Journal of Academic Language & Learning
Vol. 13, No. 1, 2019, A15-A27. ISSN 1835-5196

Association for Academic Language and Learning

The Online C-Test: Evaluating a discipline-specific academic literacy competency assessment tool

Paul Kebble

Coordinator of English Language Development, Faculty of Health Sciences, Curtin University, Perth, WA 6102

Email: paul.kebble@curtin.edu.au

Vaughan Cruickshank

Program Director – Health and Physical Education, Maths/Science, Faculty of Education, University of Tasmania, Launceston, Tasmania 7248

Email: v.j.cruickshank@utas.edu.au

(Received 8 May, 2018. Published online 23 March, 2019.)

This paper examines the use of an experimental online post-entry language assessment (PELA) tool during the first year of an undergraduate Information and Communication Technologies (ICT) course provided by an Australian university. The research was conducted in response to English language teaching staff, subject lecturers, and the Australian Universities Quality Agency (AUQA) expressing concern that students were engaging with the ICT course without appropriate English language proficiency. The response to these concerns was for the authors to implement an assessment of English language competency to establish the extent of the concerns, and to provide recommendations as how these issues could be addressed. Three forms of testing were employed, a reading/summarising exercise, a listening test, and the online discipline-specific literacy assessment tool. More specifically, this paper details the design and evaluation of the latter online assessment tool utilising the 'C-Test' procedure. The outcomes of this research were twofold. First, it identified that as many as 39% of students were likely operating with a lower level of linguistic proficiency than was considered appropriate for successful engagement with the course, which then allowed the authors to provide a list of recommendations to help alleviate this issue. Second, while extensive research suggests the C-Test is a reliable tool to gauge linguistic competency, our research found only a weak correlation between students' C-Test results and their results on a concurrent reading/writing (summarising) test. This result points to a need for more rigorous research into the concurrent validity of the C-Test and the way that it is implemented online as a measure of discipline-specific linguistic competency.

Key Words: C-Test, online application, assessment, EAL, EAP, PELA.

1. Introduction

A significant number of students with English as an additional language (EAL) wishing to develop their English language competency, for multiple motives, are present in most English speaking countries. The attractions of studying English in a country with English as its predominant

language are plentiful. Those involved in the learning and teaching of languages have consistently advocated the best way to learn a foreign language is to immerse oneself in the target language, particularly through studying abroad (e.g., Amuzie & Winke, 2009; Sasaki, 2011). Empirical research has supported these views by showing that studying in an English language environment *can* improve linguistic facets such as oral fluency (e.g., Segalowitz & Freed, 2004), syntactic development (e.g., Isabelli, 2004), and reading competence (e.g., Watson, Siska & Wolfel, 2013), though this is not inevitably the case (Birrell, 2006; Foster, 2012, p. 596). In addition to the perceived increase in English proficiency, students also choose to study abroad for reasons such as exposure to, and increased understanding of differing cultures, and occasionally as a catalyst to permanently relocating (Andrade, 2006). Indeed, many countries are benefitting from the skilled migration of overseas students who have chosen to remain in the further education host country and find employment, particularly in fields such as computing and engineering (Colebatch, 2005). The host nation therefore benefits twofold; by receiving international funds for education, and then from the learned skills and expertise of such individuals once they have qualified.

Governments in countries such as Australia, in consultation with education providers, "have developed clear national priorities and comprehensive strategies to attract a larger number of international students" (Schneider, 2000, pp. 2-3). These strategies include centralised planning, increased cooperation between government and education providers, and simplified visa and university application processes. The most obvious reason for increasing international student numbers is one of economics (Andrade, 2006; Luke, 2008), and most universities are reaping the financial rewards of such policy and strategic development. However, educators in countries such as Australia and America have expressed concern that English language proficiency (ELP) requirements for international students are being disregarded due to the monetary benefits of increased numbers of overseas students (Liu, 2016).

Academic English proficiency is a key contributor to international students' success in their further and higher education programs. For this reason, education institutions recruiting EAL or English for Academic Purposes (EAP) students require either an internationally recognised level of linguistic proficiency, usually an IELTS or TOEFL score, or success in an in-house provided EAP course, or both. The syllabi for the short-term EAP courses usually provide intensive training in: 1) the four macro skills (speaking, writing, reading, listening); 2) academic and information literacies; 3) academic communications and presentations; and 4) study of English speaking cultures, often focusing on the host country. Yet, the authors, and peers with many years of experience in the profession, have consistently perceived that a significant number of students successfully exiting from such courses are not truly linguistically equipped for successful undergraduate or postgraduate studies.

The current paper is based on the authors' practical experiences at an Australian university that provides a four-year undergraduate degree course in Information and Communication Technologies (ICT) for international students. This course has an initial year focused on discipline-specific academic English language competency development. Students were accepted into the first year program with either an academic IELTS score of 5.0 or above, with a minimum of 5.0 in each skill, or successful completion of an in-house EAP course. All students were required to reach a university assessed IELTS equivalency score of 6.0 by the end of the first academic year. Although students on the course had met the entrance requirements, English language teaching staff and subject lecturers had expressed concerns that students were being allowed to commence the course without the appropriate English language proficiencies. Consequently, a number of students were identified as struggling to deal with the linguistic challenges faced by participating in an ICT undergraduate university course delivered in English. These findings align with previous findings from Eckstein and Ferris (2018). Additionally, an audit by the Australian Universities Quality Control Agency (AUQA) evaluating student engagement and assessments indicated the same concerns. Consequently, it became the authors' professional responsibility to quantifiably assess both the levels of English of the first-year student cohort, and the linguistic methods of

course delivery. Recommendations were then to be made to the university, and hence AUQA, on ways to improve the English language competency of students during recruitment and course delivery.

