Recommended app

Designing and assessing a digital, discipline-specific literacy assessment tool

Paul Graham Kebble
National Institute of Education Nanyang Technological University, Singapore
_____
paul.kebble@nie.edu.sg

Abstract

The C-Test as a tool for assessing language competence has been in existence for nearly 40 years, having been designed by Professors Klein-Braley and Raatz for implementation in German and English. Much research has been conducted over the ensuing years, particularly in regards to reliability and construct validity, for which it is reported to perform reliably and in multiple languages. The author engaged in C-Test research in 1995 focusing on concurrent, predictive and face validity. Through this research, the author developed an appreciation for the C-Test assessment process particularly with the multiple cognitive and linguistic test-taking strategies required. When digital technologies became accessible, versatile and societally integrated, the author believed the C-Test would function well in this environment. This conviction prompted a series of investigations into the development and assessment of a digital C-Test design to be utilised in multiple linguistic settings. This paper describes the protracted design process, concluding with the publication of mobile apps.

Keywords: C-Test, language competence assessment, mobile app design.

1. Introduction

Over the past eight years, I have been involved in the design and implementation and evaluation of a series of digital language competency assessment tools utilising a unique testing format, the C-Test. The C-Test construct (Raatz, 1985a; Raatz & Klein-Braley, 1985; 2002) allows for the design of a reliable and versatile (Raatz, 1987; Rouhani, 2008) subject-specific language competency assessment tool (Alderson, 2002; Grotjahn, 1987; Grotjahn & Stemmer, 2002; Hulstijn, 2010) through the utilisation of subject-related texts. A digital C-Test, I conjectured, could be constructed to provide a linguistic assessment tool relevant to any academic or professional discipline, and highly efficient in implementation. Output of such a test would allow assessors to identify takers' levels of linguistic competency and indicate those who might require linguistic support. For example, a lecturer teaching into a 1st year Electrical Engineering subject could use the online C-Test, utilising texts selected from the 1st year Electrical Engineering program, to identify students who require subject-specific language support. An international finance company recruiting staff would be able to use a C-Test designed to determine linguistic suitability of potential employees. Equally, the C-Test could utilise texts from a year 5 syllabus and could be administered on a regular basis (bi-monthly), providing a teacher with regular assessment of language and literacy levels of her students, and hence an individual's development. My aim therefore was to design a software application which would efficiently and effectively build the test through the uploading of discipline-specific texts, and assess, collate, analyse and distribute results as to pre-specified requirements.

2. The C-Test

The C-Test construct was originally designed by (the late) Professor Christine Klein-Braley and Professor Ulrich Raatz, both of the University of Duisburg, in the 1980s and was initially utilised for assessing both German (Klein-Braley, 1985c) and English language competency. During this period, professors Klein-Braley and Raatz conducted

extensive research into the validity and reliability of the C-Test, along with many other academics and researchers, such as Professor Grotjahn and Professor Coleman, (Klein-Braley & Raatz, 1984) and found the C-Test was a highly reliable assessor of linguistic competence. In 1996, I was very fortunate to meet both Professors at a language testing conference held at the University of Portsmouth (UK) Subsequently, Prof. Raatz has kindly given full permission for the C-Test construct to be used for the basis of my online design.

The C-Test assesses linguistic competence through reduced redundancy (Klein-Braley, 1985d, 1985e, 1997; Oscarson, 1991; Raatz, 1985b, c) and the accurate restoration (Babaii & Fatahi-Majd, 2014) of a text, where interference in communication is achieved through systematic word mutilation. The C-Test is a derivation of a cloze test where every nth word, usually the 7th, is removed, however, the C-Test textual restoration functions at word, not sentence level, and is described as the 'rule of two' (Jafarpur, 1999). The C-Test is built utilising short texts, usually paragraphs taken from authentic sources (Atai & Soleimany, 2009; Khodadady, 2013; Khodadady & Hashemi, 2011; Klein-Braley, 1985a), with paragraphs obtained from either a variety of genres or one specific genre, depending on the assessment requirements (Mochizuki, A., 1994). From the beginning of the second sentence of each paragraph, the second half of every second word is removed (the rule of two), until 25 deletions are achieved, with the remaining text left intact. A mutilated text looks like this:

