Comparing a Skills-Focused English Test Against a Lexico-Fluency English Test
for International Nursing Students in Higher Education

Amanda Müller, Flinders University, Australia
Michael Daller, University of Reading, United Kingdom

Abstract

This paper expands on the rationale, context, and conclusions of a study conducted by Müller & Daller (2019) which looked at the performance of the academic version of the International English Language Testing System (IELTS) and English testing using the C-test format among international nursing students at an Australian university. The results found for the study will be reproduced, with some comments made about how to interpret the findings in a wider educational context.

Keywords: Language testing, IELTS, C-test, nursing, international students

iafor

The International Academic Forum
www.iafor.org

Background

In English-medium universities, there is a need to establish whether international students with a non-English background have sufficient language proficiency to engage in their degree. The students can either be directly tested, or establish proof of proficiency through some other means (e.g. graduating from a college that offers a course in learning English). There are some language tests are which have the specific purpose of establishing academic English proficiency (e.g. IELTS Academic, TOEFL iBT, PTE Academic, etc), and they use a skills-based format where linguistic responses are elicited within an academic-related task is tested. For example, IELTS tends to require full essays, short answers, 1-3 word cloze answers, multiple-choice selection, matching tasks, table completion, diagram labelling, comprehension responses, and so on.

In practice, for professional degrees such as nursing, the emphasis on academic English may be less appropriate because other types of English are being used. For example, in nursing there is a strong emphasis on spontaneous speech, the need to code-switch, and to fill in charts (but less so the need to write an essay at work). Thus, the requirements of the context may involve strong skills in one aspect of English that a particular test does not really concentrate on. Another means of testing language proficiency might be needed, such the C-test format. The C-test appears to measure a number of things (this itself is still up for debate), such as language fluency and expectancy grammar, and as such might yield better results for establishing communicative capacity in professional environments. On the surface, the C-test seems to be quite simple, since it only requires people to fill in the second half of every second word in the sentences presented in a paragraph of text, usually five completed in half an hour. The main task, it might be argued, is to have sufficient grasp of the elements of a sentence to anticipate the possible type of word that would be needed, recognise what aspect of grammar is involved (e.g. adjective, adverb, or noun form), and to produce the particular word (vocabulary knowledge, collocation knowledge) that would naturally follow from the letters already given in each prompt (context cues). Thus, the mechanics of the test requires the person to deal with the missing part of the word (redundancies), and use their own language experience to assist this process. Such repair of redundancies and misheard/missing input is important in a noisy, busy, and accent-diverse workplace such as a hospital.

When designing this study, it was expected that there would be a large overlap between the constructs of general English, academic English, and professional English, since both academic and professional English situations build upon general ability. It was also expected that there would be some relationship between proficiency and fluency, since you tend to become more fluent with more practice and exposure (itself which contributes to proficiency). Thus, it was with some trepidation that any comparison of two types of English test was to be made. However, the researchers were fairly confident that different aspects of language proficiency and ability were being tested - especially since the IELTS Academic test focused on academic English and the C-test was more generalist. The experience of the researchers was that gaps still existed in clinical situations among students who had reasonable IELTS scores. Building on this expectation, it was anticipated that IELTS, as a purpose-built test for academic study, would be more related to students' grades in academic subjects (non-clinical, non-practicum topics) than the C-test. It was also

expected that the C-test would have a stronger relationship to grades than IELTS for professional clinical topics which involved hospital placements and face-to-face patient simulations. This is not because the C-test is particularly geared towards clinical communication (IELTS Academic was designed specifically for Australian university study), but the C-test does seem to measure the general fluency and general language proficiency needed for professional contexts (i.e. spontaneous language performance and a broader language requirements in an environment that switch registers quickly).

Methods and Results

The full details of the study can be found in Müller & Daller (2019), but a summary will be given here. There was data collected from 49 undergraduate international nursing students (mostly Chinese language backgrounds) at an Australian university. The study paid for participant testing at an official IELTS venue (score transcripts collected from students), at the start and finish of the academic year. The researchers used a timed C-test (100 items in 5 paragraphs) at the university to gain C-test scores both at the start and finish of the academic year. The collection of grade averages was taken at the end of the year (grading of topics was completed by nursing lecturers in the course of their normal duties, independently of the researchers), and these grades were sorted according to academic versus clinical topics. Correlations were produced between these variables.

The study found that, indeed, the IELTS had a stronger relationship to academic grades than the C-test, and that the inverse was true for clinical topics where the C-test had a stronger relationship to clinical grades than IELTS. However, in both situations, the IELTS and C-test still had significant relationships to the other type of topic grade, albeit lower than the other test type. IELTS had the smallest relationship overall, in relation to clinical topics, and the C-test performed at a rate that other studies usually find as good for the relationship between IELTS and grades.

A reproduction of the table found in Müller & Daller (2019, p. 9) summarises the relationships between tests and grades:

Table 3

Correlations (Pearson) between predictor variables and dependent variables.

Variables	Academic topics	Clinical topics
IELTS	.509***, $p < .001$, n = 46	.302*, $p = .049$, n = 43
C-test	.381**, $p < .01$, n = 49	.417**, $p < .01$, n = 46

* Significant at the .05 level.
** Significant at the .01 level.
*** Significant at the .001 level.

