Predicting study success of international students

Article

Accepted Version

Daller, M. and Wang, Y. (2017) Predicting study success of international students. Applied Linguistics Review, 8 (4). pp. 355-374. ISSN 1868-6311 doi: https://doi.org/10.1515/applirev-2016-2013 Available at https://centaur.reading.ac.uk/67281/

Predicting Study Success of International Students[1]

Michael Daller and Wang Yixin

Abstract

The study success of international students in Higher Education (HE) in English speaking countries has been a major concern for both the students and the host universities. However, studies on the predictive validity of established language tests, such as IELTS, are inconclusive (for an overview see Daller & Phelan 2013). The present study explores the predictive validity of new test formats that could be used in the admissions process alongside the established tests to identify students who are at risk. In the long-term these new test formats have the potential to form the basis of a stand-alone admissions test. The formats under investigation are a gap-filling test (C-test) as test of general language proficiency and several measures of lexical richness (Guiraud, Guiraud Advanced, "D"; see also methodology). All measures were taken at the beginning of an academic year to predict the average grades at the end of the academic year (General Points Average/ GPA). In total 107 international students, mainly from China, with a wide range of subjects participated in the study. A multiple regression analysis including hierarchical models shows that the C-test and "Guiraud" have the highest predictive validity. Given the fact that these two measures can be administered quite easily in a short period of time, we conclude that they have clear advantages over IELTS in the prediction of international student's study success.

Key words: International students, Academic success, C-test, IELTS, lexical richness ("D", Guiraud's index, Guiraud Advanced)

[1] We would like to thank the anonymous reviewers of the journal and Saskia Ayşe Daller for their valuable comments on an earlier version of this article.

1. Introduction

Higher Education (HE) in English speaking countries has been internationalized rapidly in the past few decades. The number of global international mobile students in HE has risen from 0.8 million in 1975 to 4.3 million in 2011 (UK Gov., 2013). The UK attracted 435,500 international students in the academic year 2013/2014 (UKCISA, 2015), which is an increase of 3% compared with the previous academic year. Asian countries, especially China and India, constitute the main source of mobile international students. According to the UK Council of International Student Affairs (UKCISA, 2015), the biggest group of international students in the UK is Chinese students with approximately 20% (87,895) of all international students. The second largest group comes from India with 19,750 international students. Being able to speak and understand the language of instruction (English) is one of the major concerns of these international students as well as the host institutions (Pathirage, Morrow, & Walpitage 2014). In recent years, there has been a growing discussion among students, lecturers and universities about the predictive validity of established admissions tests, such as the International English Language Testing System (IELTS). Studies on the predictive validity for study success of IELTS or the Test of English as Foreign Language (TOEFL) are inconclusive (for an overview see Daller & Xue 2009, Daller & Phelan 2013) and a number of possible threats to the reliability and validity of these tests have been pointed out (see section 3). University lecturers noticed that a large number of international students are unable to understand many learning and teaching activities even if they have passed the required English language ability test (Murray, 2010; Yen & Kuzma, 2009). We therefore need alternative test formats for the prediction of international student's study success.

2. Literature review

2.1 Issues with standardised tests

The use of standardised tests, e.g. IELTS and TOEFL, as admissions test for students with a first language other than English is rapidly growing in English-speaking countries. In 2011 nearly two million candidates sat the IELTS exam worldwide, an increase of around 200,000 than the previous year (IELTS, 2011). For TOEFL no detailed numbers of test takers are available, but the test developers of TOEFL claim that more than 9,000 institutions in over 130 countries accept TOEFL scores (ETS, 2016). A number of studies investigate threats to the reliability and validity of these tests. In the present study we focus mainly on IELTS since

our data are drawn from a British context where IELTS is the predominant entrance test. Over 800 educational institutions in the UK accept IELTS (IELTS, 2016). However, the discussion below might also be relevant for other international tests, including TOEFL.

In a summary article, Uysal (2010) discusses threats that reduce the reliability and validity of IELTS. These issues include the single marking issue (Uysal, 2010), inconsistent rating behaviour (Mickan, 2003), and topics chosen that might not be relevant for future students (Kroll & Reid, 1994). Given that these sources of errors can largely reduce the reliability and validity of the test, Uysal (2010: 3) suggests "more research is necessary especially in the areas of raters, scale, task, test taker behaviour, and topic comparability to diagnose and minimize sources of error in testing writing".

In terms of validity, Moore and Morton (2005) compare the standard IELTS task 2 with a corpus of 155 university assignment tasks in terms of genre, information source, rhetorical function and object of enquiry. The results show that, although IELTS writing and university essay assignments share some similarities, there are also clear differences between the two. Moore and Morton conclude that "the type of writing the test [IELTS] elicits may have more in common with certain public nonacademic genres and thus should not be thought of as an appropriate model for university writing" (2005: 43). Recently IELTS changed the topics of the writing tasks to make them more similar to the writing requirements of universities. It is, however, unclear whether these new writing tasks are appropriate.

