



## **Re-examining text difficulty through automated textual analysis tools and readers' beliefs: the case of the Greek State Certificate of English Language Proficiency exam**

**[Επανεξετάζοντας τη δυσκολία αναγνωσιμότητας κειμένων με τη βοήθεια αυτοματοποιημένων συστημάτων ανάλυσης κειμένου και των απόψεων των αναγνωστών: η περίπτωση των εξετάσεων Αγγλικής γλώσσας του Κρατικού Πιστοποιητικού Γλωσσομάθειας]**

**Jenny Liontou**

*This article reports on an exploratory study that aimed at describing and comparing a range of linguistic features that characterize the reading texts used at the B2 and C1 level of the Greek State Certificate of English Language Proficiency exam (KPG<sup>1</sup>). Its ultimate purpose was to explore the contribution of such features to perceived text difficulty while at the same time examining the relationship between strategy use and test-takers' perceived level of reading comprehension difficulty reported in 7,250 questionnaires. Text analysis revealed significant differences between B2 and C1 reading texts for a specific number of text features such as word, paragraph and text length, readability indices, levels of word frequency and presence of words with rich conceptual content. A significant correlation was also found between B2 test-takers' perception of reading module difficulty and specific text features i.e. lexical diversity, abstract words, positive additive connectives and anaphoric references between adjacent sentences. With regard to C1 test-takers, data analysis showed that two specific text variables i.e. positive logical connectives and argument overlap, correlated significantly with readers' perception of reading module difficulty. Finally, problem-solving reading strategies such as rereading the text, guessing the meaning of unknown words and translating in mother tongue were found to correlate significantly with perceived text difficulty, whereas support-type reading strategies such as underlining or selectively reading parts of the text were less often employed regardless KPG test-takers' perception of text difficulty. The findings of this study could help both EFL teachers and test designers gain valuable knowledge regarding EFL learners' reading habits and also become more alert to the difficulty specific text features impose on the latter.*

∞

*Στο παρόν άρθρο παρουσιάζονται τα αποτελέσματα μιας μακροχρόνιας έρευνας που είχε ως στόχο τη λεπτομερή περιγραφή και σύγκριση υφομετρικών μεταβλητών των κειμένων κατανόησης γραπτού λόγου που έχουν χρησιμοποιηθεί στα επίπεδα B2 και Γ1 των εξετάσεων Αγγλικής γλώσσας του Κρατικού Πιστοποιητικού Γλωσσομάθειας (ΚΠγ).*

Απώτερος στόχος της παρούσας έρευνας υπήρξε η διερεύνηση της επίδρασης κειμενικών μεταβλητών στο βαθμό δυσκολίας αναγνωσιμότητας των συγκεκριμένων κειμένων. Παράλληλα εξετάστηκε η χρήση από τους εξεταζόμενους συγκεκριμένων στρατηγικών κατανόησης κειμένου και η σχέση των τελευταίων με το βαθμό δυσκολίας κατανόησης αγγλικών κειμένων σύμφωνα με τις απόψεις των συμμετεχόντων στις συγκεκριμένες εξετάσεις πιστοποίησης της γλωσσομάθειας όπως αυτές εκφράστηκαν σε 7.250 ερωτηματολόγια. Από την ανάλυση προέκυψαν στατιστικά σημαντικές διαφορές ανάμεσα στα κείμενα επιπέδου B2 και Γ1 όσον αφορά σε συγκεκριμένες υφομετρικές μεταβλητές σε επίπεδο λέξης, πρότασης και παραγράφου καθώς και χαρακτηριστικών λεξιλογικής πλουσιότητας και πυκνότητας. Επιπλέον, παρατηρήθηκε η ύπαρξη στατιστικά σημαντικής συσχέτισης ανάμεσα στις απόψεις των εξεταζόμενων σχετικά με τη δυσκολία αναγνωσιμότητας κειμένων και τη χρήση συγκεκριμένων στρατηγικών ανάγνωσης όπως η ανάγνωση κειμένου περισσότερες από μια φορές, η προσπάθεια υποθετικής ερμηνείας των άγνωστων λέξεων αλλά και η μετάφραση στη μητρική γλώσσα για καλύτερη κατανόηση του κειμένου. Η χρήση «υποστηρικτικών» στρατηγικών ανάγνωσης όπως η υπογράμμιση ή η επιλεκτική ανάγνωση σημείων του κειμένου υπήρξε περιορισμένη και μη σχετιζόμενη με τις απόψεις των υποψηφίων ΚΠγ για το βαθμό δυσκολίας κατανόησης συγκεκριμένων αγγλικών κειμένων. Τα αποτελέσματα της παρούσας έρευνας θα μπορούσαν να προσφέρουν στους καθηγητές της Αγγλικής ως ξένη γλώσσα αλλά και τους συγγραφείς θεμάτων για εξετάσεις γλωσσομάθειας σημαντικές πληροφορίες σχετικά με τις στρατηγικές ανάγνωσης αγγλικών κειμένων και τα υφομετρικά χαρακτηριστικά αυτών που εντείνουν τη δυσκολία κατανόησης για συγκεκριμένους αναγνώστες.