To this end, the requirement to economically assess the linguistic competence of nearly 100 students became a priority, and the authors decided that an online system of assessment would be most appropriate. This paper specifically details the design and development of an online assessment tool utilising the 'C-Test' assessment process, along with two other testing modes employing discipline-specific literacies, to assess students' linguistic proficiency in response to the previously mentioned concerns. The utilisation of three separate modes of assessment provided the researchers with the opportunity to assess the concurrent validity of the C-Test for future employment as a reliable tool for assessing discipline-specific literacy competency. Concurrent validity is said to be verified when one test correlates acceptably well with another test (O'Neill, Goffin & Gellaty, 2012), in this case, the IELTS test.

The outcome of this project was in the form of a report, accepted by both the university and the quality control body, with recommendations for improvement in pre-course assessment and EAP learning. These included the following:

1. Students with an IELTS score of 5 and above across the 4 macro-skills, or the TOEFL equivalent, should be admitted to the course and not the existing 'equivalent' IELTS score as determined by the in-house provided EAP course.
2. In conjunction with the EAP and ICT course provider, a validated C-Test be administered prior to acceptance on the course for all students.
3. Pre-university English language class size reduced from 20 to a maximum of 12 students.
4. Pre-university English language teachers and course lecturers receive appropriate professional development (PD), particularly in teaching within an EAP curriculum. This PD would focus on the development of language learning specifically for the needs of ICT students and away from the more generic EAP course materials presently provided. It would also be designed to enable language teachers and course lecturers to receive concurrent PD to encourage dialogue between both groups.
5. A professional link between the pre-university EAP course provider and the 1st year ICT course manager be established, with regular communications to discuss systems of linguistic performance and support occurring.

2. C-Test design and online build

The C-Test, a derivation of a cloze-test, is a procedure assessing linguistic competence through a psycholinguistic process referred to as 'reduced redundancy' (Klein-Braley, 1985a, 1985b, 1997; Oscarson, 1991; Raatz, 1985) and the accurate restoration of language (Babaii & Fatahi-Majd, 2014), where interference in communication is achieved through the mutilation of text. The C-Test was originally designed by Professors Raatz and Klein-Braley at the University of Duisburg, Germany, as an instrument for determining general language proficiency (2002) and used both as a pre-course placement test and as a PELA. Raatz and Klein-Braley assumed that "all language behaviour is related and thus integrative" and therefore the "validity of the C-Test can be extended to the use of the language generally (i.e. general language proficiency)" (p. 81). The C-Test presents four to six texts as discrete subject or discipline-specific paragraphs, depending on the testing requirements. The first sentence of each paragraph is complete and the second half of every second word thereafter is deleted. After 25 deletions, the remaining text remains intact, with 100 or 150 deletions in total. Micro- and macro-level cues, along with anaphoric and cataphoric referencing, are linguistic skills required to restore mutilations. The more accurate the restoration, the more proficient the restorer. Analyses of C-Tests have shown that they are capable of measuring all aspects of language knowledge, with Hastings (2002) concluding that "C-Tests tap the ability to integrate contextual information with a range of language competencies including those

involved in semantic, syntactic, morphological, lexical, and orthographic processing" (p. 53). The integration of all these features is required for language comprehension and production, therefore a test requiring the implementation of these features will be a comprehensive assessor of linguistic competency. Over the past 40 years the C-Test procedure has been extensively researched (e.g., Grotjahn, 2016) and has become widely accepted and utilised when it comes to measuring language proficiency in learners of an additional language (Linnemann & Wilbert, 2014).

An example of a mutilated text looks like this:

> There is a need for teachers to value and build on what students know and can do. There i__ also a ne__ to dev____ intellectually chall_____ and conn_____ learning opport_____ that acc____ for rap____ changing commun_____ practices. Th___ is a ne__ to bu___ on ea__ teacher's reper_____ of appro_____ to t__ teaching o_ literacy…

The original text:

> There is a need for teachers to value and build on what students know and can do. There is also a need to develop intellectually challenging and connected learning opportunities that account for rapidly changing communication practices. There is a need to build on each teacher's repertoire of approaches to the teaching of literacy…

C-Tests have been developed and validated in multiple languages including English in recent decades. Numerous empirical studies have reported strong construct validity (e.g., Baghaei & Grotjahn, 2014; Rahimi & Saadat, 2005) and reliability (e.g., Mochizuki, 1994) of the C-Test, with test-retest reliability coefficients often exceeding 0.9. These studies have additionally reported strong correlation with other English language proficiency tests such as the Standardised Test for English Proficiency (STEP) and the Comprehensive English Language Test (CELT). One author (Kebble, 1996) also conducted research for a Master's thesis measuring the C-Test for concurrent validity by measuring the C-Test against a well-validated commercially available test (the Oxford Placement Test) through the testing of 83 students. A strong correlation coefficient (Pearson's $r = 0.82$) was seen, allowing the researcher to suggest the C-Test was a valid test of general language proficiency. However, this study also found from qualitative feedback that test-takers did *not* believe the C-Test was testing their linguistic competency. This suggests a possible problem with face validity, but rather than drawing this conclusion, the researcher concluded that as the C-Test is rather different to more 'sentence' and 'lexis' focused questions often used in more traditional language tests, takers were perhaps unsure of its assessment focus and capacity. The researcher considered this was perhaps not a restrictive influence on test performance and would not therefore impact on the test-taking procedure. The conclusion was, consequently, that clearer instructions that included a brief introduction to the C-Test construct and the cognitive processing required for text completion may help alleviate the issue.

