

Examining the Impact of Text Variables on EFL Test-Takers' Reading Performance: The Case of the English National Exams in Greece

Dr Trisevgeni Liontou
Greek Ministry of Education
Greece

Abstract

The aim of the present study was to investigate the effect specific text variables have on the reading comprehension performance of EFL test-takers sitting for the Greek State Certificate of English Language Proficiency national exams. The within two levels (Intermediate-B2/Advanced-C1) analysis regarding mean reading performance and textual features present in pertinent exam texts was carried out by estimating the mean task scores per text and per examination period of a total number of 152,039 B2 and 36,517 C1 test-takers that had participated in authentic English language exams over a period of 8 years (2003-2011). From the set of 135 text variables explored in the present research with regard to 34 intermediate and 29 advanced texts, four B2 text characteristics, i.e. proportion of past tenses, frequency of words occurring in the 5th band of the BNC corpus, proportion of verbs in future tenses and words expressing agreement, and five C1 text features, i.e. syntactic structure similarity across adjacent sentences and across all sentences, percentage of word units describing affective processes, relativity concerning space, time and motion as a whole and space as a specific subcategory, were found to significantly correlate with mean reading exam scores. The findings of the study provide practical guidance to EFL teachers, material developers and test designers as to the kind of lexico grammatical features learners need to become acquainted with before sitting for a specific exam and suggestions to consider when designing EFL classroom curricula, language skills textbooks and exam papers.

Key Words: text complexity, reading comprehension, exam performance, readability, automated text analysis.

1. Introduction

The aim of the present study was to empirically investigate the effect specific text variables have on reading comprehension performance with specific reference to the Greek national exams in English for the State Certificate of Language Proficiency known with their Greek acronym as KPG exams. Although a lot of research has been conducted in the field of second language acquisition with specific reference to ways of reading and text processing strategies, Alderson (2000: 104) stressed language testers' lack of success "to clearly define what sort of text a learner of a given level of language ability might be expected to be able to read or define text difficulty in terms of what level of language ability a reader must have in order to understand a particular text". Such information would be particularly useful in providing empirical justification for the kinds of reading texts test-takers sitting for various language exams are expected to process, which to date have been arrived at mainly intuitively by various exam systems (Alderson, 2000: 104; Allen *et al.*, 1988: 164; Chalhoub-Deville & Turner, 2000: 523; Oakland & Lane, 2004: 243).

2. Literature Review

Despite the considerable advances in understanding the various aspects of foreign language acquisition and reading performance, the available research has been rather unsuccessful in clearly defining and, most importantly, in prioritizing those text features that might have a direct impact on exam performance and need to be accounted for during the text selection and item design process (Carr, 2006: 282; Prins & Ulijn, 1998: 140-1; Spadorcia, 2005: 37; Wallace, 1992: 77). Especially in relation to reading tests, it has been shown that various text variables can have a significant effect on comprehension, regardless of the employed test method (Freedle & Kostin, 1999: 5; Davies & Irvine, 1996: 173; Kozminsky & Kozminsky, 2001: 187).

For example, text structure has long been an object of both first and foreign language reading research and has been repeatedly found to affect text comprehensibility (Dreher & Singer, 2001: 98-9; Freebody & Anderson, 1983: 278; Horiba, 2000: 253; McKeown *et al.*, 1992: 79; Wylie & Mc Guinness, 2004: 510). Despite all these attempts, it is yet far from straightforward which structures learners will find more challenging in test texts, since syntactic complexity has not been operationally defined (Rimmer, 2006: 498). The Common European Framework of Reference for Languages (CEFR) has described rather than defined grammatical competence, and it has distanced itself from any theoretical model or an elaboration of how a complex sentence form can be identified (Council of Europe, 2001: 113). There is no doubt that, since the CEFR as a reference work has not been designed exclusively for English but is addressed to all European member countries, it would not have been possible to include detailed descriptions for individual languages. However, the problem for testers remains, since they do not have available a robust theory-driven construct of grammar that could be operationalized in concrete testing environments (Rimmer, 2006: 506).

While not focusing on reading comprehension difficulty, many other studies on language processing have explored a variety of text variables expected to contribute to the difficulty of reading texts. For example, Gorin (2005: 351) and Kaup (2001: 960) found that sentence negations typically increase comprehension decision time in comparison with sentences with no negations. Active constructions have also been found to be easier for students to process than passive ones (Charrow, 1988: 93; Kirschner *et al.*, 1992: 546; Silver *et al.*, 1989: 170), although there was some evidence that their frequency could be important only for beginning readers (Ulijn & Strother, 1990: 38). In addition, according to Kemper (1987: 323), embedded clauses can lead to comprehension problems, especially for older adults. Relative clauses can also contribute to difficulty, because they interrupt the flow of the main sentence, causing the reader to do more mental work, especially when multiple embedded relative clauses with no explicit markers of subordination are present in a text (Akmajian *et al.*, 1995: 197; Crain & Shankweiler, 1988: 176-8; Nation, 1982: 74). Moreover, time-related accounts that are usually marked by past tenses and past-time markers tend to make texts more difficult to understand or may even distract readers from the semantic content, especially when the latter is profoundly embedded in time-related accounts or when such accounts are predominant over the semantic content (Nagabhand *et al.*, 1993: 900). Conditionals also seem to be a source of confusion for L2 learners (Kirschner *et al.*, 1992: 546; Dufty *et al.*, 2006: 1254; De Carrico, 1986: 668).

