



UNIVERSITY of CAMBRIDGE
ESOL Examinations

Research Notes

Issue 41/August 2010

A quarterly publication reporting on research, test development and validation

Senior Editor & Editor

Dr Hanan Khalifa & Dr Ivana Vidaković

Editorial Board

Dr Nick Saville, *Director*, Research and Validation Group, Cambridge ESOL

Angela ffrench, *Assistant Director*, Assessment and Operations Group, Cambridge ESOL

Production Team

Caroline Warren, Research Support Administrator

Rachel Rudge, Production Controller

George Hammond, Design

Printed in the United Kingdom by Océ (UK) Ltd.

Research Notes

Contents

Editorial Notes	1
Insights and issues arising from the English Profile Wordlists project: Annette Capel	2
Use of words and multi-word units in Skills for Life Writing examinations: Ivana Vidaković and Fiona Barker	7
Lexis in the assessment of Speaking and Writing: an illustration from Cambridge ESOL's General English tests: Gad Lim and Evelina Galaczi	14
A mixed-method approach towards investigating lexical progression in Main Suite Reading test papers: Hanan Khalifa and Norbert Schmitt	19
A corpus-led exploration of lexical verb use in Main Suite Writing papers: Stuart Proudfoot	26
TKT: Knowledge About Language and the assessment of lexis, phonology, grammar and discourse: Evelina Galaczi and Steven Miller	31
Conference reports	37
Recent publications of interest	40

Editorial Notes

Welcome to issue 41 of *Research Notes*, our quarterly publication reporting on matters relating to research, test development and validation within Cambridge ESOL.

This issue focuses on the linguistic features of a variety of Cambridge ESOL examinations and how they differ from one proficiency level to another. In the opening article, Annette Capel draws attention to the first phase of the English Profile Programme Wordlists project, which provides lexical information for CEFR levels A1 to B2. Annette shares insights and issues arising from this project.

The next four articles look at lexical development within a suite of examinations. Vidaković and Barker investigate the lexical development of L2 learners of English using written responses to Skills for Life writing examinations across five proficiency levels. Lim and Galaczi discuss aspects of lexical knowledge in performance assessment. They focus on four key issues, namely (a) determining target lexis at particular levels, (b) controlling lexis in writing and speaking tasks at different proficiency levels, (c) features of candidate performance at different levels, and (d) designing rating scales which reflect what is being assessed. They illustrate each of these using examples from Main Suite examinations. Also using Main Suite Reading test papers, Khalifa and Schmitt investigate lexical progression across the different levels of the Suite. They used a mixed-method approach where content analysis of lexical resources in the tests was carried by expert judges, while the WordSmith software package and Tom Cobb's Compleat Lexical Tutor were used to examine lexical variation, frequency bands and lexical complexity. Still using Main Suite examinations, Proudfoot's article explores lexical verb use in the Writing test papers and highlights the areas where both lower and higher proficiency learners exhibit non-native-like performance.

All the above articles focus on (testing) knowledge *of* language. The last article in this issue by Galaczi and Miller complements them by outlining the development of an examination designed to test knowledge *about* language. KAL (Knowledge About Language) is one of the TKT (Teaching Knowledge Test) modules focusing on teachers' knowledge of concepts related to language and language use. The paper identifies the rationale behind developing the test, its construct, and the quality assurance processes underpinning its development.

We finish this issue by reporting on the conference season and events Cambridge ESOL supported. Zeynep Urkun, Chair of IATEFL Testing, Evaluation and Assessment Special Interest Group, provides an account of the SIG events at the recent IATEFL conference (Harrogate, April 2010). Neil Jones from Cambridge ESOL describes two pre-conference workshops given at the Language Testing Research Colloquium (Cambridge, April 2010). Martin Nuttall from the ALTE secretariat describes the latest events in ALTE (Rome, May 2010). Finally, Lynda Taylor provides a brief on the 15th anniversary of the *Studies in Language Testing* series and Sacha DeVelle announces the winner of the IELTS MA award.

Insights and issues arising from the English Profile Wordlists project

ANNETTE CAPEL CAMBRIDGE UNIVERSITY PRESS

Introduction

In March 2007, Cambridge University Press approached me about working on a new project within the English Profile Programme which would investigate learner vocabulary, largely through corpus-informed research, in order to grade it at four different Common European Framework of Reference (CEFR) levels (A1–B2). Initially, the *English Profile Wordlists* were seen as a resource for authors and editors, syllabus designers, teacher trainers, researchers and examination writers. Cambridge ESOL, a key English Profile partner, has actively encouraged the development of the Wordlists, viewing them as a useful adjunct to their own Vocabulary Lists.

Working with John Sinclair and Gwyneth Fox on the COBUILD project through the 1980s gave me a lot of experience of corpus-informed research, and my 15 years as a Cambridge examiner have exposed me to tens of thousands of candidate answers from all over the world, at different CEFR levels. I have worked on various Cambridge ESOL examinations as an item writer since 1990, using their Vocabulary Lists in the construction of examination tasks, and I have also written textbooks for students who are preparing for Cambridge ESOL exams. The Wordlists project has brought all these strands together in a unique way.

Originally conceived as a book and CD-ROM, the *English Profile Wordlists* project got underway in May 2007, initially focusing on the first four levels of the CEFR, A1–B2. An early sample of approximately 2% of the alphabet – the letter G – was presented at the English Profile seminar meeting of 19 July 2007. Individuals attending that meeting and other reviewers around the world commented on this sample and made invaluable suggestions as to the coverage of the Wordlists and the organisation of its entries. From this feedback and subsequent contributions from Cambridge University Press authors and editors, it soon became apparent that the confines of a book would restrict the scope of the Wordlists and it was decided that a web-delivered resource would be more appropriate.

In October 2009, some two and a half years since the inception, a crucial development stage in the project was reached, with the fully searchable electronic pilot version of the British English Wordlists available online to English Profile partners only, during an evaluation and validation phase. A public preview version of the letters D, J and K with both British English and American English Wordlists is available on the English Profile website, and feedback is also being collected from this version (visit www.englishprofile.org and go to English Profile Wordlists). The full American English version of the Wordlists is currently in preparation in New York.

Basic Users and Independent Users: the A and B CEFR levels

The macro CEFR levels A (Basic User) and B (Independent User) sub-divide into the four levels A1, A2, B1 and B2. There seem to be key differences between a Basic User and an Independent User in terms of their vocabulary development. As the CEFR describes, the Basic User encounters familiar topics and through these, starts to acquire basic lexical sets, together with the top senses of high-frequency words. Many of these words are the grammatical building blocks that enable learners to structure their language at phrase and sentence level: *a, the, some, any, and, but, if, so* and so on. There is also some learning of formulaic phrases at A1 and A2 level, for example: *See you soon, Excuse me, No thanks, Take care.*

The Independent User starts to deal with less familiar topics and thus more topic sets are added, along with additional senses of ‘known’ words. It should be noted that the B levels in particular are broad bands, usually covering more than a single year of language tuition. By the higher end of B1 and into B2, learners appear to be acquiring more ‘chunks’ at phrase level and there is gradual use of frequent collocations and certain phrasal verbs.

Various sources already exist that provide detailed information about these CEFR levels, although rather less documentation is available on the precise vocabulary that attaches to these levels. Nevertheless, the CEFR, although primarily a skills-based description, has been a good starting point for the Wordlists. The published titles in the T-series, *Waystage* (van Ek & Trim 1998a), *Threshold* (van Ek & Trim 1998b), *Vantage* (van Ek & Trim 2001), and *Breakthrough* (van Ek & Trim, now available on the English Profile website), contain much relevant detail, including a word index at Waystage and Threshold levels, ‘lexical exponents of specific notions’ at Vantage level, and ‘theme-related can-do statements and vocabulary’ at Breakthrough level.

Cambridge ESOL examinations are reliably pegged to the CEFR levels. At A1, A2 and B1 levels, item writers consult Vocabulary Lists to check whether it is permissible to test a word at a given level. These lists illustrate the span of learner understanding of vocabulary within the range of topics covered by the relevant examinations and have therefore been useful documents for the Wordlists project. The Key English Test (KET, A2) and Preliminary English Test (PET, B1) Vocabulary Lists have been in use for more than 15 years and are in part derived from the van Ek & Trim titles mentioned above. These lists are updated annually, with reference to native speaker and learner corpus evidence. Cambridge ESOL has also developed a wordlist at

A1 level for young learners – the Starters and Movers lists in the Young Learners examinations handbook (University of Cambridge ESOL Examinations 2007). However, the YLE lists have been of marginal use in the Wordlists compilation, due to their focus on a single, young age group.

Rationale

The core objective of the Wordlists project is to establish which words are commonly known by learners around the world at the CEFR levels A1 to C2, and to assign these levels not merely to the words themselves but to their individual meanings. In other words, rather than providing a syllabus of the lexis that learners should know, the Wordlists project has concentrated on verifying what they do know. This ties in with the CEFR emphasis on ‘Can Do’ statements and also reflects research being done by other English Profile partners on criterial features at each of the six levels.

What is meant by ‘know’ in this context? In early discussion at the sample stage, before the main compiling commenced, the question of possible disparity in level between receptive and productive language was raised, both internally and by some reviewers. This is an issue that has often been researched over the years (e.g. Melka 1997) and yet arguably remains intangible and unproven. Moreover, much will depend on learning styles and, indeed, classroom dynamics – the extent to which opportunities are given for productive use. In examination classes, balanced preparation of all four skills has to take place if candidates are to succeed and so it is advisable to encourage learners to actively use the words and meanings they are exposed to in textbooks and classroom teaching. In general, modern communicative classrooms provide more consistent opportunities for actively using new language than a generation ago and, especially in relation to vocabulary, the prevailing advice seems to be ‘use it or lose it’. So perhaps the gap between receptive understanding and productive use is not as wide as some people have claimed. See Melka (1997) for a fuller discussion of this area. The evidence in the *Cambridge Learner Corpus (CLC)* suggests this might be the case, and illustrations of relevant findings will be given later in this article.

Compiling methods

Most of the words and phrases covered in the Wordlists are derived in the first instance from lexicographic research into frequency carried out by Cambridge University Press, which has informed both the second and third editions of the *Cambridge Advanced Learner’s Dictionary (CALD)*, (2008). This research made use of the *Cambridge International Corpus*, a billion words of written and spoken English text taken from a very wide range of sources, in order to decide on the relative frequency of senses for words within the top 5,000–6,000 words in English. To this end, lexicographers manually counted concordance lines for these words and, according to the number of occurrences of a given sense, assigned one of three relative frequency levels to it: E, I and A (Essential, Improver and Advanced), where E represents the highest frequency of occurrence.

At the outset of the *English Profile Wordlists* project the dictionary entries for all words/senses tagged Essential (E), Improver (I) or Advanced (A) were placed in a database, and formed the starting point for the compiling process. It was thought likely that anything tagged ‘E’ or ‘I’ would fall within the four CEFR levels A1–B2, but that the ‘A’ category might well contain many words above B2 level and would need careful scrutiny and cross-checking with other sources, including other learner dictionaries and classroom materials. Following this analysis, around 55% of the ‘A’ words have been included in the Wordlists data up to B2 level. The remaining ‘A’ words will be revisited in the context of the C levels Wordlists development, which commenced in March 2010.

In the event, a few ‘I’ words have also been omitted from the Wordlists due to lack of learner evidence worldwide. For example, although the ‘I’ verb *eliminate* is fairly frequent in native speaker corpora, its use at First Certificate in English (FCE) B2 level is largely limited to L1 first language speakers. It also seems more appropriate to the C levels in terms of its register and use.

An analysis based solely on native speaker frequency does not capture certain words that are useful to learners and which have a high frequency in the language classroom. The *Cambridge Learner Corpus*, a collection of student writing which currently comprises over 40 million words with more than 150 nationalities represented, provided substantial evidence of these words. Wordlists from coursebooks and other materials for learners have also been scrutinised in order to support the inclusion of words or senses in the Wordlists. Some examples of these words are: *album, download, guidebook, haircut, questionnaire, skateboard, trainer*. Most of these additions are nouns and either represent lifestyle choices that are important to learners – downloading music or skateboarding, for example – or are words that come directly from the teaching and learning experience, as in *questionnaire*. All the examples above are listed in either the KET or PET Vocabulary List.

One further source that was consulted throughout was the *Cambridge English Lexicon* (Hindmarsh 1980). This landmark title in English Language Teaching was compiled by Roland Hindmarsh and involved many years of painstaking and detailed work. First published in 1980 (and out of print for many years), it has proved to be a very helpful guide, as it too was organised at sense level. The *Lexicon* effectively spanned all six CEFR levels, although its primary purpose was to ‘produce a lexicon worth teaching and learning at the intermediate level of FCE’, where level 1 would now correspond to A1, level 2 to A2, levels 3 and 4 to B1, and level 5 to B2. Hindmarsh (1980) then assigned levels 6 and 7 to the harder senses of the words he included within FCE level and, broadly speaking, his level 6 corresponds to C1 and level 7 to C2. It should be noted that Hindmarsh (1980) had very little computational support beyond access to basic word frequency lists and yet, where the language has not changed, his intuitions seem extraordinarily accurate and well judged.

Interestingly, the number of entries in the Wordlists pilot version is quite close to Hindmarsh’s (1980) final total of 4,500. There are currently 4,666 entries in all, and the

addition of ‘new’ words at each level which add up to this total, is given below:

Words at A1	601
New words at A2	925
New words at B1	1,429
New words at B2	1,711

These totals may be subject to some change, since the Wordlists pilot version is undergoing a validation period during 2010. English Profile partners and experts in different areas of vocabulary commissioned by Cambridge University Press are currently evaluating both the data and the electronic resource itself. As part of this validation exercise, Cambridge ESOL has been trialling use of the Wordlists on a range of exams at A1 to B2 levels. Feedback from item writers and staff is generally very positive.

Structure of a Wordlists entry

Although the Wordlists are primarily organised by conventional dictionary headword, many fixed and semi-fixed phrases, phrasal verbs and other multi-word ‘chunks’ have been dealt with prominently as separate senses, reflecting current approaches to vocabulary learning and teaching (see O’Keeffe, McCarthy & Carter 2007). In each headword, entry senses are ordered by CEFR level and within that according to their relative frequency in native speaker use.

Each entry uses reliable information from Cambridge dictionaries. Both the *Cambridge Advanced Learner’s Dictionary* (2008) and the *Cambridge Learner’s Dictionary* (2007) have been used, but examples of typical use come mainly from the latter, which targets learners at a lower level. Other information has been included from these dictionaries, including audio and phonetic pronunciations, grammar and usage information, guide words to senses, and a short definition. The dictionary examples of typical use also highlight important collocations. These examples are seen as being within B2 level, but do not necessarily reflect the actual CEFR level assigned to the word or phrase they are exemplifying. Generally, the lower level examples come first.

The majority of senses also include a learner example taken from the *Cambridge Learner Corpus*, showing typical use of the word or phrase. Any errors made by the writer that are peripheral to the use of the target word or phrase are corrected within square brackets. The learner example is presented in a box which also provides information about the examination taken by the writer, the CEFR level of that examination, and the candidate’s first language. Every effort has been made to spread the selection of learner examples across a wide range of first languages and to avoid examples of cognate use that might misrepresent the norm in terms of average ability around the world. As already mentioned, the intention is to make the Wordlists as widely applicable as possible, rather than viewing them as a Eurocentric tool. The CEFR itself is used in many different teaching and learning contexts all over the world, so any attempt to quantify vocabulary within its levels should mirror this reality.

Wherever possible learner examples have been selected at the same level as that assigned to the word or sense. This is very often the case within the B levels. However, because of the current lack of substantial amounts of data at A1 and A2 levels, it has sometimes been necessary to move up to the CEFR level above in order to find a suitable example. As the *Cambridge Learner Corpus* grows by up to 3 million words each year, further searches can be made in an attempt to rectify this and to select learner examples for any senses where they are missing.

At the same time, it has to be remembered that the *Cambridge Learner Corpus* is a corpus of *written* learner English, so spoken learner uses of words and expressions cannot be exemplified in the pilot version of the *English Profile Wordlists*. However, extensive work is now ongoing under the direction of Mike McCarthy to collect spoken learner data, consisting of both examination and other non-examination sources. It is hoped that the resultant corpus, expected to be at least 2 million words, will yield appropriate learner examples in time for incorporation into the 2011 public version of the Wordlists.

Using the English Profile Wordlists

As the Wordlists are delivered electronically, the data can be searched in any number of ways. It is recognised that different users will have different requirements and constraints. So, for example, culturally sensitive words which could give offence in materials or examination tasks – *alcohol, invade, murder* – or specific senses of a word, such as the use of the verb *drink* to mean ‘to drink alcohol’, have been highlighted in the Wordlists and can be hidden in the data.

It is possible to specify certain CEFR level(s) only (the default search runs on the entire A1–B2 data). The user can key in a particular word or phrase and check its level, or access the data alphabetically by browsing a particular letter or stretch of entries. ‘Wildcard’ searches can be conducted using the asterisk: for example, all words ending in -ing can be found by keying in ‘*ing’. In addition, the Advanced Search facility allows for restricted searches on a specific aspect or combination of aspects, including grammar, usage, topic or affixation. So, for example, the user might want to find out which nouns are known at B2 level within the topic of crime, or want to obtain a list of adjectives at B1 level containing both the prefix *un-* and the suffix *-able*.

Affixation

Affixation is an area that has been of special interest to me during the research and compilation of the Wordlists, and an early decision was taken to separate out all dictionary ‘run-ons’ (other parts of speech that are included without explanations at the end of an entry) and treat them as potential headwords in their own right provided they were sufficiently frequent in native speaker use. An experienced lexicographer, Elizabeth Walter, carried out research into native speaker frequency and produced suggestions as to the level of individual prefixes and suffixes based on corpus evidence and in comparison with the Bauer & Nation (1993) word families table. The ‘transparentness’ of meaning was a

deciding factor in determining inclusion or otherwise within the Wordlists. Two levels for affixes were arrived at within the scope of A1–B2 learners: a small group of the most key and common affixes, likely to be known from a relatively early stage of learning, such as *un-*, *-ed* to form adjectives, *-er* to refer to people; above that, a somewhat larger set of affixes likely to be known by students of intermediate level and above, for example *dis-*, *re-*, *-ful*, *-less*.

Further issues arise because the application of these affixes is not always straightforward. Take the frequent suffix *-able*, which has been given two levels in the Bauer & Nation (1993) tables: 3 and 6. On some occasions, the addition of the *-able* suffix is transparent, in other words it is easy to work out the meaning of the newly formed word, which reflects the core meaning ‘able to be’, for example *downloadable* from *download*. On the other hand *changeable* doesn’t mean that something is ‘able to change’, it means that it frequently *does* change. Similarly, *knowledgeable* falls outside the core meaning since it means ‘knowing a lot’.

To return to the prefix *dis-*, all of the words formed from it that are included in the current Wordlists are at the B levels, the majority at B2. At B1, there are the verbs *disagree*, *disappear* (two of the three senses included in the Wordlists are at B1), *dislike*, the adjective *disabled* and the noun *disadvantage*. As mentioned earlier, the frequency of words in native speaker corpora has often influenced their inclusion or otherwise in the Wordlists. So, for example, the verb *disbelieve* has been omitted, the norm seems to be to use ‘not believe’ rather than ‘disbelieve’; the noun *disapproval* has also been omitted, partly on grounds of low frequency, but also because ‘double’ affixation (*dis-* plus *-al* in this case) appears to be more challenging for learners.

There are a few interesting exceptions to this, though they usually consist of the combination of a common prefix and suffix, such as *un-* plus *-ed* or *-ly*: the adjective *unexpected*, more frequent in native speaker use than the adjective *expected*, has been given B1 level in the Wordlists due to evidence in the *Cambridge Learner Corpus* of its use worldwide, whereas *expected* barely makes B2, with few learner examples. This is possibly due to its more restricted usage, coupled with the constraint of having to come before a noun. Indeed, the *Cambridge Learner Corpus* shows that at C1 level in the Certificate in Advanced English (CAE), learners are producing this pattern much more frequently, in phrases such as *the expected amount/level/standard*. Both adjectives are tagged in CALD as ‘I’ words, though there seems to be evidence that *unexpected* is the more frequent of the two for native speakers.

An interesting parallel with native speaker frequency can be observed in relation to members of the word family for *fortunate*: there is good evidence at KET/A2 level for the adverb *unfortunately* so it is assigned to A2 in the Wordlists, whereas the adverb *fortunately* seems to belong to B1 level, for in 40 million words of CLC data, there is one cite for *fortunately* at KET, and this by a Spanish-speaking learner, illustrating cognate use; the two adjectives are much less used by learners, and then only from FCE/B2 level. This reflects the relative frequency of the words in native speaker corpora: *unfortunately* is a CALD ‘E’ word, and also gets three stars in the *Macmillan English Dictionary* (2007);

fortunately is CALD ‘I’, while the two adjectives are only ‘A’, indicating a somewhat lower frequency of use. From this example, it can be demonstrated that learners do not always meet the ‘root’ word first, and a flexible approach can be taken to word formation in the classroom, based on frequency and usefulness. Clearly, the word *unfortunately* is of practical use to learners, in a variety of spoken contexts and written genres.

Word family panels

As the above discussion shows, affixation is far from straightforward for learners and arguably needs to be focused on systematically at the B levels and above. For this reason, the Wordlists feature word family panels, which appear at the head of any entry where two or more related forms have been found to be within B2 level. So, for example, at the entry for the adverb *separately*, the family displayed is as follows:

Nouns: separation

Verbs: separate

Adjectives: separate

Adverbs: separately

This family seems to be confined to the B levels, with only two senses of the adjective known at B1 and the remaining family members belonging to B2. There are however several instances where members of word families begin to be acquired in the early stages of learning and yet related words and phrases of lower frequency are only acquired much later – the adjective *different*, for example, is an A1 word, and the noun *difference* comes in at A2 (both are CALD ‘E’ words), but the verb *differ* (a CALD ‘A’ word) is only used from B2, as are the phrases *make a (big) difference* and *make no difference/not make any difference*.

Prioritising senses: the relative difficulty of senses for the verb *keep*

The verb *keep* is a very frequent word in English, occurring within the top 300 words in the frequency listing of the *Cambridge International Corpus*. It has many senses, features in many verb–noun collocations, and combines with particles to form a number of phrasal verbs. For all these reasons, I was keen to investigate it further, so chose to tackle the letter K early on in the compiling. *Keep* is a word that shows progressive learning of its senses over the A and B levels. In the KET Vocabulary List, its use in testing is limited to two senses, as in the examples *May I keep this?* and *Keep right!* The former sense, meaning ‘to have something permanently’, is the first sense in the Cambridge dictionaries, illustrating that it is probably the most frequent for native speakers; the latter, meaning ‘to stay’ is also high up in terms of frequency (both senses are CALD E). These two senses have been assigned A2 in the Wordlists, along with the phrase *keep something in/on*, etc. which has been added from the *Cambridge Learner’s Dictionary* (2007) (here, the first CALD sense has been split to provide more support to lower level learners).

One further sense which is given ‘E’ status in CALD and is very frequent in native speaker use is ‘to continue doing something without stopping, or to do it repeatedly’. This appears in the Wordlists under the phrase *keep doing something* and has been assigned B1 level. Interestingly, there is also evidence at PET/B1 for the closely related phrasal verb *keep on doing something*. This phrasal verb is listed in the PET Vocabulary List, where *keep* itself is listed without any sense restrictions.