Through personal and published (Grotjahn, 2016) research, it was therefore the authors' belief that the C-Test was a reliable and valid form of language competency testing which could be provided online utilising digital facilities within an Online Learning Platform (OLP). However, since the C-Test was to be used to inform policy decisions, would have a non-standard format in the online environment (see Fig. 1 and related discussion), and needed to have norms established against IELTS levels, it was deemed necessary to confirm the C-Test that was developed was a valid assessment tool able to function reliably in an online environment and within a subject-specific discipline, and hence this was the goal of this research.

3. Academic linguistic assessment procedure

The linguistic competency assessment exercise was not required to focus on individual students, but rather, to provide the range of assessed competency, affording possible substantiation to the

concerns raised by both the teachers and the higher education audit. The assessment procedure included three forms of testing, the C-Test, a listening test, and a reading/writing (summarising) exercise. As such, the scores provided below were as generated by the Blackboard system and are not traceable to an individual's performance. However, the scores for each test were collated for individual takers, providing the three scores for each.

The C-Test used in this study was created using four texts based upon the topic of ICT and gleaned from the course literature. The research team, cognisant of earlier research into construct and face validity (described above) of the C-Test, decided to provide clear instructions to takers prior to their taking of the test. Along with an introduction to the functionality of the C-Test, test takers are advised to:

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.

The online C-Test was built within the Blackboard OLP utilised by the university, but the software was unable to provide the C-Test structure in the way it had originally been designed. The researcher was therefore required to provide the text, with the omissions numbered, and followed by sequentially numbered 'fill in the blank' questions, as shown in Figure 1. This format may have caused students difficulties with adequately implementing step 4 of the advice given above, an issue taken up in the Discussion (Section 5).

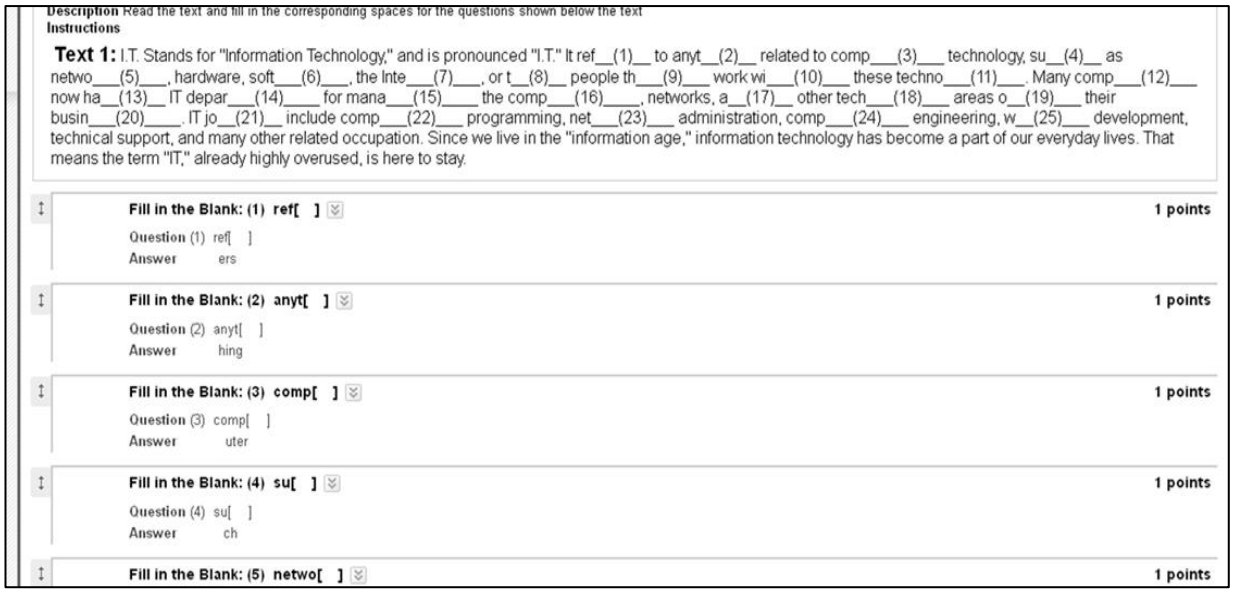

Figure 1. The C-Test in the Blackboard OLP.

For benchmarking purposes, the same C-Test was administered to a group of eight English as an additional language (EAL) students who had taken the IELTS examination within the last three months studying in a local private language college, and one English lecturer at the university. The lecturer achieved a score of 98. The local EAL students were reported by the college as operating between overall IELTS scores of 5.0 and 6.5. The range of scores from these students were from 43 to 78. These results were then aligned to the students' overall IELTS scores, which provided an acceptably approximate equivalence as indicated in Table 1. With this alignment applied to the results of the C-Test, it would suggest that students on the undergraduate ICT scoring below 40 might be identified as being at serious risk of failing to achieve an IELTS 6.0 on the

end-of-year IELTS equivalent test. This conclusion emanates from an understanding that to improve a band in IELTS requires approximately six months intensive, focused IELTS preparation study (Elder & O'Loughlin, 2003), not a first-year university course with an integral academic English program.