> *Given the continuing changes in the application of Information and Communication Technology (ICT) this survey focuses on broadband access and types and penetration rates of electronic commerce (e-commerce; the online sale of goods and services). Despite t __ growing le __ of pub __ debate conce ___ the gro ___ of e-bus ___ and e-com ___ within t __ Australian eco ___, very lit ___ data rela ___ to t ___ business u __ of th ___ facilities h __ been avai ___ to pol ___ makers o __ private enter ___. Data th ___ are avai ___ have predom ____ been gene ___ by pri ___ surveys wi ___ small samples and inconsistent definitions and scope. At best, the results of these surveys are indicative. There is a pressing need for the production of timely and comprehensive e-commerce statistics.*

The original text:

> *Given the continuing changes in the application of Information and Communication Technology (ICT) this survey focuses on broadband access and types and penetration rates of electronic commerce (e-commerce; the online sale of goods and services). Despite the growing level of public debate concerning the growth of e-business and e-commerce within the Australian economy, very little data relating to the business use of these facilities has been available to policy makers or private enterprises. Data that are available have predominantly been generated by private surveys with small samples and inconsistent definitions and scope. At best, the results of these surveys are indicative. There is a pressing need for the production of timely and comprehensive e-commerce statistics.*

Micro- and macro-level textual cues, along with anaphoric and cataphoric referencing, are specific linguistic skills required for restoration, the more accurate the restoration, the more proficient the restorer (Klein-Braley, 1996; Wedell, 1987). Micro-level textual cues refer to an individual word's construct both in terms of word composition and form, whereas Macro-level cues require understanding of syntax and function. Anaphoric referencing necessitates the test taker to refer back within the text, and cataphoric referencing, forward. With the first sentence and the remaining text left intact, the test taker is also able to contextualise the extract, providing further clues for its restoration.

Over the past 30 years, extensive research has been conducted into the C-Test procedure, its reliability and validity (Grotjahn, 1986; Khodadady, 2014; Klein-Braley, 1985b), and its function in a variety of languages. A link to a comprehensive bibliography of research is provided by Professor Rudiger Grotjahn (Grotjahn, 2014), at Ruhr-Universität Bochum:

http://www.C-Test.de/deutsch/index.php?lang=de&content=bibliografie&section=ctest

3. Initial involvement with the C-Test

I first became aware of the C-Test through Master's research conducted whilst at the University of Portsmouth, U.K. (1995-6). For my thesis, I was invited by Professors Rastall and Coleman, both at the University of Portsmouth, to assess the face, concurrent and predictive validity of the C-Test, at the time a paper test being used as the university's language assessment tool for international students entering undergraduate courses. Much research had already been conducted into the C-Test's use in higher education (Coleman, 1994a/b; Coleman, Grotjahn, Rüdiger & Raatz, 2002) with its specific use as a placement test being the focus for my research. With access to students' academic records, I was able to determine that the C-Test had no predictive capability. In terms of concurrent validity, I measured the C-Test against a well-validated commercially available test (the Oxford Placement Test) through the testing of 83 students with both. A strong correlation coefficient (Pearson's r) was seen (0.8) allowing my research to suggest the C-Test was a valid test of general language proficiency. This research also indicated that the C-Test suffered from poor face validity, through collated and analysed answers to a questionnaire, analysis showed test takers were not convinced the C-Test functioned as its purpose was described.

With the advent of the internet, I conjectured the C-Test design structure could be highly functional in, and would be especially suited to, a digital environment. My initial attempts at providing the test digitally were limited primarily to attempting a build within the Moodle and Blackboard online learning platforms. And the first time I was able to effectively use such a build was in 2011.