As already highlighted, *keep* appears in many useful phrases and collocations, and the entry in the Wordlists is quite lengthy for this reason. Users can choose to select ‘outline view’ rather than ‘full view’, to get a briefer listing of senses, with only guide words and definitions displayed. At B1 level, there is one further sense, with the guide word DELAY, as in *I’m sorry to keep you waiting*, and the phrases *keep in touch* and *keep a/ something secret*. There are four more phrases listed at B2 level, together with two other less frequent senses: the intransitive use to do with food, as in *This product will keep for three days if refrigerated* and the sense ‘to have and look after animals’. This then is a further refining of the information contained in the PET Vocabulary List, and provides evidence of which senses B1 level learners can cope with.

Finally in the *keep* entry, there are 15 senses of phrasal verbs, all but three at B2. Researching the level of phrasal verbs has been problematic: coursebooks feature them, especially from B1 level, and many are listed in other sources. Hindmarsh (1980) included a phrasal verbs appendix that runs to 10 pages. Furthermore, although they are often quite frequent in native speaker corpora, they occur relatively rarely in the *Cambridge Learner Corpus*. This could be down to a number of factors. For example, they may be more commonly spoken by learners rather than written, so it will be interesting to interrogate the spoken learner corpora in due course. For the moment, of the 14 phrasal verbs at B1 and B2 listed in the *keep* entry, only four have learner examples accompanying them. A decision on the final inclusion policy for phrasal verbs will have to be taken towards the end of the pilot phase, and it is hoped that feedback will be forthcoming in this area from users and public viewers.

The senses of the word *case*

At the February 2009 English Profile seminar, when compiling had reached the halfway stage, an interim report was delivered about insights and issues, from which this article gets its title. The lack of production of phrasal verbs within B2 was one such observation, along with the fact that the most frequent senses of a word are not always the first taught. A good example of this is the word *case*, whose most frequent sense for native speakers is ‘a particular situation or example of something’, as in the following dictionary examples:

Over fifty people were injured, in several cases seriously.

I wouldn’t normally agree but I’ll make an exception in this case.

The number of new cases of the new illness has risen.

We have lots of applications from people who want to study here and in

each case we consider the candidate very carefully.

She was suffering from an extreme case of sunburn.

However, the *Cambridge Learner Corpus* and other checking sources have shown that learners meet the lower frequency senses of ‘container’, as in *pencil case*, and ‘bag’ (the latter in British English only) earlier than other, more frequent senses. There are practical reasons in this case – pencil cases are found in most school classrooms around the world, after all. Research has also shown that concrete meanings (e.g. *pencil case*) are more easily learned than more abstract ones (e.g. *case* [=situation]), irrespective of frequency. What is perhaps more surprising is that the meaning of ‘situation’ exemplified above often seems not to be explicitly taught in coursebooks, and the sense is not currently included in the PET Vocabulary List. At the same time, Hindmarsh (1980) clearly recognised its importance in learning by allocating it the same level (level 2) as the container sense. The sense has been assigned B1 in the Wordlists for the moment, but there may be a good case for lowering it to A2.

Another insight, which arises from browsing the *Cambridge Learner Corpus*, is the influence of the learner’s first language, especially at the A levels. This can help or hinder the learner. Close cognates, as already noted, will mean earlier than average use of a word or phrase, whereas ‘false friends’ may lead to errors or inappropriate use. One interesting example in the compiling process was the phrase *in fact*. This has been assigned B1 in the Wordlists (the less frequent *in actual fact* is B2), in spite of there being evidence for the phrase at A2, i.e. being produced in the KET writing task. In consultation with reviewers, B1 was seen as more suitable, as many of the A2 cites were using the phrase wrongly. A large proportion of these learner examples were written by first language speakers of Italian, where the word *infatti* is frequently used; a common error in the *Cambridge Learner Corpus* is the writing of the phrase as a single word ‘infact’, demonstrating first language interference.

Issues arising from the compiling process

Although it has proved to be a very time-consuming project, I have become more and more convinced of the desirability of reporting CEFR level for individual senses, in order to provide more solid support for language teaching professionals such as authors, editors and examination writers. This is nothing new – I am merely following in the sturdy footsteps of Hindmarsh. However, combined with this is an increased attention to phrases and collocations, which corpus evidence can readily highlight. Further work could be done in the area of collocation and this will certainly be needed in the next stage of development, Wordlists for the C levels.

One factor of current concern is the very wide age range of learners around the world, and whether a single resource such as the Wordlists can actually report on a level for all ages. The Wordlists are recommended for anyone dealing with learners aged 11 and upwards; for young learners, a different grouping of words/senses seems inevitable and has already been developed in the ESOL Young Learners Vocabulary Lists at A1 and A2.

Coupled with this is the fact that the Wordlists focus on general English and therefore cannot really help those working in Business English, or teachers involved in Content and Language Integrated Learning (CLIL) projects. It is hoped that 'add-on' lists might be developed for these and other specific domains in due course.

Additionally, the focus of the *Cambridge Learner Corpus* on written examination data means that other learner corpora will need to be accessed in the future to verify the findings included in the Wordlists. The current development of the *Cambridge English Profile Corpus* (CEPC) is therefore an exciting one, as eventually there will be 10 million words of non-examination learner data (20% spoken and 80% written). The CEPC will also include English for Specific Purposes data.

It is obvious that a lot more work is waiting to be done, both within the current levels of the Wordlists and at the C levels. Quite apart from the inclusion of spoken learner evidence just referred to, much more research could be done on affixation and its challenges for learners; the inclusion of a selected group of headwords for certain word families based on frequency is a modest first step.

Thanks to the efforts of the highly talented and diligent electronic dictionaries team at Cambridge University Press, the *English Profile Wordlists* are an extremely attractive and powerful resource. The preview version for the letters D, J and K, together with selected entries from the A–Z pilot version known as the 'Word of the Week' feature can be viewed now. Please take the time to send in your feedback via the website, so that the Wordlists can be further improved.

References

- Bauer, L and Nation, I S P (1993) Word families, *International Journal of Lexicography* 6 (4), 253–279.
- Cambridge Advanced Learner's Dictionary* (Third Edition 2008), Cambridge: Cambridge University Press.
- Cambridge Learner's Dictionary* (Third Edition 2007), Cambridge: Cambridge University Press.
- Hindmarsh, R (1980) *Cambridge English Lexicon*, Cambridge: Cambridge University Press.
- Macmillan English Dictionary* (Second Edition 2007), Oxford: Macmillan Education.
- Melka, F (1997) Receptive vs. productive aspects of vocabulary, in Schmitt, N & McCarthy M (Eds) *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge: Cambridge University Press, 84–102.
- O'Keeffe, A, McCarthy, M J and Carter, R (2007) *From Corpus to Classroom*, Cambridge: Cambridge University Press.
- University of Cambridge ESOL Examinations (2007) *Cambridge Young Learners English Tests: Handbook for Teachers*, Cambridge: University of Cambridge ESOL Examinations.
- Van Ek, J A and Trim, J L M (1998a) *Waystage 1990*, Cambridge: Council of Europe/Cambridge University Press.
- Van Ek, J A and Trim, J L M (1998b) *Threshold 1990*, Cambridge: Council of Europe/Cambridge University Press.
- Van Ek, J A and Trim, J L M (2001) *Vantage*, Cambridge: Council of Europe/Cambridge University Press.
- Van Ek, J A and Trim, J L M *Breakthrough*, retrieved from: http://www.englishprofile.org/index.php?option=com_content&view=article&id=120&Itemid=94

Use of words and multi-word units in Skills for Life Writing examinations

IVANA VIDAKOVIĆ RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL
FIONA BARKER RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

Introduction

Launched in 2005, and set at five levels (A1–C1) of the Common European Framework of Reference (CEFR), Cambridge ESOL Skills for Life examinations are offered in three modes: Reading, Listening & Speaking and Writing. Skills for Life is the Government's strategy for improving adult literacy, numeracy and ESOL skills in England. As part of this strategy, Cambridge ESOL's Certificates in ESOL Skills for Life are based on the standards for adult literacy and the Adult ESOL Core Curriculum. They are designed for 'adult (16+) learners whose first language is not English and who are living or trying to settle in England' (University of Cambridge ESOL Examinations 2005:6). As such, they are designed for a very diverse range of learners in terms of educational and employment backgrounds, as well as aspirations, literacy levels and language learning skills.

As part of the ongoing validation of Cambridge ESOL examinations, the aim of this paper is to investigate the

validity of Skills for Life Writing examinations from a lexical perspective by ascertaining how well they differentiate between language proficiency levels. This will be done by analysing the use of individual words and lexical bundles (extended collocations) in candidates' written responses to examination questions.

Vocabulary frequency measures have been found to be insufficiently robust for capturing the differences between proficiency levels (see Lim & Galaczi 2010 in this issue, Read & Nation 2006, Schmitt 2005), possibly because single-word frequencies are not the only measure of lexical mastery. Therefore, an investigation of multi-word units, such as collocations and lexical bundles, has been called for.

Lexical bundles are recurrent word strings of three or more words. They are, unlike idioms, semantically transparent and, unlike collocations, mostly incomplete structural units (e.g. *at the end of*). Their functions and structure vary, which is why they are identified empirically,

through their frequency of recurrence. They have been called ‘important building blocks of coherent discourse’ (Hyland 2008:8) by virtue of providing ‘a discourse frame for some other new assertion’ (Biber & Barbieri 2007:283). They are preferred and conventional ways of expression and are considered to be indicators of competent language use in speech and writing (Biber & Barbieri 2007, Cortes 2004, Hyland 2008). Therefore, they are also of importance to second language (L2) learners of English. Getting grammar right is a necessary but not sufficient condition for sounding native-like, which is why ‘using well-trying expressions in appropriate places’ (Biber, Johansson, Leech, Conrad & Finegan 1999:990) would go a long way in helping L2 learners sound natural in their L2.

The functional classification of bundles varies in the literature. We have adopted the classification used by Cortes (2004) and Biber & Barbieri (2007) because it seems applicable to Skills for Life written responses. The mentioned authors distinguish between three primary functions of bundles:

- Referential bundles ‘identify an entity or single out some particular attribute of an entity as especially important’ (Biber & Barbieri 2007:271). They can be descriptive (e.g. *the size of the*), quantifying (e.g. *a large number of*), or time/place markers (e.g. *in France*).
- Discourse bundles indicate overall discourse structure the relationship between prior and coming discourse. They introduce topics (e.g. *What I want to do* is quickly run through the exercise, *ibid:271*), elaborate or clarify them (e.g. *has to do with*), express inference, contrast or comparison (e.g. *on the other hand*).
- Stance bundles convey attitudes (e.g. *to be able to*) and epistemic evaluations (e.g. *I don’t know what*) of the assertion to follow.

Nekrasova (2009) added one more class, termed ‘special conversational bundles’, which express politeness, inquiry and report (e.g. *thank you very much, I said to her*).

Several studies explored the knowledge and use of lexical bundles in the context of L2 acquisition. They either focused on advanced learners of English only or, if they examined the use of multi-word strings across proficiency levels (Read & Nation 2006), they did so in an exploratory way, without differentiating between collocations and bundles, which is why it is not clear in what ways their L2 acquisition is similar or different. Read & Nation (2006) reveal that the use of multi-word strings seems to develop and increase considerably at higher proficiency levels (IELTS Band 8). Their finding that advanced learners exhibit a fluent and native-like use of multi-word strings in a speaking task is supported by Nekrasova (2009) who found that advanced learners of English perform as well as English native speakers in a gap-filling and dictation recall task involving lexical bundles. In contrast, Granger (2001) showed that lexical bundles were not mastered by advanced learners of English who used too few native-like bundles and many foreign-sounding ones in essay writing. In addition, Schmitt, Dörnyei, Adolphs & Durow (2004) reveal that non-native speakers produced far fewer bundles correctly than English native speakers when required to repeat a stretch of

dictation under time pressure. The findings may be mixed due to differences in elicitation methods or a possibility that students’ proficiency levels across studies may not really be the same. In sum, it is still far from clear how L2 acquisition of lexical bundles unfolds and what kind of linguistic behaviour characterises each level of proficiency. This paper will, therefore, attempt to shed more light on it.

The present study examines lexical development of L2 learners of English, based on 100 written responses to Skills for Life Writing examinations across five proficiency levels (A1–C1 on the CEFR) with the aim of investigating the validity of the above examinations and informing L2 acquisition research. Our discussion focuses on frequency-based analyses of individual words and lexical bundles and also on a qualitative analysis of functions performed by bundles in L2 learners’ writing across several proficiency levels, thus filling the gap in the previous literature. We compare several methods of investigating L2 lexical development and use WordSmith Tools and other software for lexical analyses.

Research questions and methodology

Lexical use across proficiency levels will be explored through the following research questions:

1. How informative are single-word frequency counts in identifying vocabulary progression across proficiency levels of Skills for Life Writing candidates?
2. Are there differences across proficiency levels with regard to frequency of lexical bundles?
3. Does an investigation of functions performed by lexical bundles reveal differences between proficiency levels?
4. Does the use of lexical bundles distinguish better between proficiency levels than single-word frequency measures?

In order to answer the first research question, a variety of frequency-based lexical analyses were carried out: token (the total number of words used) and type (the number of different words) counts and the investigation of lexical variation/diversity, density and sophistication. All these measures were used previously (e.g. Read & Nation 2006, Schmitt 2005).

Lexical variation/diversity, measured by standardised type-token ratio (STTR) in the WordList program of WordSmith Tools, indicates a degree of lexical richness of a text based on the amount of repetition.

Lexical density is the proportion of content words (nouns, verbs, adjectives and adverbs) in a text. Since content words carry a higher information load than function words (e.g. prepositions, articles), texts with higher density figures are likely to be richer and more demanding, linguistically and cognitively. However, research has shown that higher lexical density figures could also be a product of non-native-like ‘truncated’ grammar of lower proficiency learners who may tend to omit function words thus producing texts with a higher proportion of content words (Schmitt 2005). The tool used to obtain figures for lexical density was Vocabulary Profiler within Compleat Lexical Tutor software. The same tool was employed for the analysis of lexical sophistication.

Table 1: An Overview of Skills for Life Writing papers used in this study

Task	Entry 1 (A1)	Entry 2 (A2)	Entry 3 (B1)	Level 1 (B2)	Level 2 (C1)
1	Form-filling	Form-/questionnaire-filling	A (semi-) formal letter to friends rejecting their dinner invitation	A letter of complaint	A letter of complaint
2	A message to your neighbours informing them about nearby shops	A note to an injured friend	An email/note to a friend	A letter of reference	A letter to a newspaper Editor about the controversy over a traffic ban in the city centre
3	A note to neighbours (a dinner invitation)	An email to a friend describing your trip	An essay/report for a teacher on a festival or celebration	An article for a local newspaper on the tendency nowadays to marry older	a) A report for your tutor (based on a graph provided) on the population in the UK or b) article for a magazine/newspaper on the positive and negative aspects of mobile phones

Lexical sophistication is the percentage of low-frequency or rare words in a text. The above software classifies words found in a text into four categories: the first 1,000 words, the second 1,000 words (both from West 1953), the words occurring in the Academic Word List (Coxhead 2000) and the words which are outside these three categories (off-list words). Even though it needs revision, West's (1953) wordlist covers a high percentage of fiction, non-fiction and academic texts and is, therefore, still widely used in lexical research and language teaching.

To address the second research question, the frequency lists of lexical bundles were obtained using the WordList program. Four-word clusters with at least two occurrences from each spelling-corrected file were identified and produced by the program. The qualitative analysis of lexical bundles from the functional perspective was carried out using the functional classification outlined in the Introduction.

Data

Written responses to 2008 Skills for Life Writing examinations of 20 passing candidates per entry level were analysed. There was an even spread of the 10 most common first language backgrounds of the Skills for Life candidature.

Skills for Life Writing tasks draw on authentic or semi-authentic texts and task types so that learners can use English in the way they are expected to do so in real life. There are three tasks at each level (cf. Table 1). Tasks 1 in Entry 1 and Entry 2 are not included in the current data set because they do not elicit extended writing.

Each task includes a written prompt or a written and a visual prompt, such as a map at lower proficiency levels and a graph at Levels 1 and 2.

A candidate's performance is assessed using a set of criteria with three focuses: word-focus (spelling, handwriting and the appropriateness of vocabulary), sentence-focus (e.g. grammar and punctuation) and text-focus (e.g. coherence, cohesion and organisation, task coverage, awareness of the audience and register).

Results and discussion

This section starts out with a discussion of the findings obtained through frequency analyses of individual words in our set of Skills for Life data. It then moves beyond

individual words to examine the frequency and functions of lexical bundles across five proficiency levels.

Individual words

Candidates' written responses across five Skills for Life levels were analysed to determine the total number of tokens (words), types (different words), the average word and sentence length as well as the degree of lexical variation, density and sophistication in order to address the first question (cf. the previous section).

According to the table below, the average word and sentence length increase with increasing proficiency levels. The increase in word length is small, but may reflect a greater reliance on semantically and morphologically more complex abstract content words. The increase in the average sentence length is very considerable across the levels: it is the smallest (being less than three words) between the two lowest levels, Entry 1 and Entry 2, but it ranges between four to eight words at higher proficiency levels. This is likely to indicate an increasingly more complex sentence structure at higher proficiency levels, which is illustrated below by examples from Entry 1 and Level 2 data, respectively:

Would you like to come to my house? I want make some food.

It goes without saying that the noise and the pollution level that we are now suffering, will decrease enormously.

There is a clear progression from Entry 1 to Level 2 data in terms of the total number of tokens and types. The number of tokens and types is the lowest at Entry 1 level, while it clearly grows with each subsequent level. This may be a result of task requirements (since Skills for Life tasks at

Table 2: Lexical analysis of Skills for Life candidates' written responses

Level	Average word length	Average sentence length	Tokens	Types	STTR (50)	Lexical density
Entry 1 (A1)	3.8	9.58	2,274	442	73.21	0.48
Entry 2 (A2)	3.8	12.17	3,323	541	75.31	0.45
Entry 3 (B1)	4.0	17.19	4,565	834	76.95	0.47
Level 1 (B2)	4.2	20.89	10,941	1,411	77.64	0.47
Level 2 (C1)	4.4	28.4	13,480	1,786	80.02	0.48

each higher level require lengthier responses), but this could also show that higher level learners produce longer and more varied output since they have more L2 ‘material’ at their disposal and also because they engage in more complex communicative tasks as a result of their greater linguistic ability.

Lexical variation represented as the standardised type-token ratio (STTR) shows the percentage of new types for every 50 tokens. Even though there are no guidelines as to what should be considered a large or a small difference, it is evident that the STTR increases across proficiency levels, thus indicating an increasing degree of lexical diversity in the responses produced by learners. In other words, Entry 1 candidates resort to lexical repetition most often, probably due to a limited vocabulary in their interlanguage, while Level 2 candidates do so rarely.

Lexical density, however, does not exhibit much variation across the five Skills for Life proficiency levels. The factors which may be ‘responsible’ for this monotonous picture are at least two. Firstly, as mentioned earlier, the responses at lower proficiency levels (Entry 1 and Entry 2 in this case) may be characterised by ‘truncated’ grammar since some lower level learners tend to omit function words and rely primarily on content words. An Entry 1 example illustrates this behaviour:

On the street next my flat is small market.

This is a well-known and widespread phenomenon in L2 acquisition research and teaching practice. As stated by Corder (1981:85): ‘in its early stages, interlanguage, along with all simplified codes and languages (e.g. pidgins, baby talk) exhibit strong structural similarities, such as: simple or non-existent morphology, a more or less fixed word order, a small number of grammatical function words, etc. Such systems are used in restricted communicative functions, rely heavily on situational context and are minimally redundant in information theory terms.’

Secondly, learners at Entry 1 and Entry 2 levels write about concrete topics and tend to list pieces of information, thus producing many content words in a (grammatically) simple sentence, as in an Entry 1 example below:

You can buy all type of vegetables, meats, fruits, sauces, cakes, breads, etc.

The above two tendencies explain why lexical density figures may be higher than expected for lower proficiency data, but they do not explain why those figures are very similar for learners at Entry 3 and Levels 1 and 2. Evidently, lexical density cannot differentiate well between Skills for Life levels, not even between the highest and the lowest one. In view of this, it can be suggested that lexical density is not a very useful phenomenon to analyse in the context of Skills for Life data due to the special character of lower proficiency discourse, as argued above.

Lexical sophistication is measured by a quantity of advanced/sophisticated words in a text. Since more frequent words are normally learned before the less frequent ones, it is likely that academic and other sophisticated words will be rare, if at all present, in the texts produced by lower level learners (e.g. Entry 1 and Entry 2 levels).

Table 3: Lexical sophistication in Skills for Life candidates’ responses

	Entry 1	Entry 2	Entry 3	Level 1	Level 2
K1 Words (1–1,000)					
Types	281	361	522	872	1039
Tokens	1,918	2,958	3,978	9,832	11,706
Percent	83.34%	87.98%	85.99%	89.33%	86.32%
K2 Words (1,001–2,000):					
Types	60	77	122	191	275
Tokens	231	244	327	477	751
Percent	10.16%	7.26%	7.07%	4.33%	5.54%
K1 + K2	94.5%	95.24%	93.06%	93.66%	91.86%
AWL Words (academic):					
Types	2	2	21	128	191
Tokens	2	2	27	297	554
Percent	0.09%	0.06%	0.58%	2.7%	4.09%
K1 + K2 + AWL	95.04%	95.8%	93.64%	96.36%	95.95%
Off-List Words:					
Types	87	88	141	184	254
Tokens	123	158	294	400	550
Percent	5.41%	4.7%	6.36%	3.63%	4.06%
Overall:					
Types	430	528	806	1,375	1,759
Tokens	2,274	3,362	4,626	11,006	13,561
Percent	100%	100%	100%	100%	100%

The above table shows that the largest percentage of words used by all learners belongs to the first 2,000 most frequent words in English, the first 1,000 being the most frequent by far. The percentage of the first 2,000 most frequent words slightly decreases with increasing proficiency levels. Simultaneously, the percentage of academic words (AWL) increases considerably across levels. There are only two occurrences of academic words in Entry 1 and Entry 2 data. Academic words are used slightly more often by Entry 3 learners, but still very rarely. The raw frequencies and percentages of this type of words increase considerably only in Level 1 and Level 2 data, so that at Level 2 academic words constitute about 4% of the total number of words. Here are several examples of academic words from Skills for Life data:

- Entry 1: area, available
- Entry 3: appreciate, traditional
- Level 2: controversial, nevertheless.

The above findings show that the percentage of academic words seems to be a good differentiator between the lower proficiency levels (Entry 1, Entry 2 and Entry 3) on the one hand, and the higher proficiency levels (Level 1 and Level 2), on the other. It is important to bear in mind that Skills for Life is not a test of academic English and that tasks at higher levels are not academic, apart from, perhaps, the optional task 3a) at Level 2 (cf. Table 1). They are, however, linguistically and cognitively more complex, which is expected of tasks at higher proficiency levels, and which naturally prompts the use of more semantically complex, sophisticated and, evidently, academic words. In addition, the mere fact that words in AWL (Coxhead 2000) are often used in academic texts does not mean that some of them are also not used in everyday life, in newspapers and television, for example.

As far as off-list words are concerned, their types and

tokens increase with the proficiency levels (cf. table above). Morphologically erroneous word forms (e.g. **happinnesses*) or those which are a product of negative transfer from L1/another language (e.g. **bibliothek*), as well as non-sense words (e.g. **emittion*), appear to a very small extent. The investigation also revealed that there is a progression across levels from predominantly basic everyday words referring to objects (e.g. *bakery*) and proper nouns (e.g. *Poland*) at the lower proficiency levels (Entry 1 and 2) to semantically and morphologically more complex words which refer predominantly to abstract notions, qualities and activities (e.g. *alienated*, *tranquillity*) at higher levels (Levels 1 and 2).

