

The C-test and TOEIC[®] as measures of students' progress in intensive short courses in EFL

Helmut Daller and David Phelan*

This paper investigates whether the C-test can partially replace the Test of English for International Communication (TOEIC[®]), and whether the C-test is an appropriate format for the measurement of students' progress in short intensive language courses. The TOEIC[®] is a well-established test for English as a foreign language and is used by corporations and language institutes as a selection and placement instrument. The subjects in the present study are first year, undergraduate, students from France in the UK who took a TOEIC[®] exam before and after an 11-week intensive English course with 240 hours teaching. In addition an entry and an exit C-test was administered. It was possible to show that both tests are sufficiently reliable, and that both tests measure a significant increase during the course. The significant correlation between both tests at entry and at the end of the course is an argument for the validity of the tests. The C-test, however, turns out to be much more economical. It is argued that for this reason the C-test is the more suitable test for large numbers when measuring students' progress.

1. The Test of English for International Communication (TOEIC[®])

The Japanese Ministry of International Trade and Industry asked Educational Testing Services (ETS) in the 1970s to design a test which would measure English language skills in international business settings (Gilfert, 2001). The test was first administered in Japan to 2,710 candidates on December 2nd 1979. It is now available in 39 countries (Moritosho, 2001), with 2 million tests being taken annually (TOEIC[®] website, 2003). According to Sharron (1997, p. 26) Korean and Japanese examinees account for 94% of the total number of tests taken in 1997. The TOEIC[®] User Guide (1999) states that TOEIC[®] is used by government agencies, language schools, academic institutions and more than 4,000 corporations. It is now regarded by many as the leading test of English for business purposes. Indeed TOEIC[®] has now spread from Japan all over Asia and is becoming more frequently used in Europe and South America (Gilfert, 2001). TOEIC[®] is a norm-referenced test of receptive skills and therefore has the inherent problem that it does not tell us the functional ability to use the target lan-

* **Address for correspondence:** Dr. H. Daller and D. Phelan, University of the West of England. Faculty of Humanities, Languages and Social Sciences, Frenchay Campus, Coldharbour Lane, Bristol, BS16 1QY, UK. E-Mail: helmut.daller@uwe.ac.uk and david.phelan@uwe.ac.uk

guage (Wilson, 1989). Therefore, it does not directly test productive skills. The TOEIC[®] has two equally weighted parts, listening and reading. There are four subsections in the listening and three in the reading. There is a total of two hundred multiple-choice questions. The listening lasts forty-five minutes and the reading one hour fifteen minutes. There are three scores, one for reading, one for listening and a combined total which ranges from 10 to 990. A more detailed description of the test structure is contained in Appendix 1.

Many organizations using TOEIC[®] need to be able to identify individuals with a level of English language competency that enables them to adequately perform their jobs (TOEIC[®] Can-Do Guide, 2000). Wilson (1989, p. iv) states that:

“the level of ability to use English in face to face conversation (indexed by Language Proficiency Interview (LPI) will vary relatively consistently with the level of developed English language listening comprehension.”

He also claims that:

“the TOEIC program has made it possible to develop general guidelines that permit test users to make statistically valid inferences from TOEIC scores about levels of oral English proficiency” (Wilson, 1989, p. iv).

Moritoshi (2001, p. 9) discusses a number of studies, which also claim that the test measures speaking and writing indirectly and that it is a test of general English language proficiency. High correlation with other direct measures of reading, writing, listening and speaking would suggest that TOEIC[®] also measures general language proficiency. However, as Moritoshi (2001, p. 11) points out the concurrent evidence is effectively circular, merely extending the assumption of validity to these new criteria. Years of language teaching and learning also tell us that there are cases where, for example, listening and speaking abilities are vastly different in the same person. However, these instances remain exceptions rather than the rule. The problem of the worth of correlational evidence can be overcome to some extent by ensuring the validity and reliability of the other tests used for comparison (see also Section 2 for a further discussion).

A large scale study was carried out which successfully used TOEIC[®] entry and exit scores to analyse the effects of training time and type on learners' progress (Boldt & Ross, 1998).

2. The C-test as a measure for students' progress over time

One issue that has been widely discussed is the question of the C-test's validity. That is: what exactly does it measure? According to Alderson (2002, p. 28) this is a “worrisome question”. Many researchers assume that the C-test measures

general language proficiency (see also Grotjahn, 1995). However, according to Alderson:

“The Unitary Competence Hypothesis (UCH) ... claiming that there was a unitary competence, or a general language proficiency ... is 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 (for a more detailed discussion see Daller, 1999, p. 71). The main arguments in the 1970s for such a competence were the high correlations that could be found between tests of different aspects of foreign language proficiency. Despite Alderson's claim that this hypothesis is now discredited, these high correlations were also found in later studies.

“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.” (Singleton & Singleton, 2002, p. 154).