4. Developing the digital C-Test

In 2011, I created an Online C-Test within the Blackboard learning platform when required by my Australian university (James Cook University) to assess the academic language proficiency of a large group (93) of Chinese undergraduate I.T. students studying at a Beijing university on a course provided by my university and in English medium. Through an external quality audit by the Australian Universities Quality Agency (AUQA), concern had been raised at the level of English demonstrated by some students, and my university wished to assess English language levels. Although students were required to have reached an IELTS overall 5 score, or its equivalent, to enter the course, concern had been raised that a large proportion of the 1st year cohort were not functioning acceptably in the learning and teaching medium of English. I was required to design and provide an English language competency test that could be administered efficiently, and did so using a C-Test construct in Blackboard, with four texts sourced from the 1st year IT syllabus.
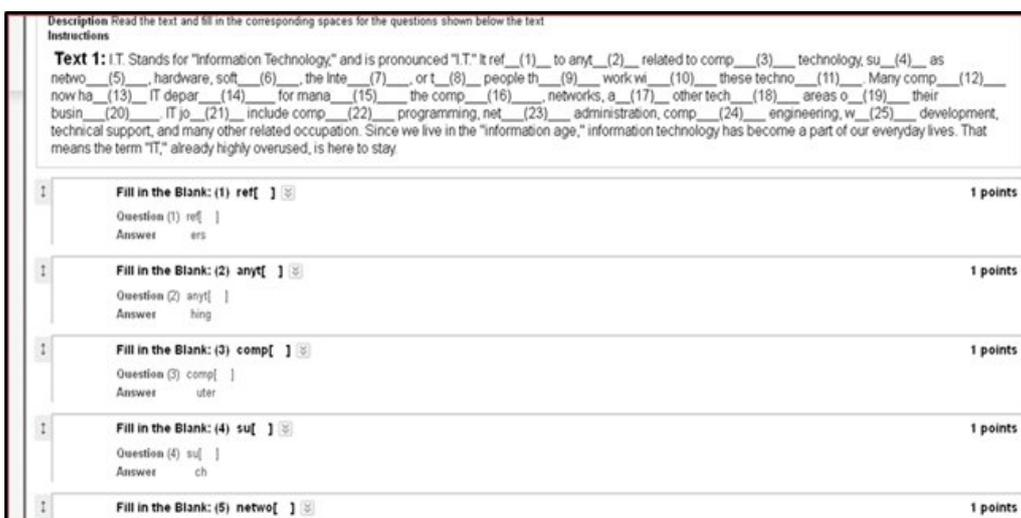
Figure 1: C-Test build in Blackboard.

As a benchmark, the same C-Test was administered to a group of eight English language students in a private language college in Cairns, Australia. The Cairns students

had IELTS scores ranging from 5 to 6.5. The range of scores from these students on the same C-Test were from 43 to 78. The C-Test results were aligned to the IELTS scores resulting in an informed approximation that scores on the C-Test >40 = IELTS 5, >50 = 5.5, >60 = 6, >70 = 6.5. When this alignment was applied to the results of the Chinese students' C-Test results the indication was that students scoring below 40 might be identified as being at risk of failing to achieve an IELTS 6 on the end-of-year IELTS equivalency test. The results indicated that this amounted to 36 (39%) students out of the total cohort.
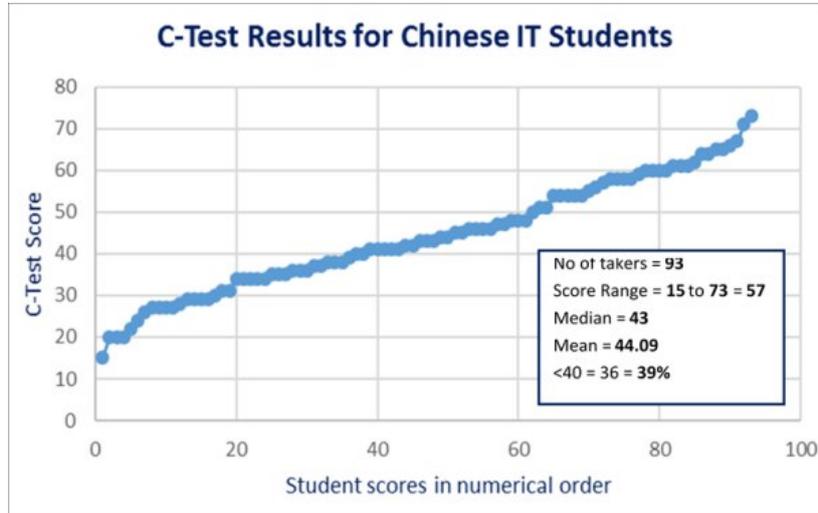

Figure 2: C-Test scores for Chinese IT students.