In brief, our findings show that Skills for Life Writing examinations differentiate between the five proficiency levels they span. They provide initial validity evidence for these examinations and also shed more light on L2 acquisition of lexis. The usefulness of certain lexical measures (or the lack thereof) was also highlighted. The next section will move beyond the level of individual words to determine if the examination of lexical bundles can also meaningfully distinguish between language proficiency levels.

Lexical bundles

Previous research highlighted the importance of investigating lexical bundles in English and in the context of L2 acquisition of English. The purpose of the following part of the study is to uncover the similarities and differences between proficiency levels as far as the use of bundles is concerned in order to address our research questions 2, 3 and 4 (cf. Research questions and methodology).

In WordSmith we used the WordList program to get 4-word bundles at each proficiency level. We focused on 4-word bundles because they are more common than 5-word strings and offer a clearer range of functions than 3-word bundles (cf. Hyland 2008:8). We calculated the frequency per million words (standardised frequency) so that we could compare our results across datasets of different sizes. Following Biber & Barbieri (2007), we used a high cut-off of 40 times per million to identify recurrent lexical bundles. In order to avoid individual writer idiosyncracies, we excluded the bundles which were used by fewer than three candidates.

The table below shows the number of tokens and types of lexical bundles in Skills for Life data after the above cut-offs have been applied. It displays the raw frequencies of tokens and types as well as standardised frequencies of tokens. As far as bundle types are concerned, 'it is not possible to directly normalise the number of lexical bundle types to a rate per million words' (Biber & Barbieri 2007:268) which is why only the raw frequencies of bundle types are provided in the table.

As Table 4 reveals, the number of tokens and types of lexical bundles increases with the increasing proficiency levels. The gradual but considerable increase of bundle tokens is also accompanied by an increase in the number of lexical bundle types. The general tendencies aside, there is an evident 'jump' at Entry 2 where candidates used more bundles than Entry 3 candidates (both in terms of raw and standardised frequency) and more bundles than Level 1

Table 4: Lexical bundles (tokens and types)

Level	Lexical bundles: tokens		Lexical bundles: types	Total words
	Raw frequency	Frequency per million		
Entry 1 (A1)	23	10,114	6	2,274
Entry 2 (A2)	142	42,732	27	3,323
Entry 3 (B1)	89	19,496	23	4,565
Level 1 (B2)	296	27,054	66	10,941
Level 2 (C1)	464	34,421	91	13,480

and 2 candidates (in terms of standardised frequency), but fewer bundle types. The high number of bundles at Entry 2 level is not an indication of exceptional mastery of lexical bundles, but is rather a result of frequent lifting of word sequences from the input by Entry 2 candidates: 63% of bundle tokens in those candidates' responses were 'borrowed'. Entry 2 candidates, being lower level learners with a very limited lexical resource, mostly lifted word strings from the input which then became recurrent lexical bundles due to a high frequency of 'lifting'. Entry 3 candidates were, on the other hand, more creative by mostly relying on their own resources in producing bundles: 74% of the bundle tokens they produced were not found in the input. Their proficiency level may not have 'allowed' them to use more bundles which were not in the input, and the level of linguistic simplicity/complexity of the tasks at Entry 3 may not have given them much chance or need to rely on word strings from the input.

In conclusion, the above findings reveal that lexical bundles are rarely used by the lowest proficiency (Entry 1) candidates. Learning (more or less conventionalised) strings of words starts developing from Entry 2 level of proficiency, but is truly productive only at Levels 1 and 2, where bundles are the most numerous and diverse.

The numbers of lexical bundle tokens and types only tell a part of the L2 acquisition story. The investigation of bundles from the functional perspective could reveal their changing roles in learners' increasingly more complex discourse across levels of proficiency.

The functions of lexical bundles

As 'important building blocks of coherent discourse' (Hyland 2008:8), lexical bundles have various functions. While discourse becomes more complex with increasing proficiency levels, it is possible that the functional role of lexical bundles also evolves, so that certain functional types can be found only at higher levels of proficiency. The functional classification of bundles we adopted categorises bundles into the following types: referential, discourse organising, stance and special conversational. Bundles which do not belong to any of the above categories were termed 'Other'.

Table 5 shows functions of the bundles in Skills for Life data at all five levels as well as the types and tokens of bundles with those functions.

It is evident that the most frequent function of lexical bundles is referential at all five Skills for Life levels. The number and variety of referential bundles increases from

Table 5: Functional categories of lexical bundles

Level	Functions: types (tokens)					Total
	Referential	Discourse	Stance	Special conversational	Other	
Entry 1 (A1)	4 (17)	–	1 (3)	1 (3)	–	6 (23)
Entry 2 (A2)	20 (111)	1 (2)	3 (15)	3 (11)	1 (3)	28 ¹ (142)
Entry 3 (B1)	12 (50)	2 (6)	6 (20)	3 (13)	–	23 (89)
Level 1 (B2)	38 (171)	11 (55)	9 (35)	7 (26)	3 (9)	68 ² (296)
Level 2 (C1)	52 (272)	13 (79)	15 (55)	4 (19)	8 (39)	92 ³ (464)

1 One bundle (*I would like to*) functions as a stance bundle in some contexts and as a discourse bundle in other contexts. This duality is, therefore, captured in the table, which is why the numbers sum up to 28 rather than 27.

2 Two lexical bundles have dual functions, which is why the number of types sums up to 68 instead of 66.

3 A lexical bundle has a dual function: it is counted as two.

Entry 1 to Level 2 (with a drop at Entry 3). Bundles with a discourse-organising function are not used at all by Entry 1 learners and are very rarely used up to Level 1. It is only at Levels 1 and 2 that learners use a greater variety of discourse bundles and with a greater frequency. In contrast, lexical bundles expressing stance exhibit a more gradual evolution than discourse bundles. They are used, even though infrequently, even at Entry 1. Their number and variety increase with increasing levels of proficiency. Special conversational bundles follow a similar pattern as far as tokens are concerned, and a similar, but a less pronounced pattern as far as bundle types are concerned. Their variety (in terms of bundle types) does not grow considerably, potentially because they constitute a very limited set in English. At levels 1 and 2, the majority of special conversational bundle types are parts of one larger bundle: *I look forward to hearing from you* (e.g. *look forward to hearing, forward to hearing from*). Below are several examples of bundles from our data:

- Referential: *in your new home* (Entry 3), *expectancy at birth for* (Level 2)
- Discourse: *I am writing to, on the other hand* (Level 1), *first of all I, would like to express* (Level 2)
- Stance: *sorry about your problem* (Entry 2), *I am very happy* (Entry 3), *would like you to* (Level 2)
- Special conversational: *thanks for your invitation* (Entry 3), *look forward to hearing* (Level 2).

The more frequent use of stance and special conversational bundles (in contrast to discourse bundles) by lower proficiency learners (Entry 1, Entry 2 and Entry 3) may indicate that they are acquired and needed in communication earlier than discourse bundles. Firstly, this may be because expressing stance and conversational messages is more important for establishing an immediate

connection with (native or non-native) users of English (at least in the Skills for Life candidates' context) than highlighting discourse structure. Secondly, the genres that lower proficiency L2 learners can engage in within the limitations of their L2 knowledge might favour certain functional types of bundles (e.g. stance and conversational) over others (e.g. discourse bundles). This may prompt an earlier acquisition of stance and conversational bundles and a later acquisition of discourse bundles. Lower proficiency candidates can normally engage in genres which involve short texts with personal messages (e.g. postcards, notes and messages), according to the CEFR descriptors (Council of Europe 2004:26), which were empirically validated and scaled to proficiency levels 'by analysing the way in which they have been interpreted in the assessment of large numbers of learners' (ibid:25). Skills for Life Entry 1 and 2 candidates are required to engage in similar genres, as shown in Table 1. At higher proficiency levels (Levels 1 and 2), the scales tip in favour of discourse structuring bundles (as shown by our findings), which is the largest category after referential bundles at those proficiency levels. The more frequent use of discourse bundles by Level 1 and Level 2 learners is due to a better language knowledge and use, and also due to a higher degree of discourse and message complexity of tasks at higher proficiency levels which require better discourse organisation. Further research could indicate if the rare use of discourse bundles at lower proficiency levels may also be due to their lower degree of perceptual salience (due to their not being complete structural units), their structural or semantic complexity, or if it is due to L1 influence (e.g. such as a rare use of discourse structuring bundles in a candidate's L1).

Conclusions

Within applied linguistics, single-word vocabulary frequency measures have not been found robust enough for tracing L2 acquisition from a lexical perspective. The investigation of multi-word units has been called for, but it is still far from clear how they are acquired in English as an L2. The present study addressed these issues by exploring both individual words and lexical bundles in L2 writing in order to: a) get a fuller and clearer picture of lexical development during second language acquisition, and b) determine if Skills for Life Writing examinations clearly distinguish between five levels of language proficiency (A1–C1).

Individual words

Our findings based on a sample of L2 writing produced by 20 candidates per Skills for Life level revealed that the average length of sentences, the number of word tokens and types as well as lexical variation consistently and clearly differentiated between candidates at different proficiency levels. Words are marginally longer, on average, at higher proficiency levels. Lexical sophistication was found to be a useful measure for distinguishing between lower (Entry 1 to Entry 3) and higher proficiency levels (Levels 1 and 2), rather than between adjacent levels. All above lexical phenomena increased in number and proportion with increasing proficiency levels indicating that:

a) lexis produced by learners clearly becomes more complex with every proficiency level, and b) WordSmith Tools and Compleat Lexical Tutor are, to a certain extent, useful software for the investigation of lexical progression. Our findings also showed that lexical density cannot differentiate well between levels in the context of Skills for Life data. This was attributed to a special nature of lower proficiency discourse which is characterised by truncated grammar and tendencies to list pieces of information, both of which increase the number of content words in a sentence, thereby increasing lexical density figures.

Lexical bundles

This study also traced L2 acquisition of lexical bundles across five proficiency levels from a frequency-based and a functional perspective. It has, thus, started to fill the gap in this domain of L2 acquisition research.

According to our findings, frequency and diversity of lexical bundles are useful differentiators between language proficiency levels. While lexical bundles are rarely used by the lowest proficiency learners (Entry 1 (A1)) their number and variety increase in the discourse of Entry 2 and 3 learners (A2 and B1 on the CEFR, respectively). However, it is not until learners reach Level 1 (B2) and Level 2 (C1) that bundles become numerous and diverse. This could indicate that, as far as writing is concerned, L2 learners of English start out by learning and using individual words which they put together by means of syntactic or other rules, rather than lexicalised routines. Learning (conventionalised) word strings starts emerging after the lowest proficiency level but becomes truly productive only at later stages of L2 acquisition which correspond to Skills for Life Levels 1 (B2) and 2 (C1). Whether or not learners at the highest examined Skills for Life level behave like native speakers in terms of the frequency and appropriateness of lexical bundles is a matter for follow-up studies.

Functional investigation of lexical bundles showed that the evolution of written discourse across proficiency levels is accompanied by the changing functional roles of lexical bundles. While bundles with a referential function are predominant in the written responses of learners at all Skills for Life levels, bundles with a discourse function are frequently used only by higher proficiency learners (Levels 1 and 2). Stance and special conversational bundles exhibit a more gradual development so that lower proficiency learners (Entry 1, 2 and 3) rely on them far more than on discourse bundles. The reason behind the early acquisition of stance and special conversational bundles and the late acquisition of discourse bundles were attributed to different communicative needs of lower and higher proficiency learners.

Overall, our findings showed that frequency, diversity and the changing discourse roles of lexical bundles clearly differentiate between language proficiency levels as instantiated in five Skills for Life levels.

The answer to the final research question is the following: frequency-based lexical analyses both at the level of individual words and the level of lexical bundles help trace the pattern of L2 lexical development and a functional analysis of bundles gives further insight into the development of lexis and discourse organisation.

Implications

This study has implications not only for practitioners in language testing and teaching but also for lexical research and L2 acquisition research.

Our findings strengthened the validity argument for Skills for Life Writing tests by showing that they clearly differentiate, at the lexical level at least, between five language proficiency levels. This has implications for language testers and language test validation in general. For Cambridge ESOL, at a practical level, the results of this study may inform future development or validation of Skills for Life rating scales and examiner training, in particular for the Writing paper, but it may also be applicable to Speaking. This study can also inform teaching practices by raising awareness of lexical bundles and their importance in the classroom, showing possible ways of analysing students' work and helping focus preparatory work for general purpose language qualifications. As far as both lexical research and L2 acquisition research are concerned, our findings have shown which quantitative and qualitative analyses are useful for exploring lexis and L2 lexical development. Moreover, they also shed some light on L2 lexical development, in particular with regard to L2 acquisition of lexical bundles, by clearly tracing progression across five proficiency levels. More studies like this are needed since most research on the L2 acquisition of multi-word sequences focuses only on the advanced proficiency level in English or does not separate collocations from lexical bundles. As the two types of sequences typically differ in perceptual salience (collocations being noun/verb/adjective phrases and bundles being mostly incomplete structural units), they may also be acquired in different ways by L2 learners of English.

References

- Biber, D and Barbieri, F (2007) Lexical bundles in university spoken and written registers, *English for Specific Purposes* 26, 263–286.
- Biber, D, Johansson, S, Leech, G, Conrad, S and Finegan, E (1999) *Longman Grammar of Spoken and Written English*, London: Longman.
- Cobb, T, Web VocabProfile [retrieved from <http://www.lexutor.ca/vp/>].
- Corder, S P (1981) *Error Analysis and Interlanguage*, Oxford: Oxford University Press.
- Cortes, V (2004) Lexical bundles in published and student disciplinary writing: Examples from history and biology, *English for Specific Purposes* 23, 397–423.
- Council of Europe (2004) *Common European Framework of Reference for Languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.
- Coxhead, A (2000) A New Academic Word List, *TESOL Quarterly* 34 (2), 213–235.
- Granger, S (2001) Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae, in Cowie, A P (Ed.), *Phraseology: Theory, Analysis and Applications*, Oxford: Oxford University Press, 145–160.
- Hyland, K (2008) As can be seen: lexical bundles and disciplinary variation, *English for Specific Purposes* 27, 4–21.
- Lim, G and Galaczi, E (2010) Lexis in the assessment of Speaking and Writing: an illustration from Cambridge ESOL's General English tests, *Research Notes* 41, 14–19.

- Nekrasova, T (2009) English L1 and L2 Speakers' Knowledge of Lexical Bundles, *Language Learning* 59 (3), 647–686.
- Read, J and Nation, P (2006) An investigation of the lexical dimension of the IELTS Speaking Test, *IELTS Research Reports* 6, IELTS Australia/British Council, 207–231.
- Schmitt, N (2005) *Lexical Resources in Main Suite Examinations*, Cambridge ESOL consultant report.
- Schmitt, N, Dörnyei, Z, Adolphs, S and Durow, V (2004) Knowledge and acquisition of formulaic sequences, in Schmitt, N (Ed.) *Formulaic Sequences: Acquisition, Processing and Use*, Amsterdam/Philadelphia: John Benjamins, 55–71.
- Scott, M (1999) *WordSmith Tools version 3*, Oxford: Oxford University Press.
- University of Cambridge ESOL Examinations (2005) *Skills for Life Handbook*, Cambridge: University of Cambridge ESOL Examinations.
- West (1953) *A General Service List of English Words*, London: Longman.

Lexis in the assessment of Speaking and Writing: An illustration from Cambridge ESOL's General English tests

GAD S LIM RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

EVELINA D GALACZI RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

Introduction

Lexical knowledge is an indispensable aspect of language knowledge, as different models of language ability indicate (Bachman & Palmer 1996, Canale & Swain 1980). Accounting for lexis at different proficiency levels, however, poses challenges for language assessment in general and for performance assessments of the productive skills in particular. One basic and obvious challenge is determining which lexical items to target at different proficiency levels (Weir 2005). Over the years, a number of wordlists have been developed to provide guidance on the range of vocabulary at different proficiency levels (e.g. Coxhead 2000, Van Ek & Trim 1990a/1998a, 1990b/1998b, 2001). However, as Khalifa & Weir (2009) note, these wordlists have by and large been developed intuitively rather than empirically. These lists have also tended not to account for the functions to which these words are related, which is an important consideration within a communicative approach to language ability (Khalifa & Weir 2009). In addition, such lists do not account for how well words are used; while a minimum of vocabulary is certainly required, quality of lexical use (i.e. appropriacy) is as important as the variety of lexis used. In fact, it is appropriacy of lexical use which typically distinguishes people as higher-level language users.

In addition to determining which lexical items are associated with specific proficiency levels, language testers are then faced with the challenge of eliciting those lexical items. In contrast to discrete-item tests, which can explicitly focus on desired items, in performance assessments the most that a test developer can do is set up tasks that provide the candidate with the opportunity to produce the lexis at the targeted level. Candidates may indeed produce the expected lexis, but they may also very well not do so, or perhaps use lower-level lexis in responding to the task.

While these challenges associated with lexical knowledge and use are indeed real, it is also the case that test makers employ a range of test development procedures which ensure that lexis is appropriately accounted for in the

assessment process. In this article we discuss such aspects of lexical knowledge in performance assessment, and focus on four key issues: determining target lexis at particular levels; controlling lexis in Writing and Speaking tasks at different proficiency levels; features of candidate performance at different levels; designing rating scales which reflect what is being assessed. We illustrate each of these using examples from Cambridge ESOL exams at different proficiency levels.

Determining lexis at different levels

Language proficiency is naturally a cline, and the idea of levels is necessarily an artificial division, devised as a common yardstick to help us communicate about the subject. It follows that there can be no 'hard and fast' rules about which lexical items are associated with particular levels, and ambiguities will necessarily be encountered. For this reason, it is invaluable to have frameworks such as the Common European Framework of Reference for Languages (CEFR), because all involved – language learners, teachers, assessors, researchers – can have a common set of broad proficiency levels to refer to and direct their efforts toward. Over time, with more work done on each level (e.g. the English Profile project) and through 'user intervention' in local contexts, each level can become better defined (Milanovic 2009:3).

The CEFR has been helpful where English lexis is concerned because part of its history has included the development of wordlists for the Waystage, Threshold, and Vantage levels (Van Ek & Trim 1990a/1998a, 1990b/1998b, 2001), which correspond to the levels A2, B1, and B2. While these wordlists have been developed intuitively, as noted earlier, they have been developed with particular functions and domains in mind, thereby contributing to their context validity, and making them useful tools and reference points for tests at different proficiency levels.

Cambridge ESOL draws from the CEFR wordlists for its A2

to B2 level exams, and documents specific lexical domains associated with each of these levels in the Item Writer Guidelines which accompany each test. In addition, each exam's wordlists are supplemented with suggested lexical items by experts (e.g. examination paper Chairs) and vocabulary items that evidence from the Cambridge Learner Corpus show to be relevant to the level. As English is an evolving language, so these wordlists also change over time, with words being added and removed as necessary. Additions and subtractions are always made by a panel of experts after due consideration of all available evidence (Ball 2002).

At the higher C levels, Cambridge ESOL exams do not employ wordlists. At these levels, the vocabulary candidates are expected to know grows exponentially, making the use of wordlists impractical and meaningless. As previously noted, quality of lexical use rather than quantity, is a better discriminator of proficiency at these levels.

Lexis in performance test tasks

In addition to determining appropriate lexis at different proficiency levels, due consideration needs to be given to task design, to ensure that test input and candidate output will be at appropriate, desired levels.

First of all, the tasks themselves need to be designed so that the target language is elicited. In Writing, for example, A2 level learners are expected to be able to write short notes or simple personal letters. Thus, Writing tasks in the Key English Test (KET)/A2 reflect these expectations; there are tasks where candidates complete a postcard, fill in a simple order form, and write a 25–35 word note to a friend using common high-frequency vocabulary. Asking candidates to write a letter of reference, for example, would clearly not be appropriate at this level. On the other hand, a candidate at Certificate in Advanced English (CAE)/C1 level should not only be able to cope with that task, the task should also be appropriately challenging so that higher-level vocabulary will be elicited. Asking a C1 candidate to write a 30-word note is unlikely to allow the candidate enough opportunities to demonstrate C1 level lexis.

Examples can be given for Speaking as well, where tasks are set to elicit the desired level of lexis. At Preliminary English Test (PET)/B1 level learners are expected to be able to use appropriate lexis when talking about familiar personal topics, whereas at Certificate of Proficiency in English (CPE)/C2 level learners should be able to use exact vocabulary to convey finer shades of meaning, even when discussing abstract topics. And so, PET candidates can be asked to speak factually (e.g. about what kinds of books they read), whereas CPE candidates are asked to speak about more abstract issues or hypothetical situations (e.g. consumerism, counterfactual situations).

Assuming the task itself has been appropriately designed, care also needs to be taken that the instructions and input are at the right level. If candidates are not able to understand the instructions, they will not be able to perform the task. To ensure that input is at the right level, Cambridge ESOL item writers use the previously mentioned wordlists for guidance and follow detailed item-writer guidelines for each level. For example, at the PET/B1 level,

the guidelines provide a list of allowable prefixes and suffixes, and state that compound words are allowed if both constituent words appear in the list and the meaning of the compound is transparent. To give another example, at the First Certificate in English (FCE)/B2 level, the guidelines allow a maximum of two words or expressions beyond the level to be used, for reasons of authenticity, so long as these words are glossed. Such guidelines help to ensure that input is at the appropriate lexical level.

It is appropriate for input vocabulary to be somewhat easier than the level being tested. Not only does it make the task accessible, it also allows the assessor to see if the candidate can produce the vocabulary at the intended level. Should vocabulary at the level already be in the input, it becomes more difficult to infer whether the candidate could independently produce the vocabulary item, or if the candidate merely borrowed it from the input.

An important aspect of the test production at Cambridge ESOL is the pretesting/trialling stage (Saville 2003). In all cases, after materials are written for Speaking/Writing tests, they are trialled prior to their use in live testing situations. Trialling ensures that the lexis used in all input materials is appropriate and accessible to candidates.

Such aspects of test development at Cambridge ESOL, i.e. the use of Item Writer Guidelines which draw on available wordlists and expert judgement, as well the trialling of new materials, contribute to the appropriately targeted lexis at different proficiency levels. A series of empirical investigations, carried out by Schmitt (2005, 2009), also serve to provide evidence supporting the progression of lexis at appropriate difficulty across proficiency levels. Word difficulty is not easy to determine, so language specialists typically use a word's frequency of occurrence as an indirect measure; in general, the more frequently occurring a word is, the earlier it is learned, and therefore, the easier it is (Schmitt, Schmitt & Clapham 2001). Such an analysis of frequency level of the task prompts for Writing can be seen in Table 1. The table presents the occurrence of lexis in Main Suite task prompts (Cambridge ESOL's core General English exams) in the first 2,000 word level (i.e. the most frequently used words), and in the Academic Word List. A general trend can be observed, where the 2,000 most frequent words are used proportionally less as one goes up the CEFR levels. At the same time, there are generally more academic words used as one goes up the levels.

Table 1: Vocabulary coverage at different Writing examination levels

Frequency level	A2/KET (%)	B1/PET (%)	B2/FCE (%)	C1/CAE (%)	C2/CPE (%)
K1 words (1–1,000)	87.15	88.50	88.46	82.39	83.89
K2 words (1,001–2,000)	6.31	6.50	4.18	7.24	4.77
K1 + K2 words	93.46	95.00	92.64	89.63	88.66
Academic Word List	0.47	0.66	1.57	4.93	5.12

The same trend is observed with a frequency analysis of the lexis in the Speaking prompts, which can be seen in Table 2. In the Speaking test, part of the input is delivered in verbal, rather than written form, which makes it more challenging for candidates. Thus, there is a greater need to make sure that vocabulary used in the input is accessible.