The probably most serious threat to the validity of IELTS is the specific test preparation that many training centers provide worldwide. This specific preparation turns IELTS from a test of English ability into a test that measures learners' test taking skills (Yu, 214). The training for the IELTS exam that is offered amongst others to Chinese students helps them to find a "short cut" to improve the candidates' scores in a short time and turns the test "into a game of skills and strategies" (Yu, 2014: 25). This is obviously a threat to the validity of IELTS as an admissions test, but it might also affect the predictive validity of this test for the study success of students who underwent intensive training with a main focus on improving their test scores rather than their general English language proficiency (Hayes, 2003).

While the ability of established, standardized tests to predict the study success of international students has been the focus of many studies, the findings of these studies are inconclusive. Some studies found significant correlations between IELTS/ TOEFL scores and academic success (Bellingham, 1993; Elder, 1993; Feast, 2002; Ferguson & White, 1992; Hill, Storch,

& Lynch, 1999; and Yen & Kuzma 2009), but others found little or no significant correlations between these tests and academic success (such as Cho & Bridgeman, 2012; Cotton & Conrow, 1998; Dooey & Oliver, 2002; Gibson & Rusek, 1992; Wongtrirat, 2010; and Yule & Hoffman, 1990). One can argue that IELTS and TOEFL tests were not developed to predict study success per se but to provide a cut-off point below which students should not be admitted to HE in English-speaking universities. However, admissions officers still make their decisions based on these test scores and assume students are likely to succeed once passing a threshold of the tests (see Dooey & Oliver, 2002: 38-41). Therefore, even if the tests were not specially developed to predict study success, this predictive aspect of test validity is still implicitly assumed.

However, there is a mathematical argument why valid tests for admissions do not necessarily have a high predictive power for study success. Any correlation between two variables is automatically very low or zero if one of these variables only has a small variance. Ferguson and White (1992), and Daller and Phelan (2013) point out that almost all studies that focus on the predictive validity of IELTS/ TOEFL are based on truncated samples for these test scores, and that the magnitude of the correlation is depressed simply because of the limited range of the test scores. The reason for this is that universities will normally not accept students below a certain cut-off point and that students themselves will not delay the start of their courses once they pass the cut-off score. For these reasons the scores of students who are admitted are quite similar with a low variance between them, and from a purely mathematical point of view, high correlations between the test scores from truncated samples and the achievement scores (e.g. Grade Points Average (GPA)) are very unlikely. Therefore, it is necessary to develop additional English language tests for the prediction of study success that overcome the issues discussed so far.

The fact that international students who pass the admission exams still struggle with their studies is the reason why many universities introduced Post-Entry Language Assessment tests (PELA) to identify students at risk. According to Dunworth (2009) more than a third of universities in Australia are currently using PELA, with more universities considering its introduction. Detailed diagnostic tests for the whole student body are time consuming and costly and not practical. In this context the approach described by Elder and Randow (2008) seems to be more practical, where first a short screening test of 20 minutes length is administered to all students, and then in a next step an in-depth diagnostic test (2 hours) is used for those students who are identified to be at risk.

2.2 Vocabulary knowledge and academic success

The concept of vocabulary knowledge is closely related to the term lexical richness. We use this term in this article as a cover term for several aspects of vocabulary use, including lexical diversity (the rate of repetition of new words) and lexical sophistication (the use of infrequent words; for a detailed discussion see Read (2000). Many studies suggest that vocabulary knowledge is closely connected with various measures of English language ability and academic success, and that inadequate vocabulary knowledge can jeopardize the study success of international students (Alderson, 2005; Daller & Xue, 2009; Daller & Phelan, 2013; Elder & Randow, 2008; Harrington & Roche, 2014a, 2014b; Laufer & Goldstein, 2004; Morris & Cobb, 2004; Roche & Harrington, 2013; Saville-Troike, 1984; Yixin & Daller, 2015).

Saville-Troike investigates the academic success of children in school settings where the language of instruction is English as a second language. She comes to the conclusion that "vocabulary knowledge is the single most important area of second language (L2) competence" (1984: 199), and that this is crucial for their academic achievement. These findings for school children are equally relevant for university students. A study with high-school students conducted by Laufer and Goldstein (2004) found that 42.6% of the participants' grades could be explained by vocabulary knowledge. According to a study by Alderson (2005), significant correlations ($r = .61 - .79$) were found between the DIALANG subtests in reading, listening, writing and grammar and vocabulary test scores (see Alderson 2005: 25-43 for the introduction and history of DIALANG).

Morris and Cobb (2004) examined the correlation between vocabulary profiles based on writing samples and the academic achievement of 122 undergraduate university students. They found that students' vocabulary profiles correlated significantly with their grades and that these profiles are therefore a useful predictor of academic achievement. Harrington and Roche conducted a series of studies between 2013 and 2014 focusing on timed Yes/No vocabulary recognition tests to detect academically at-risk students in an undergraduate English-as-a-Lingua-Franca (ELF) university programme in Oman. Vocabulary size was identified as the best predictor of GPA in their studies.