Many studies on the C-test are in line with these findings. Jafarpur (2002) finds high correlations (.65 to .80, Pearson) between a C-test and an English Placement Test (EPT, form B, for a discussion of this placement test see Corrigan et al., 1978). Hastings (2002) describes several studies that were carried out at the University of Wisconsin-Milwaukee, where on four occasions in 1991 and 1992, students in that program took the TOEFL concurrently with one of the C-tests. The correlations were between .87 and .94. The C-test has been administered in many languages other than English. Significant correlations between the C-test and the scores in other tests have also been found in these languages. Bolten (1992) compares the results of a C-test with several subtests of two German university entrance tests (PNdS and DZW). He finds correlations in the range of .63 – .92 between reading, listening, grammar and writing tests and the C-test. This is an indication that C-tests can be used as part of university entrance tests. Grotjahn and Allner (1996) look at the correlations between the scores in a German C-test and the entry exam for the “Studienkollegs” (a higher education institute that prepares students for university entry). They found high correlations between the C-test and tests of grammar and text reproduction (Grotjahn & Allner, 1996, p. 287). Grotjahn (2002) comes to the conclusion that the C-test in general correlates relatively highly with reading and writing skills.

Overall, we conclude that the C-test measures general language proficiency, and that Alderson's claim has to be modified in the sense that there is something resembling a common underlying proficiency at least in certain areas. One has to bear in mind that the C-test does not cover all areas of language proficiency. Grotjahn and Stemmer (2002, p. 125) conclude “that the C-Test is primarily

considered to measure the ability to cope receptively with context-reduced language in cognitively demanding tasks". This aspect has been used to create C-tests that focus especially on academic skills (see Daller, 1999; Daller & Grotjahn, 1999). The C-test format might therefore be especially useful as a university entrance test (together with other tests in a test battery).

Another point relevant to the present study is the question of whether or not the C-test in general can be used to measure an increase in foreign language proficiency over time, and whether the same test can be used as an entry and exit test with the same group. Coleman (2002) administered C-tests in several foreign languages to students in the last two years of secondary education and to students in higher education. The results showed that students in a higher year achieved systematically higher C-test scores when compared to students in lower years. Although this is not a longitudinal study since students were measured only once, the results indicate that there is an increase of language proficiency from year to year, and that the C-tests administered are obviously suitable measures for this gain in foreign language proficiency. Coleman also compared the C-test scores with several components of the A-level examination (equivalent to the German Abitur). He found significant correlations between these two tests ($r = -.48$ and $r = -.78^1$). The highest correlations were found between reading and writing, and between listening and the C-test (Coleman, 2002, p. 229). Huhta (1996, p. 207) found significant correlations between an EFL C-test and a university reading comprehension test but not between the same C-test and the administered listening comprehension test. The C-test that was used did not yield statistically significant differences between first-year students and students in the higher years whereas other tests (e.g. Cloze, dictation and vocabulary tests) did result in significant differences between the different years. It is therefore not completely clear whether the C-test is suitable for measuring progress over time. One has, however, to bear in mind that Huhta's study is a cross-sectional study and not a longitudinal one.

One important question in a longitudinal study is whether you can train learners in the C-test technique, and whether the experience with the entry test will automatically lead to higher scores in the exit test due to this training or possible memory effects. Bogards and Raatz (2002) carried out a study where one group received special training for the C-test and a control group did not. Both groups increased their test scores but the trained group showed a lower increase (5.4%) compared with 6.1% for the group that did not receive training. This is an indi-

¹ Due to the way of computing these correlations the original values for r were all negative. The interpretation should however be: the better the score in the A-level, the better the C-test score.

cation that special training, at least a short period of training as was the case in this study (45 minutes), has no effect on the test scores. The findings of this study are, however, an indication that taking a C-test twice will lead to increased scores. The reason for this might be a better familiarity with the test format, improved test taking strategies or a memory effect. In Bogards and Raatz' study, the interval between the first and the second test was only 4 days. One would expect that a memory effect plays a certain role over this short period of time. Over a longer period this memory effect may be less important. One might therefore assume that the increase with a longer interval between entry and exit test will be lower than 5 to 6%. Any increase above this level will be due to a real increase in foreign language proficiency. Bogards and Raatz (2002, p. 150) report on a measure of increase due to mere test repetition. An increase that is about 20 to 30% of the standard deviation of the tests (equal standard deviation assumed) can be expected when tests are repeated.

Given the discussion of recent research findings above we expect that there will be significant correlations between the C-test in the present study and the TOEIC® scores. We also expect that part of the increase between entry and exit test will be due to a repetition effect. Rates of increase that are higher than 5% will, however, be due to an increase in foreign language proficiency.