Finally, contemporary notions of lexical proficiency are premised not only on the number of words a learner knows, but also on psycholinguistic models of word knowledge and lexical networks. For example, polysemous words can be a burden to EFL users, who are often puzzled with words that they think they know, but that do not seem to make sense in a specific context (Bensoussan & Laufer, 1984: 22; Crossley *et al.*, 2010: 585; Ishii & Schmitt 2009: 9; McDonald & Shillcock, 2001: 299; Verspoor & Lowie, 2003: 568). Along with polysemy, paradigmatic relations are essential during text processing, because they allow for generalizations by means of coordination and super ordination (Chaffin & Glass, 1990: 272; Crossley *et al.*, 2009: 313; Haastrup & Henriksen, 2000: 221; Schoonen & Verhallen, 2008: 214). In addition, word concreteness has been found to influence lexical processing and overall language competence (Crossley *et al.*, 2009: 322; Crossley & McNamara, 2009: 125; Dufty *et al.*, 2006: 1253; Cacciari & Glucksberg, 1995: 291; McDonald & Shillcock, 2001: 296; Paivio, 2006: 3-4; Salsbury *et al.*, 2011: 352; Schwanenflugel *et al.*, 1997: 545). Another important feature that relates to word meaning and can affect adequate text comprehension is the presence of cognates.

As Nation (2006: 449) acknowledged, an L2 word that is a loan word or a cognate in L1 might be easier for learners to process by relating the form of the word to its meaning. Since over half of English words have derived from French, Latin or Greek (Roberts, 1965 cited in Nation, 1989: 66), maybe for Greek learners of English this could possibly mean that recognizing the meaning of parts of a cognate could make it easier for them to correctly guess the remaining part and, thus, have access to a much larger group of words. Research has also shown that, due to the unpredictable meaning of idioms, EFL users do struggle with the latter in both comprehension and production (Bulut, 2004: 106; Cieslicka, 2006: 115; Cooper, 1999: 233; Doroodi & Hashemian, 2011: 713; Saberian & Fotovatnia, 2011: 1233; Siyanova-Chanturia *et al.*, 2011: 265-6). Along with idioms, phrasal verbs, that is, any verb plus particle combination that function as a single lexical and grammatical unit, can pose additional difficulties to L2 learners, given that such verb phrases look nearly the same and the meaning of each one is not always apparent from its parts but has to be learnt by heart (Darwin & Gray, 1999: 65; Rodríguez-Puente, 2012: 72; Side, 1990: 144).

3. Aim of the Study

The aim of the present study was to add some new information to our present state of knowledge on the linguistic complexity of English texts processed by EFL users by exploring the influence of 135 text variables on comprehension within the context of the KPG language exams in English, with the ultimate purpose to identify those text variables that have a significant effect on actual exam performance across a number of examination periods.

KPG is a relatively new multi-level multilingual suite of national language examinations developed by teams of experts from the foreign language departments of the National and Kapodistrian University of Athens and the Aristotle University of Thessaloniki. The exams are administered by the Greek Ministry of Education, making use of the infrastructure available for the Panhellenic university entrance exams. Despite being in its infancy, KPG is rapidly gaining acceptance as a high-stakes exam in Greece and, because of its official recognition by the state, it can influence one's future prospects for employment and education. Exams are administered twice a year and, since November 2003, more than 300,000 test-takers have taken part in the English exams. According to the KPG specifications, the B2 level reading comprehension and language awareness paper tests candidates' ability to read for gist or for specific information long or short texts of average difficulty, and to respond to a series of tasks designed to assess their overall reading ability, their understanding of discourse and text grammar as well as their ability to make appropriate lexico grammatical choices (KPG B2 specifications¹, 2007: 3).

In addition to the above, C1 test-takers are expected to deal with more linguistically demanding texts of varied discourse, register and style, which they are likely to encounter in their social, professional or academic environment, and respond to a series of tasks designed to assess their ability to understand the overall meaning or partial meanings of these texts, to make inferences and draw conclusions, as well as understand the relationships between different parts of a single text or among various texts (KPG C1 specifications², 2007: 6-8). The level of the reading texts has been broadly defined in the Common Framework of the KPG examinations (2008: 16), according to which the B2 reading comprehension and language awareness paper is designed to test at an Independent User level the test-takers' abilities to understand the main ideas of texts of *average difficulty* on various topics, including abstract ideas or specialized information that requires some technical knowledge, whereas, at the C1 level, reading comprehension tasks are designed to test at a Proficient User level test-takers' abilities to understand texts relatively long and of a "*higher level of difficulty*".

Nevertheless, due to the absence of intelligent text processing tools, it has not yet been possible to explore the impact of more in-depth lexico-grammatical features on exam performance. By making extensive use of advanced Computational Linguistics and Machine Learning systems that rather recently became available, the current research has, thus, been designed to fill this void and further add to our present state of knowledge on EFL text difficulty in general. Moreover, given the inherent intricacy of the reading process and the fact that text difficulty cannot be estimated by a simple scale, but rather by considering a number of text and reader variables, the present research has been guided by the need to analyze the impact of several independent text variables on the reading process. In order to explore these issues, the following research question was formed:

1. Are there any statistically significant relationships between KPG test-takers' reading comprehension performance and specific lexico-grammatical features present in each set (B2, C1) of KPG English texts?

4. Methodology

Linguistic Text Analysis: The text variables analyzed in the present research were chosen for both practical and theoretical reasons. First, from the practical standpoint of comparability, it was important to establish whether particular features existed, whose presence in the KPG English language reading texts might have introduced constructive relevant variance into test scores. If this turned out to be the case, then steps could be taken to incorporate such factors into subsequent revisions of the text selection guidelines. At the same time, from a theoretical perspective, the study presented an opportunity to investigate some of the hypotheses advanced in the research literature concerning a range of variables that affect second language reading comprehension, and further explore their impact on actual exam performance.