Daller and colleagues (Daller & Xue, 2009; Daller & Phelan, 2013; Yixin & Daller, 2015) carried out a series of studies with measures of vocabulary knowledge and C-tests as a meas-

ure of general English language ability to predict the study success of international students. A combination of the vocabulary measures and the C-test score could predict between 23 and 40% percent of students' GPA. The C-test especially seems to be a good predictor of study success.

The C-test test format is a further development of the Cloze test (Taylor, 1953) by Klein-Braley (1981). While the Cloze test is based on whole word deletions, the C-test format deletes every second half of every second word. Some researchers point out that the C-test has a clear lexical focus (Little & Singleton, 1992; Stemmer, 1992) whereas others (Dörnyei & Katona, 1992; Eckes & Grotjahn, 2006) conclude from the high correlation between C-test scores and tests of other language skills that the C-test measures general language proficiency. However, these two views support rather than contradict each other since vocabulary knowledge is the basis of all language skills (Milton, 2013; Stæhr, 2008).

Dörnyei and Katona (1992) carried out a study on the C-test with 102 EFL university students and 52 secondary school students in Hungary. They compared C-test scores with scores from vocabulary, grammar, listening, and reading tests (English Language Department Test of the University of Budapest), TOEIC scores (listening and reading) and Cloze test scores. They found significant medium and high correlations between the C-test and all other tests (e.g. C-test/ TOEIC: r = .62) and therefore come to the conclusion that "The C-test proved to be a highly integrative language testing instrument, assessing general language proficiency" (1992: 202). In a similar vein, Eckes and Grotjahn (2006) compared the C-test scores of 843 learners of German as a Foreign Language with the reading, listening, writing and speaking parts of the German TestDaF (a widely accepted university entrance test in Germany). The C-test turns out to be unidimensional but also correlates highly with all other test scores, which supports the argument that the C-test measures general language ability and is therefore a good candidate for predicting the study success of international students.

In the present study we use a C-test and a writing task, and analyse the writing task with a series of measures of lexical richness (for a detailed discussion on lexical richness see Daller, Milton, & Treffers-Daller, 2007). The different measures that we use in the present study are discussed under "procedures".

2.3 English Language Proficiency

In the context of HE the notion "English language proficiency" is used in different ways. In an attempt to give a definition of this notion Murray (2010) divides it into three sub aspects, general language proficiency, academic literacy and professional communication skills. He argues that English language proficiency which includes generic skills and abilities such as listening, grammar, and vocabulary should be the focus of the assessment of international students as academic writing and professional communication skills need to be developed by all students after admission to HE. In line of this argument we focus a major generic aspect of English proficiency, vocabulary.

3. Hypotheses

The findings discussed in the literature lead us to the following hypotheses:

3.1 The IELTS scores of international students will only have a limited predictive validity for their study success as measured by GPA.

3.2 C-test scores as a measure of general language proficiency will have a higher predictive validity than IELTS.

3.3 Measures of lexical richness will have a high predictive validity for study success.

3.4 A combination of C-test and lexical richness scores will be the best predictor of study success.

4. Methodology

4.1 Participants

All participants in the present study (n = 107) were enrolled as international students at Swansea University. About 87% (94) of the participants were Chinese native speakers, and the others came from Saudi-Arabia, Korea and Japan. They took the test at the beginning of the academic year. Their average IELTS score was 6.0. The participants came from a range of subject areas, including English, Engineering, Mathematics, Media, Politics etc.. About 2/3 of the participants were undergraduates and 1/3 were Master's students. Half of the participants were gathered from the Academic Success Programme, through which free academic English lessons are provided for all international students. The participants were informed about the voluntary nature of the study and that they could withdraw from the study at any

time. They were also ensured about the confidentially of the test results and the use of their academic records.

4.2 Measures

We used a C-test and a writing task. All tests were piloted before being given to the participants. We screened the C-test with the programme "Vocabprofile" (Cobb, 2002, URL) and replaced subject specific vocabulary with more general lexical items.

4.2.1 The C-test

The C-test was adapted from an on-line resource (UAB, 2013) and the C-test used in Daller and Phelan's (2013) study. After the pilot study, a final version with five sub-texts was used (see appendix). We familiarized the participants with the C-test format by giving them three examples with solutions before we administered the actual test. In each sub-text every second half of every second word was deleted according to the classical C-test principle (see Klein-Braley, 1997). Based on the pilot study we arranged the sub-texts of the C-test according to difficulty (starting with the easiest sub-text). In total there were 100 gaps which gave a maximum score of 100. We used exact scoring, accepting only entirely correct answers.

4.2.2 The Writing task

The writing task was adapted from practice IELTS materials (Milton, Bell, & Neville, 2001). In the writing task the topic "tourism" was given and students were asked to produce a written text in the allotted time (30 minutes) with no lower or upper word limit, unlike the writing task in IELTS (40 minutes) where normally a minimum of 300 words is required. Students were also told that scores would be given based on both the quality and quantity of the written material produced. This writing task formed the basis of several measures of lexical richness (see "procedures"). In addition to the writing task we administered a short questionnaire about demographic data from the participants, and also asked for the IELTS scores the students achieved prior to admission to the university.

