Vocabulary Knowledge and Academic Success: a Study of Chinese Students in UK Higher Education

H. Michael Daller and Huijuan Xue
University of the West of England, Bristol

Introduction

The number of Chinese students undertaking international education has been increasing steadily over the past decade, and this upward trend is still ongoing (Goh, 2007; TEIU, 2008). Study failure is obviously a major concern for both international students themselves and universities in host countries. Previous attempts to relate the study success of overseas students to their English language proficiency have used scores from standardized tests such as the International English Language Test System (IELTS) or the Test of English as a Foreign Language (TOEFL). Whilst such tests are valid tools as entry tests (Rosenfeld, Oltman and Sheppard, 2004; Taylor and Falvey, 2007), they do not seem to be good predictors of academic success on their own. We therefore tried to find other measures to complement the information provided by IELTS by using data from 23 overseas students from China in the present study. Apart from their IELTS scores, we used two measures of lexical diversity (D and Guiraud) and two measures of lexical sophistication (Lexical Frequency Profile and Guiraud Advanced) based on their written essays. In addition, we employed C-tests which focus on vocabulary but also measure other aspects of foreign language proficiency (Eckes and Grotjahn, 2006). The correlations we found clearly showed that the C-tests focused on lexical sophistication rather than diversity. The present study revealed the C-test to be a powerful tool in that it allowed us to predict over one-third of the variance in the modules failed by the students during their first year in UK higher education (HE). We conclude that lexical sophistication is most closely related to overseas students’ academic success.

Second language proficiency and study success

The role of English language proficiency and its relation to the study success of international students have been the focus of many studies. Graham
(1987) reviewed several earlier studies, one of which was carried out with 2075 foreign students from 1964 to 1965 in the USA, showing that ‘English language proficiency, as measured by the institution’s own test, was one of the least predictive of the variables in the study \( (r = -0.046) \)’ (Graham, 1987, p. 509). One could argue that this raises questions about the validity of the test, for which no detailed descriptions were available. However, a number of other studies Graham reviewed similarly revealed the weak correlations between language proficiency test scores, for instance TOEFL results at entry to HE, and subsequent academic success. She drew the conclusion that there was no clear-cut relationship between English proficiency and academic success. Many other factors seemed to be involved in the study success of international students, but it was likely that a minimum threshold level of English proficiency was required before other factors became important (Graham, 1987, p. 517). Still, exactly what constitutes this minimal level remains unclear.

Patkowski (1991) carried out a study with 271 EFL students. He used a university entrance test which included reading, writing and maths subtests to predict grade point average (GPA), but only 13.7 per cent of the variance of GPA scores could be explained even when all three subtests were combined in a multiple regression. He came to the conclusion that these proficiency tests were poor predictors of academic success. Comparable findings were reported in a study of 89 students at an Australian university where Dooey (1999) investigated whether students’ IELTS scores at entry were related to the average marks they obtained in the first two semesters. The study showed that IELTS scores were generally not related to study success. Only the reading subtest revealed a moderate correlation (.396) with the marks.

On the other hand, Yule and Hofman (1990) approached the correlation between EFL proficiency and academic success from a different angle. They carried out a study with 233 international graduate teaching assistants. The dependent variable was whether or not they received positive or negative recommendations for the assignment of teaching duties after a two-year study period. The authors found that there was a difference in the average TOEFL scores at the beginning of the course between the group that were successful in getting a teaching contract at the end of the course and those who were eventually unsuccessful. The mean score of the former was 607 (SD = 39) and the mean for the unsuccessful students was 560 (SD = 25). At first sight, this difference of approximately 10 per cent did not seem to be very large, but statistically it was highly significant \( (p < .001) \) (Yule and Hoffman, 1990, p. 231). Although the authors did not report the effect size, this could be easily computed based on the reported \( t \)-value and the degrees of freedom \( (t(231) = 9.34) \) (Field, 2005). In this case, it amounted to an effect size of \( R^2 = 0.275 \).