1The full text in Greek is available online at rcel.enl.uoa.gr/kpg/docs/B2_Προδιαγραφές_Greek.pdf (Last access: 25/10/12).

2The full text in Greek is available online at rcel.enl.uoa.gr/kpg/docs/C1_Prodiagrafes_Greek.pdf (Last access: 25/10/12).

Given the inherent complexity of the reading process, it was, thus, imperative to include a comprehensive list of text features in order to minimize the risk of omitting variables that might have had a significant impact on performance. In addition, given that previous research had failed to produce a definite set of quantifiable text variables, no decision was a priori made in terms of their expected significance. In relation to text analysis a combined model based on Systemic Functional Grammar and additional text features was adopted (see Appendix 1 for the complete list of text indices). To be more specific, the presence of cohesive ties created by referencing, conjunction and lexical cohesion as well as that of nominal group structure, grammatical intricacy and lexical density was explored. Moreover, the occurrence of surface text features, such as number of words, sentences and paragraphs per text, word frequency, lexical diversity, propositional density, proportion of passive sentences, negations, phrasal verbs and idioms per text along with estimates from four well-known readability formulas, namely the Flesh Reading Ease Index, the Dale-Chall Readability Index, the Fry Readability Index and the Gunning-Fog Index, was determined.

The KPG English Reading Corpus: Thirty-five authentic B2 reading comprehension texts used between November 2003 and November 2011 examination periods and thirty C1 texts used between April 2005 and May 2012 examination periods were originally chosen for analysis. Nevertheless, at a preliminary stage of the analysis, two texts, namely the "Bridget Jones' Diary" that was used in April 2005 B2 exam and contained a total of 705 words and "Sonia Gandhi's biography" that was used in April 2005 C1 exam and contained a total of 1.032 words, were traced as outliers, due to their disproportionately high number of words, and excluded from further investigation. Thus, the main text analysis was based on 34 B2 and 29 C1 reading comprehension texts used in past KPG English language exams. For texts to be appropriate for comparisons and avoid any test-method effects only those reading passages that contained ten multiple choice reading comprehension questions with three options (A, B or C) per item were considered appropriate for further analysis. Finally, these two levels of competence were chosen for reasons of practicality since, when the research began, they were the only ones available and had attracted a great number of test-takers.

Automated Text Analysis Tools: Over the last ten years, advances in Computational Linguistics and Machine Learning systems have made it possible to go beyond surface text components and adopt more theoretically sound approaches to text readability, focusing on a wider range of "deep" text features that take into account semantic interpretation and the construction of mental models and can, thus, offer a principled means for test providers and test-takers alike to assess this aspect of test construct validity (Graesser *et al.*, 2004: 193). In the present study *Coh-Metrix 2.1*³, *Linguistic Inquiry and Word Count 2007 (LIWC)*⁴, *Vocab Profile 3.0*⁵, *Computerized Language Analysis (CLAN)* suite of programs⁶, *Computerized Propositional Idea Density Rater 3.0 (CPIDR)*⁷ and *Gramulator*⁸ were used to estimate the 135 text variables.

Reading Comprehension Mean Task Score Analysis: The relationship between exam performance and textual features within and across the two levels (B2/C1) was explored by estimating the mean task scores per text and per examination period of a total number of 152,039 B2 and 36,517 C1 test-takers that had participated in real KPG English language exams over a period of 8 years (2003-2011)⁹. These mean scores related to the specific multiple-choice reading comprehension questions included in each set of analyzed texts and provided useful information regarding internal consistency of the exam and item difficulty, whereas co relational analyses revealed significant relationships between mean reading performance and specific text variables.

3 Available online at <http://cohmetrix.memphis.edu/cohmetrixpr/index.html>

4 Available online at <http://www.liwc.net>

5 Available online at <http://www.lexutor.ca/vp>

6 Available online at <http://childes.psy.cmu.edu>.

7 Available online at <http://www.ai.uga.edu/caspr>.

8 Available online at https://umdrive.memphis.edu/pmmccrth/public/software/software_index.htm

9 The data regarding KPG test-takers' exam scores were provided by the Ministry of Education to the Director of the KPG Research Team, Prof. Bessie Dendrinou. Upon official request, the data were made available to the author of the present article, who also worked as a research assistant of the RCEl, under the specific condition that they are used confidentially and for the purpose of this research project alone.

5. Research Findings: Description & Discussion

5.1 Reading Performance & Text Features: An Across-Levels Analysis

In order to determine the contribution of the 135 text indices to mean reading performance per text but across the B2/C1 levels of competence, Pearson correlation coefficients were estimated. Data analysis showed that, when reading comprehension scores were grouped together, the mean reading performance per text of the total number of 188,556 KPG test-takers correlated weakly with five text variables, that is, Lexical Density ($r=-.266$, $p<0.05$), syntactic structure similarity across all sentences ($r=.236$, $p<0.05$), negative temporal connectives ($r=.264$, $p<0.05$) and words expressing agreement ($r=.246$, $p<0.05$) or relativity ($r=.276$, $p<0.05$).