Table 1. Equivalencies between C-Test scores and IELTS band as determined by benchmarking with eight EAL students who had recently sat an IELTS test.

C-Test range	40–49	50–59	60–69	≥ 70
IELTS band	5.0	5.5	6.0	6.5

For the listening test, a recorded 19 minute orally provided lecture was utilised, sourced from an authentic online ICT course provided by an Australian academic with a reasonably strong Australian accent. The lecture was entitled 'Information Technology in the Business World, Its Advantages, Disadvantages and Challenges.' The lecture was presented twice during the test, an initial listening, and a repeated listening 10 minutes later. Students were given 20 questions (online) and were provided five minutes prior to the first listening to read through the questions. The test duration was one hour.

The reading/writing (summarising) test provided test-takers with a 1230 word article entitled, 'How has technology changed the way we conduct business?' From this article, a 200 word summary was to be produced within one hour. Students were specifically told not to copy text, but to present relevant information in their own words. This was clearly stated in the written introduction to the test, and was emphasised orally by the invigilators. The tests were assessed by one author utilising publically available IELTS writing descriptors and through knowledge and experiential understandings as the author had previously acted as an IELTS examiner.

The language tests were provided by the authors, functioning as the University's teaching and learning development team, and were administered between 1 pm and 4.30 pm, with the C-Test first, followed by the listening and finally the reading/writing test. A short break was provided between each test. Students (N = 93) were divided into 3 groups in three adjoining computer rooms, with each room having at least two invigilators; therefore, the tests were conducted within an acceptable level of security.

4. Results of three tests

4.1. C-Test

Results from the C-Test (see Figure 2) had a range of 57, between scores of 15 and 72, with the median score at 43 and the mean score being 44.09. These results showed that 39% of takers fell below a score of 40, indicating that they were possibly functioning below an acceptable level for course success.

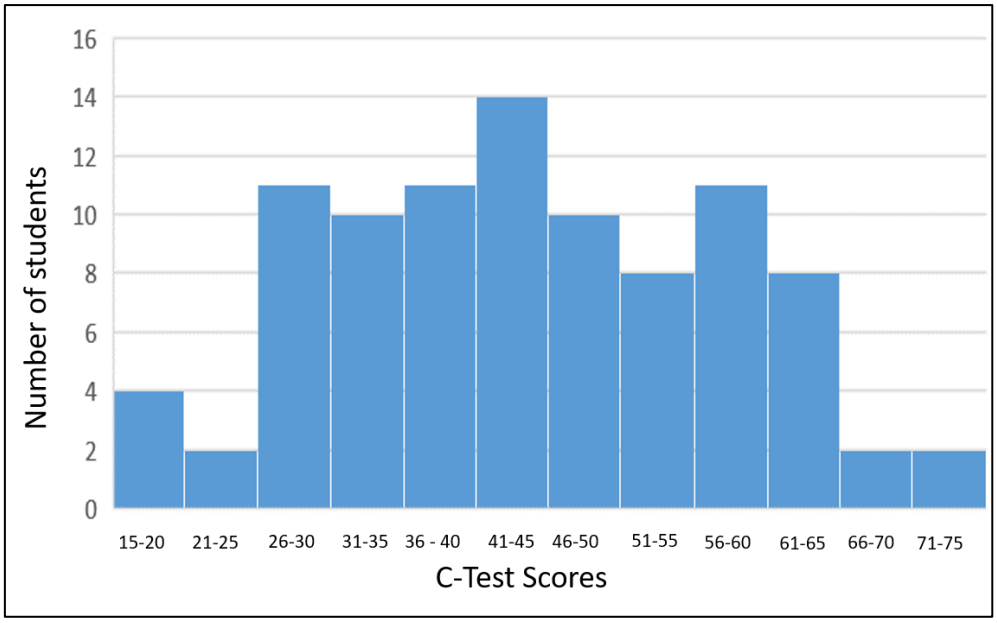

Figure 2. Distribution of C-Test Scores (maximum score = 100, *N* = 93) indicating a broad spectrum despite all students supposedly having English language competencies equivalent to IELTS scores of from 5.0 to 6.0.

4.2. Reading/writing (summarising) test

The Reading/writing (summarising) tests were assessed utilising the published IELTS descriptors and through personal experiences in IELTS testing. The range of scores for the reading/writing (summarising) test was of 5 IELTS levels Spread between 4.0 and 6.5 IELTS equivalent (see Fig. 3), with an average of 5.0. The frequency of each IELTS Score is shown in Fig. 3, with the number of scores considered under IELTS 5.0 being 33. Note that there is no score below '4' as '4' represents major copying from the original text. A 4.5 score represents major copying, but with limited originality.