Feast (2002) made a further study of IELTS scores and academic success. The subjects were 101 international students at an Australian university who were monitored over a maximum period of five semesters. The results
showed that the IELTS scores had a statistically significant relationship with study success. However, other factors, including the country of birth, were significant as well. Feast concluded that ‘Chinese born students [were] likely to be relatively more successful than non-Chinese born students with the same IELTS score, all other variables being equal’ (Feast, 2002, p. 79). Overall, this study convincingly confirmed that there was a relationship between IELTS scores and study success but that many other factors also played a role. It is important to note here that the variability in the IELTS scores was higher than in other studies with a range of 4.5–8.5 (Feast, 2002, p. 75). This meant that students were included in this study who would not have been admitted in other institutions as the minimum entry requirement is normally an IELTS score of 6.0. It was therefore not clear whether the findings could be generalized to a situation where all students meet this minimum requirement.

Brooks and Adams (2002) investigated the correlation between study success and reported foreign language use in an Australian context. They made a comparative study of 32 international students with 112 local students and found that the marks of the former were lower than those of the latter. Likewise, the reported use of English outside teaching hours was also lower for the international students. Despite the absence of information about overseas students’ first language, their lower reported use of English indicated that their first language (L1) was not English. According to the researchers, the use of the second language (L2) had some effect on students’ academic success. Factors like educational culture or lower English language proficiency might also explain the lower marks of international students, although Brooks and Adams did not attempt to include such factors in the research design.

Bayliss and Raymond (2004) reported two studies where the scores in a language proficiency test were linked to study success. The participants in the first study were 34 Chinese overseas students enrolled on a Master’s programme in Business Administration. The students were tested twice, first at the start of the programme in April 2000 and then in November 2000. The test used was the CanTEST from the University of Ottawa which included the subtests listening, reading, ‘scim-scan ability’, writing and a Cloze test. Study success was defined as the average mark (GPA) from the modules that the students took in this first part of their MA programme. Significant correlations were found between the GPAs and the April listening test \( r = .49 \), the November Cloze Test \( r = .34 \) and the November reading comprehension test \( r = .62 \). Interpretation of the data was difficult but one possibility could be that listening skills were important before and at the beginning of a programme whereas reading skills became more important later in their studies. Bayliss and Raymond reported that the majority of students had great problems with the volume of reading required.

In the second study discussed by Bayliss and Raymond (2004), the subjects were 136 students enrolled on a law course taught in French.
The students were classified as either being Francophone (n = 100) or Anglophone (n = 36). Both groups completed a French test which included subtests for listening, dictations, reading, writing and error correction, an item requiring students to identify errors in a written text. Weak to moderate statistically significant correlations were found between the academic success of the Francophone students and their scores in all subtests. This meant that all aspects of language proficiency played a role in academic success for such students who were taught in their first language. By contrast, the only significant correlation for the Anglophone students was between reading comprehension and average marks (r = .37). One should bear in mind that the two groups differed substantially in size and that other correlations for the Anglophone students might not be statistically significant because of the small size. Nevertheless, such findings once again indicated that reading comprehension was an important part of study success for international students and, as revealed in other studies in a monolingual and bilingual context, for instance studies by Qian (1999) and Ransdell (2002), reading comprehension is closely related to vocabulary knowledge.

Most of the studies reviewed so far operated with scores from tests such as IELTS or TOEFL which, as reported earlier, might be a problem because the variance in test scores of students entering HE is normally low. This might explain the inconsistency in findings, with either no significant correlations or only weak ones between study success and language proficiency in most studies. Tests such as IELTS and TOEFL include vocabulary knowledge as part of band descriptors (IELTS, 2008). However, in recent years there has been a shift in applied linguistics towards a clearer focus on vocabulary as a central part of proficiency in L1 and L2. This shift has been documented by a number of books and special issues of journals (Daller, Milton and Treffers-Daller, 2007a; Malvern, Richards, Chipere and Durán, 2004; Nation, 1990; Read, 2000; Treffers-Daller, Daller, Malvern, Richards, Meara and Milton, 2008). Various studies (Daller, 1999; Daller, van Hout and Treffers-Daller, 2003) have clearly revealed vocabulary knowledge as one of the major aspects of foreign language proficiency.