4.2.3 GPA

Students at undergraduate level are required to obtain 120 credits a year; while Master students are required to obtain 120 credits in their taught modules and another 60 in their thesis. In the present study the academic success was measured by students' overall GPA based on the average grade marks obtained from 120 module credits. For MA students the GPA was calculated from their taught sessions only, excluding the dissertation. Based on an overview of a series of studies, Bacon and Bean (2006) strongly advocate the use of GPA in educational research because of the high correlation with other educational variables such as motivation, achievement striving, team work and final academic achievement.

4.3 Procedure

The C-test and the writing task were administered in a pen-paper format. Students were given 25 minutes for the C-test and 30 minutes for the writing task. The writing task was transcribed to allow for a computerized analysis with different measures of lexical richness (see below). Spelling mistakes were corrected, abbreviations were extended and proper nouns were deleted to avoid them being counted as "infrequent words" by some of the programmes. The programme "Vocabprofile" (Cobb, 2002, URL) was used to calculate types and tokens for the writing task. Three measures of lexical richness where then calculated:

- Guiraud's index = Types/ √Tokens
- Guiraud Advanced (GA) = Advanced Types/ √Tokens
- and "D"

Guiraud's index (Guiraud, 1954) is a simple mathematical transformation of the Type-Token-Ratio (TTR), which can help overcome the problem of a systematically falling TTR with increasing text length. For the same speaker/ writer a longer text automatically has a lower TTR as words need to be repeated and the probability of occurrence of a unique new word (Type) decreases. It is therefore not possible to compare the lexical richness of texts with different lengths as they occur in natural language (Treffers-Daller, Parslow & Williams, in press). Guiraud's index is a measure of lexical diversity and it is based on the occurrence of types and tokens. In this index every type is equally weighted and no distinction is made between the use of basic (frequent) and advanced (infrequent) words by the speaker/ writer.

Guiraud Advanced (GA) has been developed to include information about the use of infrequent words in the analysis (Daller, van Hout, & Treffers-Daller, 2003). For GA we defined all types that are beyond the 2K level as advanced (based on the British National Corpus and Nation's Range Programme (Nation, 2015)).

The measure "D" follows a different approach to overcome the methodological problem of the systematically following TTR. "D" is the single parameter of a mathematical function that models the falling TTR (Malvern, Richards, Chipere, & Duran, 2004), and texts with a higher lexical richness yield a higher value for "D" than texts with a lower one. We computed "D" with the VocD command under CLAN (MacWhinney, 2000).

The measures listed so far were used as predictor variables for the academic success of students as measured by the participants' overall GPA as computed on the basis of their marks at the end of the academic year.

5. Results

5.1 Preliminary analysis

Since the participants were from different subject areas with possible different marking schemes, it could be argued that the overall GPA is not reliable and valid enough to make a comparable analysis. We therefore carried out a one-way ANOVA to compare the marks given by the different departments of our participants. The results from the ANOVA analysis showed that there was no significant difference between the marks given in the different subject areas ($F(7, 98) = 1.294$, $p = .261$). Furthermore, the internal consistency as an indication of the reliability of the C-test was analysed with Cronbach's alpha. As pointed out in the methods section, the C-test in the present study consisted of 5 sub-texts, and we used these sub-texts as super-items in the analysis. The gaps in each sub-test were not analysed individually as they are not independent because they are part of one running piece of text. The C-test in the present study yielded a value for Cronbach's alpha of .847 (5 items). The C-test is therefore sufficiently reliable to make judgments on individual participants (Meuffels, 1992: 147; but see for a critical discussion Field, 2009: 675).

5.2 Correlation analysis

In the present study we used five predictor variables (Guiraud, Guiraud Advanced (GA) and "D" as measures of lexical richness, the C-test scores as measure of general language proficiency and IELTS). The descriptive statistics of the five predictor variables is shown in Table 1. Due to a missing value, all predictor variables except the D measure are based on the full number of participants.

Table 1 Descriptive statistics for five predictor variables (IELTS, C-test, Guiraud, GA, and D), (n = 107)

	N	Minimum	Maximum	Mean	Std. Deviation
IELTS	107	5.0	7.5	6.14	.5486
C-test	107	32	91	63.27	12.796
Guiraud	107	3.61	10.5	8.17	1.01688
GA	107	.37	2.72	1.2959	.42403
D	106	40.61	133.29	75.8025	18.05201

The correlation of these predictor variables with GPA is shown in Table 2. All predictor variables correlate significantly with GPA. This is an indication of the validity of these tests as all of them had been identified on theoretical grounds to be relevant for study success. C-test scores and Guiraud show the strongest correlation with GPA. The C-test and Guiraud each explain about 21% of the average mark at the end of the academic year (R^2 = .208). The strongest correlation between two predictor variables is that between IELTS and the C-test. The shared variance between these two variables is about 45% (R^2 = .448).