Although weak, these correlations might indicate that, regardless their level of language competence, B2 and C1 test-takers as a whole scored lower in texts that were more lexically dense and higher when pertinent texts contained a higher proportion of syntactically similar sentences and lexical items showing agreement or relativity concerning space, time and motion relations. On the other hand, it is worth pointing out that, despite the fact that the above mentioned analysis is statistically acceptable, its practical usefulness for a criterion-referenced system such as the KPG exam battery is limited, since potential test-takers can certify their language knowledge per level of performance, i.e. B or C, and not in a common exam. In addition, treating B2 and C1 test-takers as one group could have deprived us of valuable information per level of competence, since moderate correlations at one level could have been partial from non-existing or even negative correlations at another level of language competence.

Bearing these limitations in mind, in the following two sub-sections, the contribution of text indices to mean reading performance per text but per level of competence is discussed, in order to gain a better understanding of the difficulty specific text features might have presented to either intermediate or advanced EFL language users.

5.2 B2 Mean Reading Performance & Text Features

In order to determine the contribution of text variables to mean reading performance per text and per level of competence, Pearson correlation coefficients¹⁰ were estimated. Data analysis showed a negative correlation between individual B2 test-takers' mean scores in each set of multiple-choice reading comprehension questions and two text variables, i.e. proportion of past tenses ($r=-.347$, $p<0.05$) and frequency of words occurring in the fifth band of the BNC corpus ($r=-.374$, $p<0.05$). Although weak, this effect might reflect B2 test-takers' difficulty in processing less frequent words compared to the first and second thousand more frequent ones of the BNC corpus, while it further supports the view that vocabulary knowledge in general and word frequency in particular could affect reading comprehension (Anderson & Freebody, 1981: 77; Crossley *et al.*, 2011: 563; Fukkink *et al.*, 2005: 71; Laufer, 1998: 256; Nation, 2001: 169; Zhang & Anual, 2008: 51), since less frequent words might not form part of B2 EFL users' repertoire at this stage of their language learning process.

The negative impact of past tenses on exam performance could also be treated as an indication of increased difficulty from the part of the readers, who might struggle to disentangle relationships between present and past events or follow a series of past actions (Nagabhand *et al.*, 1993: 900). On the other hand, the higher proportion of verbs in future tenses ($r=.380$, $p<0.05$) and words expressing agreement ($r=.417$, $p<0.05$) was found to correlate with mean reading performance scores in a significantly positive way, i.e. the higher the percentage of such features in B2 reading texts, the higher the percentage of correct responses to pertinent questions. This finding could be taken to suggest that both features facilitate text processing as they better define time relations and further clarify the intended meaning on behalf of the writer (by using expressions such as *agree, always, never, of course*). Thus, these variables appear to be related to text complexity in a more subtle way, through their positive contribution to the elucidation of apparently confusing messages.

In order to better define the relationship between text indices and mean reading performance scores on each text, an additional set of analyses was carried out with specific reference to two groups of performance, i.e. the *Low performance* group that included mean reading performance scores below 60% (N=15), and the *High performance* one that denoted mean reading performance scores above 60% (N=15), in each individual set of multiple-choice reading comprehension questions. This set of independent sample t-tests revealed statistically significant differences in mean reading performance between the two groups based on the contribution of five text features, i.e. words in the fifth frequency band of the BNC corpus (K5 words), negative causal connectives, stem overlap and density of pronouns, especially personal ones, in B2 reading texts.

10 Normality of distribution per variable was assessed using the One-Sample Kolmogorov-Smirnov Test ($p>.05$).

To be more specific, a significantly lower mean reading performance was noted when texts contained a higher percentage of low-frequency words ($t=2.383$, $df=28$, $r=0.40$, $p=0.024$), pronouns to noun phrases ($t=2.656$, $df=28$, $r=0.45$, $p=0.013$) and personal pronouns ($t=2.463$, $df=28$, $r=0.42$, $p=0.020$), whereas a significantly higher proportion of sentences that shared one or more word stems ($t=-2.218$, $df=28$, $r=0.39$, $p=0.035$) or contained more negative causal connectives ($t=-2.440$, $df=28$, $r=0.42$, $p=0.021$) was evident in those texts that test-takers received higher scores. These results could help us draw a comparative profile of B2 test-takers, given their levels of mean reading performance in actual exams. In other words, regarding their competence in processing written information and answering multiple-choice questions, B2 low-performers appeared to face difficulties when exposed to less frequent words or when the number of pronouns in a text made reference a more cognitively demanding process to them.

On the other hand, high performers seemed to take advantage of word stem repetition and even try to guess the meaning of unfamiliar words by making use of derivational rules and contextual information. They also appeared to better understand specific causal relations signaled by apposite conjunctions.

Due to the high variation in B2 mean reading performance scores per text across the 17 examination periods, with mean values ranging from 36% to 78% (Mean=60.05, SD=8.96, R=41.58), an additional classification of test-takers into three groups was performed. The *Low Performance* group included mean reading performance scores of up to 55% (N=11), the group of *Average Performance* referred to mean reading performance scores of up to 65% (N=6), whereas the *High Performance* one comprised all mean reading performance scores above 65% (N=13). To account for variances across the three groups, a One-Way ANOVA was carried out and significant differences in their mean reading performance were revealed with relation to six in-depth textual features, i.e. argument overlap across all sentences ($F(2, 27)=7.424$, $p=0.003$) and between adjacent ones ($F(2, 27)=3.865$, $p=0.033$), content word overlap ($F(2, 27)=4.805$, $p=0.016$), conceptual similarity between adjacent sentences ($F(2, 27)=4.368$, $p=0.023$) and percentage of word units describing cognitive processes as a whole ($F(2, 27)=3.926$, $p=0.032$) and tentative mechanisms in particular ($F(2, 27)=3.246$, $p=0.013$). More specifically, through Post-Hoc Multiple Comparisons Tests (i.e. Tukey HSD), a significant difference between *Low* and *Average Performance* groups was observed for texts that contained a lower percentage of argument overlap, not only between adjacent sentences, but also across all sentence pairs in a paragraph, as well as a lower proportion of content word overlap and conceptual similarity between adjacent sentences.