One study with a focus on the correlation between vocabulary and academic success in the context of Teaching English as a Second Language (TESL) was carried out by Morris and Cobb (2003). They examined the language proficiency of 122 TESL trainees and analysed the vocabulary in 300-word samples of their writing with software based on frequency lists. The frequency bands were the most common 1000 words (K1), the next most common 1000 words (K2), the Academic Word List (Coxhead, 2000) and off-list words. In addition, they took the percentage of function words into account. Quite contrary to earlier findings by Morris and Tremblay (2002) that more proficient EFL students used more function words, which seemed at first sight counter-intuitive, Morris and Cobb (2003) found a negative correlation (r = -.34) between the use of function words and the
marks that students obtained for the two obligatory courses they took. The highest correlation \((r = .37)\) was found between the students’ marks and the score that they received on the Academic Word List. The authors concluded that although this was clear evidence that vocabulary profiles could predict study success to a certain extent, they were unsuitable as a single, stand-alone measure for decisions about admission to TESL training programmes.

In summary, we conclude that there is a correlation between academic success and foreign language proficiency but that such relationships may be difficult to identify for methodological reasons. The low variance of some variables used makes it hard to obtain high correlations. In addition, many other variables apart from language proficiency seem to influence students’ academic success. The existing literature does not allow us to specify the extra-linguistic variables in detail. However, good candidates for successful linguistic predictors seem to be measures that focus on all skills, such as the average overall IELTS scores, measures of vocabulary knowledge and reading proficiency. The present study will therefore try to relate study success to the vocabulary knowledge and IELTS scores of Chinese students undertaking HE in the UK.

**Measures of vocabulary knowledge**

The measurement of lexical knowledge is a complex issue since there are various distinct aspects. Four aspects of lexical knowledge or lexical richness have been identified by Read (2000): lexical variation or diversity, lexical sophistication, lexical density and number of errors. We will focus on the first two aspects because they have been used most widely in recent research. Lexical diversity is the range of vocabulary and the avoidance of repetition and lexical sophistication is the use of sophisticated vocabulary. We include measures of lexical diversity and lexical sophistication in this study to find out which aspect of vocabulary knowledge is more important for study success.

**D as a measure of lexical diversity**

A number of previous studies have shown that lexical diversity measured by the relationship between the total number of words (tokens) and the number of different words (types) in a text, can be a good indicator of vocabulary knowledge. Johnson (1944) introduced the type-token ratio (TTR) as measure of lexical diversity and it has been used widely. TTR is, however, problematic as the probability of the occurrence of new words in a text decreases with text length. Therefore TTR is a function of text length and not suitable for comparing texts of different lengths. The index \(D\) (Malvern et al., 2004) was developed to overcome this problem. \(D\) is the single parameter in the equation for this falling TTR against token curve and allows comparison of speakers or writers irrespective of the length of the text produced. The higher the \(D\), the greater the lexical diversity of the text. It is an indication
of a combination of the vocabulary resources of the author of a text and his or her skill in deploying them in a way that minimizes repetition.

This measure has been used in research on L1 development but also to distinguish between L2 learners at different levels. Tidball and Treffers-Daller (2007) showed that it discriminated clearly between L2 learners of French at different levels and native speakers and the effect sizes (Eta²) for differences between groups in this study have a range from .61 to .67. Similarly, Daller and Xue (2007) showed that D was the vocabulary measure that discriminated most clearly out of six measures of vocabulary richness between two groups of Chinese EFL learners. Therefore, we conclude that D is a good candidate when relating vocabulary knowledge to study success of international students. In addition to D, we computed the index of Guiraud (types/square root of tokens) which has been used widely as a measure of lexical diversity (see also Treffers-Daller, this volume).

**Lexical sophistication**

A further aspect of lexical knowledge is lexical sophistication, the use of low-frequency or ‘difficult’ vocabulary. This aspect has been investigated in various educational contexts, including bilingual children (Daller et al., 2003; Vermeer, 2001). To measure lexical sophistication it is necessary to find a criterion to define the difficulty and/or frequency of words. One operationalization is based on frequency lists. In the present study, we used the program ‘Range’ (Healey, Nation and Coxhead, 2002; Nation, URL) which made use of three bands: 1K, 2K and 3K. Words not within the bands were classified as ‘Not on List’ (NoL). The first two bands (K1 and K2) were based on West’s (1953) word list, and the third, on the Academic Word List compiled by Coxhead (2000).