Although the testing process functioned well, and provided invaluable information, I was dissatisfied with the online test build in Blackboard as the user interaction was not as effective as a paper-based test would have been. Most importantly, with the Blackboard build, takers were unable to complete each word in situ (see diagram 1). Of course, there were other advantages, not using paper being one, and the automatic collation of test results and scores being the other. On returning to my university, I enquired of the I.T. department whether the C-Test could be built in another platform allowing user interaction to match a paper version. I was told it could be done, but to purchase the software and employ a programmer to build the required code would be inordinately expensive.

A move to the University of Tasmania, and meeting D. Heidermann, a senior IT manager, provided expertise in programming, and an interest in the concept. Our initial C-Test build was online and allowed paragraphs of text to be uploaded, multiple tests taken, and results to be collated and distributed. Soon after completing this package, we considered designing a mobile app that would assess language competency and instantly provide the taker with a result and related feedback. Over the ensuing year, we were able to develop and publish three discipline-specific language and literacy assessment apps, initially with Apple, and more recently as an Android app.

5. The 'IELTS Score Predictor' App

My initial design concept was for a mobile app which would provide a predicted overall score for the International English Language Testing System (IELTS). The IELTS examination is a highly reliable and very well validated English language assessment tool, providing a score for each of the macro language skills, reading writing, speaking and listening, and an averaged overall score ranging from 2 to 9, with 0.5 graduations. Undergraduate university courses usually require a 6 or 6.5 overall score, with postgraduate studies requiring 7 or 7.5. My concept was to offer a tool that would provide an overall indicator of an averaged IELTS score, along with a linguistic description of what the achieved score aligned to in terms of functionality in English. I was aware of the limitations of not providing scores for each linguistic skill, but was convinced that receiving an indication of an overall score was highly beneficial.

The app would utilise paragraphs from texts used in past IELTS papers, freely available online. Informed through the previously described use of the C-Test for achieving this outcome, I worked with D. Heidermann to design and build the software, create the C-Test paragraphs and formulate a rubric to which each score would align.

Originally, the C-Test design used four texts, however, for the IELTS predictor app I decided to use 6. Four texts would be sourced from various sections of IELTS reading tests adapted from past papers; one text would be created from a transcript of an item in the listening section; and the final text would be an imagined dialogue between a test taker and an IELTS interlocutor, achieved through many years of experience teaching and examining IELTS takers. For score alignment, the test was trialled at the Cairns language school, Queensland, Australia, (Cairns College of English) with results employed to create a table of score equivalency, shown in Table 1.

Table 1. C-Test and IELTS score equivalences													
Band Score	9	8.5	8	7.5	7	6.5	6	5.5	5	4.5	4	3.5	3	2.5
% Score	91-100	86-90.	80-85	75-79	70-74	64-69	57-63	47-56	37-46	32-36	25-31	20-24	15-19	10-14

Cognisant of my earlier research into face validity, I decided to provide test takers with a concise introduction to the C-Test construct and development, with a link to the bibliography of related research. Clear instructions were also provided to test takers prior to engagement with the C-Test, advising: 1) Read the first sentence carefully and think about the topic of the text, 2) Look at the first mutilated word, in the example above 'i___' and the word before and after. If you can repair the word from these clues, do so, 3) Sometimes you will need to think back to the first sentence, look back in the sentence, and look forward through the sentence, 4) read back through the completed sentences to check if the words sound or appear suitable.

Although the results were graduated at 0.5, the language level descriptors were provided for each integer. If a score was at 0.5, the taker would be required to engage with both higher and lower integer descriptors. Descriptors (Table 2) were written with reference to the British Council's public reference band score descriptors (http://takeielts.britishcouncil.org/find-out-about-results/ielts-assessment-criteria).

Table 2. Example of IELTS score descriptors			
7 **Good user**	You have an operational command of the language, though with occasional inaccuracies, inappropriate usage and misunderstandings in some situations. Generally, you handle complex language well and understand detailed reasoning.	You can speak and engage in conversation and discussion at length without noticeable effort or loss of coherence. You may demonstrate language related hesitation at times, or some repetition and/or self-correction, or may ask for clarification when something is not clear. You use a range of connectives and discourse markers with some flexibility. You have little problem in understanding the vast majority of language engaged with, either with other people or through media.	You are able to logically organise information and ideas and there is clear progression throughout. You can use a range of cohesive features* appropriately, although there may be some mistaken use, particularly prepositions. You are able to use and understand a sufficient range of vocabulary to allow some flexibility and precision. You are able to use less common lexical items with some awareness of style and collocation, but you may produce occasional errors in word choice, spelling and/or word formation.
6 **Competent user**	Generally you have an effective command of the language despite some inaccuracies, inappropriate usage and misunderstandings. You can use and understand fairly complex language,	You are willing to speak at length, and engage with others, although you may lose coherence at times due to occasional repetition, self-correction or hesitation. You can use a range of connectives and discourse markers, but not always	You can arrange information and ideas logically and coherently and there is a clear overall development of ideas and thoughts. You can use a range of common cohesive features* effectively, but cohesion of text within and/or between sentences