Table 2: Vocabulary coverage at different Speaking examination levels

Frequency level	A2/KET (%)	B1/PET (%)	B2/FCE (%)	C1/CAE (%)	C2/CPE (%)
K1 words (1–1,000)	86.50	95.54	97.19	95.35	95.42
K2 words (1,001–2,000)	8.10	4.31	2.66	4.27	3.13
K1 + K2 words	94.60	99.85	99.85	99.62	98.55
Academic Word List	1.45	0.76	1.15	1.91	2.27

The research shows that for texts to be easily comprehensible, learners need to know at least 95–98% of the words in those texts (Hirsh & Nation 1992, Laufer 1992, Schmitt 2009). Assuming that examinees know the first 2,000 words, which should be a safe assumption for the FCE, CAE, and CPE levels, then these three exams clearly meet this threshold. This assumption may not hold for PET, but the coverage is so high (almost 100%) it is probably safe to assume that examinees will still probably understand a sufficient percentage of the instructions. The results for KET are more unusual. An examination of task materials shows that verbal instructions are all in high-frequency vocabulary, and that the lower-frequency words all appear in the written input, which candidates can scrutinise and use as a source of scaffolding, and which is also most often glossed. In terms of vocabulary from the Academic Word List, the percentage of academic vocabulary increases as the level progresses, with CAE and CPE showing the highest percentage, as expected. (KET's higher than expected frequency of words from the Academic Word List is, in fact, a result of the overlap between some words in the KET prompts (e.g. *TV channel*, *credit card*) and their academic counterparts (e.g. *to channel*, *to credit*.)

Additional evidence about the progression in lexis across the levels can be seen in Table 3, which provides information about the new words used in each subsequent level of the Writing exams that were not in the previous level. Again, it can be seen that at the lower levels, the new words introduced in PET were mostly high-frequency words. At the highest level, a larger proportion of the new words introduced in CPE were academic and other less frequent words.

Table 3: New words at successive levels of Writing examinations

Frequency level	Words in PET not in KET (%)	Words in FCE not in PET (%)	Words in CAE not in FCE (%)	Words in CPE not in CAE (%)
K1 words (1–1,000)	79.40	61.31	55.22	49.25
K2 words (1,001–2,000)	10.55	9.55	14.93	10.55
Academic Word List	1.01	7.54	14.93	17.59
Other Words	9.05	21.61	14.93	22.61

To sum up, the lexical analysis of task prompts in both Speaking and Writing Main Suite tests suggests that the vocabulary load of these tests is well controlled and appropriately targeted for different proficiency levels.

Lexis in candidate output

Just as lexis should be accounted for in task input, its role in candidate output is an equally important area of investigation. In this respect, Schmitt (2005, 2009) analysed

candidate performance in Main Suite exams using similar measures to the ones employed in the analysis of task prompts. The rating of learner output in terms of lexis is one of the most problematic areas in vocabulary studies, largely due to the limitations of quantitative analysis programmes which can only provide quantitative frequency measures, but not estimates of appropriateness of use. One of the clear findings in Schmitt's (2005, 2009) investigations was that the type/token variation in candidate output was the most consistent (and perhaps only) indicator of progression in level, as can be seen in Table 4, which presents findings based on Speaking examinee output.

Table 4: Frequency of types and tokens in spoken examinee output

Lexical measure	A2/KET	B1/PET	B2/FCE	C1/CAE	C2/CPE
Mean number of tokens per examinee	224	400	746	720	1,228
Mean number of types per examinee	63	83	96	127	219

The findings in the table show a general trend for improvement from KET to CPE for both types and tokens per test taker. In other words, as examinees develop in proficiency, they produce both more words (tokens), and display a wider range of vocabulary by using more different words (types). This finding is in line with Iwashita, Brown, McNamara & O'Hagan's (2008) investigation of features of spoken performance at different proficiency levels, in which the authors found that an increase in level was associated with an increase in the number of tokens and types. It also supports findings reported for Main Suite written examinee output (Schmitt 2005, Shaw & Weir 2007).

An analysis of lexical variation (tokens per type) also indicates some progression across levels, but on a more limited scale. Using Speaking candidate output as an illustration, we can see in Table 5 the lexical variation in Main Suite Speaking candidate output.

Table 5: Frequency of types and tokens in spoken candidate output

Lexical measure	KET	PET	FCE	CAE	CPE
Lexical variation (Tokens per type)	3.55	4.82	7.77	5.66	5.61

A gradual increase is observed at the lower levels, which plateaus at CAE. The FCE result – an exception to the general trend – is possibly influenced by the specific task demands and prompts, and calls for a future follow-up qualitative investigation. The same finding regarding lexical variation is observed with Writing candidate output (Schmitt 2005, Shaw & Weir 2007), where the lexical variation plateaus at FCE, indicating that test takers do use progressively more lexical variation through the lower end of the suite. This measure, however, was found not to separate lexical mastery at the higher levels.

Lexical density, another useful vocabulary measure, provides the percentage of content words in a text/spoken output. As Shaw & Weir (2007) observe, greater use of content words usually corresponds to a higher information load and more 'literate' written and spoken contributions.

The lexical density analysis, seen in Table 6 for spoken output, suggests a similar profile of learner output across levels. The counter-intuitively higher percentage for KET is a result of the very tightly controlled prompts at this level, and candidates having to use the content words given in the written prompts. A parallel study, focusing on lexical density in written candidate output, produced similar findings of a flat lexical-density profile as levels increase (Schmitt 2005).

Table 6: Lexical density in spoken candidate output

Lexical measure	KET (%)	PET (%)	FCE (%)	CAE (%)	CPE (%)
Lexical density (Content words/Total)	51	46	42	44	43

Few noticeable measurable differences across levels can also be seen in Table 7, which gives an indication of the lexical resources of candidate spoken output as compared to the first 2,000 most frequently used words, and also the use of academic vocabulary in candidate output. Several scholars (e.g. Laufer & Nation 1999) have suggested that as learners advance in proficiency, the frequency profile of their vocabulary output shifts, with less high-frequency vocabulary being used and more lower-frequency vocabulary being produced. However, the candidates in this study (ranging across different levels of proficiency) produce 'relatively static frequency profiles' (Schmitt 2009:6) as they move up the levels. This can be seen in the percentage of lexis at the 1,000+2,000 frequency bands, which ranges only from 97.05% to 97.75%. KET, which counter-intuitively has a lower percentage (93.08%), is an exception. As Schmitt (2009) notes, this finding is most likely influenced by the examinees using the lower-frequency vocabulary from the written input. If this were stripped, their frequency profile would be similar to other Main Suite figures. Very similar findings have also been observed in the analysis of written output, reported in Shaw & Weir (2007) and Schmitt (2005).

Table 7: Frequency analysis of candidate spoken output

Lexical measure	KET (%)	PET (%)	FCE (%)	CAE (%)	CPE (%)
First 1,000 words	88.40	93.85	94.18	93.75	93.27
Second 1,000 words	4.68	3.20	3.43	4.00	3.91
1,000 + 2,000	93.08	97.05	97.61	97.75	97.18
Academic Word List (AWL)	.88	1.24	1.28	1.80	1.41

To sum up, the analysis of lexis in written and oral candidate output has indicated that frequency of types and tokens are robust (but perhaps limited) quantitative measures which present a strong trend across proficiency levels. Other quantitative measures, such as lexical density, lexical variation and frequency level, could not distinguish sufficiently robustly the lexical improvement which is clear from a qualitative analysis. This illustrates the limitations of the currently available quantitative analysis techniques, which are not able to discern 'goodness/appropriacy of usage' as well as a skilled rater. Frequency of single words in itself, therefore, may not be the best measure of lexical resources. In addition to frequency of use, a fundamental

measure of improving lexical proficiency is *how well* the words are used, rather than if they are necessarily of lower frequency. For example, the limitations of current lexical software packages mean that they can only 'read' correctly formatted words. Thus errors such as **differents*, **childrens*, **musics*, **bristot* must be corrected or the software will count this as off-list (e.g. very low-frequency) vocabulary. This finding naturally raises questions about the ability of automated assessment systems to provide meaningful, adequate and complex ratings of lexical resources which go beyond the mechanical frequency counts of types and tokens. It is clear that judgements of 'wellness/correctness' of use are crucial, and yet the field is still struggling to find a way of measuring such appropriacy of use in any other way than manual judgement. Skilled raters are clearly required to provide meaningful assessments of the lexical resources of examinees at different levels. As Schmitt (2009:1) writes, 'the fact that Cambridge raters take lexical appropriacy into account in their markings is reassuring, as ... it gives a better indication of the quality of vocabulary output than any automated method currently available'.

Questions still remain, therefore, whether single-word units and frequency of occurrence are adequate criteria for distinguishing between proficiency levels. The discussion above has indicated the limitations of some quantitative measures and the balance between quantity of use and appropriateness of use. As Martinez (2009) notes, single-word frequency lists are of limited validity, since the most frequent words in English 'are merely tips of phraseological icebergs'. Expanding the analysis beyond the single-word to multi-word combinations holds promise for the meaningful analysis of lexis in candidate output, as illustrated by Vidaković & Barker (2010 in this issue) in their study of lexical progression in candidates' written responses from Cambridge ESOL's Skills for Life test suite (A1–C1). As Shaw & Weir (2007) argue, at more advanced levels, the influence of collocation, phraseology, idiom and register may well be more significant in distinguishing between levels, though automated (rather than manual) measures for confirming this are not readily available. Notable exceptions are the new work of Simpson-Vlach & Ellis (2010) and Martinez (2009) whose lists are based on multi-word formulaic sequences, and would in future allow for a systematic way to assess the lexical profile of a text.

Lexis in evaluation criteria

Tasks are set up to provide candidates with opportunities to demonstrate their skills and abilities, and the role of lexis in task prompts and candidate output has so far been discussed. These performances still require trained experts to evaluate them according to pre-determined criteria. The marking criteria provide the assessor with an avenue for ensuring that a candidate's lexis meets the requirements of that level. The marking criteria are also the place where quality (i.e. appropriacy, flexibility) of vocabulary use can be specified, keeping in mind its importance in distinguishing proficiency.

In the Cambridge ESOL Writing and Speaking papers, vocabulary is one of several criteria taken into account in

evaluating candidate performance. To illustrate, the descriptors for adequate vocabulary performance in Speaking at different levels is given in Table 8. As can be seen, the descriptors address various aspects of lexical use in candidate performance. Range is accounted for, and is specified according to the types of situations and topics candidates are expected to talk about. KET and PET candidates need to have a range of vocabulary to talk about everyday situations, FCE candidates need to be able to deal not just with everyday situations, but also a range of familiar topics, whereas CAE candidates need to have vocabulary for unfamiliar topics, and at CPE candidates need to have the appropriate lexis to deal with abstract topics as well. Interactivity is also accounted for. At the two lower levels, vocabulary use is mostly related to production, whereas at the higher levels a learner needs to be able to use appropriate vocabulary in interactive situations, requiring such abilities as turn-taking, linking contributions and responding spontaneously. Finally, it will be seen that appropriateness of use is a key factor in the assessment of vocabulary, and flexibility is also required at the highest level.

Table 8: Vocabulary descriptors of adequate speaking performance at different levels

Level	Descriptor
CPE	Uses a range of appropriate vocabulary with flexibility to give and exchange views on unfamiliar and abstract topics.
CAE	Uses a range of appropriate vocabulary to give and exchange views on familiar and unfamiliar topics.
FCE	Uses a range of appropriate vocabulary to give and exchange views on a range of familiar topics.
PET	Uses a range of appropriate vocabulary when talking about everyday situations.
KET	Uses appropriate vocabulary to talk about everyday situations.

The marking criteria for Cambridge ESOL Writing papers are currently under revision, but it is expected that vocabulary will be among the criteria, and will account for quality of use in multiple ways as well (e.g. range, accuracy, use).

Future directions

Work continues on investigating the vocabulary necessary at different CEFR levels. The English Profile Programme, for example, aims to provide much more detailed reference-level descriptors for English, thus further developing descriptions of the different levels. Using corpus data, that work will also eventually identify criterial features – including vocabulary – which distinguish one level from another (Salamoura & Saville 2009).

This article has confined its discussion of lexis mostly to individual vocabulary items. However, as was noted earlier, such items alone are perhaps insufficient for distinguishing learners at different levels. Lexis does not consist of single words only. The field of language learning is increasingly paying attention to multi-word lexical bundles and formulaic sequences (e.g. Schmitt 2004, Wray 2002), which

appear to be learned and encoded in the brain as chunks rather than as individual words. Attempts are being made to define these in a useful way and to develop multi-word lists based on them (e.g. Martinez 2009, Simpson-Vlach & Ellis in press). Related to this, work is also being done on constructions (Ellis 2003, Goldberg 2003), i.e. form-meaning pairings, which will at the least challenge our thinking about the way and the level of aggregation at which to think of language use. The future of lexical research is, in a word, exciting.

References and further reading

- Bachman, L and Palmer, A (1996) *Language testing in Practice*, Oxford: Oxford University Press.
- Ball, F (2002) Developing wordlists for BEC, *Research Notes* 8, 10–13.
- Canale, M and Swain, M (1980) Theoretical basis of communicative approaches to second language teaching and testing, *Applied Linguistics* 1(1), 1–47.
- Coxhead, A (2000) A new academic word list, *TESOL Quarterly* 34, 213–238.
- Ellis, N (2003) Constructions, chunking, and connectionism: The emergence of second language structure, in Doughty, C J & Long, M H (Eds) *Handbook of Second Language Acquisition*, Cornwall: Blackwell Publishing, 63–103.
- Goldberg, A (2003) Constructions: A new theoretical approach to language, *Trends in Cognitive Science* 7 (5), 219–224.
- Hirsh, D and Nation, P (1992) What vocabulary size is needed to read unsimplified texts for pleasure?, *Reading in a Foreign Language* 8 (2), 689–696.
- Iwashita, N, Brown, A, McNamara, T and O'Hagan, S (2008) Assessed levels of second language speaking proficiency: How distinct?, *Applied Linguistics* 29 (1), 24–49.
- Khalifa, H and Weir, C (2009) *Examining Reading: Research and practice in assessing second language reading*, Cambridge: Cambridge University Press.
- Laufer, B (1992) How much lexis is necessary for reading comprehension?, in Arnaud, P J & Béjoint H (Eds) *Vocabulary and applied linguistics*, London: Macmillan, 162–132.
- Laufer, B and Nation, P (1999) A vocabulary-size test of controlled productive ability, *Language Testing* 16 (1), 33–51.
- Martinez, R (2009) *Towards the inclusion of multiword items in vocabulary assessment*, poster presented at the Language Testing Forum, Luton, November 2009.
- Milanovic, M (2009) Cambridge ESOL and the CEFR, *Research Notes* 37, 2–5.
- Salamoura, A and Saville, N (2009) Criterial features across the CEFR levels: Evidence from the English Profile Programme, *Research Notes* 37, 34–40.
- Saville, N (2003) The process of test development and revision within UCLES EFL, in Weir, C & Milanovic, M (Eds) *Continuity and Innovation: revising the Cambridge Proficiency in English Examinations 1913–2002*, Cambridge: Cambridge University Press, 57–120.
- Schmitt, N (2004) *Formulaic Sequences: Acquisition, Processing and Use*, Amsterdam: John Benjamins.
- Schmitt, N (2005) *Lexical Resources in Main Suite Writing Examinations*, Cambridge ESOL consultant report.
- Schmitt, N (2009) *Lexical Analysis of Input Prompts and Examinee Output of Cambridge ESOL Main Suite Speaking tests*, Cambridge ESOL consultant report.
- Schmitt, N, Schmitt, D and Clapham, C (2001) Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test, *Language Testing* 18 (1), 55–88.

- Shaw, S and Weir, C (2007) *Examining Writing*, Cambridge: Cambridge University Press.
- Simpson-Vlach, R and Ellis, NC (in press) An academic formulas list: New methods in phraseology research, *Applied Linguistics*.
- Van Ek, J and Trim, J (1990a/1998a) *Threshold 1990*, Cambridge: Cambridge University Press.
- Van Ek, J and Trim, J (1990b/1998b) *Waystage 1990*, Cambridge: Cambridge University Press.

- Van Ek, J and Trim, J (2001) *Vantage*, Cambridge: Cambridge University Press.
- Vidaković, I and Barker, F (2010) Use of words and multi-word units in Skills for Life Writing examinations, *Research Notes 41*, 7–14.
- Weir, C (2005) *Language testing and validation: An evidence-based approach*, Basingstoke: Palgrave Macmillan.
- Wray, A (2002) *Formulaic language and the lexicon*, Cambridge: Cambridge University Press.

A mixed-method approach towards investigating lexical progression in Main Suite Reading test papers

HANAN KHALIFA RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

NORBERT SCHMITT SCHOOL OF ENGLISH STUDIES, UNIVERSITY OF NOTTINGHAM

Introduction

When test developers construct reading tests at different language proficiency levels, they consider the interplay between text difficulty and task demands. Parameters such as the cognitive processing required or evoked by the task, familiarity with and appropriacy of response method, content knowledge as well as text length, discourse mode and linguistic (functional, grammatical, lexical) resources contribute to this interplay. In this paper, we focus on one of these parameters – lexical resources.

A number of researchers have established the important contribution syntactical and lexical knowledge makes to reading comprehension. Researchers like Perera (1984), Urquhart (1984), Weir (1993), Alderson (1993), Nuttall (1996), Shiotsu & Weir (2007) suggest that structural, lexical and conceptual difficulty strongly influence the ease with which a text can be read. Looking at a suite of internationally recognised examinations, we investigated what lexical resources were necessary to engage with the reading passages in these tests at varying proficiency levels. The tests examined, taken by at least half a million candidates worldwide on an annual basis, are Cambridge ESOL's Key English Test (KET), Preliminary English Test (PET), First Certificate in English (FCE), Certificate in Advanced English (CAE), and Certificate of Proficiency in English (CPE).

Table 1 provides a description of what the candidates are expected to be able to do in the Reading paper of each level of the Main Suite examinations.

Methodology

A mixed-method approach was used to investigate lexical resources in the above-mentioned examinations. A group of 10 expert judges provided content analysis of the examinations. The judges were all experienced item writers, were familiar with the examinations, have an MA in Applied Linguistics and have taught English as a Foreign Language at one point in their career. The content analysis was based on the *Manual for Relating Language Examinations to the CEFR*

Table 1: Expectations of reading ability at each level of Main Suite

KET	KET candidates are expected to be able to understand the main message, and some detail, of a variety of short factual texts: for example, signs, notices, instructions, brochures, guides, personal correspondence and informative articles from newspapers and magazines. They should also have strategies for dealing with unfamiliar structures and vocabulary.
PET	PET candidates are expected to be able to understand public notices and signs; to read short texts of a factual nature and show understanding of the content; to demonstrate understanding of the structure of the language as it is used to express notions of relative time, space, possession, etc; to scan factual material for information in order to perform relevant tasks, disregarding redundant or irrelevant material; to read texts of an imaginative or emotional character and to appreciate the central sense of the text, the attitude of the writer to the material and the effect it is intended to have on the reader.
FCE	FCE candidates are expected to be able to deal with both short and long texts, from the following sources: newspaper and magazine articles; reports; fiction; advertisements; correspondence; messages; informational materials. They will be able to locate specific information and detail and recognise opinion and attitude. They will also understand the purpose, main idea, tone and gist of the text, and be able to recognise the structure of a text and follow its development.
CAE	CAE candidates are expected to be able to deal with both short and long texts, from the following sources: newspapers and magazines; journals; books (fiction and non-fiction); promotional and informational materials. They will be able to understand the detail, tone, purpose, main idea and implication of the texts, and the opinions and attitudes expressed in them. They will recognise text organisation features such as exemplification, comparison and reference, and they will also understand how texts are structured and be able to follow text development.
CPE	CPE candidates are expected to be able to understand in detail a range of texts, both short and long, from the following sources: books (fiction and non-fiction); non-specialist articles from journals, magazines and newspapers; informational materials. They will be able to recognise the purpose and main ideas and details of the texts and the opinions and attitudes expressed in them. They will also recognise the structure of certain texts and be able to follow their development.

(Council of Europe 2003 – pilot version) as well as reviewing available documentation and resources. For example, test specifications, guidelines for item writers, handbooks for teachers, and published wordlists were examined.

Another stage of the study involved examining current practices followed by Cambridge ESOL, e.g. the use of

Table 2: Expert judges' analysis of the lexical resources required at each level of Main Suite examination

Lexical resources					
	Overview of nature of lexis	More specific aspects of lexis	Lexical resources	Type of use based on	What this means in practice: analysis of six papers per level
KET	Restricted to common items which normally occur in the everyday vocabulary of native speakers.	Lexis appropriate to simple personal requirements, e.g. nationalities, hobbies, likes and dislikes.	Waystage 1990 (van Ek & Trim 1998a) and other high-frequency or otherwise appropriate words from corpus evidence.	Mainly literal use.	<ul style="list-style-type: none"> • set of notices • set of sentences on theme of university life • two turn dialogues • longer dialogue: about renting a room • article about child violinist • factual article about badgers • vocabulary definitions relating to places in town
PET	General vocabulary sufficient for most topics in everyday life.	Lexis appropriate to personal requirements, e.g. nationalities, hobbies, likes and dislikes.	Threshold 1990 (van Ek & Trim 1998b) and other high-frequency or otherwise appropriate words from corpus evidence.	Mainly literal use.	<ul style="list-style-type: none"> • set of notices, emails and memos • descriptions of people to match to descriptions of museums • informational text about Short Story Society • interview with a new young TV star • encyclopaedic article about grass
FCE	Good range of vocabulary. Topics are addressed in detail and with precision.	General lexis as appropriate to specified topics relating to everyday life.	Vantage (van Ek & Trim 2001) and other high-frequency or otherwise appropriate words from corpus evidence.	Literal + some inferential evaluative/synthesis/analytical use. Light fiction may be used.	<ul style="list-style-type: none"> • popular fiction text – first person account of vet's life • piece of journalism – article about woman downhill mountain bike racer • 4 texts about collectors of different items
CAE	Broad range of vocabulary including idiomatic expressions and colloquialisms as well as language relating to opinion, persuasion and ideas.	Candidates should be challenged by complexity of expression rather than arcane subject matter and specialist vocabulary.	Vocabulary appropriate to specific contexts demonstrating mastery of a particular domain. Lexical appropriacy determined by the professional judgement of item writing team supplemented by pretesting information.	Literal/inferential evaluative/synthesis/analytical use. Fiction may be used.	<ul style="list-style-type: none"> • 3 texts relating to aspects of scientific research (competition instructions; extract from novel; opinion article) • magazine article about honeymoon canoeing on Zambesi • article about how people are taught to have good TV presence • number of reviews of crime novels
CPE	Very wide range of vocabulary including idiomatic expressions and colloquialisms as well as language relating to opinion, persuasion and abstract ideas.	Candidates should be challenged by complexity of expression rather than arcane or specialist lexis but at this level they are expected to be able to cope with the full range of general topics that might interest an educated native speaker.	As for CAE but with a range and appropriateness of vocabulary which an educated native speaker might be expected to understand. Texts that can be understood in real world professional contexts by native speakers.	Literal/inferential evaluative/synthesis/analytical use. More complex fiction may be used.	<ul style="list-style-type: none"> • 3 articles on different themes as basis for lexical items – (dealing with architecture, shopping in Europe and cosmetic dentistry) • 4 articles on aspects of advertising and publicity (articles from textbook, newspaper/magazine articles and a novel) • review of jazz album • comparison of US and UK weather forecasting (extract from book)

corpora. The development of corpora and the application of corpus linguistic tools have made it easier to derive more empirically grounded wordlists for use in pedagogy and assessment contexts. These can be used to help validate and improve existing wordlists, as well as create new wordlists sometimes with a specific level/domain focus. Within Cambridge ESOL, corpus studies have been used to inform test revision projects (e.g. CPE, see Weir & Milanovic 2003), devise new test formats (Hargreaves 2000), and

create or revise test writer and candidate wordlists (see Ball 2002, Barker 2004).