The output of the program gave the percentage of words from each list and allowed us to create a lexical frequency profile (LFP) for each text. The output also allowed us to compute Guiraud Advanced (advanced types/square root of tokens). We classified as advanced all types that were not in the K1 and K2 list. This measure has been used successfully in the description of the academic profile of bilinguals (Daller et al., 2003). The measure Guiraud Advanced has also recently been used in a study based on 55 essays written by young adults in their first language (Wray, A., Mollet, Fitzpatrick, Wright and Wray, N., in preparation). The exact details are not yet published but a first result is that there are strong correlations between a verbal IQ test and Guiraud Advanced (personal communication from Eugene Mollet). The test used by Wray et al. was the verbal IQ subtest of the Multidimensional Aptitude Battery MAB-II (Sigma Assessment Systems, URL).

**The C-test as a measure of vocabulary knowledge**

From the literature review above we conclude that academic success can be predicted to a certain extent with standardized tests, such as IELTS, and that reading proficiency and vocabulary knowledge seem to be important factors
for study success in international education. A test format closely associated with reading proficiency and vocabulary is the C-test. This test format is a further development of the Cloze test but instead of deleting whole words only the second half of every second word is deleted. It has been used in more than 200 L1 and L2 studies (Grotjahn, 2007) and in recent research on L1 attrition (Opitz, 2008).

C-tests consistently yield significant correlations with all other aspects of language proficiency, including oral proficiency in various studies (Eckes and Grotjahn, 2006; Sigott, 2004, 2006). For example, Arras, Eckes and Grotjahn (2002) reported a Spearman correlation of .64 between a C-test and a ‘simulated oral proficiency interview’ in a study of 145 learners of German. These significant correlations with all aspects of language proficiency led many researchers to the conclusion that the C-test was a test of general language proficiency (Cohen, Segal, and Weiss, 1985; Grotjahn, 1992, 1995; Klein-Braley, 1985a, b; Raatz, 1985). However, the validity of the C-test and what it measures has been widely discussed. Alderson viewed its validity as a ‘worrisome question’ (2002, p. 28) and argued that ‘claiming that there was a unitary competence, or a general language proficiency … [was] now generally discredited’ (Alderson, 2002, p. 21).

The concept of a unitary competence underlying different skills in a foreign language was developed in the 1970s (Daller, 1999). The main argument for such a competence was the high correlations that could be found between tests of different aspects of foreign language proficiency. Despite Alderson’s claim that this hypothesis was now discredited, these high correlations were still found in later studies. As Singleton and Singleton (2002) noted:

High correlations have been found between sets of scores from tests purporting to measure grammatical knowledge and sets of scores from tests purporting to measure lexical knowledge, and there has been little success in attempts to demonstrate that ‘grammar tests’ and ‘vocabulary tests’ tap fundamentally distinct aspects of linguistic knowledge. (p. 154)

C-test scores are found to be correlated with scores in the four classical skills (reading, writing, listening and speaking) in many studies rather than with tests of vocabulary. However, a closer look at published research findings seems to confirm that the C-test operates largely at the lexical level. Vocabulary is relevant to all sub-skills; therefore the concept of general language proficiency can perhaps be reinterpreted, at least to some extent, as vocabulary knowledge.

Little and Singleton (1992) administered C-tests in French and German to university students (L2 learners), and analysed the items that posed particular difficulties. They viewed the test as ‘an instrument with a clear lexical focus’ (p. 175) and concluded that in filling C-test slots the subjects gave ‘priority to a ready lexical solution over morphosyntactic and more
general semantic issues’ (p. 188). Stemmer (1992) also carried out a study on the C-test with reading aloud protocols. She drew the conclusion that the informants operated predominantly within one meaning unit, whilst higher processing strategies at a macro level were less involved. This strongly suggests that the C-test does indeed have a lexical focus. Sigott (2004) administered decontextualized C-test items (containing truncated words only) and a C-test in its canonical form to 60 university students of English in Austria. He confirmed that students with a higher English proficiency operated at the lexical level when filling in C-test gaps. Students with a lower proficiency depended more on contextual information than those with a higher proficiency. In Sigott’s study, English proficiency was measured with the Oxford Placement Test (OPT) which showed a high correlation with the C-test results ($r = .83$). The OPT was meant to measure the foreign language proficiency of the students in relation to their potential academic success. The high correlation between the OPT and C-test scores indicated that the latter may also be a good candidate to measure academic success (see also below).