Table 2 Correlation (Spearman) between variables in the present study (n = 107)

	GPA	IELTS	C-test	Guiraud	GA	D

GPA	-	.377**	.457**	.457**	.201*	.281**
IELTS		-	.670**	.338*	.301**	.051 ns
C-test			-	.433**	.369**	.113ns
Guiraud				-	.599**	.575**
GA					-	.279**
D						-

* = significant at the .05 level; ** = significant at the .01 level

We also investigated whether the differences between the correlation coefficients are statistically significant, which would give us some indication of the predictive power of the different predictor variables[2]. Apart from the trivial difference between the non-significant and the significant correlation coefficients there are no significant differences between the correlation coefficients in Table 2. Therefore, we need a more detailed analysis to gain more insight into the predictive validity of the different predictor variables.

5.3 Regression analysis

First, we carried out a multiple regression with GPA as the dependent variable and Guiraud, GA, IELTS, D and C-test scores as independent variables (method: Enter). This led to a significant model ($F (5, 100) = 10.608$, $p < .001$) which explained 34.7 % (R^2) of the variance of GPA. Significant variables in this model are the C-test (Std. Beta = .330, $p < .01$) and Guiraud (Std. Beta = .296, $p < .05$). IELTS, D and GA are not significant in this model. There is no indication for multicollinearity (all values for tolerance are > .02 and all values for VIF < 5).

To investigate the contribution of each predictor variable further we carried out several hierarchical multiple regressions. First we were interested in the unique contribution of IELTS towards the predictive power of a regression model. We carried out a hierarchical regression with all proficiency measures as independent variables (the C-test, Guiraud, Guiraud Ad-

[2] We used the software available under the following URL for these computations: http://www.danielsoper.com/statcalc3/calc.aspx?id=104

vanced and D) in block 1 and IELTS in block 2 and GPA as the dependent variable (cases excluded listwise). The variables in block 1 (the C-test, Guiraud, GA and D) produce a significant model (ANOVA, F (4,101) = 13.046, p < .001) that predicts 34.1 % (R^2) of GPA. Significant predictor variables in this model are: the C-test (Std. Beta = .402, p < .001) and Guiraud (Std. Beta = .288, p < .05). GA and D are not significant in this model. There is no indication for multicollinearity (all values for tolerance are > .02 and all values for VIF < 5). Introducing IELTS in the second block in this hierarchical multiple regression did not lead to a significant increase in R^2. This means that IELTS does not make a significant unique contribution when added to the model.

Secondly, we were interested in the unique contribution of the C-test towards the predictive power of a regression model as this test format has been successfully used in various studies on predicting study success (see literature review). We carried out a hierarchical regression with all proficiency measures including IELTS but not the C-test in block 1 and added the C-test in block 2. The variables for block 1 (IELTS, Guiraud, GA and D) produce a significant model (ANOVA, F (4, 101) = 10.712, p < .001) that predicts 29.8 % (R^2) of GPA. Significant predictor variables in this model are: IELTS (Std. Beta = .293, p < .01) and Guiraud (Std. Beta = .412, p = <.01). GA and D are not significant in this model. We introduced the C-test in the second block of this hierarchical multiple regression, which led to a significant change to R^2 (ΔR^2 = .049, F change (1,100) = 7.454, p < .01). Model 2 is significant (ANOVA, F (5,100) = 10.608), p < .001) and explains 34.7 % (R^2) of the variance of GPA. Significant variables in model 2 are the C-test (Std. Beta = .330, p < .01) and Guiraud (Std. Beta = .296, p < .05), but IELTS is not longer a significant variable in the second model, nor is D and GA. There is no indication for multicollinearity (all values for tolerance are > .02 and all values for VIF < 5).

Since Guiraud was a significant predictor in the two hierarchical multiple regression so far, we carried out a similar computation as above with Guiraud as added variable in the second block. Block 1 led to a significant model (F (4, 101) = 11.665, p < .001) which explains 31.6 % (R2) of the variance of GPA. Significant variables in this model are the C-test (Std. Beta = .412, p < .01) and D (Std. Beta = .338, p < .001). IELTS and GA are not significant in this model. There is no indication for multicollinearity (all values for tolerance are > .02 and all values for VIF < 5). We introduced Guiraud in the second block in this hierarchical multiple regression, which led to a significant change to R^2 (ΔR^2= .031, F change (1,100) = 4.68, p < .05). Model 2 is significant (ANOVA, F (5,100) = 10.608, p < .001) and explains 34.7 % (R^2)

of the variance of GPA. Significant variables in model 2 are the C-test (Std. Beta = .330, $p < .01$) and Guiraud (Std. Beta = .296, $p < .05$). IELTS, D and GA are not significant in this model.[3]

We also computed two hierarchical multiple regressions with the same variables as in the previous computations but with D or GA entered in block 2. There was no significant change in R^2 when these variables where entered in the second block. Overall, we come to the conclusion that only the C-test and Guiraud make a single significant contribution when added in the second block in a hierarchical multiple regression.