	particularly in familiar situations.	appropriately. You have few problems understanding sympathetic communicators, but occasionally need to ask for repetition or clarification.	may be faulty or mechanical. You may not always use referencing clearly or appropriately. You are able to use a satisfactory range of vocabulary and you attempt to use less common vocabulary, but with some mistakes. You sometimes make errors in spelling and/or word formation, but they do not cause too many problems. You understand much of general texts, but often need a dictionary or to ask for clarification if reading academic or scientific texts.

6. The 'How good is my English?' Apps

These two apps, although fundamentally the same, were designed either to be used by those learning and studying English as a 2nd or other language (E4L2) and those who use English as a first language (E4L1). The E4L1 and E4L2 C-Tests use 8 paragraphs of increasingly complex language taken from multiple sources, ranging from children's literature to academic texts, from newspaper articles to non-fiction prose. Although the C-Test is the same, the output score and descriptors differ. The E4L2 test score aligns with language learning levels widely used by English language learning course books and schools (Table 3), whilst the E4L1 results align with globally recognised education systems and levels (Table 3). Score alignment to descriptors in E4L1 provides takers with a basic language proficiency explanation for their specific level, as shown below (Table 4, whilst E4L2 takers are provided with descriptors more appropriate for language learners (Table 5).

Table 3. Scores and levels for E4L1 and E4L2 C-Tests

C-Test results for E4L2 takers		C-Test results for E4L1 takers	
Percentage %	**Level**	**Percentage %**	**Level**
0-10	Beginner	0-29	Below year 6
11-29	Elementary	30-40	Year 7-8
30-45	Pre-intermediate	41-55	Year 9-10
46-59	Intermediate	56-64	Year 11-12 & Diploma
60-69	Post-intermediate	65-77	Undergraduate Degree
70-79	Pre-advanced	78-87	Master's Degree
80-89	Advanced	88-100	Doctoral Degree
90-100	Near-native like user		

Table 4. Descriptor for post-intermediate level in E4L2 C-Test feedback

Score of 60-69 - Post-intermediate level:

Generally, you have an effective command of the language despite some inaccuracies, inappropriate usage and misunderstandings. You can use and understand fairly complex language, particularly in familiar situations. You are willing to speak at length, and engage with others, although you may lose coherence at times due to occasional repetition, self-correction or hesitation. You can use a range of connectives and discourse markers, but not always appropriately. You have few problems understanding others, and occasionally need to ask for repetition or clarification. You can arrange information and ideas logically and coherently and there is a clear overall development of ideas and thoughts. You can use a range of common cohesive features* effectively, but cohesion of text within and/or between sentences may be faulty or mechanical. You may not always use referencing clearly or appropriately. You are able to use a satisfactory range of vocabulary and you attempt to use less common vocabulary, but with some mistakes. You sometimes make errors in spelling and/or word formation, but they do not cause too many problems. You understand much of general texts, but often need a dictionary or to ask for clarification if reading academic or scientific texts.

Note: * requires reader to refer to post-script note.

Table 5. Descriptor for undergraduate degree level in E4L1 C-Test feedback

Score of 65-77 - Undergraduate degree level:

You have a good operational command of the language, though with occasional inappropriate usage and misunderstandings in some unfamiliar situations. Generally, you handle complex language well and understand detailed reasoning. You can speak and engage in conversation and discussion at length without noticeable effort or loss of coherence. You use a wide range of connectives and discourse markers with flexibility. You have little problem in understanding the vast majority of language engaged with, either with other people or through media, however, you may find some academic texts more difficult to engage with and will need to read multiple times to have a clear understanding. You are able to logically organise information and ideas and there is clear progression throughout. You can produce quality text that meets most academic and professional requirements.

