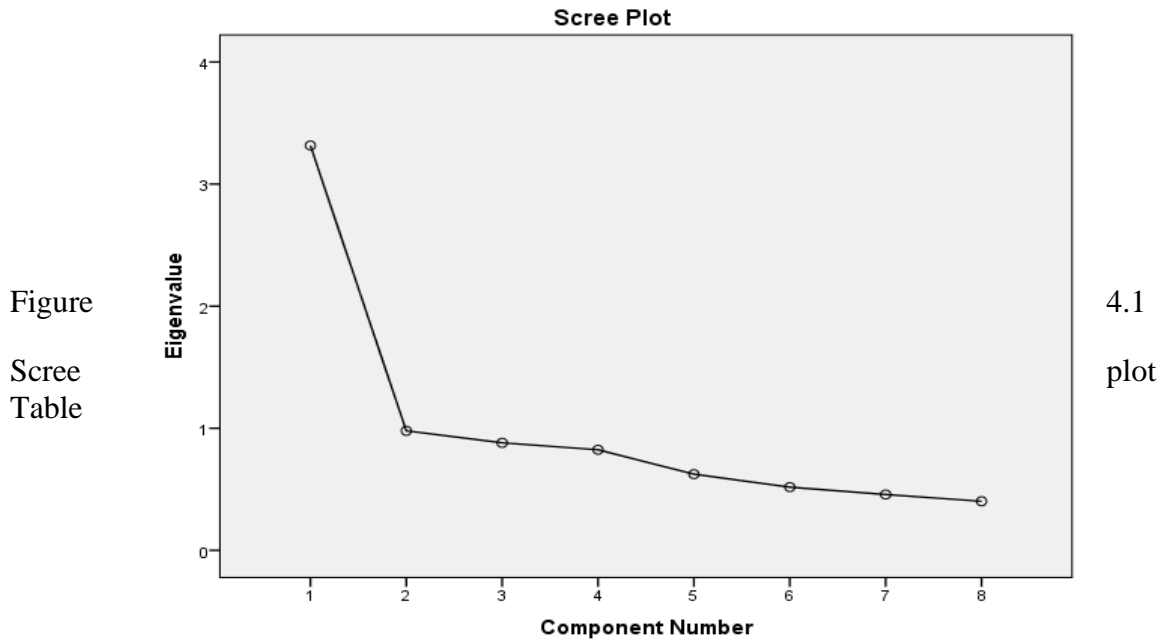
Independent variable	Beta	T	P	Part correlation
		-.04	.96	
S.Test1	.38	4.4	.00	.30
S.Test2	.06	.82	.41	.05
Standard1	.32	3.96	.00	.27
Standard2	.27	3.39	.00	.23

4.6 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-Olkin value was .83, exceeding the recommended value of 0.60 (Kaiser 1970, 1974) and Bartlett's test of

Sphericity (Bartlett 1954) reached statistical significance, ($p=0.00$) supporting the factorability of the correlation matrix.

Principal components analysis showed the presence of one component with eigenvalue exceeding 1, explaining 41%, of the variance. An inspection of the screeplot revealed one clear break after component one.



4.4 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 .40 and we can name the factor as reading comprehension factor 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 the 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 4.4: *Factor loadings for the variables of the study*

	Component
	1
S.Test1	.684
S.Test2	.496
Standard1	.743
Standard2	.705
text1	.405
text2	.590
text3	.678
text4	.762

4.7 Results and Discussions

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-Yazdi, 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 composite scores of various language tests is also needed.

5. Conclusions

5.1 Summary of Findings

The goal of this study was to determine the differences between standard C-test and speeded one and then select the prior one for checking learners' language proficiency and improving language learning and teaching in general. Besides, the researcher intended to increase all teachers' awareness regarding the great usability of different types of C-test in assessing learners' levels or general proficiency.

In order to address the first research question SPSS software was applied and data analysis displayed a significant relationship between speeded C-test and reading comprehension.

Obtained results also showed that there is a highly significant relationship between standard C-test and reading comprehension. Additionally, in order to determine the most appropriate type of C-test, regression analysis was run. According to this operation, both types of C-tests could measure reading comprehension ability but standard C-test proved to be a better predictor of reading comprehension.