To put it simply, the variance in scores between *Low* and *Average* performers seems to reflect their different levels of ability in processing more semantically and conceptually loaded texts, with the former falling behind when such a loading of information was beyond their level of competence. Words expressing cognitive processes (i.e. *know*, *think*, *consider*) and especially those showing uncertainty (i.e. *maybe*, *perhaps*, *guess*) also appeared to have an impact on mean reading performance, with *Average* performers dealing with them in a more efficient way than *Low* ones. In relation to *High* performers, i.e. those scoring above 65% in each individual set of pertinent reading comprehension questions, statistical analysis showed a significant mean score difference with *Average* ones regarding the same six variables. This further supports the view that KPG test-takers who got higher mean scores in each set of reading comprehension questions, were able to perform more complex decoding processes across sentences that were dense with ideas, and could, also, draw sound conclusions despite the presence of tentative words.

5.3 C1 Mean Reading Performance & Text Features

In order to explore the presence of significant relationships between C1 test-takers' mean reading performance scores and pertinent linguistic features, Pearson correlation coefficients were once again estimated and five text variables, i.e. syntactic structure similarity across adjacent sentences ($r=.353$, $p<0.05$) and across all sentences ($r=.414$, $p<0.05$), percentage of word units describing affective processes ($r=-.432$, $p<0.05$), relativity concerning space, time and motion as a whole ($r=.475$, $p<0.05$) and space as a specific subcategory ($r=.434$, $p<0.05$), were identified for their strong connection with mean scores in each set of reading comprehension questions. Although medium, these correlations could be interpreted as indicative of the difficulty more in-depth textual characteristics pose to advanced EFL users, who appear to be at a disadvantage when having to interpret syntactically complex sentences or disentangle space and time relations that are poorly defined within a given context. Moreover, the finding that the higher the proportion of words describing positive and negative emotions the lower the percentage of correct responses in C1 reading comprehension questions might reflect KPG test-takers' endeavour to interpret more personal information from the part of the writer.

This attempt could have been further hindered by the increased number of adjectives used in describing such processes and the confusion that might have been caused when fluctuations between positive and negative emotions took place within the same text. Finally, contrary to the significant relationship between word frequency and mean reading comprehension performance that was observed at the B2 level, C1 test-takers seemed to have developed their lexical repertoire to a greater extent, so as to be able to cope with less frequent words present in texts of this level.

In order to better define the text features that might have accounted for significant differences in the mean reading performance of *Low* (up to 65%, N=13) and *High* (above 65%, N=16) performers, a new set of independent sample t-tests was carried out. The results of this analysis showed that texts in which test-takers got lower mean reading comprehension scores were characterized by their significantly higher percentage of propositional idea density ($t=2.281$, $df=27$, $r=0.40$, $p=0.031$), along with a significantly lower proportion of syntactic similarity across sentences ($t=-2.632$, $df=27$, $r=0.45$, $p=0.014$) and words describing relativity ($t=-2.192$, $df=27$, $r=0.39$, $p=0.037$), especially space relations ($t=-2.351$, $df=27$, $r=0.41$, $p=0.026$). This finding comes in agreement with correlations mentioned above and can help us better explain significant variations in the degree of language competence among C1 test-takers, with *High* performers being able to process more compact in ideas and complex in structure texts, as well as being more efficient in interpreting abstract cognitive processes and space restrictions.

A final set of analysis involved the identification of finer differences across three levels of mean reading performance scores per text, i.e. *Low* performers with mean reading scores below 65% (N=9), *Average* performers with mean reading scores ranging from 65% to 70% (N=9), and *High* performers with mean reading scores above 70% (N=11). The One-Way ANOVA revealed significant differences among the three groups in their mean reading performance scores per text in relation to three specific text variables, that is, proportion of positive additive connectives ($F(2, 26)=3.457$, $p=0.047$), logical operators ($F(2, 26)=3.681$, $p=0.039$) and words describing the space dimension ($F(2, 26)=5.348$, $p=0.011$). The set of Post-Hoc Multiple Comparisons Tests (i.e. Tukey HSD) showed that *High* performers appeared to be more successful than *Average* ones in processing texts with more additive connectives and logical operators, while the latter seemed more efficient in following space descriptions than *Low* performers. In other words, this finding complements the results mentioned earlier that, at the C1 level, *High* performers appeared to be more competent in handling dense texts by processing cohesive devices in a more efficient way, whereas for *Low* performers comprehension might have been impeded by their limited ability in decoding the spatial and temporal relationships between events.

6. Conclusion

From the list of 135 text variables analyzed in the present research, four B2 text characteristics, i.e. proportion of past tenses ($r=-.347$, $p<0.05$), frequency of words occurring in the fifth band of the BNC corpus ($r=-.374$, $p<0.05$), proportion of verbs in future tenses ($r=.380$, $p<0.05$) and words expressing agreement ($r=.417$, $p<0.05$), and four C1 text features, i.e. syntactic structure similarity across sentences ($r=.414$, $p<0.05$), percentage of word units describing affective processes ($r=-.432$, $p<0.05$), relativity concerning space, time and motion as a whole ($r=.475$, $p<0.05$) and space as a specific subcategory ($r=.434$, $p<0.05$), were found to significantly correlate with mean reading performance scores. Although these results are suggestive of the association between text complexity and reading performance, the lack of high correlations between mean reading comprehension scores and text indices might indicate the interference of additional factors, such as reader and item variables, in the reading process, the exact contribution of which warrants further investigation.