**Keywords:** text difficulty, readability, lexical complexity, reading strategies, test-taking strategies

---

## Introduction

Over the last sixty years readability or text difficulty has been an area of concern for all those who need to establish the appropriacy of a given text for any pedagogic purpose (Brabham and Villaume, 2002; Fry, 2002; Hatcher, 2000; Mesmer, 2005). It has also been recognized that providing test-takers with texts that are too difficult or too easy to process can affect their performance and contaminate test results (Bachman, 2002; Carr, 2006; Bailin and Grafstein, 2001; Kozminsky and Kozminsky, 2001). A range of reader factors that affect the reading process have also been recognized including motivation, background knowledge and previous reading experience (Brantmeier, 2005; Drucker, 2003; Keshavarz *et al.*, 2007; Krekeler, 2006; Rupp *et al.*, 2006; Salataci and Akyel, 2002). However, as these variables are essentially beyond the control of the researcher, it is facets of the text that have received the most attention (Leong *et al.*, 2002; Parker *et al.*, 2001). Particularly in relation to reading strategies<sup>2</sup> a number of studies have used introspective methods (e.g. Anderson *et al.*, 1991; Cohen & Upton, 2007; Nikolov, 2006; Storey, 1997) or questionnaires (e.g. Brand-Gruwel *et al.*, 1998; Nevo, 1989; Phakiti: 2003a, 2003b; Purpura, 1997; Zimmerman & Pons, 1986) to examine the processes employed by test-takers and provide evidence for the construct validity of an examination. However, given the lack of sufficient research evidence regarding the relationship between text difficulty and employed strategies, more validation studies in this area are needed for both test designers and test-takers to become aware of the nature of such processes and their contribution to exam performance.

## Literature review

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) stresses 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". Fulcher (1997) also draws on the importance of text difficulty or text accessibility as a crucial but much neglected area in language testing. For him test developers need to be aware of the range of factors that make texts more or less accessible in order to be able to select reading texts at appropriate levels for inclusion into the reading sub-tests of their examinations. He further points out that research in this area is particularly pertinent because text difficulty is re-emerging as an area of great concern not only in language teaching and materials writing but also in the testing community.

Despite the considerable advances that have been made in exploring and understanding the various aspects of foreign language acquisition and reading performance, the available research has been rather unsuccessful in clearly defining those text features that have a direct impact on text complexity (Davies and Irvine, 1996; Dale and Chall, 1995). Consequently, for reasons of practicality, many researchers are still resorting to readability formulae or their own experience for assigning reading levels to texts (Juan, 2006; Ko, 2005; Kobayashi, 2002; Oakland and Lane, 2004; Trites and McGroarty, 2005). However, many researchers have pointed to the serious limitations of readability formulae and stressed the need for a more in-depth analysis of text features (Lehner, 1993; Shokrpour, 2004; Spadorcia, 2005). Weir (2005) further acknowledges that although the Common European Framework of Reference for Languages (CEFR) attempts to describe language proficiency through a group of scales composed of ascending level descriptors, it fails to provide specific guidance as to the topics that might be more or less suitable at any level of language ability or define text difficulty in terms of text length, content, lexical and syntactic complexity. Alderson *et al.* (2004: 13) also stress that many of the terms in the CEFR remain undefined. They further argue that "difficulties arise in interpreting it [i.e. the CEFR] because it does not contain any guidance, even at a general level, of what might be simple in terms of structures, lexis or any other linguistic level".