The view that the C-test is to a large extent a vocabulary test is also supported by other correlational studies. Grotjahn and Stemmer (1985) carried out a study with 115 students of French, adopting a C-test and the Bochum Diagnostic Test for French (BDTF). The highest correlations ($r = .63$) were found with the two subtests ‘pronouns’ and ‘verbs’ of the BDTF. Klein-Braley (1985b) used the Duisburg Diagnostic Test for English (DELTA) and a C-test over a period of four years with students ($N = 202$) in the English department. She found high correlations between all subtests of DELTA and the C-test. In a factor analysis the highest loading on the first factor was found for the vocabulary subtest (loadings between .83 and .94). Furthermore, Cohen et al. (1985) reported C-test studies carried out in Hebrew. They concluded that the test encouraged micro-level processing and that ‘students who did not understand the macro-context could still mobilize their vocabulary skills’ (p. 125).

There are also clear indications that C-test scores are related to academic success. In a study with 358 pupils at secondary schools in Germany, Klein-Braley (1985a) showed that C-test scores in the first language were directly linked to grade and school type. Pupils who attended school types that lead to HE (Gymnasium) consistently had higher C-test scores than those that prepared for vocational training (Realschule, Hauptschule). This finding is supported by a study carried out with 75 pupils (mean age 10.8) by Raatz (1985) who found a correlation of .51 between the C-test scores and a test of non-verbal intelligence (figure completion and figure sequence completion tasks). Furthermore, Coleman (1994) employed a C-test to investigate the language proficiency of students enrolled at foreign language degrees in the UK. His study showed that the highest correlation between a C-test in French and five subsets of A level exams was with the subset ‘reading and writing’ ($r = .78$).
Overall, we conclude that the C-test has a specific focus on vocabulary and is potentially a good predictor of academic success. The specific focus on vocabulary might be precisely the underlying cause for the high correlations between C-test scores and test scores in all other aspects of language proficiency, including the important aspect of reading in a foreign language. Since they have to fill in gaps that are created randomly (every second word is truncated), test-takers also encounter different words and, depending on the text, infrequent words. The C-test format therefore taps into lexical diversity as well as sophistication. In addition, we conclude that there is a relationship between C-test scores and IQ scores which will also contribute towards the usefulness of this test format for predicting academic success.

The present study tested the following hypotheses:

1. There is a relationship between academic success and language proficiency.
2. Vocabulary knowledge is an important aspect of study success in a foreign language and can be used to predict academic success.
3. C-test scores are an indication of vocabulary knowledge and are therefore related to academic success.

An additional research question was whether lexical diversity or lexical sophistication was more important for academic success.

Subjects and Data Collection

The subjects in the present study were 23 Chinese students attending a British university and following a one-year taught course for postgraduates in a business school. The first data collection took place in China in February 2004, half a year before the students came to Britain. A C-test was administered under controlled conditions as part of the selection process. The students also obtained IELTS scores in China. Those with scores between 5.5 and 6.0 attended an eight-week pre-sessional English language programme in the British university before starting their postgraduate studies in September 2004. These students took the C-test at the beginning and end of the programme. Some with a score above IELTS 6.5 attended a three-week pre-sessional course and others did not. Before the start of their Master's programmes, the same C-test was administered to these students again and they were also asked to write an essay on 'The Internet'. They were informed this time that the test was to investigate their English proficiency and would not affect starting their studies. The essays were then transcribed into the CHAT format which allowed the computation of D and other measures with the CLAN software (MacWhinney, 2000). Spelling mistakes were corrected but the texts were not lemmatized. The number of types, tokens and D were computed. Students’ IELTS scores and previous
learning history in China were obtained as additional background information (e.g. whether they had already been awarded academic degrees or not before taking UK HE, and whether they switched subjects in Britain).

Predictor and dependent variables

The following predictor variables were used in the present study:

- The IELTS scores prior to entry into UK HE;
- Whether or not the student had been awarded a Bachelor’s or similar degree in China before coming to the UK;
- The values for D, Guiraud, LFP and Guiraud Advanced obtained from the essays written by the students;
- C-test scores from February 2004 and September 2007.

We operationalized study success with two variables:

- Whether a student failed at least one module in the first year or passed all modules at the first attempt;
- The number of modules failed in the first year.

Results

Predictor variables

The students’ IELTS scores are listed in Table 11.1. It should be noted that these scores were obtained in China before their pre-sessional course. According to the teachers on this course, the IELTS scores typically rose by half a band in an eight-week course.

Students’ educational background is shown in Table 11.2. The variables were whether the students had studied the same subject at home.