3. The present study

3.1. The students

Two groups were involved in the present study, one with 14 students and one with 18 students. The first course took place in 2002, the second in 2003. Both courses received exactly the same input, that is two hundred and forty hours of teaching over a period of eleven weeks. Teaching materials and teachers were similar for both courses. Also the background of the students was quite similar. They were all students from a French business institute who took the intensive EFL and business course in the second term of their first year. The initial TOEIC® tests were administered 12 weeks before departure to the UK and immediately after returning to France. These tests were carried out by the Chauncey Group International in Paris. After the initial TOEIC® and before the start of the intensive course students received some English language input in France. The students received 3 hours of teaching per week for 7 weeks, researching aspects of Britain and giving oral presentations. The students also completed a self-study grammar course of 15 hours. The first group consisted of 10 female and 8 male students between the ages of 18 and 25. One of the female students had to withdraw due to illness. The second group consisted of 8 male

and 6 female students again between the ages of 18 and 25. In the second group one of the female students followed a different course from the others due to her superior level of English. The level of the first group was in the range 530 to 745 while the second group was between 485 and 715 (total TOEIC[®] scores at entry = listening + reading scores). All students were of an upper-intermediate level with the second group being slightly lower than the first. All the students were from comfortable socio-economic backgrounds and were attending NEGOCIA –Academie Commerciale Internationale (ACI), an international business school affiliated to the Paris Chamber of Commerce and Industry, France. The students were all studying for an undergraduate degree in international business, marketing and negotiation.

3.2. The course

The 11-week course in the UK represented the second term of the first year of their 3-year degree programme in International business. The objective of the 11-week course was to acquire an excellent command of the English language and to increase cross-cultural awareness and receive an introduction to the business environment of the UK. The students were required to study the English language, UK culture and civilisation, the UK business environment and to complete a project. All courses were taught in English by native speakers and the 240 hours was made up of 9 different elements (see Appendix 2 for further details). The first group attended from 8 January to 22 March 2002 and the second from 7 January to 21 March 2003. All students were placed in university accommodation, sharing with other students whose common language was English.

3.3. The C-test

There were six texts with twenty gaps in each. The texts were taken from the website of number 10 Downing Street (www.number-10.gov.uk). The texts were about general and cultural aspects of the UK. When marking the tests the missing words had to be spelt correctly. Only the words contained in the original texts were accepted as correct.

4. Results

4.1. Reliability of the C-tests and test of normal distribution

We administered the same C-test to the students in the first week of the course and in the last week. There was a time difference of 11 weeks between the two test sessions. As mentioned above one student withdrew from the course and another received different input. We therefore decided to exclude these two students. All computations below are based on the remaining 30 students.

The C-test entry has a reliability of .836 (Cronbach's alpha)² which is an astonishing result since the C-test has not been pre-tested. The C-test exit (which is exactly the same test as the C-test entry) yielded only .65 (Cronbach's alpha). As will be explained below this difference is mainly due to the fact that a homogenization of the group has taken place. The spread of the scores is lower in the exit test as reflected by the lower standard deviation (for the influence of the standard deviation on the reliability see also Rietveld & van Hout, 1993, pp. 200ff.). This is a general problem when using the same test as an entry and exit test for a course and cannot be easily solved. There is no consensus in the literature about the lower limit of an acceptable reliability. Nunnally (1978) suggests .7 as the lower limit. According to Klein-Braley (1991, p. 78) and Meuffels (1992, p. 147) a value for Cronbach's alpha of at least .9 is necessary for decisions on individuals whereas .8 or even .6 will be sufficient for research purposes. In the light of this we decided to use both test results for our further computations. The value for Cronbach's alpha is based on the correlations between the items. We therefore made further investigations into these correlations as shown in Tables 1 and 2.

Table 1: Correlation between the 6 texts of the entry test (Pearson)

	Text 1	Text 2	Text 3	Text 4	Text 5
Text 1	–				
Text 2	.352	–			
Text 3	.356	.511**	–		
Text 4	.507 **	.639**	.494**	–	
Text 5	.593 **	.660**	.630**	.601**	–
Text 6	.336	.324	.285	.189	.372*

* $p \leq .05$, ** $p \leq .01$ (two-tailed)

² All computations were carried out with SPSS version 10.

Table 2: Correlation between the 6 texts of the exit test (Pearson)

	Text 1	Text 2	Text 3	Text 4	Text 5
Text 1	–				
Text 2	.356	–			
Text 3	.099	.424*	–		
Text 4	–.102	.241	.388*	–	
Text 5	.353	.349	.162	.261	–
Text 6	.280	.052	.103	–.025	.474**

* $p \leq .05$, ** $p \leq .01$ (two-tailed)

The SPSS option “scale if item is deleted” indicates that the entry test would get a slightly higher Cronbach’s alpha if text 6 were deleted (.847 instead of .836) whereas a deletion of any item in the exit test would decrease the value for Cronbach’s alpha. Although text number 4 shows negative correlations in the exit test, the deletion of text number 4 would lead to a slight decrease of Cronbach’s from .650 to .649. We therefore decided to use the unaltered test and include all items in the following computations.

As earlier studies have shown (Grotjahn, 1996, p. 101; Grotjahn, 2002, pp. 91f.) the distribution of C-test scores can significantly differ from a normal distribution. It is important to investigate the appropriate model (distribution) especially with relatively few cases. A check as to whether or not we can assume a normal distribution can be made by drawing a probability plot. The data (measured scores) are compared with the expected scores of the model, in this case a normal distribution. If a normal distribution is the appropriate model, the measured scores will lie approximately on a straight line through the origin. The probability plot for the entry C-test is shown in Figure 1.