Considering all the above, in the present research it is assumed that a better measure of text complexity can be achieved by using systemic functional grammar as a basis for the text analysis, since such a model of language can help analyze prose in a manner that classifies ideas according to their role in conveying the total meaning of the passage and further show how relationships among ideas account for the overall coherence of a text (Freebody and Anderson, 1983). To this end, the framework for the analysis of text cohesion in this paper has been largely based on Halliday & Hasan's work (1976) and further supplemented with features proposed by Kintsch and van Dijk (1978). At this point, it is worth mentioning that Halliday & Hasan's model of text cohesion has been applied by a number of researchers in general discourse analysis, but its applicability and usability in examining foreign language reading comprehension text difficulty remains greatly unexplored.

## Aim and methodology

The aim of the research presented in this paper has been twofold: a) to delineate and compare a range of linguistic features that characterize the reading texts used at the B2 (Independent User) and C1 (Proficient User) level of the Greek State Certificate of English Language Proficiency in order to explore their contribution to text difficulty and b) to examine whether specific reading and test-taking strategies are related to text complexity or

test-takers' perceptions of text difficulty. The stimulus for such an investigation has been the need for empirical evidence to supplement the to date mainly intuitive selection by item writers<sup>3</sup> of reading texts to be used at the B2 and C1 reading sub-tests of the specific exam. At this point it should be emphasized that the lack of research based evidence on the way levels of difficulty are assigned to test texts does not exclusively characterize KPG exams, but is a rather common feature among various well-established and long-administered exam systems that fail to provide sound evidence of their text selection processes (Bachman *et al.*, 1988; Chalhoub-Deville and Turner, 2000; Fulcher, 2000). These two levels were also 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.

Although it is beyond the scope of the present paper to provide a detailed description of the KPG English Language exam it is worth mentioning that each test consists of four sub-tests designed to assess the following competencies: a. reading comprehension and language awareness, b. writing and mediation, c. listening comprehension and d. speaking. According to the KPG specifications, at the B2 level of the reading comprehension and language awareness subtest candidates are required to skim through, scan or read closely longer or shorter texts of average difficulty and respond to a total of 75 items (reduced to 50 since November 2007) of various types designed to assess their overall reading skills, their knowledge of discourse and text grammar as well as their ability to make appropriate lexicogrammatical choices (KPG B2 specifications<sup>4</sup>, 2003). At the C1 level candidates are asked to skim through or read carefully longer and 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 total of 75 items of various types (reduced to 60 since November 2007) designed to assess their ability to understand the overall meaning or partial meanings of these texts, to make reasoned inferences and draw conclusions as well as understand the relationships between different parts of a single text or among various texts (KPG C1 specifications<sup>4</sup>, 2005). The level of the reading texts has been broadly defined in the Common Framework of the KPG examinations, applicable to all languages<sup>5</sup>, according to which "the B2 Level exams are designed to test at an Independent User level the candidates' abilities to use English in order 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 "the C1 Level exams are designed to test at a Proficient User level the candidates' abilities to understand texts relatively long and of a *high level of difficulty*" (Common Framework of the KPG examinations, 2003, p.6). However, it has not yet been possible to define, based on empirical evidence, the readability level of texts and the specific lexicogrammatical features that could be more appropriate to the intended audience i.e. prospective B2 or C1 test-takers. 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. In order to explore these issues, the following research questions have been formed:

1. Are there any statistically significant differences between the B2 and C1 test texts with regard to specific lexicogrammatical features?
2. Is there a relationship between test-takers' impressions of text difficulty and the specific lexicogrammatical features estimated for each set (B2, C1) of KPG texts?
3. What reading and test-taking strategies do B2 and C1 KPG test-takers report using when responding to a multiple-choice reading comprehension test?
4. Is there a relationship between reported strategies and test-takers' perceived level of text difficulty?