7. Conclusion

The IELTS predictor was published in February 2017 and by August had had close to 2000 downloads. Being able to track, through Apple statistics, where downloads have occurred is valuable, to date with approximately 80% of downloads in the Asia Pacific region, 9% in the Americas, and 5% in Europe, with China leading by far on a country by country basis. The 'How good is your English' app has had somewhat fewer downloads in its 3 months presence. All three apps will be published in Android form by the end of 2017, and all are free to download. All apps require further research, particularly in terms of score alignments, as only preliminary appraisal has been conducted prior to the apps build and publishing. My programming colleague and I will now concentrate on further refining the initial online C-Test software, and wish also to engage extensively in research around its effectiveness and practicality, particularly within diverse learning, academic and professional settings, and I would be delighted to engage with anyone who might be interested in conducting further research within their specific setting.

Downloads - 26 Feb to 26 Aug - top 10 Territories	
Territory ▼	**Units** ▼
China	1318
USA	118
India	45
Australia	34
UK	33
Saudi Arabia	11
Thailand	8
Nigeria	6
Philipines	5
France	5

Figure 3. Downloads of The IELTS Score Predictor.

References

Alderson, J.C. (2002). Testing proficiency and achievement: principles and practice. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Eds.), *University language testing and the C-test* (pp. 15-30). Bochum: AKS-Verlag.

Atai, M. R. & Soleimany, M. (2009). On the effect of text authenticity & genre on EFL learners' performance in C-tests. *Pazhuhesh-e Zabanhaye Khareji* 49, 109-123. Available from https://jor.ut.ac.ir/m/article_27676_b6d42cffb07714a23474fe66363c8eee.pdf.

Babaii, E. & Fatahi-Majd, M. (2014). Failed restorations in the Ctest: Types, sources, and implications for C-test processing. In Rüdiger Grotjahn (Ed.), *Der C-Test: Aktuelle Tendenzen/ The C-Test: Current trends* (pp. 263-276). Frankfurt am Main: Lang.

Coleman, J. A. (1994a). Degrees of proficiency: Assessing the progress and achievement of university language learners. *French Studies Bulletin*, 50, 11-16.

Coleman, J. A. (1994b). Profiling the advanced language learner: The C-Test in British further and higher education. In Rüdiger Grotjahn (Ed.), Der CTest. *Theoretische Grundlagen und praktische Anwendungen* (Vol. 2, pp. 217- 237). Bochum: Brockmeyer. Available from http://www.c-test.de/deutsch/index.php?lang=de&section=originalia

Coleman, J. A., Grotjahn, Rüdiger & Raatz, Ulrich. (Eds.). (2002). *University language testing and the C-test.* Bochum: AKS-Verlag.

Grotjahn, R. (1986). Test validation and cognitive psychology: some methodological considerations. *Language Testing,* 3(2), 159-185.

Grotjahn, R. (1987). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In Rüdiger Grotjahn, Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Taking their measure: The validity and validation of language tests* (pp. 219-253). Bochum: Brockmeyer. Available from http://www.c-test.de/deutsch/index.php?lang=de&section=originalia

Grotjahn, R. (2014). The C-Test bibliography: version January 2014. In Rüdiger Grotjahn (Ed.), Der C-Test: Aktuelle Tendenzen/*The C-Test: Current trends* (pp. 325-363). Frankfurt am Main: Lang.

Grotjahn, R. & Stemmer, B. (2002). C-Tests and language processing. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Eds.), *University language testing and the C-test* (pp. 115-130). Bochum: AKS-Verlag. Available from http://www.c-test.de/deutsch/index.php?lang=de&section=originalia

Hulstijn, J. H. (2010). Measuring second language proficiency. In Elma Blom & Sharon Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 185-199). Amsterdam: Benjamins. [C-Test: 191-193]

Jafarpur, A. (1999). What's magical about the rule-of two for constructing C-tests? *RELC Journal*, 30(2), 86-100.

Khodadady, E. (2013). Authenticity and sampling in C-Tests: A schema-based and statistical response to Grotjahn's critique. *The International Journal of Language Learning and Applied Linguistics World*, 2(1), 1-17.