It should also be noted that, for both levels, discrimination of reading comprehension competence among *Low*, *Average*, and *High* performers was achieved by assessing their ability to cope with more in-depth lexical, semantic, and syntactic features. At the same time, given their weak to mediocre impact on exam performance, some of the linguistic factors that might worth EFL teachers' special attention are past tenses, syntactic complexity, and space and time relations. More specifically, the negative impact of past tenses on B2 exam performance could be an indication of the difficulty such tenses present to intermediate EFL users, who might struggle to disentangle relationships between present and past events and could benefit from extra practice on these grammatical features. Additional parts of the lexicon that appeared to present difficulties to B2 test-takers and might worth some extra practice are phrasal verbs, negations, embedded questions, and logical operators.

On the other hand, advanced EFL users might need greater exposure to and training on better defining space and time relations and processing syntactically complex sentences, since these two specific language aspects were found to negatively correlate with exam performance. Finally, although the relationship between specific text features and mean task scores in pertinent reading comprehension questions was examined, it did not fall within the scope of the present research to include a more detailed analysis of individual task characteristics and their possible effect on actual performance. In other words, the correlation between lexico grammatical complexity and average reading comprehension performance per text was investigated, but a more detailed task and item analysis should form part of future research.

References

- Akmajian, A., Demers, R., Farmer, A. & Harnish, R. (1995). *Linguistics: An introduction to language and communication*(4th edition). Cambridge, MA: The MIT Press.
- Alderson, C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Allen, D., Bernhardt, B, Berry, T. & Demel, M. (1988). Comprehension and text genre: an analysis of secondary school foreign language readers. *The Modern Language Journal*, 72, 2, 163-172.
- Anderson, R. & Freebody, P. (1983). Vocabulary Knowledge. In R. Ruddell, M. Ruddell & H. Singer (Eds.), *Theoretical Models and Processes of Reading* (pp. 343-371). Newark: International Reading Association.
- Bensoussan, M. & Laufer, B. (1984). Lexical Guessing in Context in EFL Reading Comprehension. *Journal of Research in Reading*, 7, 1, 15-32.
- Bulut, T. (2004). Idiom Processing In L2: Through Rose-Colored Glasses. *The Reading Matrix*, 4, 2, 105-116.
- Cacciari, C. & Glucksberg, S. (1995). Understanding idioms: do visual images reflect figurative meanings? *European Journal of Cognitive Psychology* 7, 3, 283-305.
- Carr, N. (2006). The factor structure of test task characteristics and examinee performance. *Language Testing*, 23, 3, 269-289.
- Chaffin, R. & Glass, A. (1990). A comparison of hyponym and synonym decisions. *Journal of Psycholinguistic Research*, 19, 4, 265-280.
- Chalhoub-Deville, M. & Turner, C. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS and TOEFL. *System*, 28, 4, 523-539.
- Charrow, V. (1988). Readability vs. comprehensibility: a case study in improving a real document. In A. Davison & G. Green (Eds.), *Linguistic complexity and text comprehension* (pp. 85-114). Hillsdale, New Jersey: Lawrence Erlbaum.
- Cieslicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, 22, 2, 115-144.
- Cooper, T. (1999). Processing of idioms by L2 learners of English. *TESOL Quarterly*, 33, 2, 233-262.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Crain, S. & Shankweiler, D. (1988). Syntactic complexity and reading acquisition. In A. Davison & G. Green (Eds.), *Linguistic complexity and text comprehension* (pp. 167-192). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Crossley, S. & McNamara, D. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18, 2, 119-135.
- Crossley, S., Greenfield, J. & McNamara, D. (2008). Assessing Text Readability Using Cognitively Based Indices. *TESOL Quarterly*, 42, 3, 475-492.
- Crossley, S., Salsbury, T. & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, 59, 2, 307-334.
- Crossley, S., Louwse, M., McCarthy, P. & McNamara, D. (2007) A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal*, 91, 1, 15-30.
- Crossley, S., Salsbury, T. & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, 60, 3, 573-605.
- Crossley, S., Salsbury, T., McNamara, D. & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28, 4, 561-580.
- Darwin, C. & Gray, L. (1999). Going after the Phrasal Verb: An Alternative Approach to Classification. *TESOL Quarterly*, 33, 1, 65-83.