In relation to text analysis, a range of lexicogrammatical features was automatically measured through computer programs such as Coh-Metrix 2.0, Web VocabProfile 3.0, AceReader Pro Deluxe, TextAnalyzer 2.0 and CLAN. To be more specific, Coh-Metrix 2.0 was

used to measure syntactic complexity and the frequency of particular syntactic classes along with text abstractness and conceptual similarity across sentences and paragraphs of the same text (see variables V1-V10, V24-V28, V31-V69 in Table 1). In addition, more surface text features that have been reported to contribute to text difficulty at a word and sentence level were estimated using Web VocabProfile 3.0 (see variables V17-V22, V29-V30 in Table 1), AceReader Pro Deluxe (see variables V11-V13, V16 in Table 1), TextAnalyzer 2.0 (see variables V14-V15 in Table 1) and CLAN (see variable V23 in Table 1). All in all, 24 B2 reading texts used between November 2003 to November 2010 examination periods and 24 C1 texts used between April 2005 to November 2010 examination periods were analyzed with regard to sixty-nine text variables. For the texts to be appropriate for comparisons a specific set of criteria was followed during the selection process i.e. only those reading comprehension texts that contained ten multiple choice questions with three options (A, B or C) per item were considered appropriate for further analysis.

In order to collect information regarding KPG test-takers' profile and reading strategies as well as their perceptions of text and task difficulty, a paper-and-pencil survey was conducted by the Research Centre for English Language Teaching Learning and Assessment (RCEL) at the University of Athens in cooperation with the Greek Ministry of Education. Taking into consideration test-takers' feedback -and in a way treating them as "judges" of reading texts- is a crucial part of the research, since they were the actual readers who had to interact with the texts in order to perform a set of tasks and demonstrate a successful performance in the specific exams. So, their feelings and opinions about the texts are believed to be of great importance since they can give us more in-depth information about text difficulty from the reader's point of view. To this end a questionnaire was administered on the day of the exam to all KPG test-takers sitting for the B2 and C1 English language exams in the May and November 2006 and 2007 and May 2008 examination periods<sup>6</sup>. To date, 7,250 questionnaires from five examination periods have been analyzed of which 4,750 referred to the B2 level and 2,500 to the C1 level. For the survey sample to be appropriate for statistical analysis and as representative as possible of the target population, a decision was made for its size to be at least 10% of the total number of test-takers taking each exam, with a minimum of 500 participants per examination period. In order to ensure a balanced geographical distribution and avoid any variation in response rates due to urbanization, 50% of the participants was randomly drawn from the five most densely populated<sup>7</sup> cities in Greece i.e. Athens, Thessaloniki, Patras, Larisa and Herakleion, Crete, 25% from rural areas and the remaining 25% from the Greek islands. Once test-takers finished their exam and before leaving the examination room, they were kindly requested to fill in a questionnaire and anonymously express on a five-point Likert scale their agreement or disagreement with a variety of statements regarding text and task difficulty, topic and genre familiarity, topic preference and text lexical complexity. Apart from the Likert scaling system, respondents also answered dichotomous questions in order to provide more general information regarding their reading and test-taking strategies.

## Results and discussion

Once the analysis of text characteristics per level was completed, independent sample t-tests<sup>8</sup> were carried out in order to explore and further determine the significance of existing differences between the B2 and C1 level texts and thus answer the first research question (*Are there any statistically significant differences between the B2 and C1 test texts with regard to specific lexicogrammatical features?*). As can be seen in Table 1, among the range of analyzed text features, significant differences were found for a specific number of text variables including word (V2-V5), paragraph (V8) and text length (V1), readability indices (V11-V16), levels of word frequency (V17, V25) and Latent Semantic Analysis<sup>9</sup> cosines for adjacent sentences (V67) and paragraph-to-paragraph units (V69). In other words, texts used

at the C1 level included a significantly ( $p < 0.05$ ) higher number of words in text (tokens) than their B2 counterparts (V1). They also contained a significantly higher number of unique words (V2), which could have made comprehension comparatively more difficult since more content words would need to be decoded and integrated within the same discourse context. Moreover, C1 texts were characterized by significantly longer words in terms of syllables per word (V3), average number of characters per word (V4) and average number of syllables per 100 words (V5) as well as a significantly higher number of words per sentence (V6) and sentences per paragraph (V8), all of which could have contributed to overall text difficulty. It is notable that despite the serious limitations of readability formulae, in the present research a significant difference was revealed between B2 and C1 texts in relation to the six employed readability indices i.e. the Flesch Reading Ease Index (V11) and the Flesch-Kincaid Grade Level Index (V12), the Dale-Chall Grade Level Index (V13), the Spache Grade Level (V14), the Gunning's Fog Index (V15) and the Fry Readability Graph (V16), all of which rated B2 texts as less difficult than those used at the C1 level. This finding supports the view that despite their apparent simplicity, readability formulae do seem to come in some agreement with KPG test-designers' perception of text difficulty and could be of practical usefulness to them during the text selection and validation process. Finally, C1 texts were found to consist of a significantly lower percentage of the one thousand most frequent English words of the BNC frequency list (V17), a feature that could lead to increased text difficulty. Most importantly, the fact that C1 texts were characterized by significantly less conceptually similar sentences (V67) and paragraphs (V69) could have made the comprehension process more demanding since readers had to process and decode a higher number of ideas. Against our expectations no significant differences were found regarding additional text features especially in relation to syntactic and semantic complexity and text abstractness (V23-V66). This may be taken to suggest that more explicit text differences across levels could be drawn should test designers become more alert to such features as verb and noun hypernym levels (V33-V34), word concreteness (V26-V27), anaphoric reference (V54-V55) and content word overlap (V66) and take them into account during the text selection process.