So far, the non-hierarchical and the hierarchical multiple regressions yielded two potential predictor variables: C-test and Guiraud. In order to find out how much variance of GPA these two predictor variables alone explain, we carried out a multiple regression (Enter) with only these predictor variables.[4]

This led to a significant model ($F (2, 104) = 20.871$, $p < .001$) which explains 28.6 % (R^2) of the variance of GPA. (Std. Beta for the C-test = .361, $p < .001$; and for Guiraud Std. Beta = .298, $p < .01$; Beta for the constant = 28.792). There is no indication for multicollinearity (all values for tolerance are > .02 and all values for VIF < 5). The unstandardized beta value for the C-test is .215 and for Guiraud it is 2.236. This means that the 28.6% of the average mark of the students at the end of the academic year can be predicted with the following regression line:

GPA = 28.792 + .215 x C-test scores + 2.236 x Guiraud's index.

This is a remarkable result given the fact that this prediction of GPA at the end of the academic year was made with test scores that were obtained eight months earlier at the beginning of the academic year.

6. Discussion and conclusion

Our findings support Hypotheses, which states that IELTS scores will only have a limited predictive validity for the study success of students. Although the IELTS scores show a sig-

[3] The explained variance for this hierarchical model and the previous one is the same as all variables entered up to block 2 are similar.

[4] Although IELTS was not part of this computation, we used the same 107 participants who were part of the earlier computations to ensure comparability.

nificant correlation with GPA, they predict only 12% of the average marks. This is much less than the predictive power of the C-test and Guiraud, which each predict 21% of the marks. Therefore the data provide support for hypotheses 2 and 3. One has to bear in mind that the C-test is a much shorter test (25 minutes) than IELTS and can easily be administered and marked objectively. Hypothesis 3 states that vocabulary knowledge is an important factor for study success. This is supported by the fact that all measures of lexical richness used in the present study correlate significantly with GPA, with Guiraud's index as the best predictor variable explaining almost 21% of the average mark. Hypothesis 4 states that a combination of the C-test scores and lexical richness measures would be the best predictor for study success. To investigate this in detail we carried out a series of multiple regressions. The results from several hierarchical regressions show a more fine-grained picture and are consistent with the results of the non-hierarchical regression analysis. Together, they can yield more robust results when compared with non-hierarchical multiple regressions because they reveal the unique contribution made by each variable to explaining the variance in the dependent variable.

Overall, we conclude that the best predictor variables in this study are the C-test and Guiraud. In a series of hierarchical multiple regressions both make a unique contribution to the explained variance of study success, while all other variables, including IELTS, did not. We conclude that vocabulary knowledge and general language proficiency are two key predictors of study success, and that these two aspects of language ability can be measured effectively with the measure Guiraud and tests based on the C-test format. This is in line with the findings of Dörnyei and Katona (1992) and Eckes and Grotjahn (2006) regarding the validity of the C-test for the measurement of general language proficiency as discussed in the literature review. Our findings show that this test of general language proficiency together with Guiraud as measure of vocabulary knowledge is a good predictor of study success.

7. Limitations and future directions

Our study was not intended to investigate the validity of IELTS scores as a cut-off point in the admissions process. We intended to find ways of predicting study success of students who passed the cut-off point and were admitted to university. Our test battery is therefore a post entrance language test to identify students at risk. Further studies are needed to decide a score range below which students are at risk. In this context, the predictive validity of IELTS for

students who pass this cut-off point is very limited. We show that a combination of two easy to administer tests, the C-test and a writing task can predict the study success of international students. Although this test battery predicts 30% of the students GPA at the end of the course, still 70% remains unexplained. Other factors such as motivation, academic literacy and professional communication skills in L1 and L2 (see Murray, 2010) might also play a role. A further factor that might be important is the ability of the students to adapt to a different culture under a different educational system. This factor needs to be taken into account in future studies.

A limitation of the present study is the fact that the students took part on a voluntary basis. The tests administered were probably seen as low stake tests and some students might not have taken them as seriously as, say, the IELTS exams. Nevertheless, the fact that despite this potential threat to the validity of the tests battery 30% of the GPA could be predicted is encouraging. Further studies should use the test battery as a compulsory post entry exam, making it a high stake test with a potentially higher predictive validity.

A further limitation of the present study is the fact that mainly Chinese participants were tested. Whether the test-battery has a similar predictive validity with other language groups needs to be investigated in further studies. When the results of further studies are available, the findings can be used for the development of a stand-alone entry exam, including the present test battery. In this context it is important to note that the two measures for study success suggested here can easily be administered within one hour and scoring can be automated, making human ratings and potentially subjective judgments unnecessary.

Appendix

English Language Ability Test

*Name:_____ *Student number: _____

*Age:_____ *Gender:_____

*IELTS score: _____ *Department: _____

*Major: _____ *Nationality:_____

*Have you taken pre-sessional language courses? _____

A. Yes B. No

*Which degree are you doing from September 2013? _____

A. Undergraduate

B. Postgraduate---Master

C. Postgraduate---Ph.D

D. Other (Please specify)_____

Task 1: Please fill in the missing letters. You have 25 minutes in total. Roughly the second half of the word is deleted. Be aware of the inflection such as *third person singular, plural, past tense, etc*. Only words with correct spelling can score. There are 100 points available with the whole test, with 20 points available from each passage. Thanks for your interest in this test and your contribution will be highly appreciated.