- Davies, A. & Irvine, A. (1996). Comparing test difficulty and text readability in the evaluation of an extensive reading programme. In M. Milanovic & N. Saville (Eds.), *Performance Testing, Cognition and Assessment: Selected Papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem* (pp. 165-183). Cambridge: Cambridge University Press.
- DeCarrico, J. (1986). Tense, aspect, and time in the English modality system. *TESOL Quarterly*, 20, 4, 665-685.
- Doroodi, S. & Hashemian, M. (2011). The Relationship between Reading Comprehension and Figurative Competence in L2 Learners. *Theory and Practice in Language Studies*, 1, 6, 711-717.
- Dreher, M. & Singer, H. (2001). Friendly Texts and Text-Friendly Teachers. *Theory into Practice*, 28, 2, 98-104.
- Dufty, D., Graesser, A., Louwerse, M. & McNamara, D. (2006). Assigning Grade Levels to Textbooks: Is it just Readability? In R. Son (Ed.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 1251-1256). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Freebody, P. & Anderson, R. (1983). Effects of vocabulary difficulty, text cohesion and schema availability on reading comprehension. *Reading Research Quarterly*, 18, 3, 277-294.
- Freedle, R. & Kostin, I. (1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, 16, 1, 2-32.
- Fukkink, R., Hulstijn, J. & Simis, A. (2005). Does training in second-language word recognition skills affect reading comprehension? An experimental study. *The Modern Language Journal*, 89, 1, 54-75.
- Gorin, J. (2005). Manipulating processing difficulty of reading comprehension questions: the feasibility of verbal item generation. *Journal of Educational Measurement*, 42, 4, 351-373.
- Graesser, A., McNamara, D., Louwerse, M. & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers*, 36, 2, 193-202.
- Haastrup, K. & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics*, 10, 2, 221-240.
- Horiba, Y. (2000). Reader Control in Reading: Effects of Language Competence, Text Type, and Task. *Discourse Processes*, 29, 3, 223-267.
- Ishii, T. & Schmitt, N. (2009). Developing an Integrated Diagnostic Test of Vocabulary Size and Depth. *RELC Journal*, 40, 1, 5-22.
- Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Memory & Cognition*, 29, 7, 960-967.
- Kemper, S. (1987). Life-span changes in syntactic complexity. *Journal of Gerontology*, 42, 3, 323-328.
- Kirschner, M., Wexler, C. & Spector-Cohen, E. (1992). Avoiding Obstacles to Student Comprehension of Test Questions. *TESOL Quarterly*, 26, 3, 537-556.
- Kozminsky, E. & Kozminsky, L. (2001). How do general knowledge and reading strategies ability relate to reading comprehension of high school students at different educational levels? *Journal of Research in Reading*, 24, 2, 187-204.
- KPG B2 Level Exam Specifications (2007). The full text in Greek is available online at: rcel.enl.uoa.gr/kpg/docs/B2_Prodiagrafes_Greek.pdf (Last access: 23/07/13).
- KPG C1 Level Exam Specifications (2007). The full text in Greek is available online at: rcel.enl.uoa.gr/kpg/docs/C1_Prodiagrafes_Greek.pdf (Last access: 23/07/13).
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19, 2, 255-271.
- McDonald, S. & Shillcock, R. (2001). Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing. *Language and Speech*, 44, 3, 295-323.
- McKeown, M., Beck, I., Sinatra, G. & Loxterman, J. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly*, 27, 1, 79-93.
- McNamara, D., Cai, Z. & Louwerse, M. (2011). Optimizing LSA measures of cohesion. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 379-400). New York: Routledge.
- McNamara, D., Louwerse, M., McCarthy, P. & Graesser, A. (2010). Coh-Metrix: Capturing Linguistic Features of Cohesion. *Discourse Processes*, 47, 4, 292-330.

- Min, H. & McCarthy, P. (2010). Identifying Varietals in the Discourse of American and Korean Scientists: A Contrastive Corpus Analysis Using the Gramulator. In H. Guesgen & C. Murray (Eds.), *Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference* (pp. 247-252). Menlo Park, California: The AAAI Press.
- Nagabhand, S., Nation, P. & Franken, M. (1993). Can Text be too Friendly? *Reading in a Foreign Language*, 9, 2, 895-907.
- Nation, P. (1982). Getting information from advanced reading tests. In J. Heaton (Ed.), *Language Testing* (pp. 73-76). London: Modern English Publications.
- Nation, P. (1989). Dictionaries and language learning. In M. Tickoo (Ed.), *Learners' Dictionaries: State of the Art*(pp. 65-71). Singapore: Southeast Asian Ministers of Education Organization.
- Nation, P. (2001c). How many high-frequency words are there in English? In M. Gill, A. Johnson, L. Koski, R. Sell & B. Wårvik (Eds.), *Language, Learning and Literature: Studies Presented to Håkan Ringbom* English Department Publications 4 (pp. 167-181). Åbo Akademi University: Åbo.
- Nation, P. (2006). Second language vocabulary. In K. Brown (Ed.), *Encyclopaedia of Language & Linguistics* (pp. 448-454).Oxford: Elsevier.
- Oakland, T. & Lane, H. (2004). Language, Reading, and Readability Formulas: Implications for Developing and Adapting Tests. *International Journal of Testing*, 4, 3, 239-252.
- Paivio, A. (2006). Dual Coding Theory and Education.Draft chapter for the conference on *Pathways to Literacy Achievement for High Poverty Children* (pp. 1-19). The University of Michigan School of Education, September 29-October 1.
- Prins, E. & Ulijn J. (1998). Linguistic and cultural factors in the readability of mathematics texts: the Whorfian hypothesis revisited with evidence from the South African context. *Journal of Research in Reading*, 21, 2, 139-159.
- Rimmer, W. (2006). Measuring grammatical complexity: the Gordian knot. *Language Testing*, 23, 4, 497-519.
- Rodríguez-Puente, P. (2012): The Development of Non-compositional Meanings in Phrasal Verbs: A Corpus-based Study. *English Studies*, 93, 1, 71-90.
- Saberian, N. & Fotovatnia, Z. (2011). Idiom Taxonomies and Idiom Comprehension: Implications for English Teachers.*Theory and Practice in Language Studies*, 1, 9, 1231-1235.
- Salsbury, T., Crossley, S. & McNamara, D. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, 27, 3, 343-360.
- Schoonen, R. & Verhallen, M. (2008).The assessment of deep word knowledge in young first and second language learners. *Language Testing*, 25, 2, 211-236.
- Schwanenflugel, P., Stahl, S. & McFalls, E. (1997). Partial Word Knowledge and Vocabulary Growth during Reading Comprehension. *Journal of Literacy Research*, 29, 4, 531-553.
- Side, R. (1990). Phrasal verbs: Sorting them out. *English Language Teaching Journal*, 44, 2, 144-152.
- Silver, N., Phelps, G. & Dunlap, W. (1989). Baddeley's grammatical reasoning test: Active versus passive processing differences re- examined. *Language Testing*, 6, 2, 164-171.
- Siyanova-Chanturia, A., Conklin, K. & Schmitt (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers.*Second Language Research*, 27, 2, 251-272.
- Spadorcia, S. (2005). Examining the text demands of high-interest, low-level books. *Reading & Writing Quarterly*, 21, 1, 33-59.
- Verspoor, M. & Lowie, W. (2003). Making sense of polysemous words. *Language Learning*, 53, 3, 547-586.
- Wallace, C. (1992). *Reading*. Oxford: Oxford University Press.
- Wylie, J. & McGuinness C. (2004). The interactive effects of prior knowledge and text structure on memory for cognitive psychology texts. *British Journal of Educational Psychology*, 74, 4, 497-514.
- Zhang, L. & Anual, S. (2008). The Role of Vocabulary in Reading Comprehension: The Case of Secondary School Students Learning English in Singapore. *RELC Journal*, 39, 1, 51-76.