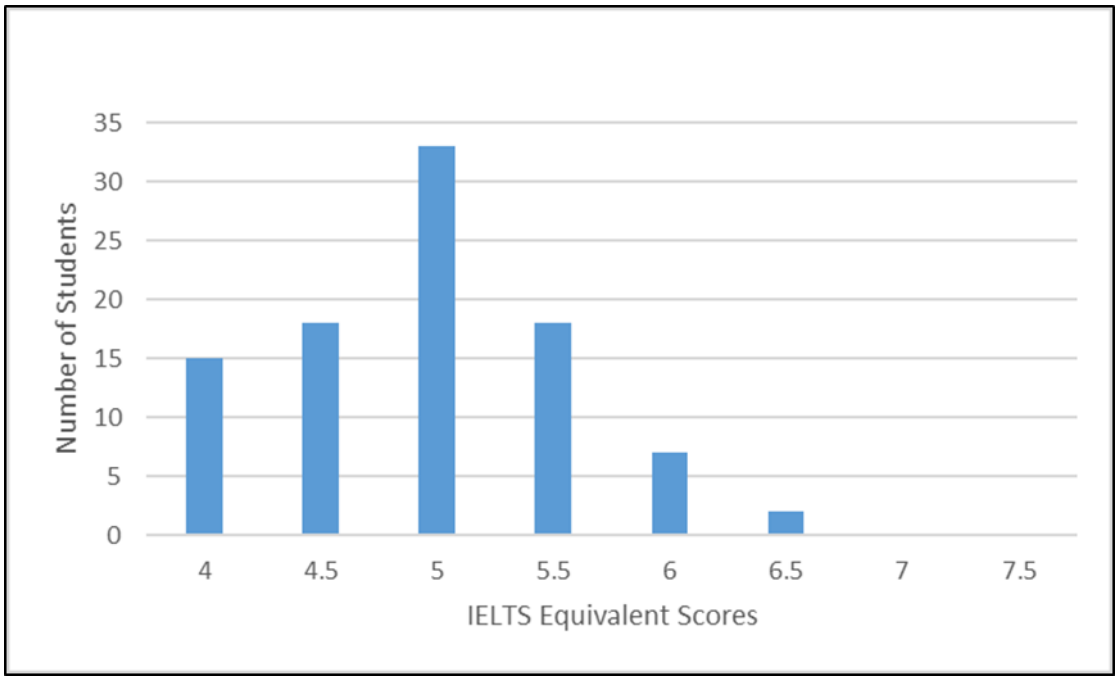

Figure 3. Distribution of Reading/Writing Test Scores.

4.3. Listening Test

Without being able to conduct an IELTS concurrent validity exercise, the authors initially considered halving the overall listening score as an IELTS equivalence on the basis that the IELTS range is 1 to 9 and the listening test had 20 questions, i.e. a score of 8 would = IELTS 4, 15 would = 7.5 and anything over 18 would be IELTS 9.0. The authors, however, accept that this equivalency has limited reliability, and as such the IELTS results have restricted relevance. Surprisingly, the range of scores was between 1 and 18, so the researchers decided that any score under 8 (IELTS 4.0) would be considered a 4. The number of students scoring each IELTS Equivalent on the listening test is shown in Fig. 4.

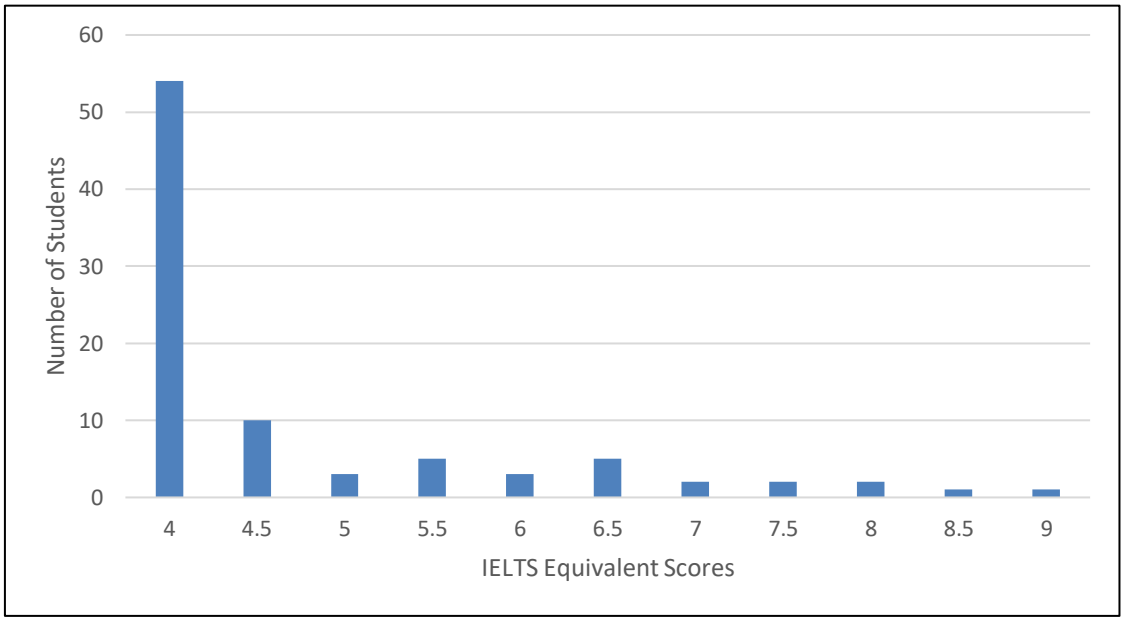

Figure 4. Distribution of Listening Test Scores.

5. Discussion of Test Results

Table 2 provides the scores from each test as IELTS score equivalents. The numbers within the table show the total number of students within each category. Although no correlation can be inferred from these figures as the students scoring in each band for each test are not necessarily the same students, the table suggests a large number of course participants were potentially functioning at an academic linguistic level requiring intervention and support.