A further stage of the study was the use of WordSmith software and Tom Cobb's Compleat Lexical Tutor (see www.lextutor.ca). The analysis was based on a set of six past papers per examination. The papers were equivalent in terms of measurement characteristics and came from the 2003–06 test administrations. Data files were obtained from the Cambridge ESOL item bank.

In examinations, such as KET and PET, where reading and writing are measured within the same test paper, Writing tasks were removed so that only Reading passages remained. Similarly, question letters, numbers and rubrics were removed keeping the reading passages and multiple-choice question options. The analysis focused on lexical variation, frequency bands, and lexical complexity.

Results and discussion

Content analysis

The expert judges' analysis of the lexical resources required by candidates to deal with the Reading papers in the examination suite is provided in Table 2.

On examining the above table, some key points emerge with regard to lexical development throughout the examinations. As candidates progress up the levels, the lexical demands they face generally increase. This is shown primarily through the number and complexity of the lexical items they are required to understand (see also Table 1). Another point is that as candidates advance, they are gradually expected to deal with increasingly subtle uses of the language of feelings and ideas. Fiction inevitably requires a broader receptive vocabulary and this is introduced from FCE onwards; more abstract texts are presented to candidates at CAE and CPE levels while lexis at the lower levels (KET and PET) is restricted to everyday, literal and factual language. A related point is that at the higher levels, candidates are required to handle a much larger number of texts in the exam than at lower levels covering a wider range of genres with increasing levels. A further point is that at KET, PET and FCE levels, there are documents which help support decisions as to the appropriacy of specific lexical items, mainly based on *Waystage*, *Threshold*, and *Vantage* lexical lists. The *Waystage* and *Threshold* lexical lists stem from a relatively constrained set of notions and functions, and as such provide a coherent guideline to work from. However, the lexical exponents at the *Vantage* level are much less principled, and are regarded as examples of appropriate lexis, rather than specifications as suggested in the *Vantage* document itself.¹

The exponents listed here are not presented as a defined lexical syllabus, nor even as 'recommended exponents'. They represent stimuli which maybe found useful by those involved in the development of theme-related ability to *Vantage*. ... In accordance with its intended role the list presented here is to a large extent open-ended. The majority of the lexical items contained in it are listed as members of open classes, to be reduced, expanded, or otherwise altered as may best suit the needs and interests of the learners (Van Ek & Trim, 2001:120).

Thus at CAE and CPE levels, the professional judgement of Cambridge ESOL item writers and test developers plays the main role in informing decisions about lexical suitability. This judgement is supported by the use of corpora and of pretesting. Cambridge ESOL in collaboration with Cambridge University Press has been building corpora since the early

1990s. The Cambridge Learner Corpus (part of the Cambridge International Corpus) includes over 30 million words of written learner English. This corpus together with the British National Corpus (BNC) which includes 100 million words of written and spoken native speaker data are used to validate KET/PET vocabulary lists. It is worth noting that wordlists derived from learner corpora relate to learner production, while wordlists derived from the BNC relate more directly to learner comprehension, such as that required by the Reading and Listening papers of Main Suite examinations. The KET/PET lists are updated on an annual basis by the addition and removal of words using a corpus-based approach, with suggested additions to the wordlists being collated and the frequency of these words being obtained by reference to the previously mentioned corpora (see Ball 2002, Barker 2004). There is also a project to create a learner production wordlist as part of the English Profile Programme (see Capel 2010 in this issue).

Lexical variation

Table 3 lists type-token information, as calculated by the Vocabulary Profiler (VP) English version 2.6 software available on the Compleat Lexical Tutor website. This software is a version of the venerable Vocabulary Profiler first developed by Paul Nation in the early 1990s.

Table 3: Type-token analysis of Main Suite Reading passages

Lexical characteristics	KET	PET	FCE	CAE	CPE
Type-token analysis					
Tokens (words in text)	1,310	3,962	17,332	21,895	19,601
Types (different words)	483	1,184	3,404	4,773	4,664
Type-token ratio	.37	.30	.20	.22	.24
Tokens per type	2.71	3.35	5.09	4.59	4.20
Lexical density (content words/total)	.51	.55	.50	.52	.52

The total number of words at the KET and PET levels is much lower than that at the other three levels. They are, therefore not comparable either with each other or with the three other levels. Although type-token ratios are influenced by token size, the number of tokens in the FCE/CAE/CPE levels is close enough to make comparison reasonable. With this in mind, a couple of points are worth noting:

- The ratio between types and tokens in FCE, CAE, and CPE is very similar. Across the Reading passages sampled, each type was repeated between 4.2 and 5.1 times. Thus, in terms of how many different words (types) candidates must understand in the Reading passages, there does not seem to be any progression through the upper end of the suite. Note that this applies to a number of reading passages combined, and in any single examination, the repetition per reading would be less.
- The number of lexical (content) words in relation to function (grammatical) words, appears to be constant, at about 50%. This mirrors the nature of language (a large percentage of function words are necessary to 'organise' language), and so it is not feasible to increase lexical difficulty by simply increasing the percentage of content words.

¹ It must be said that the lexical requirements have never been established for any of the levels of the CEFR. The wordlists given in the *Waystage*, *Threshold*, and *Vantage* books were not derived in an empirical manner, and the CEFR specifications give little or no concrete guidance about what vocabulary is necessary to reach each level.

Frequency analyses

The frequency of the words in Main Suite Reading passages were analysed using three different methods.

1,000/2,000/ Academic categories

The first method was with the classic Vocabulary Profiler (VP), which divides word frequency into four categories: 1st 1,000 words in English (K1), 2nd 1,000 words in English (K2), academic vocabulary according to the Academic Word List (Coxhead 2000)², and any remaining words not on any of the previous three lists (off-list). This VP version highlights high-frequency lexis, and so is useful in illustrating how the different levels of the examinations differ in their concentrations of basic lexis. Table 4 lists frequency information according to this analysis method.

Table 4: Classic Vocabulary Profiler analysis of Main Suite Reading passages

	KET (%)	PET (%)	FCE (%)	CAE (%)	CPE (%)
K1 words (1–1,000)	86.95	81.22	82.24	77.67	77.98
K2 words (1,001–2,000)	5.04	8.81	6.65	6.12	6.32
Academic Word List (AWL) words	.61	2.45	3.30	4.58	4.33
Off-list words	7.40	7.52	7.81	11.63	11.37

Table 4 demonstrates the following:

- The KET level is clearly the easiest (in terms of lexical requirements) in that it has a high percentage of first 1,000 vocabulary (≈87%), and a relatively low percentage of off-list words (essentially >2,000 frequency band). The KET Reading passages also have a very low percentage (<1%) of academic words (as defined by the Academic Word List).
- The PET level is probably the next easiest, and although it has a slightly lower percentage of 1,000 word vocabulary than the FCE level, it has a higher percentage of 2,000 words. It also has a slightly lower percentage of off-list words than FCE. On balance, the PET level is slightly easier than the FCE level.
- Both the CAE and CPE Reading tests have fewer high-frequency words and more off-list words than FCE, making them more difficult. However, the VP analysis shows little difference between the CAE and CPE levels.
- There is a fairly clear progression in the number of words from the AWL occurring through the examination suite, with the exception of CAE and CPE, which have similar percentages of AWL vocabulary. The percentages of AWL vocabulary in the CAE and CPE (≈4.5%) lie somewhere between what one might expect to find in general English texts and what one might expect to find in academic texts. For comparison, Coxhead (2000) reports that the words on the AWL made up about 10% of the tokens in her main academic corpus, 8.5% of a second academic corpus, and 1.4% of a corpus made up of fiction texts.

² The AWL is not primarily based on frequency, and so the academic category should not be seen as the follow-on frequency level from the first two.

- Overall, there is a reasonably clear progression through the first four levels of the suite in terms of high-frequency/off-list/academic vocabulary, but not between CAE and CPE.

Six-level BNL categories

The second frequency analysis was based on the experimental Bare Naked Lexis (BNL, Neufeld & Billuroğlu 2007) frequency categories, as calculated by the Compleat Lexical Tutor. The categories are based on a revised and expanded version of the General Service List (GSL). Background information on BNL is available at www.editthis.info/thebnl/Main_Page. Since the BNL is based on the GSL, it also highlights high-frequency vocabulary.

Table 5: BNL analysis of Main Suite Reading passages

Frequency analysis (BNL)	KET (%)	PET (%)	FCE (%)	CAE (%)	CPE (%)
BNL-0	49.08	46.10	50.75	48.42	48.68
BNL-1	33.85	33.36	29.01	27.26	27.17
BNL-2	5.00	7.36	6.10	5.24	5.56
BNL-3	3.00	3.77	3.70	4.22	3.69
BNL-4	.38	1.82	2.00	2.18	2.29
BNL-5	.85	.97	1.20	1.75	1.92
BNL-6	.46	.77	.53	.92	.90
Off-list	7.38	5.85	6.71	10.01	9.79

Given that the six levels of the BNL focus on the most frequent 2,700 word families of English, it is easiest to compare the percentages of vocabulary which appear in the off-list category, i.e. >2,700, and so are of relatively lower frequency. FCE has slightly more of this vocabulary than PET, and both CAE and CPE have more than FCE. Again CAE and CPE have similar amounts of this lower-frequency vocabulary.

KET Reading texts have a somewhat higher percentage of off-list words than either PET or FCE Reading passages. However, this is probably an artefact of having very short reading passages. Even these short Reading passages need to have contextualisation, utilising words and proper nouns such as *Africa*, *America*, and *Maria*. Since the Reading passages are short, these appear to occur at a relatively high rate. These kind of words also occur in higher level Reading passages, but the longer length of the Reading passages tends to lower the percentage in which they occur. As many of these words are place names (e.g. Scotland), which are likely to be already known, they do not necessarily add to the vocabulary burden. In short, the fact that the KET level has an apparently high degree of off-list words should probably not be considered problematic, or indicative of a lack of lexical progression.

Overall, if we disregard the KET level, the BNL analysis indicates a lexical progression through the suite, except for the CAE and CPE levels, which are similar. It is worth mentioning here that the BNL is new and experimental, and so standardised interpretations of the category results have not yet been developed, but this should not affect the above analysis, as it only looked at off-list words, which are of relatively low frequency.

BNC-20 frequency levels

The frequency analysis tool with the finest degree of gradation currently available is the BNC-20K software available on the Compleat Lexical Tutor website, and it is this method we used for our third frequency analysis. It gives the percentage of occurrence of texts in each of the 20 most frequent 1,000 bands. The criterion corpus is the BNC.

Table 6 lists the results of this analysis. Note that the different wordlists and word parsers underlying the VP, BNL, and BNC-20 programs lead to slightly different coverage percentages being reported (e.g. the VP figure for KET 1,000 is 86.95%; the BNC-20 figure is 89.31%).

Table 6: BNC-20 analysis of Main Suite Reading passages

Frequency levels (BNL)	KET (%)	PET (%)	FCE (%)	CAE (%)	CPE (%)
K1	89.30	84.73	84.17	78.67	78.95
K2	5.04	8.63	7.75	8.53	8.45
K3	.69	2.32	2.57	3.30	3.71
K4	1.22	.83	1.25	2.29	2.25
K5	.69	.43	.82	1.26	1.13
K6	.08	.08	.36	.85	.87
K7	.15	.05	.18	.67	.54
K8	0	.20	.28	.50	.45
K9	.08	.20	.09	.34	.36
K10	0	.10	.09	.32	.33
K11	0	.15	.05	.24	.31
K12	0	0	.08	.21	.22
K13	0	0	.07	.16	.21
K14	0	0	0	.11	.13
K15	0	0	.01	.04	.06
K16	0	0	0	.04	.04
K17	0	0	0	.01	.03
K18	0	0	.02	.03	.02
K19	0	0	0	.01	.05
K20	0	0	.01	0	.03
Off-list	2.75	2.27	2.19	2.42	1.88
Tokens per family (on-list)	3.54	4.66	8.42	7.45	6.65
Types per family (on-list)	1.28	1.37	1.59	1.56	1.53

Table 6 illustrates a number of points:

- At the K1 level (most frequent 1,000 word families in English), KET has the highest percentage, then PET and FCE with similar percentages, followed by CAE and CPE with similar percentages.
- At the K2 level, PET, CAE and CPE have similar percentages ($\approx 8.5\%$), with FCE and KET having lower percentages.
- At the K3 level, KET has dropped sharply down to .69%, PET and FCE have about 2.5%, and CAE and CPE have 3.3% and 3.7% respectively.
- In terms of overall frequency, KET clearly has the highest percentage of high-frequency vocabulary. Mirroring the results from the VP analysis, PET and FCE Reading passages have quite similar frequency distributions.
- CAE and CPE clearly have lower-frequency vocabulary than the FCE. However, the two levels have extremely similar distributions all the way down the frequency chart.

- The off-list percentages are similar among the five levels of the examinations. However, given that the off-list words indicate a >20,000 frequency band in this analysis, they are largely made up of proper nouns, and so there is no real difference between the levels in this respect.

Perhaps an easier way to appreciate the vocabulary loads is to consider a cumulative chart. Table 7 is the cumulative version of Table 6. When interpreting this table, it is useful to note that even small differences in percentage coverage (e.g. the difference between 95% and 96% coverage) can make a big difference in the ease of reading.

Table 7: Cumulative BNC-20 analysis of Main Suite Reading passages

	KET (%)	PET (%)	FCE (%)	CAE (%)	CPE (%)
K1	89.30	84.73	84.17	78.67	78.95
K2	94.34	93.36	91.92	87.20	87.40
K3	95.03	95.68	94.49	90.50	91.11
K4	96.25	96.51	95.74	92.79	93.36
K5	96.94	96.94	96.56	94.05	94.49
K6	97.02	97.02	96.92	94.90	95.36
K7	97.17	97.07	97.10	95.57	95.90
K8	97.17	97.27	97.38	96.07	96.35
K9	97.25	97.47	97.47	96.41	96.71
K10	97.25	97.57	97.56	96.73	97.04
K11	97.25	97.72	97.61	96.97	97.35
K12	97.25	97.72	97.69	97.18	97.57
K13	97.25	97.72	97.77	97.34	97.78
K14	97.25	97.72	97.77	97.45	97.91
K15	97.25	97.72	97.78	97.49	97.97
K16	97.25	97.72	97.78	97.53	98.01
K17	97.25	97.72	97.78	97.54	98.04
K18	97.25	97.72	97.80	97.57	98.06
K19	97.25	97.72	97.80	97.58	98.11
K20	97.25	97.72	97.81	97.58	98.14
Off-list	100	100	100	100	100

Table 7 shows that:

- KET has the highest percentage of K1 words, and since this is by far the best known band by learners, this indicates the relative lexical ease of the KET level. This advantage also sustains through the K2 level.
- If we include the K3 level, then PET and KET have similar percentages of coverage, and this does not change through the rest of the frequency bands. This means that candidates who know mainly words in the 0–2,000 frequency bands should find KET Reading passages easier than PET Reading passages, but if they know more vocabulary than this, they should find little difference in lexical difficulty between the two examination levels.
- A similar situation exists between PET and FCE levels, but here the threshold of equal coverage occurs at about the 6,000 frequency band.
- A comparison of the cumulative coverage figures between FCE and CAE levels shows that CAE readers must know words at the 11,000 frequency band or beyond to have a similar degree of coverage ($\approx 97\%$) as FCE readers would have with vocabulary at the 6,000 frequency band. Thus, CAE level seems clearly more difficult in terms of lexis than FCE.

- Once again, the analysis shows the close similarity of lexis between the CAE and CPE.
- It is interesting to note what frequency level of vocabulary is necessary to reach the 95% coverage level suggested by Laufer (1988). This is the percentage Laufer suggests as necessary for learners to understand the gist of a text and perhaps be able to infer the meaning of unknown words in the text. In KET, PET, and FCE levels, learners would need to know the words in the K1–K3 bands. For CAE and CPE, this goes up to include the K1–K6 bands which suggests that candidates will need knowledge of many more words to fully engage with the CAE and CPE texts.
- However, it is probably more useful to use a higher criterion (97%), which is closer to that suggested by the more current research (98%) carried out by Nation (2006), and supported by an in-depth study of the coverage-comprehension relationship carried out by Schmitt, Jiang & Grabe (2010). Using this higher coverage criterion, we find that KET and PET Reading passages would require knowledge of words at the K5 level, moving up to the K6 level for FCE Reading passages, and to the K10 level for CAE and CPE Reading passages. This suggests that for true ease in reading the passages (at least in lexical terms), candidates require a large vocabulary, even at the lower levels, but especially so at the higher levels.

Lexical complexity

The above analyses, based on lexical frequency and lexical variation, go some way in indicating the lexical load of the various examination levels. However, the limitations of such analysis methodologies are obvious. The crux of what makes vocabulary difficult for learners is its complexity, made up of a wide variety of factors, including but not limited to the following factors (see Laufer 1997, Schmitt 2010):

- the similarity or dissimilarity to a learner's L1
- the morphological/phonological complexity
- regularity of spelling
- the number of words in the L2 which have similar spellings to the target word
- amount of register marking
- amount of polysemy
- whether lexemes are individual words or multi-word units (note that the analyses contained here describe only individual word forms)

Frequency of occurrence can only be an indirect indication of this complexity. What is needed is a direct measure of this complexity, but unfortunately such a standardised measure does not currently exist. There are many facets to knowing a word (depth of knowledge), and it is not clear whether any single one can represent quality of word knowledge, or whether this requires a battery of tests to obtain a reliable measurement.

One of the elements of knowing a word is knowing the various members of a word's family (e.g. *crazy*, *craziness*, *craze*, *crazily*). Although Schmitt & Zimmerman (2002) found that learners usually did not know all of the related word family members for the individual words they knew,

in many cases the different word family members have very similar forms, and should be relatively transparent. Below are some of the word families at the 2,000 level from the easiest (KET) and most difficult (CPE) suite levels.

2,000 level word families with multiple members in KET Reading passages

animals animals animals animals
centuries century
mountain mountain
states states states states
swim swimming swimming swimming
teeth teeth
theatre theatre theatre theatre theatre
weather weather weather weather

2,000 level word families with multiple members in CPE Reading passages (beginning with 'A' only)

above above above
advance advancing
affair affairs
agenda agendas
aim aim aiming
alarming alarmingly
alter alternative alternative
among among among among amongst
analyse analysing analysis analysis
appealing appealing
arrived arrived arriving
aspect aspect aspects aspects
assessed assessment
assurance assured
attached attached attachments
attempt attempt attempt attempt attempts
awarded awards

In many cases the exact word forms are repeated, and this serves to lower the lexical load. In many other cases, different members of the word family are repeated, but an examination of the two lists reveals that most of these seem to be easily comprehensible if a learner knows one of the word family members. For example, if one knows *assessment*, then *assessed* is likely to be relatively transparent; the same is true of *alarming* and *alarmingly*. Following this reasoning, having more members per word family should lighten the vocabulary load. At the bottom of Table 6 above, the word family statistics are reported. The number of tokens per word family is quite low for KET and PET, but this is probably due to the relatively low number of words in the passages in general. The FCE, CAE, and CPE figures are more comparable, and we see that the number of tokens per family decreases as the level increases. That is, there is less repetition of word family members which are related to each other. We also see that number of different types per family is stable at the higher suite levels, so the vocabulary load of recognising different word family members stays about the same through the higher levels.

Conclusion

The type of analyses undertaken here can help identify the lexical load of the reading passages at the various suite levels. Frequency analysis shows progression across the

levels with the exception of CAE and CPE. CAE and CPE have shown similar distributions in the frequency charts. It is worth pointing out here that the similarities existing between CAE and CPE are not surprising since the selection of lexical items at the CAE and CPE levels are largely based on the judgements of experienced item writers and test developers. However, research into intuitions of frequency has generally shown that it is very difficult to make fine-grained distinctions of frequency at the low-frequency levels (cf. Schmitt & Dunham 1999). The professional judgements are supported by reference to corpus-based frequency information, but at the lower-frequency levels, this information can be disproportionately influenced by the topic and texts included in a particular corpus. Taken together, these factors make it relatively more difficult to obtain frequency figures which are robust enough to differentiate the highest levels of proficiency, as the CAE and CPE aim to do.

While it is difficult to specify *which* words are necessary for any particular language use context, vocabulary research has been more successful at specifying what size of vocabulary is necessary to achieve certain language aims. Around 2,000–3,000 word families should supply the bulk of the lexical resources required for basic everyday conversation (Adolphs & Schmitt 2003). About 3,000 word families is the threshold which should allow learners to begin to read authentic texts, probably with teacher support. Based partly on Laufer's (1988) research, it was formerly thought that knowledge of around 5,000 word families would provide enough vocabulary to enable learners to read a wide variety of authentic texts without lexical problems. However this was based on 95% coverage of texts, but now the consensus is moving toward a view that closer to 98% coverage is necessary for ease of reading which would require a larger vocabulary: something in the area of 8,000–9,000 word families (Nation 2006; Schmitt, Jiang & Grabe 2010). Of course many words will still be unknown, but this level of knowledge should allow learners to infer the meaning of many of the novel words from context, and to understand most of the communicative content of the text. Beyond this, for a wide L2 English vocabulary, a size of 10,000 word families is the figure most often cited (Hazenberg & Hulstijn 1996). It is important to note that these sizes are approximations, and the ability to accomplish the things in English also depends on many other factors, including speaking and reading skills, background knowledge, and strategy use. However they do provide 'rules of thumb' which may prove useful for test developers to keep in mind (see Schmitt 2008 and 2010 for more detailed discussions of vocabulary requirements).

To conclude, word frequency seems to be the best criteria readily available at the moment, but this can only be a general guide. Hopefully further research into the depth of vocabulary knowledge will suggest the means to grade vocabulary in a more contextualised manner (e.g. appropriacy of use), but this remains in the future.