Figure 1 shows that a normal distribution can be assumed for this test. Similar probability plots can be obtained for all six texts. They all suggest a normal distribution with the exception of the TOEIC[®] exit test where the straight line does not go through the origin as is shown in Figure 2.

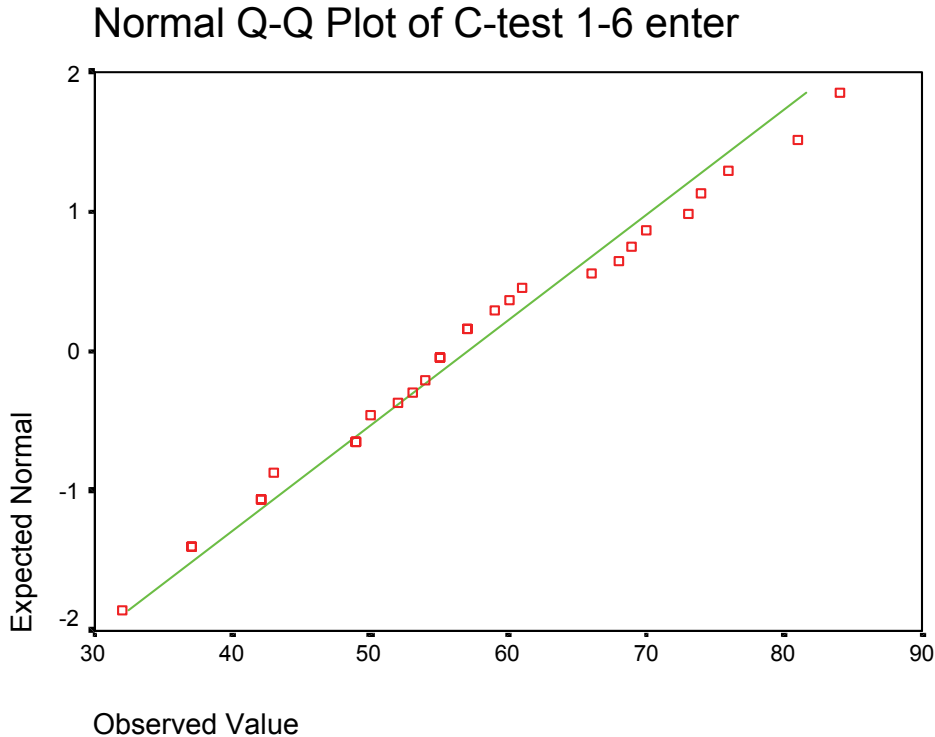


Figure 1: Probability plot for the C-test entry

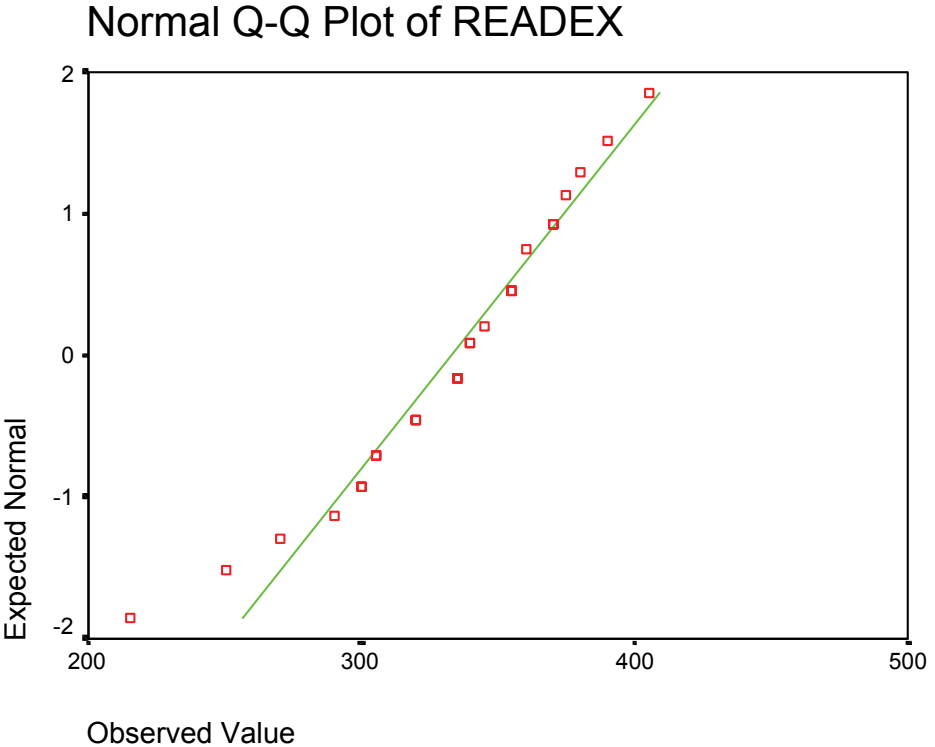


Figure 2: Probability plot for the TOEIC® reading exit

We also carried out two tests of normality for all tests. These are the Kolmogorov-Smirnov test modified after Lilliefors and the Shapiro-Wilk-Test. The latter is computed by SPSS for samples smaller than 50, as is the case in the present study. This type of exploratory data analysis can be carried out with the “explore” command from SPSS. The results are displayed in Table 3.