Another factor to be considered is the reliability of the purported IELTS scores. Although the concept is beyond this research, there is evidence (Wray & Pegg, 2009) that EAL students are achieving higher IELTS scores than their true competency levels through various means. The fact the AUQA audit had already focused upon the perceived low levels of English language competence demonstrated within this specific course suggests limited correlation between recorded IELTS scores and linguistic competence.

Table 2. Number of students' IELTS equivalency scores for each test.

Test & IELTS Score	<5	5	5.5	6	6.5	7+
C-Test	36	25	16	14	2	
Summary Writing	33	33	18	7	2	
Listening	64	3	5	3	5	7

The listening scores were particularly low, with the researchers considering this anomaly due to either the listening test being too complex, or the students' limited experience of listening to accented English being spoken, as the orator of the lecture had a reasonably distinctive accent. When comparing the listening scores with the C-Test scores (Fig. 5), no evident correlation was observed. The researchers therefore considered the C-Test was either an unreliable form of assessment for audio receptive skills or that the listening test itself was poorly designed. In either case, the researchers believe further research is required to investigate the assessment of listening skills and how this may be aligned to the C-Test construct.

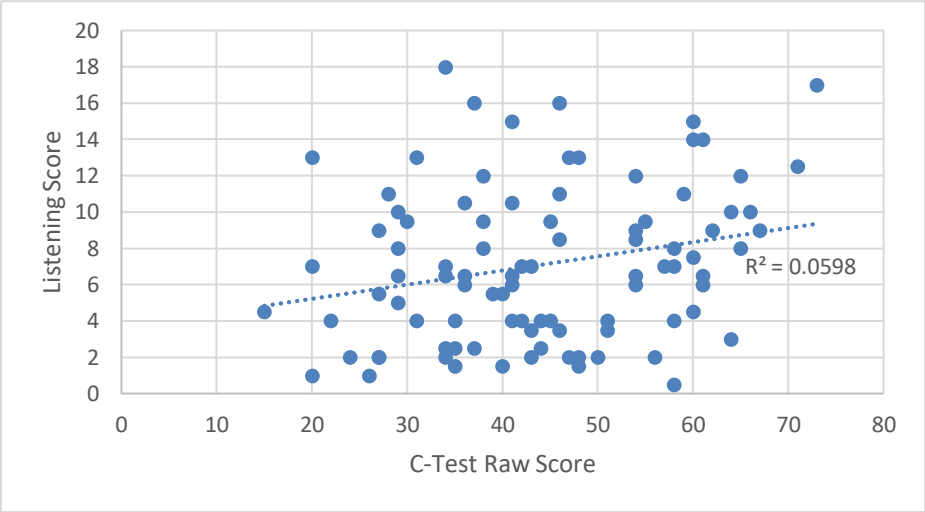

Figure 5. No relationship of any importance was found between the C-Test and Listening scores.

The correlation between students' C-Test score and their reading and writing test IELTS score (see Figure 6) was much weaker than expected (Pearson's r = 0.38). Consequently, it was concluded that these two tests provide limited aligned evidence in support of a concurrently valid indication of academic literacy competency.

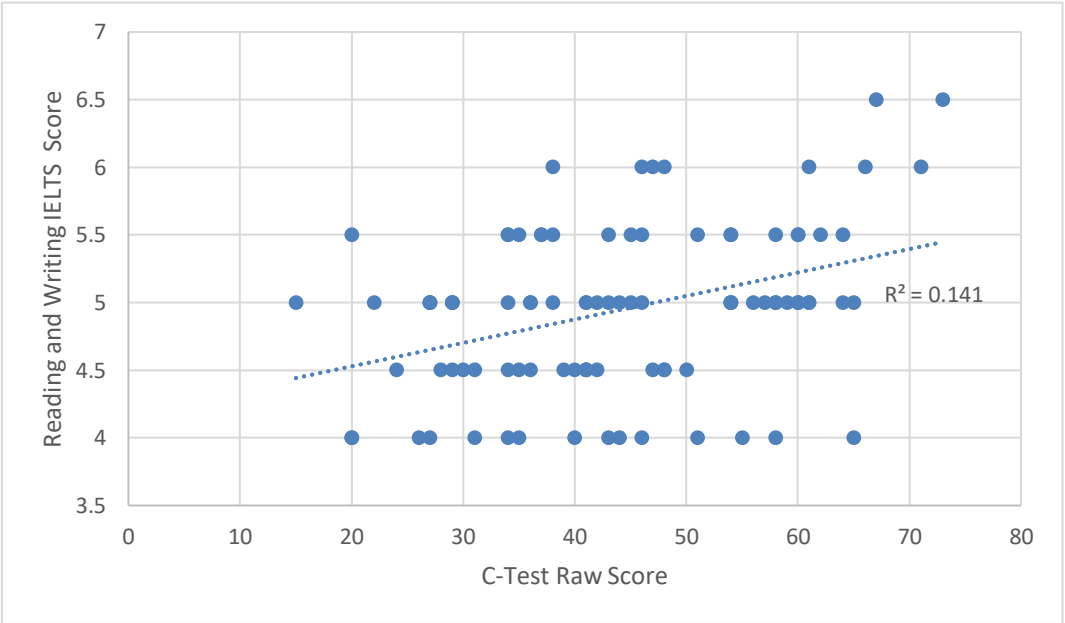

Figure 6. Reading/writing scores versus C-Test raw score revealing a weak correlation.