<table>
<thead>
<tr>
<th>IELTS score</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>1</td>
</tr>
<tr>
<td>6.0</td>
<td>10</td>
</tr>
<tr>
<td>6.5</td>
<td>9</td>
</tr>
<tr>
<td>7.0</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obtained degree in China</th>
<th>Subject switched</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>9</td>
</tr>
<tr>
<td>No</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obtained degree in China</th>
<th>Subject switched</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>20</td>
</tr>
<tr>
<td>No</td>
<td>3</td>
</tr>
</tbody>
</table>
or whether they switched subjects in Britain, and whether they had been awarded a Bachelor’s degree in China.

It was surprising that most students had chosen to take a different subject (a Master’s in Business) instead of pursuing what they had studied previously. This could, however, be explained as a clear change in career perspective by the students.

Figure 11.1 shows the distribution of D scores. Apart from one outlier all lay between 48 and 110 with a median of 80. The standard deviation of 20.56 indicated a good spread of the scores, revealing the potential for this measure to be a good predictor if the range of vocabulary usage was indeed related to individual differences in academic achievement in the students from China.

The results for both C-tests illustrated in Table 11.3 and Figure 11.2 reveal the spread of two test scores transformed into percentages. Both C-tests turned out to be highly reliable: Cronbach’s alpha was .873 for February 2004 and .876 for September 2004. The correlation between the two tests was significant and moderately strong (r = .604, p = .008, N = 18).

![Figure 11.1 Spread of D scores](image)

**Table 11.3 C-test results**

<table>
<thead>
<tr>
<th>C-test</th>
<th>Number</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb. 2004</td>
<td>21</td>
<td>30.83</td>
<td>71.67</td>
<td>55.27</td>
<td>13.26</td>
</tr>
<tr>
<td>Sept. 2004</td>
<td>20</td>
<td>43.13</td>
<td>83.75</td>
<td>58.81</td>
<td>10.81</td>
</tr>
</tbody>
</table>
Table 11.3 shows a slightly higher mean for September 2004. However, it can be seen from Figure 11.2 that there is little, if any, difference in the median scores and overall the difference was not significant (paired \( t \)-test, \( t = 1.42, \) d.f. \( = 17, p = .173 \)). Note that there were only 18 paired test scores because not all students did both tests.

**Modules failed**

As mentioned earlier, we operationalized study success in two ways. Firstly, we divided the group into those who passed all modules at the first attempt and those who had at least one failed module. Only 9 out of 23 students passed all modules at their first attempt and 14 students failed one or more modules (see Table 11.4). Failure seemed to be a serious issue for the cohort. More than half failed one module and over a quarter of the students failed four or more modules while taking taught courses in their first year.

![Figure 11.2](image-url)  
*Figure 11.2*  
The spread of C-test scores

<table>
<thead>
<tr>
<th>Number of failed modules</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>9</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Predicting academic success

In order to investigate the contribution that variables make to predicting study success, we computed the correlations between variables and the number of failed modules. Two variables (IELTS and C-test February 2004) yielded significant results with a one-tailed test. The Spearman correlation between IELTS scores and the number of failed modules was $-0.382 (p = 0.036, N = 23)$. An even stronger negative correlation was found between the C-test scores from February 2004 and the number of failed modules ($-0.565, p = 0.004, N = 21$).

Figure 11.3 illustrates the relation between C-test and the number of failed modules, with a linear line of best fit included. The data point for the student who failed eight modules was not included in the graph since no C-test data were available.

No other variable correlated significantly with the number of failed modules. However, some intercorrelations were significant and gave some insight into the specific focus of such variables (see Table 11.5). The only variable from the Lexical Profile Analysis to enter into significant correlations was the number of types not on the list (types NoL) and the number of tokens not on the list (tokens NoL). The C-test scores are also included in Table 11.5 to investigate the lexical focus of the test.

As can be seen from the table, $D$ and Guiraud were significantly intercorrelated which is in line with expectations, as both measures focus on lexical diversity. Guiraud Advanced and types NoL were also significantly

![Figure 11.3](image-url)
correlated which is an indication that they both measure the same aspect of vocabulary knowledge, in this case lexical sophistication. The C-test was significantly correlated with these two measures of lexical sophistication but not with the two measures of diversity. This is a clear indication that the C-test is sensitive to lexical sophistication rather than just the range of one's vocabulary.