References

- Adolphs, S and Schmitt, N (2003) Lexical coverage of spoken discourse, *Applied Linguistics* 24 (4), 425–438.
- Alderson, J C (1993) The relationship between grammar and reading in an English for academic purposes test battery, in Douglas, D and Chappelle, C (Eds), *A New decade of Language Testing Research*, Washington: TESOL, 203–214.
- Ball, F (2002) Developing wordlists for BEC, *Research Notes* 8, 10–13.
- Barker, F (2004) Using Corpora in Language Testing: Research and validation of language tests, *Modern English Teacher* 13 (2), 63–7.
- Capel, A (2010) Insights and issues arising from the English Profile Wordlists project, *Research Notes* 41, 2–7.
- Council of Europe (2003) *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEF), Manual: Preliminary Pilot Version*, DGIV/EDU/LANG 2003, 5, Strasbourg: Language Policy Division.
- Coxhead, A (2000) A New Academic Word List, *TESOL Quarterly* 34 (2), 213–238.
- Hargreaves, P (2000) How Important is Collocation in Testing the Learner's Language Proficiency?, in Lewis, M (Ed.) *Teaching Collocation – further developments in the Lexical Approach*, Hove: Language Teaching Publications.
- Hazenberg, S and Hulstijn, J H (1996) Defining a minimal receptive second language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics* 17 (2), 145–163.
- Laufer, B (1988) What percentage of text-lexis is essential for comprehension?, in Lauren, C & Nordmann, M (Eds), *Special language: From humans to machines*. Clevedon: Multilingual Matters, 316–323.
- Laufer, B (1997) What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words, in Schmitt, N (Ed.) *Vocabulary: Description, Acquisition, and Pedagogy*, Cambridge: Cambridge University Press.
- Nation, ISP (2001) *Learning Vocabulary in Another Language*, Cambridge: Cambridge University Press.
- Nation, ISP (2006) How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63 (1), 59–82.
- Neufeld, S and Billuroğlu, A (April 2007) *Bare Naked Lexis*, presentation given at the AAAL Conference, Costa Mesa, California.
- Nuttall, C (1996) *Teaching Reading Skills in a Foreign Language*, London: Heinemann.
- Perera, K (1984) *Children's Writing and Recording*, Oxford: Basil Blackwell.
- Schmitt, N (2008) Instructed second language vocabulary learning, *Language Teaching Research* 12 (3), 329–363.
- Schmitt, N (2010) *Researching Vocabulary: A Vocabulary Research Manual*, Basingstoke: Palgrave Macmillan.
- Schmitt, N and Dunham, B (1999) Exploring native and non-native intuitions of word frequency. *Second Language Research* 15 (4), 389–411.
- Schmitt, N Jiang, X, and Grabe, W (in press, scheduled 2010) The percentage of words known in a text and reading comprehension, *Modern Language Journal*.
- Schmitt, N and Zimmerman, CB (2002) Derivative word forms: What do learners know? *TESOL Quarterly* 36 (2), 145–171.
- Shiotsu, T and Weir, C J (2007) The Relative Significance of Syntactic Knowledge and Vocabulary Breadth in the Prediction of Second Language Reading Comprehension Test Performance, *Language Testing* 24 (1), 99–128.
- Urquhart, A H (1984) The Effect of Rhetorical Ordering on Readability, in Alderson, J C and Urquhart, A H (Eds) *Reading in a Foreign Language*, London: Longman, 160–75.
- Van Ek, J and Trim, J (2001) *Vantage*, Cambridge: Cambridge University Press.
- Weir, C J (1993) *Understanding and Developing Language Tests*, New York: Prentice Hall.
- Weir, C J and Milanovic, M (Eds) (2003) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*, Studies in Language Testing 15, Cambridge: UCLES/Cambridge University Press.

A corpus-led exploration of lexical verb use in Main Suite Writing papers

STUART PROUDFOOT ASSESSMENT AND OPERATIONS GROUP, CAMBRIDGE ESOL

Introduction

This exploratory study, based on an assignment carried out as a requirement of the author's MA in Applied Linguistics and TESOL at the University of Leicester, seeks to analyse the use of lexical verbs and the problems this area causes for learners of English. Our analysis of corpus data in this complex and wide-ranging lexico-grammatical area will illustrate non-native-like verb use, while mother tongue influence and idiosyncracies of English will explain why it occurs. It is then proposed that test developers, teachers and ultimately learners can benefit from harnessing such findings.

Area of investigation

The following study will be restricted to an analysis of lexical (full/main) verb use, though occasional overlap with auxiliary verbs, modal verbs, tense, aspect and voice may be unavoidable. In addition, the findings below are confined in scope to the following areas:

- **Genre:** the written scripts of General English examination candidates.
- **Source:** the Cambridge Learner Corpus (CLC)¹ is a collection of candidates' written examination scripts taken from a variety of Cambridge ESOL examination suites from 1993 onwards. The following study will be informed by a subset of the CLC, the set of five General English examinations also known as the 'Main Suite', ascending in proficiency level: Key English Test (KET), Preliminary English Test (PET), First Certificate in English (FCE), Certificate in Advanced English (CAE) and Certificate of Proficiency in English (CPE). At the time of writing, this subset contained over 20 million words.
- **Level:** KET, the lowest-level examination in our subset, is set at A2 level on the Council of Europe's Common European Framework of Reference, putting it above the level of beginners and the weaker elementary learner. The five Main Suite levels are kept together in the following analysis as they are a well-defined group of examinations of general language proficiency. Examples given in the following analysis are taken from a variety of levels, as it is candidates' mother tongues, and not their level, which are of primary interest in this study.
- **Mother tongue (L1):** while no L1 is omitted from the corpus, L1 representation in our subset and hence our analysis reflects the candidature in Main Suite

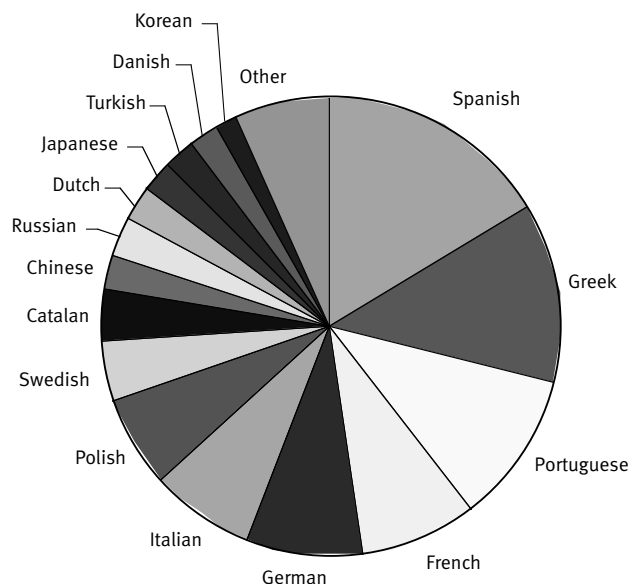


Figure 1: Proportion of Words per L1 (total number of words in sub-corpus: 20,705,087)

examinations. The number of words per L1 and proportion of the total corpus (see Figure 1) is heavily weighted in favour of European language groups, though it must be noted that Latin American learners account for a large percentage of the Spanish- and Portuguese-speaking candidature.

Hypotheses and methodology

We proceed from the assumption that differences between L1 and L2 are a source of difficulty for learners of English, and that this leads to non-L2-like use of lexical verbs. We also expect to find examples of L1 influence in the writing of learners at high as well as low levels of L2 proficiency, on the basis that learners' L1 backgrounds are likely to have a greater effect on their production of lexical verbs in L2 in the more stressful and time-limited examination context.

Since this study is exploratory in nature, it does not rely on frequencies of errors at each level, but aims to identify and explain the sources of non-native-like verb use by learners from a variety of L1 backgrounds.

The following facets of lexical verb use will be analysed in turn:

1. Transitivity – whether verbs take direct or indirect objects.
2. Prepositions and verbs – e.g. dependent prepositions and stranding.
3. Phrasal verbs – verbal constructions comprised of verb+preposition/adverb.

¹ The CLC (www.cambridge.org/gb/elt/catalogue/subject/custom/item3646603/Cambridge-International-Corpus-Cambridge-Learner-Corpus/?site_locale=en_GB) is updated biannually and forms part of a larger collection of corpora called the Cambridge International Corpus, which also contains spoken data and native speaker data.

4. Catenative verbs – verbs linking with a following dependent verb.
5. Infinitives – the base, or dictionary form, of verbs.
6. Distribution – allowable uses and nuances of verbs in a particular language.
7. Reflexive verbs (e.g. *wash oneself*) and psych verbs (e.g. *miss someone*).

The catalogue of scripts within the electronic corpus can be searched for all instances of a word or phrase, each presented in the context of the surrounding text. The lists produced are called concordances, and those run on our subset provide the main basis for further investigation. Where particularly relevant, concordances are restricted to a particular language group, and individual scripts consulted in more depth. To preserve authenticity, the wording of scripts is not altered in any way, which means that typographical errors are shown alongside those of grammar. However, for the purposes of illustration, the use of bold face is employed to highlight certain features.

Results

Transitivity

A feature of English verbs which can challenge learners is transitivity. Some verbs take an object (transitive verbs), two objects (ditransitive verbs) or do not take an object at all (intransitive verbs). According to Huddleston (1984:191), ‘a great many verbs ... occur in both transitive and intransitive clauses – certainly a majority of those that can take an object can also occur without’, a situation he describes as ‘something of a peculiarity of English, for languages generally make a sharper division within the lexicon between transitive and intransitive verbs’. One such ergative verb (taking both patterns) is *marry*:

They **married** last summer.
He **married her** within months.

The first use of *marry* is intransitive (though one could infer the unstated object *each other*); the second is transitive. Chang (2001:316) states that, frequently, verbs which would be used transitively in English are used as intransitive verbs in Chinese, and vice versa. This is also true of other languages (e.g. Spanish and Portuguese), which results in L1 influence on L2 performance, as illustrated by examples from the CLC below:

Spanish PET I went at the University and he **married with** his girlfriend.
Portuguese PET My wife have bought my clothes since I **married with** her.

The above L2 structures consist of a verb followed by a prepositional phrase, rather than a direct object. In essence *marry* is being used intransitively, with the above examples analogous to *I played with my friends* (intransitive) as opposed to *I played football* (transitive).

The Spanish example could be rendered in L1 as follows, with the preposition *con* separating the verb and its object:

Él se casó **con** su novia.
[He married **with** his girlfriend.]

Ditransitive verbs such as *give*, *owe* and *teach* take two objects, a direct object (O^d) and an indirect object (Oⁱ), indicated here:

He gave a promotion **to** her.
(O^d) (Oⁱ)

Problems arise when word order is reversed, as in Standard English the indirect object no longer takes a preposition:

He gave **to** her a promotion.
(Oⁱ) (O^d)

Some candidates have not mastered this L2 rule, as shown in the examples below:

French	FCE	... he was glad with her work and for that he gave to her a promotion ...
Tamil	KET	... it was very great party and I gave to her a small mobile ...
Portuguese	FCE	Last Christmas, your mother gave to her a watch ...

As our selection of examples shows, ditransitivity is potentially problematic even for learners at higher proficiency levels (e.g. FCE (B2) learners).

Prepositions and verbs

We will now consider the use of verbs in tandem with prepositions which either follow or precede a verb. As Partington (1998:80) acknowledges, ‘one major source of error in non-native language is the area of verb plus prepositional colligation (defined as the collocation of a lexical and a grammatical item)’.

Dependent prepositions are a likely source of error when a different preposition is used in L1. The verb *depend* is used with *of* rather than *on* in many European languages, including Portuguese, a speaker of which produces the following in their FCE script:

But depends **of** where you live.
[Mas depende **de** onde você vive.] (Portuguese)

Where verbs such as *like* take a dependent preposition in an L1, this is often produced in L2 at both low and high levels of proficiency, as these Portuguese speakers’ scripts show:

KET I love going out to restaurants, I like **of** all kinds ...
CAE ... and even make you like **of** natural sciences.

The second example could be rendered in Portuguese in the following way, with the dependent preposition *de* following the verb *gostar*:

... e até te convenço a gostar **de** ciências naturais.

The usage of verbs following prepositions in English is clear cut, according to Petrovitz (2001:173), who states that infinitives ‘cannot serve as the objects of prepositions, while gerunds are commonly found in these syntactic environments’. This does not preclude student error, particularly for speakers of Romance languages, as shown in examples below of sentence-initial *Before to*:

French	FCE	Before to write a next article ...
French	CPE	Before to leave France for England ...
Italian	CAE	Before to use these equipment ...
Spanish	PET	Before to start I was afraid ...

The same pattern of non-native-like use among Romance L1 speakers has also been attested for the preposition *after* in our corpus data.

Andersson & Trudgill (1992:113) note that it is acceptable ‘in English and related languages such as Swedish to end sentences with a preposition’ (e.g. *What did you talk about?*). This phenomenon, known as preposition stranding, is ungrammatical in Latin-derived and Slavic languages. It can, therefore, lead to written (and more markedly, spoken) language which may appear over-formal or stilted to the native speaker, as in this Polish FCE student’s informal letter:

You were asking me **at what time** will I arrive so my answer is flight from Warsaw will arrive on 11 o clock at your airport ... What about my apperance? Well now I’ve got long blond hair, mustage and contact lenses with blue colour. There is no doubt that you will recognise me ... What clothes should I wear there? You must answer me thus quickly! What about presents from Poland ... See you soon. Love,

Though register is a somewhat grey area, it could be argued that a native speaker would be less likely to use the preposition *at* in this way in an informal letter to a friend. Where it is used, it would more likely be placed after the verb *arrive*, stranded from the rest of the adverbial phrase.

The above cases have shown how the use of lexical verbs in association with prepositions can challenge even relatively proficient learners.

Phrasal verbs

Space limitations prevent an exhaustive analysis of the problems caused to learners by phrasal verbs. As lexical verbs, however, they fall within the scope of this study, so some of the basic points are outlined below.

Side (1990:145) notes that ‘since teachers and/or course books usually give definitions of phrasal verbs, students will stick to and use the Latinate definition rather than the Anglo-Saxon phrasal verb ... the Latinate word is easier to learn, particularly if it is related to a word in the students’ own language, and seems to make more sense.’ This can lead to the overuse of more formal Latinate verbs, and a lack of confidence in the use of phrasal verbs. For example, despite sprinkling his response liberally (and effectively) with phrasal verbs (e.g. *went out*, *turned out*), this Polish FCE student used a Latinate verb and felt the need to hedge their bets:

All the lights went out ... The firemen soon arrived at the fire and **extinguished it (put it down)**. It turned out that only kitchen where the fire had started was damaged.

Gołębiowska (2001:176) reveals that ‘there are no phrasal verbs in Polish, so a Polish learner is more likely to say *return* than *give back*, *write* than *write down*’. This is the case for most of the L1s represented in our analysis. It should be acknowledged, however, that the avoidance of phrasal verbs has far more impact on learners’ spoken English than it does in writing, where they are less appropriate in text types such as reports or letters of application.

Turning to accuracy, particle placement is another source of error in the domain of phrasal verbs. Where transitive 2-word phrasal verbs are separable, and where a

Table 1: Positioning of phrasal verbs and objects

Separated with noun	Separated with pronoun	Unseparated with noun	Unseparated with pronoun
turn the TV off	turn it off	turn off the TV	*turn off it
drop my sister off	drop her off	drop off my sister	*drop off her

pronominal object is used, the phrasal verb must be separated, rendering the combinations marked with an asterisk (*) non-native-like (cf. table above).

The CLC provides evidence of candidates struggling with pronominal objects in combination with phrasal verbs:

Greek FCE Then I lestened the telephone rang and I ran to **pick up it**.
 German PET I would **pick up it** on Friday.
 Czech KET May I **pick up it** at 6 p.m.
 Portuguese KET I’ll go **pick up it**.

Phrasal verbs are mostly non-productive, to the detriment of these two Dutch FCE students’ scripts:

The teacher said that I had to **put off** my blouse.
 But when I arrived home and when I **put off** my clothes I realized ...

The generation of this incorrect form is an intelligent and logical application of the knowledge within these students’ intergrammars that clothes can be *put on*.

Even though they are mostly non-productive, Side (1990:146) notes that ‘new phrasal verbs are constantly being invented, especially in slang’. Maxwell (2006) also points to recent media-led examples such as *to sex up* (a report), which she defines as ‘to change something in order to make it seem more interesting, exciting, or important’. Creative uses of the particle component can generate phrasal verbs with opposite meanings, as this 2004 BBC website news story extract shows: ‘The government has been accused of “sexing down” a draft report on alcohol misuse to prevent the study damaging the case for extending pub opening hours’. Such creativity complicates the learner’s task even further. This, along with restrictions in terms of pronominal object placement, and availability of Latinate verbs, makes phrasal verbs a troublesome area for learners.

Catenative verbs

These are main verbs that take non-finite (non-tensed) clauses as their complement, but in different, often fixed patterns. These rules contribute to what Salem (2007:213) calls ‘word-sensitive’ errors such as *enjoy to speak*, arising from ‘violation of a word-intrinsic restriction of “enjoy” not to be followed by an infinitive’. A parallel example in our sub-corpus is this instance of non-native-like use of *avoid* by an Italian CPE candidate:

everybody should avoid **to** repeat the same mistakes
 [ognuno dovrebbe evitare **di** ripetere gli stessi errori] (Italian)

This student may be using the full infinitive as it mirrors L1 with a preposition in place of *to* and then the base (dictionary) form of the subsequent verb.

Shepherd (2001:121) asserts that there is ‘no equivalent in Portuguese for the substantival use of the *-ing* form (gerund)’, which is true of many languages, leading to

overuse of the infinitive. An additional complicating factor for L2 learners is that some verbs (e.g. *stop*) take both infinitive and gerund complements, but in different contexts. The following are concordances of *stop* from the written responses of two Portuguese-speaking FCE candidates:

In the way, he wanted to **stop to drink** something and we took some beers ...
you should stop to smoke, and **stop to drink** too much, because I know ...

The first example is standard, *to stop* (a journey in order) *to drink something*, whereas the second example is non-native-like, where the intention requires the gerund, *to stop* (the activity of) *drinking*. The nuance in meaning for some verbs is extremely subtle even for high-proficiency students, as this Polish CAE candidate's script shows:

The school, however, promised individual tutors on demand. I **tried finding** them several times to solve my language problems but I didn't manage.

Context here makes it clear that the tutors were not found, but this should be indicated by the infinitive, i.e. *tried to find*, as the gerund indicates that they were found.

Causative constructions such as *have+object+past participle* feature what Huddleston (1984:215) calls an 'intervening' noun phrase between *have* and the past participle. Non-use of this complex formation leads to the unlikely scenario that only two of these FCE candidates have *had their hair cut* by a professional:

French FCE And I **cut my hair**.
Italian FCE I haven't really changed, I only have **cut my hair** a little bit ...
Catalan FCE Now I'm taller than before and I had **cut my hair** ...

Catenative verbs are, evidently, a source of error which can lead to semantic ambiguity or even serious miscommunication in L2.

Infinitives

The use of infinitives also seems to be an area of difficulty for L2 learners of English (of both high and low proficiency). In many languages, an equivalent of 'for' is used with verbs to express purpose. Duguid (2001:82) explains that the 'Italian infinitive of purpose structure uses *per (=for)*', as in: '**She went out for to buy .../for buy .../for buying ...*'. This feature is one which commonly leads to transfer, judging by the scripts of candidates from 18 L1 backgrounds who produced examples of *for+to+verb* in contexts clearly denoting purpose. Turkish is one of these:

CAE Did I study economics **for to look** after children.
FCE We also can make private organisation **for to help** to people.
PET She came here **for to learn** English.

Papaefthymiou-Lytra (2001:137) asserts that Greek learners 'may replace the infinitive after *to* by a past tense form (an overgeneralisation or transfer from Greek)', and a Greek CPE student from our sub-corpus bears this out:

Immediately she telephoned her husband who named was John **to told him** the good news.

Languages such as Portuguese use a bare infinitive and not a full infinitive after adjectives. This transfers into use of English, as in the case of this CAE candidate, a speaker of

Brazilian Portuguese:

Is **difficult** here in Brazil **see** a woman who earns more than a man in the same job. Is **easy see** a woman working as a secretary, a house keeper ...

It should be acknowledged, however, that the communication of meaning is unlikely to be seriously impeded by errors in the use of infinitives such as those examined above.

Distribution

Where L1 verbs have more than one equivalent in English, with different shades of meaning or different distributions, this can lead to L2 errors. Efstathiadis and King (1972:166) provide this example from Greek:

deny
a'rume <
refuse

Greek candidates' CAE and CPE scripts from our corpus illustrate the pitfalls of such one-to-many distributions:

She **denied** the tissue another passenger gave her.

The Greek Prime Minister **denied** their demand to walk through the greek territory.

At first I **denied** to take part but something inside me was ...

A brief study of concordances of the verb *win* reveals further transfer problems:

French CPE He didn't accept that she **wins** more money than him.
Greek FCE Liverpool, which is my favourite team, **won** Manchester United!

In the CPE case, the learner may be assuming that *win* mirrors the French verb *gagner*, which collocates with income as well as games or competitions. In the FCE example, the learner has not accounted for the fact that, though transitive, *win* cannot take the defeated party as its direct object; this is the role of the verb *beat*. Conversely, one could not *beat* a competition, only *win* one.

The classic Fitikides study aid (2000:89) lists numerous similar examples for which we can find evidence in our corpus, such as *remember/remind*, *borrow/lend*, *steal/rob*, and *accept/agree*:

Portuguese KET ... the most precious things I own because **remember** me of my favourite grandmother.
Czech PET So can you **borrow** me the bicycle please?
Portuguese FCE ... get people's attention while the others **stole** the bank.
Danish FCE ... I had **accepted** to take care of their dog ...

The examples above highlight the problem that asymmetric lexical distributions pose for learners at both lower (KET/PET) and higher (FCE/CAE/CPE) proficiency levels.

Reflexive verbs and psych verbs

Swan (1995:485) states that a 'common use of reflexive pronouns is to talk about actions where the subject and object are the same person', e.g. *I cut myself*. Some languages utilise reflexive pronouns or first-person object pronouns for many more verbs than does English, and this transfers into L2 as this concordance shows:

Czech PET My dear teacher, I **apologise myself** ...
German PET ... first of all I want to **apologise myself** for missing ...

Portuguese FCE I **apologised myself** ...

Catalan FCE ... and who was wrong was me, so I had to **apologize myself** ...

Where English speakers say *I apologise*, French has *Je m'excuse* and German *Es tut mir leid*, among other expressions.

This can lead to possible confusion, as in this Italian learner's CPE text:

Liza went downstairs and ... found Bob waiting for her. They didn't say anything but **kissed themselves**. The day after they left for Hawaii for a fifteen days holiday.

L2 conventions would offer *kissed each other*, or merely *kissed*, as opposed to the use of an object pronoun in Italian – *si baciarono* – which, transferred into English, indicates the less romantic possibility that the girl kissed herself and the boy kissed himself.

Such confusions extend to object pronouns and verbs of psychological state (or 'psych verbs') in the case of Italian, which 'commonly expresses the idea of 'liking' with the equivalent of 'to please' (*piacere*). This means that the Italian subject corresponds to the English object, and vice versa', according to Duguid (2001:82). The same is true of another 'psych verb' – *miss*: 'the Italian for I miss you is *Tu mi manchi* – literally *You are lacking to me*' (Duguid 2001:82). Greek learners also struggle in this area, as the second of these examples shows:

Italian PET My new class **likes me** so much.

Greek FCE Dear John, I am writing to tell you how much you **miss me**.

L1 influence in this area, then, can lead to serious ambiguities in meaning, even for relatively proficient learners.

Conclusion

Returning to our initial assumptions, it has been shown that transfer from L1 does indeed impact upon L2 production of lexical verbs, and illustrations of this have been found for learners at both low (KET, PET) and high (FCE, CAE, CPE) levels of proficiency. The possibility exists, however, that the pressure of the examination situation may cause more slips. Learners may be more inclined to mentally translate from L1 rather than think in L2. Further studies may seek to compare candidate examination performance with that of the same candidates performing similar tasks in the classroom.

It has been shown that the use of learner corpora can confirm the intuitions of language professionals. This knowledge can be applied to inform teaching in both multilingual and monolingual settings. Indeed, teachers may encourage learners to utilise native speaker corpora in data-driven learning. Mishan (2004:223) explains the choice between a 'bottom-up' approach where 'evidence is examined, patterns are perceived, and generalisations are formed', and a "top-down" approach of hypothesis-experiment-conclusion'.

From a language testing point of view, such an approach can be used to build upon our understanding of what non-native learners of English bring to a particular lexico-

grammatical domain that native speakers do not. This may inform the writing or editing of test items, e.g. in identifying potentially strong distracting (but wrong) options for multiple-choice questions.

At syllabus level, it is potentially useful to distinguish errors made at particular levels (which corpora such as the CLC facilitate), giving a picture of what a student's writing looks like at a certain level. Corpora of spoken English may also make this possible for Speaking exams. Compilers of examination wordlists may draw upon corpora data in seeking to target expressions which trouble learners, with the aim of creating some positive washback.