Here are some examples to help your understanding:

1. I li__ to go with you.

Answer: like

2. I wo____ love to go with you.

Answer: would

3. They are teen_____.

Answer: teenagers

A. Learning to write

I was four when I started to learn to write. My grandfather started to teach me before I went to school. I remember th____ I always fo____ the cap_____ letters mu___ easier t__ write th___ the sm___ letters. I reme____ that on___ we sta____ writing i___ school w__ were n__ allowed t__ use pe___, we h___ to u___ pencils un___ we bec___ really go__ at writing. I can write a few characters in Chinese now, but not very many.

B. Teenagers

It is clear that not all teenagers respond in the same way to peer group pressure. For exa_____, young peo_____ in t_____ early ye_____ of seco_____ school a_____ more lik_____ to fe_____ under pres_____ to we_____ the sa_____ clothes a_____ listen t_____ the sa_____ music a_____ the re_____ of t_____ peer gr_____. By t_____ time th_____ reach middle or late adolescence, however, young people are more able to stand up against such influences.

C. Sleep room

One in three Japanese suffer from sleep disorders, which has prompted technologists to build a sound- insulated capsule – the Suimin (Sleep) Room. Users start by sit_____ up i_____ bed i_____ front o_____ a screen sho_____ a river win_____ through for_____. Soft mu_____ plays i_____ the backg_____, along wi_____ the so_____ of wa_____ and bird_____. After a f_____ relaxing min_____, the lig_____ dim, t_____ screen go_____ blank, t_____ music fades and the bed reclines into a sleeping position.

D. Geography

The UK is located on a group of islands known as the British Isles, which lie between the Atlantic Ocean and the North Sea, northwest of France. At i_____ widest t_____ UK i_____ 300 mi_____ across a_____ 600 mi_____ from No_____ to So_____. It sha_____ a sin_____ land bor_____ with the Irish Repu_____. Despite i_____ relatively sm_____ size t_____ UK boa_____ incredibly var_____ and of_____ very beau_____ scenery, fr_____ the mountains and valleys of the North and West to the rolling landscape of the South, and from downland and heath to Fens and marshland.

E. Record employment

Latest employment figures show that there are 28.2 million people in work. Work & Pensions Secretary said this showed the UK labour market has coped well so far with the current international economic uncertainty. He said: "Employment cont_____ to ri_____, with th_____ month's fig_____ showing a rec_____ 28.2 mil_____ people i_____ work. Th_____ are 65,000 mo_____ people i_____ work th_____ last qua_____ and 252,000 mo_____ than la_____ year. Alth_____ both meas_____ of unempl_____ have ri_____ slightly, th_____ are st_____ significantly lower than they were a year ago."

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*: A&C Black.

Bacon, D. R., & Bean, B. (2006). GPA in research studies: An invaluable but neglected opportunity. *Journal of Marketing Education, 28*(1), 35-42.

Bellingham, L. (1993). The relationship of Language Proficiency to Academic Success for International Students. *New Zealand Journal of Educational Studies, 30*(2), 229-232.

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT scores to academic performance: Some evidence from American universities. *Language Testing, 29*(3), 421-442.

Cobb, T. (2002). The Web vocabulary profiler. *Computer Software*. Retrieved August 16, 2013, from http://www. er. uqam. ca/nobel/r21270/textools/web_vp. html.

Cotton, F., & Conrow, F. (1998). An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania. *IELTS Research Reports, 1*, 72-115.

Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*: Cambridge University Press Cambridge.

Daller, H., van Hout, R., & Treffers-Daller, J. (2003). Lexical Richness in the Spontaneous Speech of Bilinguals. *Applied Linguistics, 24*(2), 197-222. Retrieved from http://applij.oxfordjournals.org/content/24/2/197.abstract

Daller, H., & Xue, J. (2009). English proficiency and academic success: a study of Chinese students in UK higher education. In R. Brian, H. Daller, D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The Interface Between Theory and Application* (pp. 79-193): Palgrave.

Daller, M. H., & Phelan, D. (2013). Predicting international student study success. *Applied Linguistics Review*, *4*(1), 173-193.

Dooey, P., & Oliver, R. (2002). An investigation into the predictive validity of the IELTS Test as an indicator of future academic success. *PROSPECT-ADELAIDE-*, *17*(1), 36-54.

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, *9*(2), 187-206.

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*(3), 290-325.

Elder, C. (1993). Language proficiency as a predictor of performance in teacher education. *Melbourne Papers in Language Testing*, *2*(1), 72-95.

ETS. (2016). Educational Testing Service. TOEFL. Retrieved February 2, 2016, from http://www.ets.org/toefl

Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal*, *3*(4), 70-85.

Ferguson, G., & White, E. (1992). A predictive validity study of IELTS –a report commissioned by UCLES: Institute of Applied Linguistics, University of Edinburgh.

Field, A. (2009). *Discovering statistics using IBM SPSS statistics*: Sage.

Gibson, C., & Rusek, W. (1992). *The validity of an overall band score of 6.0 on the IELTS test as a predictor of adequate English language level appropriate for successful academic study*. (Unpublished Masters of Arts (Applied Linguistics) thesis), Macquarie University, New South Wales.

Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*: Presses universitaires de France.