Appendix 1

Text Variables List Basic Text Information

- V1. No. of words in text (tokens)
- V2. Syllables per word
- V3. Words per sentence
- V4. Average number of characters per word
- V5. Average number of syllables per 100 words
- V6. No. of sentences (T-Units)
- V7. Sentences per paragraph
- V8. Average number of Sentences per 100 words
- V9. Text Genre (Narrative ≠ Expository)
- V10. Sum 1 to 6-Letter Words
- V11. Sum 7 to 14-Letter Words
- V12. 1-Letter Word
- V13. 2-Letter Word
- V14. 3-Letter Word
- V15. 4-Letter Word
- V16. 5-Letter Word
- V17. 6-Letter Word
- V18. 7-Letter Word
- V19. 8-Letter Word
- V20. 9-Letter Word
- V21. 10-Letter Word
- V22. 11-Letter Word
- V23. 12-Letter Word
- V24. 13-Letter Word
- V25. 14-Letter Word

Word Frequency Indices

- V26. K1 Words (1-1000)
- V27. K2 Words (1001-2000)
- V28. K3 Words (2001-3000)
- V29. K4 Words (3001-4000)
- V30. K5 Words (4001-5000)
- V31. Academic Word List
- V32. Anglo-Sax Index
- V33. Greco-Latin Cognates
- V34. Log freq. content words
- V35. Log min. freq. content words

Readability Indices

- V36. Flesch Reading Ease Index
- V37. Flesch-Kincaid Grade Level
- V38. Gunning-Fog Index
- V39. Fry Readability Graph
- V40. Dale-Chall Grade Level Index

Lexical Richness Indices

- V41. vocd-D (Lexical Diversity)
- V42. MTLD (Lexical Diversity)
- V43. MAAS (Lexical Diversity)
- V44. Apax Legomena
- V45. Dis Legomena
- V46. Entropy

- V47. Relative Entropy
- V48. Lexical Density (content words/total)
- V49. Types per Text
- V50. Tokens per Type
- V51. Families per Text
- V52. Tokens per Family
- V53. Types per Family
- V54. Verb Density
- V55. Adverb Density

Text Abstractness Indices

- V56. Concreteness content words
- V57. Min. concreteness content words
- V58. Noun hypernym
- V59. Verb hypernym

Syntactic Complexity Indices

- V60. Higher level constituents
- V61. Noun Phrase incidence
- V62. Modifiers per Noun Phrase
- V63. Words before main verb
- V64. Negations
- V65. Passive sentences
- V66. Syntactic structure similarity (adjacent sentences)
- V67. Syntactic structure similarity (across paragraphs)
- V68. Syntactic structure similarity (within paragraphs)
- V69. Conditional operators
- V70. All connectives
- V71. Pos. additive connectives
- V72. Pos. temporal connectives
- V73. Pos. causal connectives
- V74. Pos. logical connectives
- V75. Neg. additive connectives
- V76. Neg. temporal connectives
- V77. Neg. causal connectives
- V78. Neg. logical connectives

Cohesion & Coherence Indices

- V79. Causal cohesion
- V80. Causal content
- V81. Intentional content
- V82. Temporal cohesion
- V83. Spatial cohesion
- V84. Logical operators

Referential & Semantic Indices

- V85. Anaphoric reference
- V86. Adjacent anaphoric reference
- V87. Argument overlap
- V88. Adjacent argument overlap
- V89. Stem overlap
- V90. Adjacent stem overlap
- V91. Content word overlap
- V92. Pronoun ratio
- V93. Personal pronouns

- V94. Impersonal pronouns
- V95. Propositional Idea Density
- V96. LSA for adjacent sentences
- V97. LSA for all sentences
- V98. LSA for all paragraphs

Psycholinguistic Processes

- V99. Affective Processes (Total)
- V100. Positive Emotions (Total)
- V101. Negative Emotions (Total)
- V102. Cognitive Mechanisms (Total)
- V103. Insight
- V104. Cause
- V105. Discrepancy
- V106. Tentative Words
- V107. Certainty
- V108. Agreement (Assent)
- V109. Inhibition
- V110. Inclusive Words
- V111. Exclusive Words
- V112. Perception (Total)
- V113. See