Table 3: Tests of normality for the entry and exit tests

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	<i>df</i>	<i>p</i>	Statistic	<i>df</i>	<i>p</i>
TOEIC Reading entry	.128	30	.200 ⁺	.958	30	.377
TOEIC Listening entry	.121	30	.200 ⁺	.950	30	.263
C-test entry	.099	30	.200 ⁺	.976	30	.171
TOEIC Reading exit	.117	30	.200 ⁺	.948	30	.225
TOEIC Listening exit	.154	30	.066	.950	30	.259
C-test exit	1.080	30	.200 ⁺	.971	30	.607

⁺ This is a lower bound of the true significance.

^a Lilliefors significance correction

There is only one case, the TOEIC[®] listening exit test, where there is an almost significant value for the Kolmogorov-Smirnov test ($p = .066$). The Shapiro-Wilk test, however, shows a clear picture. The p -values for all tests are well above significance level, which means that we cannot reject the assumption that the data are normally distributed. We therefore conclude that the data are (roughly) normally distributed with the possible exception of the reading exit test.

4.2. Results: the scores of the students

There is a significant correlation of .775 between the C-test scores in the entry test and the scores in the exit test (Pearson, two-tailed, $p \leq 0.001$). This means that those who achieved high scores in the entry test also scored well in the exit test and vice versa as illustrated in Figure 3.

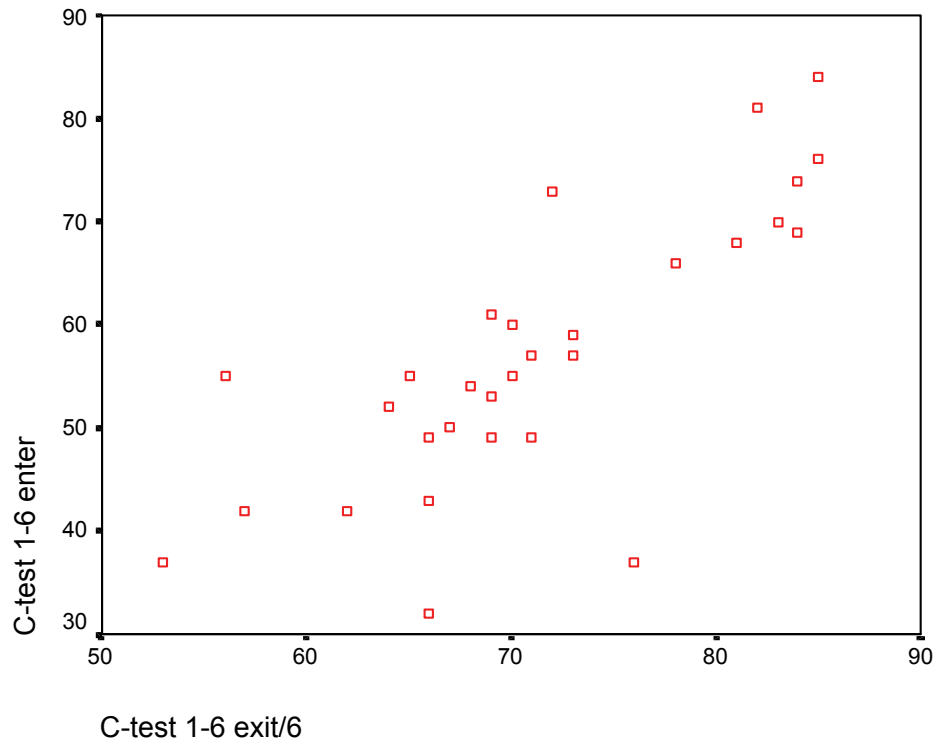


Figure 3: C-test scores in the entry and exit test

There is also a significant correlation of .473 between the entry and exit TOEIC® reading test (Pearson, two-tailed, $p = .008$) and a significant correlation of .667 between the entry and exit TOEIC® listening test (Pearson, two-tailed, $p \leq .001$). This is in line with the expectations since students with a high proficiency in one of the tasks at the beginning of the course are also expected to show a high proficiency in this task at the end of the course. The main results for the three entry and the three exit tests are shown in Table 4.