Khodadady, E. (2014). Construct validity of C-Tests: A factorial approach. *Journal of Language Teaching and Research*, 5(6), 1353-1362. Available from http://www.academypublication.com/issues/past/jltr/vol05/06/19.pdf

Khodadady, E. & Hashemi, M. (2011). Validity and C-Tests: The role of text authenticity. *Iranian Journal of Language Testing*, 1(1), 30-41. Available from http://www.ijlt.ir/portal/files/404-2011-01-01.pdf

Klein-Braley, C. (1985a). 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. (1985b). C-Tests and construct validity. In Christine Klein-Braley & Ulrich Raatz (Eds.), Fremdsprachen und Hochschule 13/14: Thematischer Teil: *C-Tests in der Praxis* (pp. 55-65). Bochum: AKS-Verlag.

Klein-Braley, C. (1985c). C-Tests as placement tests for German university students of English. In Christine Klein-Braley & Ulrich Raatz (Eds.), Fremdsprachen und Hochschule 13/14: Thematischer Teil: *C-Tests in der Praxis* (pp. 96-100). Bochum (Ruhr-Universität): AKS-Verlag

Klein-Braley, C. (1985d). Reduced redundancy as an approach to language testing. In Christine Klein-Braley & Ulrich Raatz (Eds.), Fremdsprachen und Hochschule 13/14: Thematischer Teil: *C-Tests in der Praxis* (pp. 1-13). Bochum (Ruhr-Universität): AKS-Verlag.

Klein-Braley, C. (1985e). Tests of reduced redundancy – theory. In Viljo Kohonen & Antti J. Pitkänen (Eds.), Language testing in school. AFinLA Yearbook 1985 (pp. 33-48). Tampere: AFinLA.

Klein-Braley, C. (1996). Towards a theory of C-Test processing. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 3, pp. 23-94). Bochum: Brockmeyer. Available from http://www.c-test.de/deutsch/index.php?lang=de&section=originalia

Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: An appraisal. *Language Testing*, 14(1), 47-84.

Klein-Braley, C. & Raatz, U. (1984). A survey of research on the C-Test. *Language Testing*, 1(2), 134-146.

Mochizuki, A. (1994). C-Tests: Four kinds of texts, their reliability and validity. *JALT Journal*, 16(1), 41-54.

Oscarson, M. (1991). Item response theory and reduced redundancy techniques: Some notes on recent developments in language testing. In Kees de Bot, Ralph B. Ginsberg & Claire Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 95-111). Amsterdam & Philadelphia: Benjamins [C-Test: pp. 106-107].

Raatz, U. (1985a). Better theory for better tests? *Language Testing*, 2, 60- 75.

Raatz, U. (1985b). Tests of reduced redundancy – the C-test, a practical example. In Viljo Kohonen & Antti J. Pitkänen (Eds.), *Language testing in school. AFinLA Yearbook 1985* , No. 41 (pp. 49-62). Tampere: Association Finlandaise de Linguistique Appliquée. [long version of Raatz, 1985e]

Raatz, U. (1985c). Tests of reduced redundancy – the C-Test, a practical example. In Christine Klein-Braley & Ulrich Raatz (Eds.), Fremdsprachen und Hochschule, 13/14: Thematischer Teil: *C-Tests in der Praxis* (pp. 14-19). Bochum: AKS-Verlag

Raatz, U. (1987). C-Tests: What do they measure, how do they measure it and what can they be used for? In Christine Schwarzer & Bettina Seipp (Eds.), *Trends in European educational research* (pp. 38-44). Braunschweig: Universität

Raatz, U. & Klein-Braley, C. (1985). How to develop a C-Test. In Christine Klein-Braley & Ulrich Raatz (Eds.), Fremdsprachen und Hochschule, 13/14: Thematischer Teil: *C-Tests in der Praxis* (pp. 20-22). Bochum: AKSVerlag.

Raatz, U. & Klein-Braley, C. (2002). Introduction to language testing and to C-Tests. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Eds.), *University language testing and the C-test* (pp. 75-91). Bochum: AKS-Verlag. Available from http://www.c-test.de/deutsch/index.php?lang=de&section=originalia

Rouhani, M. (2008). Another look at the C-Test: A validation study with Iranian EFL learners. The *Asian EFL Journal*, 10(1), 154-180. Available from [http://www.asian-efl-journal.com/March_2008_EBook.pdf

Wedell, M. (1987). The C-test - current relevance, text difficulty and student strategies. *Language Testing Update*, 4, 23. [summary of Wedell, 1985].