	B2 N=24		C1 N=24		t (48)	Adj. sig. (2-tailed)
	Mean	SD	Mean	SD		
V1. No. of words in text	417.75	111.273	590.42	148.949	-4.550	.000
V 2. No. of different words in text (types)	223.96	49.074	298.58	55.013	-4.959	.000
V3. Syllables per word	1.539	.093	1.611	.084	-2.804	.007
V 4. Average number of characters per word	4.700	.2309	4.913	.2346	-3.091	.003
V 5 Average number of syllables per 100 words	150.986	8.106	158.917	9.106	-3.187	.003
V 6. Words per sentence	18.102	3.652	20.607	6.300	-1.685	.099
V7. No. of sentences	23.92	9.325	29.08	11.225	-1.735	.090
V8. Sentences per paragraph	3.346	1.530	5.704	3.404	-3.095	.004
V9. Average number of sentences per 100 words	5.988	1.666	5.071	1.105	2.245	.030
V10. No. of paragraphs	8.17	3.559	7.83	3.130	.345	.732
V 11. Flesch Reading Ease	58.205	8.969	48.530	9.133	3.703	.001
V 12. Flesch-Kincaid Grade Level	9.574	1.750	11.143	1.458	-3.374	.002
V 13. Dale-Chall Grade Level	9.521	1.913	11.750	2.658	-3.334	.002
V 14. Spache Grade Level	4.829	.700	5.488	.896	-2.835	.007
V 15. Gunning's Fog Index	10.241	2.314	13.390	2.743	-4.297	.000
V 16. Fry Readability Graph	9.17	1.949	11.38	2.163	-3.716	.001
V 17. K1 Words (1-1000)	79.781	5.217	74.573	4.857	3.579	.001
V 18. K2 Words (1001-2000)	8.024	2.257	8.817	1.880	-1.323	.192
V 19. K3 Words (2001-3000)	2.902	1.576	3.930	1.764	-2.127	.039
V 20. K4 Words (3001-4000)	1.986	1.423	2.437	.907	-1.307	.198