We also wanted to predict the variable ‘Fail’ which put the students into two categories: those who passed all modules and those who had at least one fail at the first attempt. We used a logistic regression with ‘IELTS scores’ initially as the only predictor variable. This model was significant ($\chi^2 = 8.416$; d.f. = 3; $p = .038$; d.f. was 3 was because there were four IELTS levels in the study: 5.5, 6.0, 6.5 and 7.0). This statistical procedure produced beta values for each ‘IELTS score’ separately, and the highest negative beta value was obtained for IELTS score 5.5 ($B = -38.56$), followed by score 6.0 ($B = -20.7$), 6.5 ($B = -19.5$) and score 7.0 ($B$ set to zero). The lower the IELTS score, the more likely it was, therefore, that the student failed a module. The same computation was carried out with the C-test scores but no significant model could be obtained.

Overall, the present study showed that failure of at least one module could be predicted by IELTS scores. The lower the IELTS scores, the higher risk of failing at least one module. IELTS scores also explained about 11 percent of the variance of the number of failed modules.

Interestingly, none of the measures taken in September 2004 in the UK before the students started their studies, including the C-test, predicted their academic success. One possible explanation could be that the C-test taken in February 2004 in China had been administered under strictly controlled conditions, with several British and Chinese staff monitoring, and the test was taken as part of the admissions procedure. By contrast, the students knew in September 2004 that they had secured a place at a British university and that tests and essays would not affect their further study.

Table 11.5: Intercorrelations between lexical measures (Pearson)\(^a\)

<table>
<thead>
<tr>
<th></th>
<th>C-test</th>
<th>D</th>
<th>Tokens NoL</th>
<th>Types NoL</th>
<th>Guiraud</th>
<th>Guiraud Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-test</td>
<td>0.135</td>
<td>0.393</td>
<td>0.522**</td>
<td>0.167</td>
<td>0.613***</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>0.236</td>
<td>0.475**</td>
<td>0.564***</td>
<td></td>
<td>0.275</td>
</tr>
<tr>
<td>Tokens NoL</td>
<td></td>
<td>0.737***</td>
<td>0.422</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Types NoL</td>
<td></td>
<td>0.369</td>
<td>0.696***</td>
<td>0.518**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guiraud</td>
<td></td>
<td>0.275</td>
<td>0.696***</td>
<td>0.518**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guiraud Advanced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) N = 19 for all intercorrelations with the C-test, otherwise N = 21

** Significant at the .05 level;
*** Significant at the .01 level.
This might have impacted on their motivation and reduced the validity of the measures (for an in-depth discussion on test-taker motivation and test validity see Nation, 2007). The relatively modest correlation between the two C-tests could be a further indication of decreased motivation, although firm conclusions were not possible with the existing data set.

**Conclusion**

The present study shows that it is possible to predict study success and failure before overseas students come to the UK. English proficiency tests such as IELTS and a C-test are useful predictor variables. The most powerful predictors for the number of failed modules are the C-test scores obtained more than half a year before the students came to the UK. Almost 40 per cent of the variance in the number of failed modules during the first-year taught course of Master’s programmes can be predicted from their C-test scores. In addition, we used measures of lexical diversity (D and Guiraud) and measures of lexical sophistication (Guiraud Advanced and LFP) derived from students’ essays. The present study reveals that the measures of lexical sophistication correlate significantly with each other and so do the measures of lexical diversity, which supports the validity of the measures used. However, the C-test correlates significantly only with measures of lexical sophistication (and not with measures of lexical diversity) which is an indication that knowledge of infrequent words may be related to academic success in a foreign language. The highest correlation between the C-test and any measure of vocabulary richness was found with Guiraud Advanced. As a recent study has shown, Guiraud Advanced appears to be related to verbal intelligence. This and earlier findings on the relationship between the C-test format and intelligence scores lead us to the conclusion that lexical sophistication and (verbal) intelligence both play a role in C-test completion. Therefore this test format is an excellent tool to predict study success.

The present study provides implications for admissions procedures. Language tests which tap into lexical sophistication and verbal intelligence can be developed to measure overseas students’ ability to study in a foreign language before they embark on international education.

**Note**

1. We are very grateful for the comments on an earlier draft of this chapter made by Brian Richards and an anonymous reviewer. We would like to thank Paul White for his advice on statistics.