The researchers expected there would be a stronger correlation than these results displayed, particularly when examining the extensive research that has previously been conducted, including personal research. The researchers deliberated on the possible explanations for this weak correlation and concluded that the selection of texts might be a primary cause. As previously explained, these texts were extracted from course materials, however, we were unaware of whether the used texts had already been presented, and hence concepts engaged with in the test were clearly understood. Expecting test-takers to clearly comprehend unengaged-with topics, including specific lexis, would put them at a distinct disadvantage. It was therefore felt the test designers should have consulted more rigorously with the course lecturers to collaboratively select the most appropriate texts.

Another possible consideration was to have used texts gleaned directly from past IELTS papers, if they were offered for general distribution. ICT topic-specific texts could have been collated, giving the C-Test a discipline-specific focus, but in essence, the C-Test would have been formulated through the use of IELTS texts designed for purpose.

6. Considerations emanating from research project

An initial consideration emanating from this research project was the very limited cohort of IELTS/C-Test scores available in the initial alignment process. The researchers considered that accumulating a far greater number of IELTS/C-Test results would have provided a far more accurate alignment, however, time and appropriate participant constraints meant this was unachievable before the assessment project was required to be undertaken. Also, the eight test participants who engaged in the initial alignment process between the C-Test and IELTS results did not necessarily have an ICT background, again potentially skewing the accuracy of the alignment outcome.

Another consideration was that as nothing specific was aligned to the outcomes of these tests, a proportion of students may not have tried particularly hard, knowing the outcome had no bearing on their course grades. Also, fatigue may have played a part as the testing was conducted at the end of the day, with the writing being presented as the last component of three hours of testing. This would possibly also have contributed to a proportion of students not giving their best efforts.

Also, many students may have found the layout of the C-Test shown in Fig. 1 challenging. Not being able to put the text completions into the text itself may have hampered students' ability, and perhaps even willingness, to fluently read through their answer choices and check for accuracy and correctness.

These finding indicated that the C-Test in this case did not effectively and reliably provide an assessment of students' discipline-specific literacy competency, even though the expectations were somewhat higher. The authors accept this specific exercise in concurrent validation between the C-Test and an overall IELTS score is only preliminary and requires further research to be able to recommend the implementation of the C-Test as a reliable form of either a screening test, placement test, or a PELA. Although this research has provided limited evidence, previous research suggests there is potential for the C-Test to be utilised as a discipline-specific academic literacy and language competency assessment tool that can be delivered effectively online, if the preparation and design is more systematically engaged with. Therefore, an online C-Test has the potential to provide a cost-effective and practical system of linguistic assessment if its reliability and validity can be proven.

The Pearson's correlation coefficient between the C-Test and the listening test produced a value of $r = 0.24$, which is also a weak positive correlation. The researchers acknowledge the lack of a prior test validation exercise, and accept the results for the listening test are inconclusive.

6. Conclusion

Although it is beyond the scope of this paper, the results obtained through the above tests justified the concerns raised by both the staff and AUQA that a sizeable percentage of students did not have the prerequisite English language competency to successfully engage with their ICT undergraduate course. Although the results appear to show a third of the cohort of first year ICT students functioning below the expected and required linguistic levels of the course, the results cannot be accepted as conclusive evidence of such. However, the testing has shown the validity of the trend of a significant number of students being accepted onto academic courses with a lower level on academic English competency than their IELTS scores would suggest, initially reported by language teachers and content lecturers, and subsequently indicated by the quality control body. The process of EAP competency development and assessment prior to the ICT course must be examined for its appropriateness. The research does not indicate how many students in this cohort had entered with IELTS scores or through attending the pre-university academic English program, although the researchers were reliably informed it was approximately a 60/40 split in favour of pre-course attendance.

Although limited concurrent validation was provided prior to the use of the C-Test through the comparison exercise with EAL students who already had an IELTS score, the Pearson's Correlation Coefficient comparison with the summary writing test has not been able to provide further validation to this version of the C-Test. The researchers accept additional validation is required and that validation is an on-going process and not a "one off occurrence". What can be gleaned from this research, though, is the process of designing an effective C-Test for a particular cohort requires extensive planning, including an appropriate form of validation. However, the authors suggest these results have shown some, albeit limited, potential that an online C-Test can function appropriately as a tool for assessing discipline-specific literacies within a given genre, whether this be general, academic, or discipline-specific literacies.

The researchers accept the IELTS examination is highly reliable and well validated, and believe it is the most appropriate expansive assessment tool commercially available. However, it is also expensive for students to take multiple times. The development of an inexpensive online testing tool capable of reliably providing an overall indication of academic literacy competency within specific academic disciplines would be highly beneficial. Although this particular research has not provided the expected outcomes for the reasons discussed, the authors suggest previously described research indicates a C-Test adaptation might be an appropriate assessment format to provide at least a part of an efficient and economic online discipline-specific linguistic assessment tool, and hence future research attempting to address the issues identified with the used testing procedures will be pursued.

References

Amuzie, G. L., & Winke, P. (2009). Changes in language learning beliefs as a result of study abroad. *System, 37*(3), 366-379. https://doi.org/10.1016/j.system.2009.02.011

Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal of Research in International Education, 5*(2), 131-154. http://dx.doi.org/10.1177/1475240906065589

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

Baghaei, P., & Grotjahn, R. (2014). Establishing the construct validity of conversational C-Tests using a multidimensional Rasch model. *Psychological Test and Assessment Modeling, 56*(1), 60-82.