V 21. K5 Words (4001-5000)	.960	.701	1.137	.675	-.891	.378
V 22. Academic Words	4.228	2.488	4.894	1.688	-1.084	.284
V 23. Lexical Diversity	107.401	22.468	103.511	20.243	.630	.532
V 24. Log freq. content words	2.253	.154	2.143	.145	2.537	<b>.015</b>
V 25. Log min. freq. content words	1.162	.349	1.109	.298	.571	.571
V 26. Concreteness content words	376.891	23.404	386.434	16.312	-1.639	.108
V 27. Min. concreteness content words	177.63	19.763	174.71	15.909	.563	.576
V 28. Lexical Density	.552	.042	.554	.030	-.233	.817
V 29. Anglo-Sax Index:	73.410	4.948	72.812	5.389	.401	.691
V 30. Greco-Lat/ Fr-Cognates	26.589	4.948	27.187	5.389	-.401	.691
V 31. Negations	6.081	4.534	6.479	4.963	-.290	.773
V 32. Passive sentences	.118	.085	.173	.080	-2.315	.025
V 33. Noun hypernym	4.905	.455	5.284	1.979	-.914	.365
V 34. Verb hypernym	1.559	.148	1.741	.650	-1.338	.188
V 35. Causal cohesion	.834	.467	.881	.385	-.375	.709
V 36. Temporal cohesion	.827	.102	.816	.082	.389	.699
V 37. Spatial cohesion	.488	.062	.493	.085	-.241	.811
V 38. Causal content	50.594	13.273	47.180	9.732	1.016	.315
V 39. Conditional operators	1.617	2.084	1.080	1.362	1.055	.297
V 40. Intentional content	13.371	4.709	13.754	7.035	-.222	.825
V 41. All connectives	73.432	16.441	71.115	11.411	.567	.573
V 42. Pos. additive connectives	37.086	14.287	31.761	11.496	1.423	.162
V 43. Pos. temporal connectives	6.821	5.055	9.076	3.951	-1.721	.092
V 44. Pos. causal connectives	21.521	5.954	20.610	6.369	.512	.611
V 45. Pos. logical connectives	19.278	7.065	19.283	5.351	-.003	.998
V 46. Neg. additive connectives	12.132	7.281	9.501	4.826	1.475	.147
V 47. Neg. temporal connectives	.390	1.098	.424	.922	-.119	.906
V 48. Neg. causal connectives	1.111	1.906	.412	.762	1.667	.106
V 49. Neg. logical connectives	12.390	6.738	10.117	4.679	1.357	.181
V 50. Logic operators	43.114	13.120	37.504	10.736	1.621	.112
V 51. Syntactic structure similarity adjacent	.080	.020	.084	.019	-.682	.499
V 52. Syntactic structure similarity all_01	.078	.014	.078	.018	-.124	.902
V 53. Syntactic structure similarity all_02	.088	.015	.085	.024	.452	.654
V 54. Anaphor reference	.145	.097	.117	.076	1.125	.266
V 55. Adjacent anaphor reference	.285	.170	.288	.140	-.067	.947
V 56. Adjacent argument overlap	.445	.118	.487	.136	-1.139	.261
V 57. Adjacent stem overlap	.408	.146	.467	.173	-1.272	.210
V 58. Argument overlap	.368	.153	.407	.118	-.995	.325
V 59. Stem overlap	.324	.144	.379	.140	-1.339	.187
V 60. Noun Phrase incidence	273.873	19.515	281.133	23.049	-1.178	.245
V 61. Pronoun ratio	.210	.109	.170	.071	1.476	.147
V 62. Personal pronouns	57.375	30.424	48.342	21.944	1.180	.244
V 63. Modifiers per Noun Phrase	.879	.176	.959	.152	-1.679	.100
V 64. Higher level constituents	.726	.042	.708	.032	1.683	.099
V 65. Words before main verb	4.239	1.407	5.233	2.358	-1.773	.083
V 66. Content word overlap	.071	.022	.078	.018	-1.172	.247
V 67. LSA sentence adjacent	.149	.057	.208	.078	-2.946	<b>.005</b>
V 68. LSA sentence all	.155	.059	.182	.083	-1.322	.193
V 69. LSA paragraph	.235	.141	.321	.149	-2.034	<b>.048</b>

Table 1. Results of Independent samples t-tests between B2 and C1 KPG reading texts

In order to answer the second research question (*Is there a relationship between test-takers' impressions of text difficulty and the specific lexicogrammatical features estimated for each set (B2, C1) of KPG texts?*) Pearson correlation coefficients were estimated for all B2 and C1 KPG texts included in the questionnaires and data analysis showed a significant correlation between B2 test-takers' perception of reading module difficulty and lexical diversity ( $r = .862$ ,