Table 4: Descriptive statistics of the three sets of tests (Diff = difference as percentage of the entry score)

	<i>n</i>	Mean	Std. Error	Std. Dev.	Diff. in %
TOEIC Reading entry	30	269.00	5.39	29.55	23.72 %
TOEIC Reading exit	30	332.83	7.54	41.29	
TOEIC Listening entry	30	335.67	11.84	64.84	18.42 %
TOEIC Listening exit	30	397.5	8.83	48.35	
C-test entry	30	56.97	2.43	13.30	24.92 %
C-test exit	30	71.17	1.59	8.71	

All tests show an increase in the mean score. There is also a decrease in the standard deviation and the standard error for the TOEIC® Listening test and the C-tests. Interestingly there is an increase in these values for the reading test. A possible explanation is that all the students were placed in living accommodation where English was spoken and hence were exposed to spoken English outside the classroom. This might be the reason for the homogenization of the listening skills throughout their stay in England. Such an effect is less likely with reading skills where apart from the classroom setting students do not have to read. Some students considerably improved their skill in this area while others showed less progress. A test of significance (paired *t*-test) shows that all differences are highly significant³ (cf. Table 5).

Table 5: Paired *t*-tests for the difference between entry and exit tests

	Paired Differences	St. Dev.	Std. Error	<i>t</i>	<i>df</i>	<i>p</i> (2-tailed)
TOEIC Reading	-63.83 ⁴	37.73	6.89	-9.27	29	.000
TOEIC Listening	-61.83	48.57	8.87	-6.97	29	.000
C-Test	-14.20	8.56	1.56	-9.09	29	.000

For all tests the increase of the scores (see Table 5) is far more than a standard deviation in the entry/exit test (see Table 4). As discussed in Section 2, a maximum of 20 – 30% of the increase might be due to a repetition factor. This means that most of the increase can be explained in terms of a gain of language proficiency.

³ Given the high level of significance there is no need to adjust the *p*-value for multiple measures.

⁴ SPSS automatically computes the score “entry – exit” and not “exit – entry”, which leads automatically to negative values when there is an increase in the test scores; e.g. the difference between TOEIC® Reading entry and exit is -63.83. This means that the students scored 63.83 points higher on average in the exit test.

Table 6: Correlations between the six tests (Pearson)

	TOEIC Reading enter	TOEIC Listening enter	C-test enter	TOEIC Reading exit	TOEIC Listening exit	C- test exit
TOEIC Reading enter	–					
TOEIC Listening enter	.272 <i>p</i> = .146	–				
C-test enter	.483 <i>p</i> = .007**	.455 <i>p</i> = .011*	–			
TOEIC Reading exit	.473 <i>p</i> = .008**	.355 <i>p</i> = .054	.612 <i>p</i> = .000**	–		
TOEIC Listening exit	.381 <i>p</i> = .038*	.667 <i>p</i> = .000**	.555 <i>p</i> = .001	.557 <i>p</i> = .001**	–	
C-test exit	.253 <i>p</i> = .177	.368 <i>p</i> = .045	.775 <i>p</i> = .000	.495 <i>p</i> = .005**	.415 <i>p</i> = .023*	–

** Correlation is significant at the .01 level (2-tailed).

* Correlation is significant at the .05 level (2-tailed).

Interestingly, there is no significant correlation between the TOEIC® reading and listening at entry but there is at the exit point (see Table 6). The C-test shows a significant correlation between both parts of the TOEIC® at the entry and at the exit points. The correlation with the reading text is slightly higher in both cases but the findings do not indicate that the C-test is only a reading test.

As mentioned earlier, there are indications that the scores for the TOEIC® listening exit are not normally distributed. We therefore double checked our findings and carried out non-parametric tests for the differences between the exit and entry tests and the correlations (Wilcoxon sign rank test). The results support the findings already reported. All differences between entry and exit tests are highly significant ($p < .001$).

5. Conclusions

The present study shows that the C-test can be used to measure students' progress in short term intensive courses. The same C-test can be used as an entry and as an exit test. The TOEIC[®] test shows a comparable increase in test scores between the two test points. We conclude that both tests are suitable measures for students' progress over time. The C-test, however, is far more economical than the TOEIC[®]. The present study also shows that even a pre-tested C-test as the entry test does not automatically lead to a highly reliable result with the same group 11 weeks later. Due to homogenization in the group the exit test has a lower reliability than the entry test. For this reason, C-tests that are used as entry and exit tests need careful pre-testing at both testing points.

Both the listening and the reading subtests of the TOEIC[®] show significant correlations with the C-test. The correlation between the C-test and the reading scores are, however, slightly higher than the correlation between the C-test and the listening scores. This is a further indication that the C-test measures general language proficiency. As both listening and reading skills are important in a university context we conclude that the C-test has a high predictive validity for academic success in a foreign language. We therefore suggest that the C-test should be part of a university entrance test together with other tests that measure skills beyond the scope of the current study. Further studies are necessary to investigate the predictive value of the C-test for students' progress during their academic career in a foreign language environment

