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.

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.

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* (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.	• 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	• 800 words approx in total • 250 words maximum per single text • 4 texts per paper
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.	• 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	• 1,600 words approx in total • 550 words maximum per single text • 5 texts per paper
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.	• 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	• 2,000 words approx in total • 700 words maximum per single text • 3 texts per paper
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.	• 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	• 3,000 words approx in total • 1,100 words maximum per single text • 6 texts per paper
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 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.	• 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)	• 3,000 words approx in total • 1,100 words maximum per single text • 9 texts per paper

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.
- 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.

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.

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 (≈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 (≈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 inference 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.

Bali, 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.