To answer the fourth research question regarding the C- test reliability, statistics showed that the reliability of standard C-test was more than speeded C-test. In other word, standard C-test measures what it is expected to measure.

The fifth research question considered the structure of speeded C-test, standard C-test and reading comprehension test as making a factor. In this regard, it was found that these three tests loaded on one factor so it could be concluded that speeded C-test and standard C-test are proper means of measurement for reading comprehension. Additionally, standard C-test had a higher loading than speeded C-test on the single extracted factor from exploratory factor analysis. Therefore, it was concluded that standard C-test is a better measure of learners' reading comprehension ability.

5.2 Conclusion

The aim of this study was to compare the psychometric characteristics of speeded and standard C-test. Thus, 100 undergraduate English students from Islamic Azad University and Tabaran Institute of Higher Education in Mashhad, Khorasan Razavi, Iran, participated in this study. The results showed that both speeded C-test and standard C-test could measure reading comprehension ability; however, standard C-test could be regarded as a better predictor for this goal. Moreover, there was not any significant difference between speeded C-test and standard C-test, therefore teachers had better employ standard C-test if there is no time limitation.

In case of time constraints, employing speeded C-test is highly suggested 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 level.

5.3 Suggestions for Further Studies Research

In order to fill in the possible gaps in this area of research the following suggestions for future researches are recommended:

- At first, in the present study, age and gender of participants were not taken into account. Moreover, EFL learners were selected randomly and they had different levels of language proficiency. Consequently, further studies are needed to take the possible effects of these variables into account.
- Second, in this study the researcher only investigated the relationships between speeded C-test, standard C-test, and reading comprehension ability of Iranian EFL learners. It is engrossing to conduct a study investigating the relationship between speeded C-test, standard C-test and other language skills.
- Next, the context of this study was chosen out of only two Iranian universities in Mashhad; it is recommended that the replication of this study be done in other universities, schools, and language institutes as well, even with a larger population.
- Finally, other means of assessing language proficiency could be checked and compared as well. As teachers are responsible to be dynamic, they ought to be up to date and aware of the best methods for any types of activities.

5.4 Pedagogical Implications

The implications of the study are threefold. On the one hand, it could be assumed beneficial for novice teachers to become more familiar with various types of measuring reading comprehension and general language proficiency. In addition, more explanations were provided for each type of C- test in order to make the topic easier and more understandable for all teachers and teaching learners.

On the other hand, the findings of this study could be quite fruitful for further evaluative assessments in this area. Besides, it was proved that time constraints and personal point of view could be determining factors in choosing any type of C- tests in measuring general language proficiency or reading comprehension. In other words, any teacher should be assessed based on the situation and setting in which evaluation occurs.

Additionally, curriculum writers also benefit from this study as they could provide more proper materials based on learners' needs and levels. In this regard, if students are able to understand the content with a greater speed, curriculum writers could provide a greater volume of materials in the same time constraint.

Moreover, learners are also advantaged in the way that they will be divided to different groups based on their abilities. This way, learners will be more interested in learning new points and each group will receive the amount of knowledge they need.

REFERENCES

- Ajideh, P., & Mozaffarzadeh, S. (2012). C-test vs. Multiple-choice Cloze Test as Tests of Reading Comprehension in Iranian EFL Context: Learners' Perspective. *Canadian Center of Science and Education*, 143, 5(11), doi:10.5539/elt.v5n11p143
- Alderson, J. C. (1979). "The cloze procedure and proficiency in English as a foreign language." *TESOL Quarterly*, 13(2), 219-227.
- Aguado, K., Grotjahn, G., & Schlak, T. (2005). Erwerbssalter und Sprachlernerfolg-Theoretischen und methodologische Grudnlagen siens empirischen Forschungsprojekts. *Zeitschrift für Fremdsprachenforschung*, 16(2), 275-294.
- Azimi, M. (2016). The Relationship between Anxiety and Test-Taking C-Test and Cloze-Test. *Malaysian Online Journal of Educational Sciences*, 4(1), 30-42.
- Baghaei, P. (2008a). The effects of the rhetorical organization of texts on the C-Test construct: A Rasch modelling 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 aktuellen Forschung/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 Verlag Dr 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.

- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bormuth, J. R. (1969). Factor Validity of Cloze Tests as Measures of Reading Comprehension Ability. *Reading Research Quarterly*, 4(3), 358-365. doi:10.2307/747144
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL quarterly*, 32(4) 653-675.
- Carroll, J. B. (1964). Language and thought. *Reading Improvement*, 2(1), 80.
- Carroll, J.B. (1987). Review of Klein-Braley and Raatz. C-tests in der praxis. *Language Testing*, 4, 99-106.
- Coleman, J. A., Grotjahn, R., & Raatz, U. (2002). *University language learning and the C-Test*. AKS-Verlag.
- Douglas, D. (1998). In L. F. Bachman & A. D Cohen (Eds.), *Interfaces between second language acquisition and language testing research*, Cambridge University Press, 141–153.
- Eckes, T., & Baghaei, P. (2015). Using testlet response theory to examine local dependency in C-Tests. *Applied Measurement in Education*, 28, 85–98.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23(3), 290-325.
- Ellis, S. (2009). The development and validation of a Korean C-Test using Rasch Analysis. *Language Testing*, 26(2), 245-274.
- Farhady, H., & Keramati, M. N. (1996). A text-driven method for the deletion procedure in cloze passages. *Language Testing*, 13, 191-207.
- Fulcher, G. (2015). *Re-examining language testing: A philosophical and social inquiry*. Routledge.
- Grotjahn, R. (1986). Test validation and cognitive psychology: some methodological considerations. *Language Testing*, 3, 159- 185.
- Grotjahn, R. (1987). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In R. Grotjahn, C. Klein-Braley & D.K. Stevenson (Eds.), *taking their measure: The validity and validation of language tests*, 219-253. Bochum: Brockmeyer.
- Grotjahn, Rüdiger, Klein-Braley, Christine & Raatz, Ulrich. (2002). C-Tests: An overview. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Hrsg.), *University language testing and the C-Test*, 93–114. Bochum: AKS-Verlag.
- Hinofotis, F. B. (1980a). Cloze as an alternative method of ESL placement and proficiency testing. In J. W. Oller, Jr. & K. Perkins (Eds.), *Research in Language Testing*, 121-128. Rowley, MA: Newbury House.
- Hulstijn, J. H. (2001). *Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity* (pp. 258-286).
- Johnson, K. (1996). *Language Teaching and Skill Development*. Oxford: Blackwell.
- Karimi, N. (2011). C-test and Vocabulary Knowledge. *Language Testing in Asia*, 1(4), 7.
- Katona, L., & Dornyei, Z. (1993). The C- Test: A Teacher Friendly Way to Test Language Proficiency. *Forum*, 31(2), 35.
- Khodadady, E., & Hashemi, M. (2011). Validity and C-Tests: The role of text authenticity. *Iranian Journal of Language Testing*, 1(1), 30-41.
- 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. (1985). A cloze-up on the C-Test: a study in the construct validation of authentic tests. *Language Testing*, 2(1), 76-104.
- Klein-Braley, C. (1996). Towards a theory of C-Test processing. In Rudiger Grotjahn, editor, "Der C-Test. *Theoretische Grundlagen und praktische Anwendungen* 3, 23–94. Brockmeyer, Bochum.
- Klein-Braley, C. & Raatz, E. (1984). A survey on the C test-1. *Language Testing*, 1 (2), 134-146.
- Kondo-Brown, K. (2005). Differences in language skills: Heritage language learner subgroups and foreign language learners. *The Modern Language Journal*, 89, 563-581.
- Köberl, J., & Sigott, G. (1994). Adjusting C-test difficulty in German. *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*, 2, 179-192.
- Lin, W., Yuan, H., & Feng, H. (2008). Language reduced redundancy tests: A reexamination of cloze test and C-test. *Journal of Pan-Pacific Association of Applied Linguistics*, 12, 61-79.
- Ohnmacht, F. W., Weaver, W. W., & Kohler, E. T. (1970). Cloze and closure: A factorial study. *The Journal of Psychology*, 74(2), 205-217.
- 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. (1982). The C-Test-A modification of the cloze procedure. In T. Culhane, C. Klein-Braley, & D.K. Stevenson (Eds.), *Practice and Problem in Language Testing IV. Proceeding of the Fourth International Language Testing Symposium of the Interuniversitare Sprachtestgruppe* (pp. 113-138). Colchester: University of Essex press.
- Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. *Cognition and second language instruction*, 288.
- Salehi, M., & Bagheri Sanjareh, H. (2013). On the comparability of C-test and cloze test: A verbal protocol approach. *English for Specific Purposes World*, 14, 39.
- Saeedi, M., Tavakoli, M., Rahimi Kazerooni, S., & Parvaresh, V. (2011). Do c-test and cloze procedure measure what they purport to be measuring?: A case of criterion-related validity. *International Journal of Human and Social Sciences*, 6, 2-99.
- Sasaki, M. (2000). Effects of cultural schemata on students' test taking processes for cloze tests: a multiple data source approach. *Language Testing*, 17(1), 85–114.
- Segalowitz, N., Trofimovich, P., Gattbonton, E., & Sokolovskaya, A. (2008). Feeling affect in a second language: The role of word recognition automaticity. *The Mental Lexicon*, 3(1), 47-71.
- Saeedi, M., Tavakoli, M., Rahimi Kazerooni, S., & Parvaresh, V. (2011). Do C-Test and Cloze Procedure Measure what they Purport to be Measuring? A Case of Criterion-Related Validity. *World Academy of Science, Engineering and Technology International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 5(2), 190- 199.
- 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 South California.
- Spolsky, B. (1969). *Reduced redundancy as a language testing tool*. Paper presented to the Language Testing section of the second International Congress of Applied Linguistics, Cambridge, England.