References

- Alderson, Charles. (2002). Testing proficiency and achievement: principles and practice. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Eds.), *University language testing and the C-test* (pp. 15–30). Bochum: AKS-Verlag.
- Arras, Ulrike, Eckes, Thomas & Grotjahn, Rüdiger. (2002). C-Tests im Rahmen des 'Test Deutsch als Fremdsprache' (TestDaF): Erste Forschungsergebnisse. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 175–209). Bochum: AKS-Verlag.
- Boldt, Robert F. & Ross, Steven J. (1998). *Scores on the TOEIC (Test of English for International Communication) test as a function of training time and type* (TOEIC[®] Research Report Number 3). Princeton, NJ: The Chauncey Group International.
- Bolten, Jürgen. (1992). Wie schwierig ist ein C-Test? Erfahrungen mit dem C-Test als Einstufungstest in Hochschulkursen Deutsch als Fremdsprache. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 2, pp. 193–203). Bochum: Brockmeyer.
- Bogards, Sandra & Raatz, Ulrich. (2002). Sind C-Tests trainierbar? In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 157–174). Bochum: AKS-Verlag.

- Coleman, James A. (1994). Profiling the advanced language learner: the C-Test in British further and higher education. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 2, pp. 217–237). Bochum: AKS-Verlag.
- Coleman, James A. (2002). The European Proficiency Survey: an overview of findings. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Eds.), *University language testing and the C-test* (pp. 201–206). Bochum: AKS-Verlag.
- Corrigan, A., Dobson, Barbara, Kellman, E., Spaan, Mary. W. & Tyma, S. (1978). English Placement Test (Form B). Ann Arbor: Testing and Certification Division, University of Michigan.
- Daller, Helmut. (1999). *Migration und Mehrsprachigkeit. Der Sprachstand türkischer Rückkehrer*. Frankfurt a. M.: Lang.
- Daller, Helmut & Grotjahn, Rüdiger. (1999). The language proficiency of Turkish returnees from Germany: An empirical investigation of academic and everyday language proficiency. *Language, Culture and Curriculum*, 12(2), 156–172.
- Daller, Helmut, Treffers-Daller, Jeanine, Ünalı-Ceylan, Aylin & Yıldız, Cemal. (2002). The development of a Turkish C-Test. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Eds.), *University language testing and the C-test* (pp. 187–199). Bochum: AKS-Verlag.
- Gilfert, S. (2001). 'A Review of TOEIC' The Internet Journal (1/11/01). (www.aitech.ac.jp/iteslj/Articles/Gilfert-TOEIC.html)
- Grotjahn, Rüdiger. (1995). Der C-Test: State of the Art. *Zeitschrift für Fremdsprachenforschung*, 6(2), 37–60.
- Grotjahn, Rüdiger. (1996). 'Scrambled' C-Tests: Untersuchungen zum Zusammenhang zwischen Lösungsgüte und sequentieller Textstruktur. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 3, pp. 95–125). Bochum: Brockmeyer.
- Grotjahn, Rüdiger. (2002). 'Scrambled' C-Tests: eine Folgeuntersuchung. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 83–115). Bochum: AKS-Verlag.
- Grotjahn, Rüdiger & Allner, Burkhard. (1996). Der C-Test in der Sprachlichen Aufnahmeprüfung an Studienkollegs für ausländische Studierende an Universitäten in Nordrhein-Westfalen. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 3, pp. 279–335). Bochum: Brockmeyer.
- Grotjahn, Rüdiger and Stemmer, Brigitte. (2002). C-Tests and language processing. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Eds.), *University language testing and the C-test* (pp. 115–130). Bochum: AKS-Verlag.
- Hastings, Ahley J. (2002). In defense of C-testing. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 11–29). Bochum: AKS-Verlag.
- Huhta, Ari. (1996). Validating an EFL C-test for students of English philology. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 3, pp. 197–229). Bochum: Brockmeyer.
- Jafarpur, Abdoljavad. (2002). A comparative study of a C-test and a cloze test. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 31–51). Bochum: AKS-Verlag.
- Klein-Braley, Christine. (1991). Ask a stupid question....: Testing language proficiency in the context of research studies. In Kees de Bot, Ralph B. Ginsberg & Claire Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 73–94). Amsterdam: Benjamins.
- Meuffels, Bert. (1992). *Methods and techniques of empirical research*. Amsterdam: SICSAT (Stichting Internationaal Centrum voor de Studie van Argumentatie en Taalbeheersing).