As expected, the study found there was overlap between the two types of English test, with a correlation of .68, so a reasonable proportion of the variance between the scores of the two tests was shared (see Müller & Daller 2019 for more detail of this). It would

have been suspicious if there was no overlap, because many aspects of language proficiency theoretically overlap. However, there were also differences detected between the two tests, and the tests reveal that different language constructs were tapped into.

Conclusion

Language proficiency is clearly important for international students who wish to study in an English-medium university, but do not have English as their first or native language. This study highlighted that it is important to keep the purpose and gains of each language test in mind when selecting which test to use. Educators, policy makers, and institutions would do well to note this.

In regards to the relationship between language proficiency and grades, it is very interesting the grades were correlated to any language score at all. These grades were given by nurse educators to students for their academic and clinical nursing performance, and not for language ability. However, language ability appears to be an important correlated variable. It may be the case that particular levels of language ability impacts on the students' ability to engage in the educational process. It may also be the case that the level of language skill that the examiner detected in the person being assessed determined how the content of the assessment was marked (e.g. the number of errors made in speech or writing negatively affects grades). Either way, if language skill is an important factor in understanding or communicating educational concepts, we are left with a situation of what is enough English? It is possible that higher levels of language proficiency (general or academic) will mean that there is no longer a significant negative impact on grades (a ceiling effect) because language proficiency has been fully met for the task. Conversely, there is probably a point where insufficient language proficiency entirely impedes understanding of academic content (a floor effect). Where these floor and ceiling levels are located are not yet known with certainty, but they are important factors that must be researched further. Academic attainment, at least in this study, seemed to be related to English ability of some kind (as evident for both tests) and more needs to be known about the relationship of language to the university experience and outcomes.

Some observations can be made about the strength of the results found in this study. Our study had a good range among the variables. The range of IELTS scores was 5.5-7.5 (out of 9) and grade averages (12-89 out of 100). One reason is that not all students had taken a test to enter university (which usually requires them to have IELTS 6.5), and instead many entered through non-tested pathways. It is speculated in the article that the variability found among the cohort allowed more robust findings. This variability gave us a better picture of how the elements were correlated, especially for English ability (many studies on IELTS have cohorts with little variation in English ability). This points to the need for educational research to consider the influence of sampling in their projects, and to consider the possibility that a null result might simply be from poor sampling or statistical understanding. While there was a skew towards Chinese participants, it had the positive attribute of sampling a strong English as a Foreign Language population (where incidental English exposure is low). This skew towards Chinese participants might also be a weakness, in terms of not being representative of international students generally, but admittedly, Australia gets about 32% of its international students from China.

The wider utility of language tests is another thing to consider here. Might a slightly lower performing test do a good enough job? The focus is usually on gaining the most accurate result, with repeated verification of each version of the test costing a lot of money, and this is passed on to the consumer. To do the tests themselves are often also time-consuming, costing anything from AUD $300-$600 and taking a whole day to complete (traditionally, two days for IELTS). The C-test, however, costs nothing beyond the time needed to: find a natural text; apply the C-test format to it; and double-check that the test makes sense (i.e. avoid some proper nouns, jargon, acronyms). It takes half an hour to do, with immediate results (i.e. you don't have to wait a few days, like the other tests). The C-test format is fairly reliable, and seems to be quite robust across texts. For those who wish to create their own C-test, a good summative guide can be found in Hood (1990). Weighing accuracy and gain against the tests, the C-test format seems to be a possibility worth considering for a various number of purposes in a higher education setting, and possibly beyond, especially given how cheap it is to create and the short amount of time to do, with an immediate result. Returning to this study, was the IELTS result worth the cost needed to get the higher correlation to grades when it performs worse in a clinical setting? This is a decision for the policy makers. The C-test seems more versatile for the purpose, in this type of degree. Overall, this study shows that policy makers need to be clear about the type of English they need to test for, and then use the appropriate means to do so, taking into account the full range of cost/benefit involved.

Another point can be made that extends on the previous paragraph, and this relates to test use in different sectors of society. In Australia, tests like PTE and IELTS are used for migration and nursing registration, yet these are not necessarily a good match of test to purpose. This study shows that there is a shift in the type of language ability required for different communicative situations. Admittedly, in terms of migration, PTE and IELTS Academic should capture the general English required, but maybe the bar is set too high for migrants who come to menial jobs with no academic component. Refugees are also subject to these tests, perhaps unfairly. These tests also may not capture the types of language proficiency needed for professional communication. Ideally, different types of English test should be used for the purpose they were built for, and cautiously extended beyond this. In the case of the health professions, the Occupational English Test was built for this purpose; IELTS was not, it causes one to wonder why IELTS and other academic English tests continue to be used to establish English proficiency for nursing and health registration.

In conclusion, this study revealed some interesting results regarding test appropriateness and the link between English and grades, and hopefully, the discussion of the reasoning behind the study and the relationship this study has to practice will spur others to commence research into similar issues. Language testing is often high stakes and affects large numbers of people in different ways, and a better understanding and application of such tests is desirable.

References

Hood, M.A.G. (1990). The C-Test: A Viable Alternative to the Use of the Cloze Procedure in Testing? In L.A. Arena (Ed.) *Language Proficiency: Defining, Teaching, and Testing* (pp. 173-189). Boston: Springer.

Müller, A. & Daller, M. (2019). Predicting international students' clinical and academic grades using two English language tests (IELTS and C-test): a correlation research study. *Nurse Education Today*, 72, 6-11. doi: https://doi.org/10.1016/j.nedt.2018.10.007.