Weaver, W. W., & Kingston, A. J. (1963). A factor analysis of the cloze procedure and other measures of reading and language ability. *Journal of Communication, 13*(4), 252-261.

Appendixes

APPENDIX I

The speeded C-test

Birthday

Many children have a birthday cake with candles on their birthday. In so..... countries, li.....England a..... Scotland, th..... is ano..... custom, t..... There peo..... spank o..... hit t..... child o..... his/her birt..... This m..... hurt a lit....., but th..... say i..... is ve..... lucky f..... the ch..... The ch..... must ne..... cry. T..... custom sa..... that i..... you c....., you wi.....cry all year.

The reason for birthday spank is to make the bad spirits go away. The harder you spank the better it is.

Laughter

Some people say that laughter is the best medicine. Scientists a..... beginning t..... agree wi..... this. Th..... are stud..... laughter seri..... and a..... finding i.....is rea..... good f.....us.

S..... what hap..... when w..... laugh? We..... use fif..... different mus..... in o..... face, a..... laughing i..... good f.....every or..... in o..... body. Wh..... we la....., we bre..... quickly and exercise the face, shoulders, and chest. Our blood pressure goes down, and our circulation gets better.

APPENDIX II

The standard C-test

Money

When we think of money, we think of coins and paper bills. That i..... what mo..... is to..... But i..... the pa..... people us..... many thi..... in pl..... of mo..... . Some coun..... used co....., other coun..... used sa....., tobacco, t....., or sto..... . Today th..... are st..... some pla..... in t..... world th..... do n..... use pa..... money. O..... place i..... the isl..... of yap in the Pacific Ocean. On the island of Yap, people use the heaviest money in the world-Yap stones.

Right brain or left brain?

What do Leonardo da Vinci, Paul McCartney, Napoleon, and John McEnroe have in common? They were all left-handed. Today about 15 percent of the population is left-handed. But why are people left-handed? The answer is in the way the brain works.

The brain has two halves—the right hemisphere and the left hemisphere. The right half controls the left side of the body, and the left half controls the right side of the body. So right-handed people have a strong left brain, and left-handed people have a strong right brain.