$p < 0.05$ ). This implies that the wider the range of vocabulary displayed in a text the more difficult its processing became for the B2 candidature. Apart from lexical diversity, it was found that positive additive connectives ( $r = -.646$ ,  $p < 0.05$ ) and anaphoric references between adjacent sentences ( $r = -.530$ ,  $p < 0.05$ ) negatively correlated with perceived module difficulty. This could be due to the fact that both of these variables relate to the semantic and syntactic complexity of a text and their absence could lead to information presented in a less straightforward and thus more difficult way to decode. Moreover, perceived text vocabulary difficulty was found to correlate significantly with the mean hypernym value of nouns ( $r = -.835$ ,  $p < 0.05$ ). This means that B2 test-takers found it difficult to decode abstract words with fewer hypernym levels possibly due to their lack of language competence or limited exposure to abstract words. Finally, the analysis showed that the higher the proportion of logical operators ( $r = .709$ ,  $p < 0.05$ ) the more lexically demanding the text became for B2 test-takers. This finding is in agreement with the fact that texts with a high density of logical operators<sup>10</sup> are considered difficult for most readers to process since such operators are often used to link rather long sequences of events. With regard to C1 test-takers, data analysis showed that two specific text variables i.e. positive logical connectives ( $r = -.789$ ,  $p < 0.05$ ) and argument overlap ( $r = -.794$ ,  $p < 0.05$ ) correlated significantly with their perception of reading module difficulty. In other words, the lower the proportion of positive logical connectives or the proportion of sentences in a paragraph that shared one or more arguments, the more syntactically and semantically complex the text became for the majority of C1 test-takers. Finally, at the C1 level, reported text lexical complexity was found to correlate significantly with word frequency i.e. K5 words ( $r = .666$ ,  $p < 0.05$ ), word abstractness<sup>11</sup> ( $r = .706$ ,  $p < 0.05$ ) and concreteness content words<sup>12</sup> ( $r = .752$ ,  $p < 0.05$ ), all of which relate to lexicogrammatical complexity and may be more difficult for EFL learners to master.

Regarding the third research question i.e. *what reading and test-taking strategies do B2 and C1 KPG test-takers report using when responding to a multiple-choice reading comprehension test*, statistical analysis showed that a limited number of reading and test-taking strategies were more frequently employed by KPG test-takers when sitting for the B2 exam (Table 2). To be more specific, the most frequent strategies employed by B2 test-takers were “*trying to guess the meaning of unknown words*” which ranked first in the frequency list with a mean of almost 71% followed by “*reading the text more than once*” (68%) and “*translating words in L1 to better understand their meaning*” (65%). The next strategy that emerged was a test-taking one i.e. “*reading first the questions and then the text*” with almost 53%, followed by “*combining information from different parts of a text*” with a mean of 48%. “*Underlining parts of a text*” was chosen by less test-takers (almost 32%) whereas “*selectively reading parts of the text*” and “*keeping notes while reading*” ranked rather low in the list with mean values of 17% and 16% respectively. It is interesting that the strategy running first in the frequency list for the C1 level is similar to the one reported by most B2 test-takers i.e. “*trying to guess the meaning of unknown words*”. However, as we move down the list, a noticeable difference can be traced in comparison with the B2 level since “*combining information from different parts of the text*” ranked third with a mean value of 58% followed by “*translating words in L1 to better understand their meaning*” with 53%. Moreover, “*underlining parts of the text*” appeared to be more often employed by C1 test-takers with a mean of 44% compared with the 31% of the B2 level. Interestingly, a test-taking strategy such as “*reading first the questions and then the text*” was less often employed by C1 test-takers with a mean of almost 40% in comparison with their B2 counterparts who reported more frequent use of it with a mean of 53%. Similarly to the B2 level, “*selectively reading parts of the text*” and “*keeping notes while reading*” ranked low in the list with a mean value of 11% and 10% respectively.

When checking for significant differences across the employed strategies between the two levels it became apparent that although the same nine strategies were frequently employed by both B2 and C1 test-takers, their levels of frequency in certain cases varied significantly



(Table 2). The independent samples t-test showed that three specific strategies i.e. “reading items first”, “translating words in L1 to better understand their meaning” and “selectively reading parts of the text” were significantly more often employed by B2 level test-takers who seemed more prone to problem-solving strategies. On the other hand, C1 test-takers appeared to employ a mixture of problem-solving and support reading strategies with “reading the text more than once” being the strategy used at a significantly higher proportion. This was possibly due to the fact that C1 texts were significantly longer than their B2 counterparts and test takers would need to read them more than once to understand the presented information. It is worth mentioning at this point that the frequent use of similar strategies at both levels maybe due to the fact that, since C1 test-takers have progressed from the B2 level, they make frequent use of strategies they have already mastered. Moreover, the use of more reading (S2-S7, S9) than test-taking strategies (S1, S8) seems to support the view that KPG test-takers were to some extent actively involved in the reading process and tried to understand the text rather than simply answer items by taking shortcuts and resorting to test-taking strategies. However, more cognitively demanding strategies such as “combining information from different parts of the text” or “selectively reading parts of the text” ranked rather low in the list, which raises the question of whether and to what extent test-takers have been exposed to such strategies and would be able to make use of them, if needed.