Harrington, M., & Roche, T. (2014a). Identifying academically at-risk students in an English-as-a-Lingua-Franca university setting. *Journal of English for Academic Purposes, 15*, 37-47.

Harrington, M., & Roche, T. (2014b). Word recognition skill and academic success across disciplines in an ELF university setting. *Association for Language Testing and Assessment of Australia and New Zealand (ALTAANZ): Papers in Language Testing, 3*(2), 76-99.

Hayes, B. M. (2003). *IELTS preparation in New Zealand: An investigation into the nature of the courses and evidence of washback* (Doctoral dissertation). Retrieved from http://hdl.handle.net/10063/751.

Hill, K., Storch, N., & Lynch, B. (1999). A comparison of IELTS and TOEFL as predictors of academic success. *IELTS Research Reports*, *2*, 52-63.

IELTS. (2011). IELTS Annual Review. Retrieved October 25, 2015, from https://www.ielts.org/.../8330_3Y03_IELTSAnuualReview2011_web.pdf

IELTS. (2016). Retrieved February 12, 2016, from https://takeielts.britishcouncil.org/choose-ielts/who-accepts-ielts/recognising-organisations?

Klein-Braley, C. (1981). *Empirical investigations of cloze tests: An examination of the validity of cloze tests as tests of general language proficiency in English for German university students.* Universitat Duisburg.

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

Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats, and cautions. *Journal of Second Language Writing*, 3(3), 231-255.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436.

Little, D., & Singleton, D. (1992). The C-test as an elicitation instrument in second language research. *Der C-Test. Theoretische grundlagen und praktische anwendungen, 1*, 173-192.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2): Psychology Press.

Malvern, D., Richards, B., & Chipere, N. Dur· n, P., 2004. Lexical Diversity and Language Development: Quantification and Assessment: Palgrave Macmillan, Hampshire, England.

Meuffels, B. (1992). *Methods and Techniques of Empirical Research: An Introduction*: International Centre for the Study of Argumentation (SICSAT).

Mickan, P. (2003). What's your score? An investigation into language descriptors for rating written performance. *International English Language Testing System Research Reports, 5*, 125-157.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis*, 57-78.

Milton, J., Bell, H., & Neville, P. (2001). *IELTS Practice Tests*. Newbury.

Moore, T., & Morton, J. (2005). Dimensions of difference: a comparison of university writing and IELTS writing. *Journal of English for Academic Purposes, 4*(1), 43-66.

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System, 32*(1), 75-87.

Murray, N. (2010). Conceptualising the English language needs of first year university students. *The International Journal of the First Year in Higher Education, 1*(1), 55.

Nation, P (2015). *Range Programme*. Retrieved January 8, 2015, from http://www.victoria.ac.nz/lals/about/staff/paul-nation

Pathirage, D. N. A., Morrow, J. A., Walpitage, D. L., & Skolits, G. J. (2014). Helpfulness of ESL courses for international students studying in the United States. *International Education*, *43*(2), 25.

Read, J. (2000). *Assessing vocabulary*: Cambridge University Press Cambridge.

Roche, T., & Harrington, M. (2013). Recognition vocabulary knowledge as a predictor of academic performance in an English as a foreign language setting. *Language Testing in Asia*, *3*(1), 1-13.

Saville‐Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly*, *18*(2), 199-219.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139-152.

Stemmer, B. (1992). An alternative approach to C-test validation. Der C-Test. Theoretische Grundlagen und praktische Anwendungen, 1, 97-144.Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. Journalism quarterly, 30, 415-433.

Treffers-Daller, J., Parslow, P. & Williams, S. (2016). Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: how measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, amw009.

UAB. (2013). Universitat Autònoma de Barcelona. Retrieved June 15, 2013, from http://wuster.uab.es/ctestpractice/

UK, G. (2013). UK Government Report: International education-global growth and prosperity, an accompanying analytical narrative. Retrieved September 16, 2015, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/340601/bis-13-1082-international-education-accompanying-analytical-narrative-revised.pdf

UKCISA. (2015). UK Council of International Students Affairs. Retrieved September 15, 2015, from http://institutions.ukcisa.org.uk//Info-for-universities-colleges--

schools/Policy-research--statistics/Research--statistics/International-students-in-UK-HE/

Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT journal*, *64*(3), 314-320.

Wongtrirat, R. (2010). *English Language Proficiency and Academic Achievement of International Students: A Meta-Analysis*: ERIC.

Yen, D., & Kuzma, J. (2009). Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students. *Worcester Journal of Learning and Teaching*(3), 1-7.

Yixin, W., & Daller, M. (2015). Predicting Chinese Students' Academic Achievement in the UK. *Learning, Working and Communicating in a Global Context, 217*.

Yu, Q. (2014). *Various Items Causing IELTS Test-Taker's Low Performance in Mainland China: an International Joint Education Program Solution.* Paper presented at the 2014 International Conference on Global Economy, Finance and Humanities Research (GEFHR 2014). Atlantis Press: Paris, France.

Yule, G., & Hoffman, P. (1990). Predicting success for international teaching assistants in a US university. *TESOL Quarterly*, 227-243.