- Moritoshi, Paul. (2001). The Test of English for International Communication (TOEIC®): necessity, proficiency levels, test score utilisation and accuracy. The University of Birmingham (Internal publication) www.cels.bham.ac.uk-resources-essays-Moritoshi5.pdf
- Nunnally, Jum C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Rietfeld, Toni & van Hout, Roeland. (1993). *Statistical techniques for the study of language and language behaviour*. Berlin: Mouton de Gruyter.
- Sharron, R. H. (1997). 'TOEIC® today and TOEIC® tomorrow'. In TOEIC® steering committee (Ed.), *The 58th TOEIC® seminar*. Princeton, NJ: Educational Testing Service. Quoted by Moritoshi (2001).
- Singleton, David & Singleton, Emer. (2002). The C-Test and L2 lexical acquisition/processing research. In James A. Coleman, Rüdiger Grotjahn & Ulrich Raatz (Eds.), *University language testing and the C-test* (pp. 143–168). Bochum: AKS-Verlag.
- TOEIC® website, www.toeic.com; date of access 26/06/03
- TOEIC® Can-Do Guide (2000). The Chauncey Group International. www.toeic.com; date of access 2/04/03.
- TOEIC® User Guide (1999). www.toeic.com; date of access 3/05/03.
- Wilson, Kenneth M. (1989). *Enhancing the interpretation of a norm-referenced second-language test through criterion-referencing: A research assessment of experience in the TOEIC testing context* (TOEIC Research Report Number 1). Princeton, NJ: Educational Testing Service.

Appendix 1

Structure of the TOEIC®

Listening

1. Photographs – 20 questions
2. Question-response – 30 questions
3. Short conversations – 30 questions
4. Short talks – 20 questions

Reading

5. Incomplete sentences – 40 questions
6. Error recognition – 20 questions
7. Reading comprehension – 30 questions

Listening (45 minutes)

1. Photographs – 20 questions

Four sentences heard once, the examinee must choose the sentence which best matches the photograph.

2. Question-response – 30 questions

A question is heard together with three possible responses. The examinee must choose the correct response.

3. Short conversations – 30 questions

Examinees listen to a short dialogue and then read a question and four possible answers, they have to identify the correct answer.

4. Short talks – 20 questions

A short monologue followed by two or three written questions. The examinee must choose the most appropriate answer.

Reading (75 minutes)

5. Incomplete sentences – 40 questions

Sections 5 and 6 are designed to measure the ability to recognize language that is appropriate for standard written English.

The examinee reads a sentence with one word or phrase missing and has to choose from four alternatives.

6. Error recognition – 20 questions

Examinees read a sentence with four words underlined. They have to choose which one is correct.

7. Reading comprehension – 30 questions

Short reading passages followed by two to five questions, each has four options.

Appendix 2

English Language and Business Training

Courses:

1. Reading and Writing Skills	33hours
2. Speaking and Listening skills	33 hours
3. Business in the media	33 hours
4. UK Culture	27 hours
5. Introduction to Marketing	17 hours
6. Sales Management	16 hours
7. Research Tutorials	45 hours
8. Company visits	20 hours
9. Cultural Visits	16 hours
Total	240 hours

Appendix 3

1. 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 its widest the UK is 300 miles across and 600 miles from North to South. It shares a single land border with the Irish Republic. Despite its relatively small size the UK boasts incredibly varied and often very beautiful scenery, from 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.

2. UK Passport Service

A new passport office that has opened in London will help the UK Passport Service provide a much better service to customers who need a passport urgently. The new office runs on an appointment-only basis, removing the need for a lengthy wait before being seen. The new building, Globe House replaces the Petty France office, which after 50 years of continuous service, has now closed its doors. The London Passport Office has the capacity to issue 5000 passports weekly.

3. Record employment

Latest employment figures show that there are 28.2 million people in work. Work & Pensions Secretary Alistair Darling said this showed the UK labour market has coped well so far with the current international economic uncertainty. Mr Darling said: “Employment continues to rise, with this month’s figures showing a record 28.2 million people in work. There are 65,000 more people in work than last quarter and 252,000 more than last year. Although both measures of unemployment have risen slightly, they are still significantly lower than they were a year ago.” The latest claimant count figures, for the month on Dec 13 2001, show a rise of 3,200 on the previous month. At 963,500 claimants, it remains 70,000 lower than this time last year.

4. Government consults on plans to modernise animal welfare

Plans to review, modernise and simplify outdated laws on animal welfare have been announced by the Government. Animal welfare groups, local authority representatives, courts, police and industry are to be consulted in what will be a far reaching review drawing together the environmental and industrial concerns of animal welfare. The Department for the Environment, Food and Rural Affairs (DEFRA) wants to hear views on the existing 11 Acts of Parliament governing the welfare of pets and farm animals.

5. Government is going in right direction on crime says Blunkett

The chances of being a victim of crime are at their lowest level for 20 years – despite the worrying increases in mobile phone robbery – Home Secretary David Blunkett said today in a speech to a residents group in Sheffield.

This is being achieved through massive Government investment not just in the police service but in education, employment, neighbourhood and community renewal, the development of community partnerships and record levels of spending on crime reduction. In his first major speech of 2002 Mr Blunkett said the Government is delivering on its promises on crime, speeding up youth justice and targeting persistent offenders.

6. School for heroes

The internationally acclaimed Fire Services College in Gloucestershire is widely regarded as the best fire college in the world. The site, a former airfield which once launched Wellington bombers for raids on Germany, was recently transferred from the Home Office to the Fire Services. And such is its reputation that people come from as far away as China and Trinidad to sharpen up their leadership skills. Even fire chiefs from the other side of the world have attended as College students.