Birrell, B. (2006). Implications of low English standards among overseas students at Australian universities. People and Place, 14 (4), 53-64. Retrieved from https://search.informit.com.au/fullText;dn=332480460154352;res=IELAPA

Colebatch, T. (2005). 'Foreign students settle'. Retrieved from http://www.theage.com.au/news/Immigration/Foreign-students-settle/2005/01/21/1106110942999.html?oneclick=true

Eckstein, G., & Ferris, D. (2018). Comparing L1 and L2 texts and writers in first-year composition. *TESOL Quarterly*, *52*(1), 137-162. http://dx.doi.org/10.1002/tesq.376

Elder, C. & O'Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS International English Language Testing System. (IELTS) Research Reports 2003: Volume 4. Retireved from https://search.informit.com.au/documentSummary;dn=909050898723873;res=IELHSS

Foster, G. (2012). The impact of international students on measured learning and standards in Australian higher education, *Economics of Education Review*, *31*, 587-600, http://dx.doi.org/10.1016/j.econedurev.2012.03.003

Grotjahn, R. (2016). The electronic C-Test bibliography: Version 2016. Retrieved from https://www.researchgate.net/publication/295693869_The_electronic_C-Test_bibliography_version_2016

Hastings, A. (2002). Error analysis of an English C-Test: Evidence for integrated processing. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 53–66). Bochum, Germany: AKS-Verlag.

Isabelli, C., (2004). Study abroad for advanced foreign language majors: Optimal duration for developing complex structures. In: H. Byrnes & H. Maxim (Eds.). *Advanced foreign language learning: A challenge to college programs* (pp. 114–130). Boston, MA: Heinle.

Kebble, P.G. (1996). Assessing the Concurrent, Face and Predictive validity of the C-Test employed as an undergraduate placement test. University of Portsmouth Library.

Lee, G., & Wallace, A. (2018). Flipped learning in the English as a foreign language classroom: Outcomes and perceptions. *TESOL Quarterly*, *52*(1), 62-84. https://doi.org/10.1002/tesq.372

Liu, D. (2016). Strategies to promote Chinese international students' school performance: resolving the challenges in American higher education. *Asian-Pacific Journal of Second and Foreign Language Education*, *1*(8), 1-15. https://doi.org/10.1186/s40862-016-0012-9

Klein-Braley, C. (1985a). Reduced redundancy as an approach to language testing. In C. Klein-Braley & U. 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. (1985b). Tests of reduced redundancy – theory. In V. Kohonen & A. Pitkänen (Eds.), *Language testing in school. AFinLA Yearbook* 1985 (pp. 33-48). Tampere: AFinLA.

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

Luke, A. (2008). TESOL in the corporate university. *TESOL Quarterly*, *42*(2), 305-313. https://doi.org/10.1002/j.1545-7249.2008.tb00125.x

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

O'Neill, T. A., Goffin, R. D., & Gellatly, I. R. (2012). The Knowledge, skill, and ability requirements for teamwork: Revisiting the Teamwork-KSA test's validity. *International Journal of Selection and Assessment*, *20*(1), 36-52.

Oscarson, M. (1991). Item response theory and reduced redundancy techniques: Some notes on recent developments in language testing. In K. de Bot, R. Ginsberg & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 95-111). Philadelphia, PA: Benjamins.

Raatz, U. (1985). Tests of reduced redundancy – The C-test, a practical example. In C. Klein-Braley & U. Raatz (Eds.), *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis* (pp. 14-19). Bochum (Ruhr-Universität): AKS-Verlag.Raatz, Ulrich & Klein-Braley, Christine. (2002). Introduction to language testing and to C-Tests. In J. Coleman, R. Grotjahn & U. Raatz (Eds.), *University language testing and the C-Test* (pp. 75–91). Bochum, Germany: AKS-Verlag.

Rahimi, M., & Saadat, M. (2005). A verbal protocol analysis of a C-Test. *Iranian Journal of Applied Linguistics*, *8*(2), 55-85.

Sasaki, M. (2011). Effects of varying lengths of study-abroad experiences on Japanese EFL students' L2 writing ability and motivation: A longitudinal study. *TESOL Quarterly*, *45*(1), 81-105. https://doi.org/10.5054/tq.2011.240861

Schneider, M. (2000). *'Others' open doors. How other nations attract international students: Implications for US educational exchange'*. Retrieved from http://exchanges.state.gov/iep/execsummary.pdf

Segalowitz, N. & Freed, B.F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish at home and study abroad contexts. *Studies in Second Language Acquisition 26*(2), 173–199. https://doi.org/10.1017/S0272263104262027

Wray, A. & Pegg, C. (2009). The effect of memorized learning on the writing scores of Chinese IELTS test-takers. In Thompson, P. (2009). *International English Language Testing System (IELTS) Research Reports.* Volume 9, British Council and IELTS Australia, 191-216. Retrieved https://search.informit.com.au/documentSummary;dn=070412442610334;res=IELHSS> ISBN: 9781906438517.

Watson, J. R., Siska, P., & Wolfel, R. L. (2013). Assessing gains in language proficiency, cross-cultural competence, and regional awareness during study abroad: A preliminary study. *Foreign Language Annals*, *46*(1), 62-79. https://doi.org/10.1111/flan.12016