As for the vagaries of the lexical verb in English, exacerbated by the problems of L1 transfer, acquisition by learners is not a straightforward process, in great part due to the unique development of the English language, which the Danish scholar Otto Jespersen compared to an English park, 'laid out *seemingly* [my italics] without any definite plan, and in which you are allowed to walk everywhere according to your own fancy' (Jespersen 1956:16).

References and further reading

- Andersson, L-G and Trudgill, P (1992) *Bad Language*, Penguin Books: London.
- BBC News (2004, June 7) *Government "sexed down" alcohol report*. Retrieved from <http://news.bbc.co.uk/1/hi/programmes/panorama/3767247.stm>
- Cambridge ESOL/Cambridge University Press (2009) *Cambridge Learner Corpus* (commercial corpus), Cambridge: Cambridge University Press.
- Chang, J (2001) Chinese Speakers, in Swan, M and Smith, B (Eds) *Learner English* (2nd ed.), Cambridge: Cambridge University Press, 310–324.
- Coe, R (2001) Speakers of Spanish and Catalan, in Swan, M and Smith, B (Eds) *Learner English* (2nd ed.), Cambridge: Cambridge University Press, 90–112.
- Duguid, A (2001) Italian Speakers, in Swan, M and Smith, B (Eds) *Learner English* (2nd ed.), Cambridge: Cambridge University Press, 73–89.
- Efstathiadis, S and King, P (1972) Some Lexical and Structural Errors made by Greek Learners of English, *ELT Journal* 26 (2), 159–167.
- Fitikides, T J (2000) *Common Mistakes in English* (6th ed.), Harlow: Pearson Education Limited.
- Golebiowska, L (2001) Polish Speakers, in Swan, M and Smith, B (Eds) *Learner English* (2nd ed.), Cambridge: Cambridge University Press, 169–178.
- Huddleston, R (1984) *Introduction to the Grammar of English*, Cambridge: Cambridge University Press.
- Jespersen, O (1956) *The Growth and Structure of the English Language* (9th ed.), New York: Doubleday and Co.
- Maxwell, K (2006, March) How new phrasal verbs develop, *MED Magazine* 36. Retrieved from <http://www.macmillandictionary.com/med-magazine/March2006/36-Phrasal-Verbs-New.htm>
- Mishan, F (2004) Authenticating corpora for language learning: a problem and its resolution, *ELT Journal* 58 (3), 219–227.
- Papaefthymiou-Lytra, S C (2001) Greek speakers, in Swan, M and Smith, B (eds) *Learner English* (2nd ed.), Cambridge: Cambridge University Press, 129–144.
- Partington, A (1998) Patterns and Meanings, *Studies in Corpus Linguistics* 2, Amsterdam/Philadelphia: John Benjamins B.V.

- Petrovitz, W (2001) The sequencing of verbal-complement structures, *ELT Journal* 55 (2), 172–177.
- Salem, I (2007) The lexico-grammatical continuum viewed through student error, *ELT Journal* 61 (3), 211–219.
- Shepherd, D (2001) Portuguese speakers, in Swan, M and Smith, B (Eds) *Learner English* (2nd ed.), Cambridge: Cambridge University Press, 113–128.
- Side, R (1990) Phrasal verbs: sorting them out, *ELT Journal* 44 (2), 144–152.
- Swan, M (1995) *Practical English Usage* (2nd ed.), Oxford: Oxford University Press.

TKT: Knowledge About Language and the assessment of lexis, phonology, grammar and discourse

EVELINA D GALACZI RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

STEVEN MILLER ASSESSMENT AND OPERATIONS GROUP, CAMBRIDGE ESOL

Introduction

Since its inception in 2005, the Cambridge ESOL Teaching Knowledge Test (TKT) suite has provided a range of tests which have proved popular with both governments and institutions that require accessible and flexible ways to assess English language teachers' knowledge of teaching. From the outset, the TKT modules have been seen by Cambridge ESOL as part of a platform for professional development which would provide a route into the English language teaching profession for new teachers, and career development opportunities for experienced teachers. Five years later, TKT has resulted in the development of a range of courses, reflecting a variety of teaching circumstances that in turn offer increased opportunities for teachers' professional development. Since its launch, more than 150,000 entries have been made for all modules in the TKT suite by teachers working in over 75 countries (see Harrison 2007 and Valazza 2008 for more on TKT adoption, recognition and impact).

The TKT: Knowledge About Language (KAL) test was launched in 2008 as one of the modules in the TKT suite, with a focus on the knowledge of concepts related to language and language use. The test is aimed at teachers of English who are both native and non-native speakers of English, and since its launch it has seen entries from teachers working in both state and private contexts, who typically possess a minimum of two years' experience, and who work with students ranging from primary age to adults. This article outlines the development of KAL, and in it we consider the rationale behind developing the test, the test construct and the quality assurance processes that underpinned its development.

Identifying a need for KAL

An examination of the Cambridge ESOL Teaching Awards, carried out as part of the DELTA review in 2006, revealed

extensive assessment and certification of the practice of English language teaching in a wide variety of contexts. The range of available assessments included course-based qualifications for teachers, such as the Certificate in English Language Teaching to Adults (CELTA), the Certificate in English Language Teaching to Young Learners (CELTYL) and the Diploma in English Language Teaching to Adults (DELTA), as well as tests for teachers, namely the TKT examinations. These awards spanned initial teacher training through to ongoing professional development and offered flexible and accessible means by which teachers could ensure that their teaching practice was well-informed.

One aspect of the 2006 review focused on the teaching awards that provided assessment of the subject matter knowledge required by English language teachers – TKT Module One, CELTA, CELTYL and DELTA (given in Table 1). One of the main conclusions to emerge was that no teaching award focused solely on subject matter knowledge and no award catered for teachers whose language level was minimally at a CEFR B2 level. Such a conclusion appeared to be at odds with Cambridge ESOL's flexible, modular approach to Teaching Awards aimed at meeting the evolving needs of teachers through expanding the Teaching Awards framework and giving teachers the widest possible choice when selecting the qualification that would best enhance their professional development.

The development of KAL would seek to address this conclusion by providing a test focused specifically on the subject matter knowledge required by English language teachers, and a minimum of CEFR B2 level would be needed to access the test. It would, as such, be a step beyond the initial TKT modules which provide a 'basic T1' language teaching qualification (see Rossner 2009 for a useful 'Profiling grid' which involves broad categories of teacher development, mapped against levels of professional development), and would target teachers at the 'T2' level (Rossner 2009:6). The development of KAL would also seek to extend the principles of flexibility and accessibility

Table 1: Cambridge ESOL Teaching Awards and tests for teachers

	TKT Module 1	CELTA	CELTYL	DELTA
Syllabus areas	<ul style="list-style-type: none"> • knowledge of terms and concepts common in English language teaching • knowledge of factors underpinning the learning of English • knowledge of available pedagogic choices 	<ul style="list-style-type: none"> • knowledge of learners and teachers and the learning and teaching context • language analysis and awareness • planning and resources • teaching skills and classroom management • development of language skills • resources and materials 	<ul style="list-style-type: none"> • knowledge of learners and teachers and the learning and teaching context • language analysis and awareness • planning and resources • teaching skills and classroom management for young learners • development of language skills • resources and materials for young learners 	<ul style="list-style-type: none"> • knowledge of language acquisition and language teaching • knowledge of approaches and methodologies • knowledge of language systems and learners' linguistic problems • knowledge of language skills and learners' problems • knowledge of resources • knowledge of assessment concepts and terminology
Teaching experience	not essential	not essential	not essential	recommended
Course attendance	not required	yes	yes	not required
Previous qualifications/training	not essential	qualifications which allow access to higher education	qualifications which allow access to higher education	an initial teaching qualification
Suggested language level	minimum CEFR B1	minimum CEFR C2/C1	minimum CEFR C2/C1	minimum CEFR C2/C1

embodied by existing TKT awards. Previous qualifications, previous teaching experience and course attendance would not be required by those taking the exam. Similarly to other TKT awards, KAL would cater for the widest possible range of teaching circumstances, would seek to be relevant to teachers regardless of their students' age group, would lend itself to the development of preparation courses and would increase opportunities for professional development. In line with other TKT modules, it would use an objective multiple-choice format.

Further to the findings of the DELTA review undertaken by Cambridge ESOL in 2006, evidence from the teacher training sector was taken into consideration in the planning phase, and an increasing recognition that subject matter knowledge was key to teacher professionalism was noted (Andrews 2003). Widespread anecdotal evidence from those involved in delivering English language teacher training suggested that a lack of subject matter knowledge was too often apparent in classrooms. It seemed that significant numbers of both native speaker and non-native speaker teachers of English, with differing levels of experience and training, were not always able to determine the suitability and purpose of the language content of teaching materials in relation to their learners. In addition, at times they appeared unable to field unanticipated language questions from their learners. This anecdotal evidence was also borne out in the literature on language awareness (see Andrews 2008, Wright & Bolitho 1993:292).

The KAL construct

One of the key issues when developing a test is a definition and description of the construct (i.e. latent knowledge or ability) it aims to assess. In the case of KAL, the construct clearly relates to teachers' knowledge about language.

Defining knowledge about language

Although teacher language awareness is an area of study that is still very much in development, with a lack of

consensus regarding the definition of concepts and common terms of reference (Andrews 2008), it is also clear that some agreement has emerged about the core components of the knowledge base in language teaching. Grossman's (1990) four components of teaching knowledge serve as a useful starting point in delineating the KAL construct. The author distinguished between:

- general pedagogic knowledge (general principles of teaching and learning which are applicable across subject disciplines)
- subject matter knowledge (the understanding of the facts, concepts and terminology of a subject discipline)
- pedagogic content knowledge (the representation of the subject matter through examples, analogies and procedures, to make it more comprehensible to the students)
- knowledge of context (knowledge of educational aims, students and other contextual factors, which would inform the application of the other three types of knowledge).

Grossman's framework was later applied to English language teaching in the work of Tsui & Nicholson (1999), who further developed the framework with specific reference to English as a foreign/second language. Tsui & Nicholson's framework served as guidance when developing the test specifications and syllabus for TKT Modules 1 to 3 (Ashton & Khalifa 2005, Ashton & Galaczi 2009, Harrison 2007, Spratt 2005). The initial three TKT Modules, therefore, are based on a broader syllabus encompassing a range of the core competencies for teaching, which includes knowledge about the subject and general theoretical knowledge about teaching, as well as subject-specific knowledge about teaching. This core knowledge is generally accepted by the wider academic and educational community and forms the basis of many English language teacher training courses.

The knowledge area which had the most obvious

implications for KAL was Grossman's (1990) 'subject-matter knowledge'. This knowledge area was further developed by Tsui & Nicholson (1999) for English language teaching and sub-divided into four distinct categories: phonology, lexis, grammar and discourse. Although, as Andrews (2008:293) notes, much current research is 'pushing the boundaries' of currently accepted ideas of what comprises subject-matter knowledge – e.g. Walsh's (2003) proposal that an understanding of interactional processes is a crucial component of knowledge about language – there is now widespread support that the core subject matter knowledge components are the four-partite distinction proposed by Tsui & Nicholson (1999). These sub-components of subject-matter knowledge form the basis of the KAL syllabus, and will be returned to later in the article.

Language proficiency and language awareness

Language teachers play an unusual role, in that they need to develop three competencies simultaneously, as Edge (1988) has noted: language user, language analyst, and language teacher. The first role, the user, is related to the teacher's language proficiency and their role as a model for the students; the second role refers to the teacher's metalinguistic awareness about the language systems; the third role taps into the teacher's general and specific pedagogic knowledge about procedures and processes in an English language classroom. Across these roles there is, therefore, an intertwining of language proficiency (knowledge of language) and subject matter knowledge (knowledge about language).

The focus of the KAL examination is the second of these roles, i.e. the teacher as language analyst who possesses meta-knowledge about language systems, but KAL also uses English as a medium. In order to access the content of KAL, teachers need a level of English of at least CEFR level B2. The choice of this proficiency level as a minimum was felt to be appropriate by the test developers, as it situates KAL teachers' minimum language competence at the higher end of an 'independent' user, in CEFR terms, and also at the T2/T3 level in Rossner's (2009) grid. The B2 level of KAL items also enhances the flexibility and accessibility of Cambridge ESOL's Teaching Awards, as it bridges the divide between the other TKT modules (at B1), and the DELTA Module 1 examination (at C1/C2).

Quality assurance and KAL

A substantial part of the quality assurance underlying KAL is based on building a case for its validity, supported by relevant evidence. Such an approach to quality assurance is part of an evidential approach to validity (Messick 1989, Weir 2005) where claims about the validity or usefulness of a test need to be based on evidence and the weakness/strength of the claims explored. The evidence is, in turn, based on data, which is collected through qualitative and quantitative methods (Saville 2003). Weir's (2005) socio-cognitive framework for test validation provides a useful and systematic approach to providing validity evidence for a test, with its focus on test-taker characteristics: cognitive, context, scoring, consequential, and criterion-related validities. Various stages of the development of

the KAL test have provided validity for and evidence about the test.

The development cycle of KAL drew on several quality assurance procedures. At the initial stage of the test development process, the draft test specifications and syllabus content areas were all compiled following consultations with content experts (LKT draft specifications October 2006, December 2006a, December 2006b, February 2007), and were closely mapped onto widely accepted views in the academic and professional literature regarding English language subject matter knowledge (Tsui & Nicholson 1999). External consultants were asked to report on the overall concept of the test, the coverage of subject matter knowledge it provided, the ordering of this coverage, the accessibility of the sample tasks to their intended audience of English language teachers with a minimum English language level equivalent to B2 level on the CEFR, and the tasks themselves (Bell 2007, Oakes 2007).

The reports that this process generated then led to further drafts and resulted in the identification of specific features of subject matter knowledge which would inform the syllabus areas of the test.

A trialling stage also provided evidence for the soundness of the test. This is what we turn to now.

The KAL trial took place in December 2007 in 10 locations around the world (China, Australia, Russia, Estonia, Hungary, Uruguay, Brazil, Argentina, Peru, and the UK), prior to the exam's launch. A principal aim of the trial was to investigate how well the test performed and to gather stakeholder feedback, which would in turn inform subsequent stages in the development of the test. The trial participants (n = 297) provided a range in terms of age, gender, teaching qualifications held, teaching experience, workplace and language ability. They were a varied and, more importantly, representative sample of the anticipated KAL candidature. The information collected through the trial included responses to KAL items and tasks, an English language test, a candidate background questionnaire, and a feedback questionnaire for candidates and administrators.

One of the areas of investigation during the trial focused on the psychometric attributes of the items and tasks (i.e. their scoring validity). A Classical and Rasch analysis indicated that the trial test had a reliability of 0.93 (Cronbach's alpha), which indicated high internal consistency of the test. Indices of item misfit (as indicated by the Rasch analysis) signalled that a small number of tasks were misfitting (the majority associated with two specific tasks), and were a signal that the tasks in question had to be scrutinised – both for content and format – in the final development stage of the test.

Another area of investigation focused on the relationship between candidate background variables and performance on the test. The overall performance of candidates in the different language ability groups showed that in general, the mean KAL score increased as language proficiency increased. This is not surprising since, as noted earlier, there is an interconnection between knowledge *of* a language and knowledge *about* a language. An important finding, however, was the wide range of marks within each

proficiency level (seen in Figure 1). More specifically, in the B2 group, approximately two thirds of the candidates achieved a KAL mark of 50 (out of 80) or above, whereas in the highest (C2) group, a small proportion of candidates received a mark below 49. This was seen as an encouraging finding which provided context validity evidence for the test, as it indicated that, while language proficiency is naturally part of the KAL construct, simply being proficient in English did not lead to a higher score. Alternatively, it also indicated that lower language proficiency was not a barrier for a candidate to perform at a satisfactory level, and that candidates at B2 – the level needed to access the test – had performed adequately. The extended feedback from the trial participants also supported this finding – a large majority (91.3%) had said that the language of the test was at the right level.

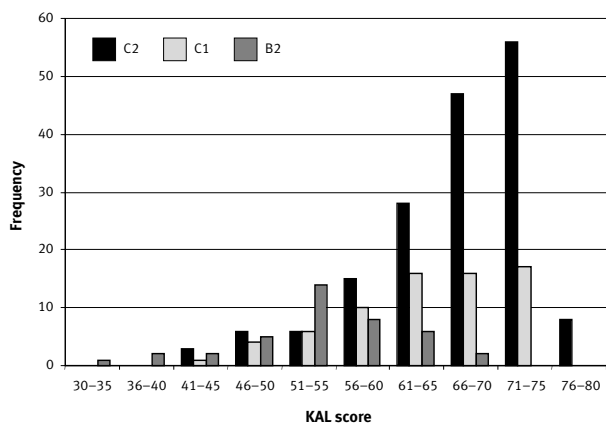


Figure 1: Spread of marks, KAL trial

Another background variable which had some effect on performance on KAL was the length of teaching experience. In general, candidates who had two years or more of teaching experience performed better than those who had one or less years of experience, as seen in Figure 2. The difference in scores between the least experienced and most experienced group was found to be relatively small however, which is a reflection of the construct underlying this test – KAL is a test which focuses on knowledge of the linguistic systems of grammar, lexis, phonology and discourse, and not on teaching practice.

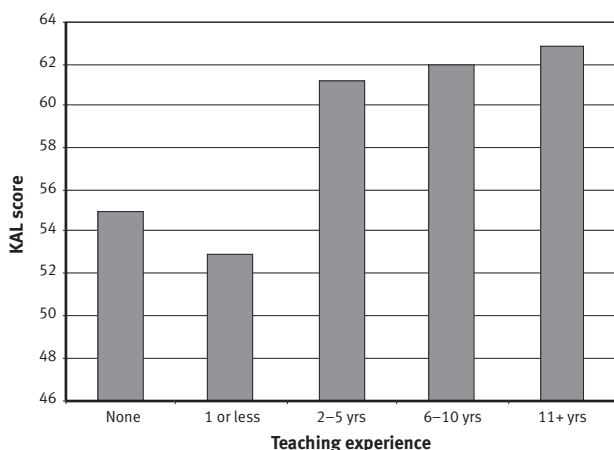


Figure 2: KAL score and teaching experience

In addition to quantitative information about the psychometric qualities of the items, and the role of background variables such as language ability and teaching experience, the trial provided qualitative feedback from the candidates and administrators about the test. One of the themes running through the feedback was the use of terminology in the test. A large part of the extended feedback indicated that candidates felt that terminology was used differently in different contexts and teaching situations. A few examples:

'The terminology used in the test is a bit different ... from what we used to learn in university.'

'Phonological terminology was different from what I had studied at university.'

'Is "multi-word" verb the same thing as "phrase verb" or "prepositional verb"? Maybe there must be a section before the questions – you explain what you mean by "multi-word verb"?''

Exposure to different terminology and frameworks is generally considered an unavoidable aspect of membership of the wider English language teaching community. The lack of professional consensus regarding certain terms which was evident in the qualitative feedback highlighted the importance of providing a gloss for some of the terms employed in the test, in order to avoid test bias.

Other candidates and administrators voiced the opinion that some of the terminology used in the test was not applicable to teaching situations, as seen in the following feedback:

'There is a specific terminology in the test which I don't consider relevant for language teaching such as anaphoric, cataphoric and exophoric.'

'Such discourse feature as referencing is studied at universities and generally has no practical application in everyday teaching.'

'I know most of the terminology. I only didn't know some terms. I don't work on all the areas with my students because I am working with beginners at the moment.'

Such comments highlighted the need for the aims of the test to be made clear, namely that there is a distinction between knowing a concept/term and using that concept in the classroom. Such knowledge of concepts and terminology contributes to the creation of a professional teaching community with a degree of consensus over what language areas a teacher needs to know. For example, the medical profession involves the knowledge of terminology which is not necessarily used during a consultation with a patient, but which is nevertheless part of the knowledge base of a medical doctor. However, some of the candidate and administrator feedback did, in fact, comment on the need for teachers to be familiar with relevant terminology when describing language systems and structures. For example:

'It was challenging but it tests what a good teacher must know.'

'I think that the test focuses the attention on the areas which I normally don't look at. It helped me to focus and re-evaluate my knowledge.'

'As I had never studied the terminology used in the test I found it quite challenging. This is not bad at all as I will have this as a goal for my future teaching career.'

The quality-assurance steps we have outlined in this section contributed to providing evidence for the validity of the test and its fitness for purpose. The test specification and syllabus which were produced as a result of the test development process are given in the following section.

KAL syllabus and sample tasks

As noted earlier, the KAL syllabus taps into four core areas of subject-matter knowledge: lexis, phonology, grammar and discourse. The following section illustrates these four core areas with sample tasks and related commentary.

Lexis

In terms of lexis, the KAL syllabus covers a wide range of features, which include the following:

- different types of meaning: denotation, connotation, figurative, contextual (e.g. situational and collocational)
- sense relations: synonyms, antonyms, hyponyms, whole/part, lexical sets, false friends
- word formation: morphemes, affixation, compounds, types of abbreviations, acronyms, spelling rules, word families
- lexical units: collocations, idioms, fixed phrases, formulaic phrases
- lexico-grammatical features: grammatical function of words and morphemes
- register: domain-specific vocabulary; degrees of formality in vocabulary and their contexts of occurrence.

The following sample items illustrate one of the ways in which the assessment of lexis is addressed in KAL:

Figure 3: Sample KAL lexis task. Source: KAL Handbook for Teachers (2010:10)

A teacher is completing a teacher development quiz on lexis.

For questions 1–6, look at the extract from a novel. Read the questions about the lexis in the extract from a novel and three possible answers listed A, B and C.

Choose the correct answer.

Mark the correct letter (A, B or C) on your answer sheet.

- 1 Which line in the text contains a compound?
A line 3 B line 6 C line 7

- 2 Which line in the text contains a noun suffix?
A line 6 B line 8 C line 12

Extract from a novel

Mary turned away, feeling relieved, and looked out of the window at the quiet street below, bathed in bright sunshine, with the BGR company building in the distance. She was surprised to discover that she'd no wish to leave yet, even though she was also rather embarrassed to be there. Through the window, she watched an old woman dressed in a heavy coat and wearing a scarf, despite the heat. She was walking along the street with an elderly sad-looking dog with a bow round its neck. It was soothing to gaze out at the woman's slow movements – to watch her unfastening her heavy front gate, closing it behind her with fussy precision, and then, halfway to her front door, bending with difficulty to pull up a weed from the narrow bed that ran along the entire length of her front path. As she did so, the dog waddled towards her unsteadily and licked her wrist. The lady and her dog went indoors, and the street was empty again. It could be any Saturday.

line 1
line 2
line 3
line 4
line 5
line 6
line 7
line 8
line 9
line 10
line 11
line 12

Phonology

The following features of this subject area are included in the test:

Segmental features:

- phonemes: the different phonemes of English and their places and manner of articulation, e.g. bilabial plosive; voiced or unvoiced consonant; front or back vowels, etc.
- word stress: primary, secondary stresses and weak syllables; contrastive word stress and changing word stress
- standard word patterns in word stress e.g. words ending in *sion/tion*
- the role of segmentals in conveying meaning e.g. full vs. weak forms, contrastive word stress.

Suprasegmental features:

- sentence stress: primary and secondary stresses in a sentence; emphatic and contrastive stress.
- connected speech: linking, assimilation, elision, intrusion (liaison), weak forms, contraction
- intonation: pitch direction and range; common intonational patterns such as for 'yes/no' questions; the grammatical, attitudinal and interactive functions of intonation
- the role of suprasegmentals in conveying meaning.

The sample task below provides an example of the assessment of features of connected speech within this content area.

Figure 4: Sample KAL phonology task. Source: KAL Handbook for Teachers (2010:12)

A teacher has identified some sentences in a recording that contain certain features of connected speech.

For questions 21–28, match the phonemic transcriptions of the sentences with the features of connected speech that they contain listed A–D.

Mark the correct letter (A–D) on your answer sheet.

You will need to use some of the options more than once.

Features of connected speech

- A intrusion (adding an extra sound)
B weak form of a vowel
C assimilation (a sound changing towards a neighbouring sound)
D elision (omission) within a consonant cluster

Phonemic transcriptions

- 21 /nɒt ət ɔ:l/
Not at all.
22 /dəʊm pleɪ/
They don't play football there.

Grammar

The features of grammatical knowledge included in the KAL test syllabus are:

- the role of context in determining meaning and use of grammatical forms
- word classes and the functional role of words in a sentence
- grammatical roles performed by words in a sentence

- the form, meaning and use of structural patterns
- the noun phrase: different types of nouns, determiners, adjectives, noun phrase structures, nominalisation
- the verb phrase: verb types, verb patterns, mood, modality, time and tense, aspect, hypotheticality, adverbials, the passive, reported speech
- the sentence: simple sentences, complex sentences.

The task below shows one approach to testing verb patterns within this syllabus area:

Figure 5: Sample KAL grammar task. Source: KAL Handbook for Teachers (2010:14)

A teacher has designed a worksheet on verb patterns for his students and is writing the answer key.

For questions 51–57, look at the verbs and the three grammatical patterns listed **A**, **B** and **C**.

Two of the patterns can follow the verb. One of the patterns CANNOT follow the verb.

Mark the pattern (**A**, **B** or **C**) which CANNOT follow the verb on your answer sheet.

- 51** allow
A + object + to + infinitive **B** + object + *ing* form **C** + *ing* form
- 52** suggest
A + *ing* form **B** + that + clause **C** + object + to + infinitive

Discourse

Finally, the features of discourse included within the KAL syllabus are:

- coherence: what it is and how it is achieved in writing and speaking
- grammatical cohesion: linking words; anaphoric, cataphoric and exophoric reference; substitution and ellipsis; parallelism
- lexical cohesion: lexical chains, discourse markers; text structuring
- register/appropriateness: the relationship of register to audience, message and setting; degrees of appropriateness: formal, neutral, informal
- recognising register: choice of vocabulary, grammar, layout, genre, functions
- general distinctions between written and spoken English, e.g. sentences and utterances, complexity and simplicity of grammar and lexis, cohesion, register, organisation of discourse
- genres: notion of audience; features of common genres for writing and speaking
- distinctive features of spoken genres, e.g. turn-taking, feedback and adaptation to audience, incomplete utterances
- differences between pragmatic and semantic meaning
- different ways of expressing a range of written and spoken functions using grammatical and lexical means
- the significance of phonology and context in determining coherence and pragmatic meaning.

The task below provides an illustration of a task addressing knowledge of discourse.

Figure 6: Sample KAL discourse task. Source: KAL Handbook for Teachers (2010:16)

A teacher has prepared an exercise on the functions of discourse markers for his class and is writing the answer key.

For questions **74–80** look at the underlined discourse markers and the three possible functions listed **A**, **B** and **C**.

Choose the function which matches the discourse marker.

Mark the correct letter (**A**, **B** or **C**) on your answer sheet.

- 74** Di: Mum, why can't we go on holiday this summer?
 Mum: Well, for a start, we can't afford it.
- A** introducing the first point
B rejecting an argument
C introducing a new topic
- 75** Katya: Don't forget to buy some bread on your way home.
 Tomas: I won't. Oh and by the way, I invited Jon for the weekend. Is that OK?
- A** linking similar things
B introducing a contrast
C changing the focus

Future steps

The performance of the TKT: Knowledge About Language (KAL) test is part of an ongoing operational and research agenda which ensures that the test meets established principles of good practice. One of the key areas of investigation is the relationship between subject-matter knowledge and performance in the classroom. As Andrews (2003, 2008) argues, subject-matter knowledge about language systems is just one part of a teacher's language awareness, and we must distinguish between knowledge and the ability to use that knowledge appropriately in the classroom. A research project currently underway at Cambridge ESOL is an investigation into the relationship between performance on TKT and performance in the classroom, and the impact of the TKT suite on professional development. The findings of such research will go on to inform all Cambridge ESOL Teaching Awards, provide valuable information about the KAL test and the intertwining of the different teacher roles, that of language teacher, language analyst and language user.

References and further reading

- Andrews, S J (2003) Teacher Language Awareness and the Professional Knowledge Base of the L2 Teacher, *Language Awareness* 12, 81–95.
- Andrews, S J (2008) Teacher Language Awareness, in Cenoz, J and Hornberger, N H (Eds) *Encyclopedia of Language and Education* (Volume 6, Second Edition), New York: Springer, 287–298.
- Ashton, M and Khalifa, H (2005) Opening a new door for teachers of English: Cambridge ESOL Teaching Knowledge Test, *Research Notes* 19, 5–7.
- Ashton, M and Galaczi, E (2009) *Developing tests for teachers: Insights from the Teaching Knowledge Test*, paper presented at American Association of Applied Linguistics Conference, Denver, Colorado, 21–24 March 2009.
- Bell, J (2007) *Review of Language Knowledge Test (LKT) draft specifications and sample tasks*, Cambridge ESOL consultant report.
- Edge, J (1988) Applying linguistics in English language teacher training for speakers of other languages, *ELT Journal* 42 (1), 9–13.

- Grossman, P (1990) *The Making of a Teacher: knowledge and teacher education*, New York: Teacher's College Press.
- Harrison, C (2007) Teaching Knowledge Test update – adoptions and courses, *Research Notes* 29, 30–32.
- Language Knowledge Test (LKT) draft specifications version 1* (October 2006), internal Cambridge ESOL report.
- Language Knowledge Test (LKT) draft specifications version 2* (December 2006a), internal Cambridge ESOL report.
- Language Knowledge Test (LKT) draft specifications version 3* (December 2006b), internal Cambridge ESOL report.
- Language Knowledge Test (LKT) draft specifications version 4* (February 2007), internal Cambridge ESOL report.
- Messick, S (1989) Validity, in Linn, R L (Ed.), *Educational Measurement*, 3rd edn, NY: Macmillan, 13–103.
- Oakes, S (2007) *Review of Language Knowledge Test (LKT) draft specifications and sample tasks*, Cambridge ESOL consultant report.
- Rossner, R (2009) Developing common criteria for comparison and assessment in language teacher education, *Research Notes* 38, 4–14.
- Saville, N (2003) The process of test revision and development within UCLES EFL, in Weir, C J and Milanovic, M (Eds), *Continuity and Innovation: Revising the Cambridge Proficiency in English examination 1913–2002*, Studies in Language Testing 15, Cambridge: UCLES/Cambridge University Press, 57–120.
- Spratt, M (2005) *The construct of TKT*, Cambridge ESOL consultant report.
- Tsui, A B M and Nicholson, S (1999) Hypermedia database and ESL teacher knowledge enrichment, *Journal of Information Technology for Teacher Education* 8 (2), 215–237.
- University of Cambridge ESOL Examinations (2010) *Teaching Knowledge Test Knowledge About Language (KAL) handbook for teachers*, Cambridge: UCLES.
- Valazza, G (2008) Impact of TKT on language teachers and schools in Uruguay, *Research Notes* 34, 21–26.
- Walsh, S (2003) Developing interactional awareness in the second language classroom through teacher self-evaluation, *Language Awareness* 12 (2), 124–142.
- Weir, C (2005) *Language testing and validation. An evidence-based approach*, Basingstoke: Palgrave Macmillan.
- Wright, T and Bolitho, R (1993) Language Awareness: a missing link in language teacher education?, *ELT Journal* 47(4), 292–304.

Conference reports

IATEFL pre-conference report: Testing, Evaluation and Assessment SIG (TEA SIG) in Harrogate 2010

The 44th annual IATEFL conference and exhibition took place in Harrogate, UK, from 7–11 April 2010 and was attended by around 3,000 delegates. The first day of the conference was made up of 11 Pre-Conference Events (PCE) organised by IATEFL Special Interest Groups (SIGs) which are professional development days.

The PCE held by IATEFL Testing, Evaluation and Assessment SIG (TEA SIG), sponsored by Cambridge ESOL, was entitled 'Assessing Reading'. Its aim was to explore issues related to assessing reading and concentrate on the following:

- How can test designers best approach the testing of different levels of reading comprehension?
- How can test designers minimise test bias and maximise reliability?
- Can 'reading' be broken down into easily assessable skills?
- What are the essential steps one should follow when designing a reading test?
- What are some of the 'ideal' task types?
- How does one give feedback to test takers on their performance?

The day's talks covered the following areas:

Keith Morrow from *English Language Teaching Journal*

In his talk titled 'Testing Reading: What's the problem?', Keith looked at the two fundamental concerns of what we

mean by 'reading' and how we can test it, by examining the content and design of some current EFL reading tests. He also gave examples of how demands of reliability in assessing reading may hamper attempts to 'really' test reading and concluded by suggesting that alternative approaches to assessment might actually be better.

Christine Coombe from Dubai Men's College, UAE

In her talk 'A Step-by-step Approach to Reading Assessment', Christine gave a very practical outline of the important steps that need to be followed in order to develop reading tests that accurately measure students' general reading comprehension and determine students' competency regarding specific reading sub-skills. She stated that because reading is regarded as such an important skill, it is often given considerable weighting when it comes to student assessment, saying that it was essential that reading tests were a valid and reliable measure of students' reading ability.

Hanan Khalifa from Cambridge ESOL, UK

Hanan started by providing an overview of how theories of reading comprehension have evolved since the 1960s and how they have impacted on language teaching and assessment. In her talk 'Assessing reading comprehension ability', she proposed a framework of processing levels that aims to assist test developers in establishing the cognitive validity of their L2 reading tests at the different levels in the Common European Framework of Reference. She then gave examples of the applications of this model to internationally recognised examinations, such as Cambridge ESOL exams, to show how the reading processes

may differ from CEFR levels A2 to C2 within an assessment context.

Neil Anderson from Brigham Young University, Utah, USA

Neil's talk brought a different dimension to assessment of reading as he emphasised the importance of developing self assessment strategies for learners in his talk 'Self Assessment Strategies of Reading Performance to Keep the Flame of Motivation Burning'. He emphasised how teachers' motivational practice in the classroom matters, and that we as language teachers need to take appropriate steps to keep the flame of motivation burning in order to help our learners develop self assessment strategies of reading. That way, they can turn moments of perceived failure into moments of learning.

Neus Figueras from Catalan Ministry of Education/EALTA, Spain

The last speaker of the day, Neus Figueras, started by saying that although language testers today have a variety of frameworks to choose from when writing their own specifications for testing reading, including the descriptor scales in the CEFR, they still had to struggle with definitions and terminology issues when describing reading operations, text and task characteristics and text and task difficulty. In her talk 'Testing Reading: from operationalization to difficulty estimation', Neus said finding a way of reconciling estimated difficulty as perceived by expert judges and real difficulty as evidenced by pilot tests still remained a challenge.

The day ended with a panel discussion.

Testing, Evaluation and Assessment (TEA) SIG programme in Harrogate, April 2010

This year the TEA SIG programme was on the third day of the conference where the programme included selected talks to reflect the wide variety of research and applications in the areas of testing, evaluation and assessment. As usual, the TEA SIG programme in Cardiff attracted a wonderful mixture of 'regulars' and new faces. The interests of TEA SIG members were reflected in the content of the talks which covered a wide and fascinating range of topics:

- Assessing a new writing task type: Short Answer Responses (Zeynep Urkun)
- Developing speaking assessment scales across languages and levels (Ann Humphry-Baker)
- SIMTEST: a web-based component in an EFL placement test battery (Mick Sumbling & Pablo Sanz)
- Empowering teachers as testers (Christine Coombe)
- Assessing a learner's proficiency in spoken English (Sue Davies).

Once again, thanks to our speakers, we will be able to publish articles in a forthcoming issue of the TEA SIG newsletter.

As part of the programme, TEA SIG held its Open Forum, where, once again, we had a mixture of long-standing TEA SIG members and some new ones. After going over the finances, the past and future events of the SIG, we

brainstormed topics for possible TEA SIG conferences and managed to draw up a list of exciting possibilities. The TEA SIG Committee is now looking forward to an equally active year in 2010. Please visit <http://tea.iatefl.org/> to learn more about future TEA SIG events.

Two workshops on the Common European Framework of Reference: LTRC April 2010

Two 1-day pre-conference workshops on the Common European Framework of Reference (CEFR) were held on 12 and 13 April.

The first day was on *The CEFR: its purposes, origins, development and current status*; the second on *The CEFR: alignment and standard-setting issues*. They were coherent with the conference theme of *investigating levels, domains and frameworks in language assessment*.

Presenters included several authorities on the CEFR: John Trim and Brian North, two of its authors, Neus Figueras and Norman Verhelst, co-authors of the *Manual for relating language examinations to the CEFR*, and Sylvie Lepage, who has co-ordinated standard-setting conferences on behalf of the Council of Europe. Cambridge ESOL presenters included Neil Jones, Nick Saville, Angeliki Salamoura and Szilvia Papp.

Workshop One began with a presentation by John Trim on the origins and purposes of the CEFR. He traced the progression from Threshold Level and Vantage to the CEFR. He explained the addition of the common reference levels as a 'vertical dimension' to the Framework beside the 'horizontal' taxonomy of language use and competences, as a useful tool to enable learning space to be mapped or profiled, even if simply.

Brian North presented the CEFR as 'a heuristic model, a metalanguage to help people to exchange expertise and to relate their assessments to each other', in order to help overcome the 'Tower of Babel' problem of lack of common standards.

Neil Jones conducted two workshop exercises inviting participants to consider the way proficiency levels are defined in the CEFR. One exercise focused on the statement in the CEFR that 'level C2 ... is not intended to imply native-speaker or near native-speaker competence'. On analysing the C2 level 'I can' statements from CEFR Table 2, participants found that descriptions of some skills – writing in particular – suggested a level far beyond many native speakers, and thus that the C2 levels refer to aspects of educated or academic competence. This recalled Cummins' distinction between CALP (cognitive academic language proficiency) and BICS (basic interpersonal communicative skills); it was suggested that for some learning contexts it would be useful to separate these competences more clearly, and that development of CALP might equally be addressed at levels below the C levels.

Nick Saville introduced Reference Level Descriptions, an area of work promoted by the Council of Europe to produce linguistic descriptions of progression for each European language. This new generation of reference descriptions is based on the six CEFR level descriptors: it is a case of

identifying the forms of a given language (words, grammar, etc.), mastery of which corresponds to the communicational, socio-linguistic, formal and other competences defined by the CEFR. Nick focused on the *English Profile* programme, a collaboration to produce a detailed corpus-based linguistic description for English, articulated with the help of sophisticated computational linguistic analyses.

The workshops stressed the importance of referring specific learning contexts to the CEFR (rather than applying the CEFR to contexts in a standardised and rigid manner). Szilvia Papp presented a case study on exams for younger learners. She described a project to develop a set of calibrated 'Can Do' statements appropriate to this age group(s), selecting from a bank of descriptors used in various implementations of the European Language Portfolio. The presentation stimulated a number of questions: how would teachers or learners use such statements, and in what ways might this have a formative effect? What skills should 'Can Do' statements for the classroom include – should they be primarily functional, linguistic or learning-skills focused?

Brian North reviewed the range of criticisms which the CEFR has attracted, in relation to the style and organisation of its text, the view of language learning presented, the underlying descriptive theory, its use as a language policy document or tool, the notion of a common metric to which to relate tests, and the status of the descriptors. Some of these criticisms were interpretable as misunderstandings as to the purpose or scope of the CEFR. A few indicated ways in which the CEFR may need complementation, e.g. Weir's call for a fuller treatment of cognitive processes, and his statement that 'The CEFR may never have been intended to provide a comprehensive account of (contextual parameters at different levels) but language testers need to, however difficult this may prove.'

This review set the scene for a final round table discussion, addressing questions posed by participants.

Workshop Two had a more practical focus on linking examinations to the CEFR. The morning focused on performance skills, and the afternoon on standard setting for objectively marked tests.

Brian North provided an introduction to the approach and organisation of the *Manual for relating language examinations to the CEFR*, and then conducted a couple of exercises to illustrate the process of familiarising users with the CEFR scales and levels.

Sylvie Lepage and Brian then led a session based on video examples of spoken performance calibrated in a multilingual rating conference organised by CIEP on behalf of the Council of Europe in 2008. Participants rated two such examples against the CEFR levels adopted. The DVD of CIEP examples, and a Cambridge ESOL DVD of examples at five CEFR levels, were also provided to participants in their workshop pack.

Neil Jones presented the outcomes of a comparative study conducted during the 2008 conference, in which multilingual rankings (i.e. comparative judgements) were elicited alongside the ratings (i.e. absolute judgements). Both kinds of data produced comparable and highly correlated ability scales. This confirmed that rankings offer an attractive approach to cross-language alignment which

does not depend on judges sharing an absolute understanding of CEFR levels.

In the afternoon Neus Figueras conducted two sessions, the first of which introduced the process of specification proposed in the *Manual for relating language examinations to the CEFR*, and of profiling the components of a test in relation to the CEFR.

The second session concerned training for standard setting of objective tests, and the standardisation of judgements. Norman Verhelst's presentation complemented this practical introduction, reviewing standard-setting procedures from a more theoretical perspective.

Both workshops were fully booked, with most participants signing up for both days. This seems to reflect a considerable interest in the CEFR, and a recognition in the language testing community that it is a topic to be engaged with seriously. This was indeed the message that the workshops sought to convey. The CEFR challenges us to address the twin goals of achieving validity within a specific learning context, and comparability within a general proficiency framework. Reconciling these only partially reconcilable goals requires engagement with the nature of the link between the process of language learning and its outcomes in terms of communicative language ability. This, it was proposed, can only be good for language testing.

38th ALTE meeting and conference, Rome, Italy

ALTE has recently completed a successful week of activities in Rome centring around its biannual meeting and conference. The week began with a 2-day Introductory Course in Assessing Writing which was run by Gad Lim, Research and Validation Group, Cambridge ESOL. The course covered a number of issues related to the assessment of writing, including a model of writing ability and a framework for testing writing, choosing test formats and tasks, and the rating process, and attracted participants from several countries including Italy, Germany, Denmark, Sweden and Spain as well as two colleagues from Assessment and Operations Group, Cambridge ESOL.

On 11 May, ALTE co-hosted a 1-day event on language assessment for migration with the Centre for Evaluation and Language Certification (CVCL) from the University for Foreigners in Perugia, the ALTE member for Italian. This one-day event brought together a wide range of practitioners, consultants, educationalists and policy-makers, including Fiorella Perotto from the European Commission and Philia Thalgott from the Council of Europe, Maria Rosa from the Italian Ministry of the Interior, Giuseppe Silveri from the Italian Ministry of Labour, Professor Elana Shohamy, Piet van Avermaet, as well as ALTE members and affiliates, and several local stakeholders to exchange ideas and experiences related to language assessment for integration from an Italian and international perspective.

The meeting and conference took place from 12–14 May and a number of Cambridge ESOL colleagues and consultants ran workshops and/or gave presentations, including Luna Filipović, Dittany Rose, Angeliki Salamoura, and Nick Saville. ALTE was also greatly honoured that Dr John Trim, one of the authors of the CEFR, attended the

week's events to run a workshop with Nick Saville, and give a plenary presentation on the development of the Reference Level Descriptions (RLDs). Dr Trim also chaired a very successful round table discussion on the RLDs and assessment at the end of the conference day.

A series of foundation courses on language testing presented by Annie Broadhead and Jacky Newbrook, Consultants to Cambridge ESOL, rounded off a very productive week of activities.

Registration is now open for the ALTE summer testing courses which will take place in Bilbao from 20–24 September, and from 27 September–01 October. These courses are being hosted by ALTE's Basque member, the Basque Government. Lynda Taylor and Cyril Weir will run the first week's course – The ALTE Introductory Course in Language Testing; and Hanan Khalifa and Ivana Vidaković will run the second week's course – The ALTE Introductory Course in Testing Reading.

From 10–12 November, ALTE will hold its 39th meeting and conference at the Charles University in Prague. The theme of the conference will be 'Fairness in Language

Testing', and this will be part of a week of events related to fairness and the ALTE minimum standards, beginning with a 2-day programme of auditor training. The conference will then be followed by a 2-day course on Structural Equation Modelling run by Ardeshir Geranpayeh, which will launch ALTE's Tier 3 professional development courses.

Looking ahead to 2011, the ALTE 4th International Conference will take place from 7–9 July at the Jagiellonian University in Kraków, Poland. The theme of the conference will be 'The Impact of Language Frameworks on Assessment, Learning and Teaching: policies, procedures and challenges', and the plenary speakers will be Professor Lyle Bachman, Professor Giuliana Grego Bolli, Dr Neil Jones, Dr Waldemar Martyniuk, Dr Michaela Perlmann-Balme and Professor Elana Shohamy. The Call for Papers is now open and will run until the end of January 2011. We encourage you to submit a proposal and further details are available on the following link: www.alte.org/2011/call-for-papers.htm

For further information about all of ALTE's events and activities, please visit the ALTE website at: www.alte.org.

Recent publications of interest

Studies in Language Testing

This year sees the *Studies in Language Testing* (SiLT) series celebrate its 15th anniversary, with over 30 volumes now available.

Studies in Language Testing is an important publishing collaboration between University of Cambridge ESOL Examinations and Cambridge University Press, both non-teaching departments of the University of Cambridge. The series began in 1995 with two main purposes in mind. The first was to profile in the public domain the test development, validation and research activities conducted by Cambridge ESOL for its wide range of language proficiency examinations. The second was to support and promote work in the fields of applied linguistics and language assessment, and to enable the language testing community to access and benefit from research which makes a contribution to the field but which might not otherwise reach publication, including PhDs of calibre and selected conference papers. No end point for the series was envisaged at the time, but neither did we think that 15 years later there would be over 30 titles and that more than 200 academics would have either written books in the series or contributed papers to edited volumes. Not surprisingly, the series has become an indispensable resource for anyone interested in new developments and research in language testing and assessment. Individual volumes are relevant to a wide range of testing and assessment stakeholders including researchers, academic lecturers, postgraduate students and educational policy-makers, as well as teachers, teacher trainers, curriculum designers and textbook writers.

To celebrate 15 years of successful publishing activity a special *Studies in Language Testing* publication has been produced – the *15th Anniversary Edition*. This slim volume contains an informative Introduction chronicling the development and growth of the series, including the rationale for each new title added to the range. The anniversary edition also includes a brief content description of each title together with a full set of the Series Editors' notes that normally appear at the front of each volume.

SiLT's 15th anniversary was celebrated as part of the Social Programme at the 2010 Language Testing Research Colloquium (LTRC) hosted in April this year by Cambridge ESOL (see www.CambridgeESOL.org/LTRC2010/index.html); conference delegates were served with slices of a specially prepared birthday cake during the afternoon refreshment break and they received a free copy of the 15th Anniversary Edition.

More information on the SiLT series is available at: www.CambridgeESOL.org/what-we-do/research/silt.html

IELTS MA Award Winner 2009

The IELTS Master's Award 2009 was presented at the 2010 Language Testing Research Colloquium (LTRC) by Nick Saville (Cambridge ESOL) on behalf of the IELTS partners. The winner was Ruslan Suvorov, from Iowa State University, USA, for his submission and its significant contribution to the field of language testing. Details of his dissertation, entitled 'Context visuals in L2 listening tests: The effectiveness of photographs and video vs. audio-only format' were published in *Research Notes* 39. Ruslan's supervisor was Associate Professor Volker Hegelheimer.