	B2 Level		C1 Level		t (48)	Adj. sig. (2-tailed)
	Mean	SD	Mean	SD		
S1. Reading first the questions and then the text.	.527	.010	.395	.005	22.051	.000
S2. Underlining parts of the text	.316	.044	.445	.034	-6.525	.000
S3. Trying to guess the meaning of unknown words.	.709	.074	.776	.050	-1.488	.187
S4. Translating words in L1 to better understand their meaning	.651	.050	.534	.026	4.119	.006
S5. Selectively reading parts of the text	.177	.015	.112	.030	3.743	.016
S6. Reading the text more than once	.683	.107	.772	.039	-2.419	.032
S7. Combining information from different parts of the text.	.488	.082	.584	.039	-2.113	.079
S8. Answering items without reading the text	.109	.010	.120	.033	-.631	.552
S9. Keeping notes while reading	.163	.021	.107	.009	3.322	.080

Table 2. Results of Frequencies and Independent samples t-tests comparing B2 and C1 KPG test-takers' reading strategies

Finally, in order to answer the fourth research question i.e. *if there is a relationship between reported strategies and test-takers' perceived level of text difficulty*, Pearson correlations were calculated between the valid percent of test-takers' responses to questions regarding text and item difficulty and the mean values of various reading strategies (Table 3)<sup>13</sup>. The data analysis revealed that test-takers were using different strategies given their perceptions of text and item difficulty. Regardless of their level of language competence they seem to have processed text in the following way: the closer to their expectations the text was the less they resorted to guessing and combining information from different parts of the text. If the text seemed easier than expected they refrained from translating, rereading, guessing the meaning of unknown words or combining information and opted for more test-taking strategies such as reading the items first or even selectively reading parts of the text to answer the questions. However, when the text seemed more difficult than expected, test-takers focused more on the text instead of reading the items first and started making use of specific strategies such as guessing, combining information, underlining parts of the text or reading the text more than once. A similar approach was observed in relation to perceived item difficulty since the more demanding a question became the more times test-takers read

the text, tried to guess the meaning of unknown words and combined information from different parts of the text. These findings could help both EFL teachers and test designers gain valuable knowledge regarding learners' ways of processing a reading text. They could also make them alert to the fact that test-takers seem to have been more familiar with and have at hand a limited set of strategies which they repeatedly employ when dealing with a more linguistically complex text, possibly because they had been more exposed to them through their coursebooks.

	Text as difficult as expected by test-takers	Text less difficult than expected by test-takers	Text more difficult than expected by test-takers	Items as difficult as expected by test-takers	Items less difficult than expected by test-takers	Items more difficult than expected by test-takers
Translating words in L1		-.845				
Underlining parts of the text			.641		-.657	
Rereading the text		-.738	.545	-.515	-.677	.636
Guessing unknown words	-.819	-.909	.956	-.899		.950
Combining information	-.872	-.903	.881	-.875	-.935	.870

Table 3. Pearson correlations between text/item difficulty & employed strategies

## Conclusion

The present study aimed at providing a detailed description of text features characterizing the reading texts used at the B2 and C1 level of the KPG English language exams while at the same time exploring the textual differences between them. The comparison identified significant differences between the two exams for a specific number of text variables including paragraph and text length, readability indices, levels of word frequency and presence of words with rich conceptual content as well as estimates of LSA cosines for paragraph-to-paragraph units. The absence of differences on many other of the employed syntactic, referential and semantic measures could be of practical usefulness to KPG test providers who may wish to add such a list of text features to their text selection guidelines for texts used at different levels of the KPG exams to be more clearly distinguishable.

This study also provided useful insights into the reading comprehension process of Greek learners of English while shedding light on the effect specific text and reader variables have on the reading outcome and the extent to which they interfere with text difficulty. Until now, these aspects remained unexplored in the context of the KPG English language exams. However, since KPG test-takers made their own interpretations to questionnaire items and reported strategy use to the best of their conscious knowledge, we must be wary of the limitations of the obtained data. As in other studies, at best, these data indicated trends in strategy use. It should also be noted that other strategies that were not included in the questionnaire might have been employed or even that the reported ones might have been used more or less often than test-takers indicated.

On the other hand, the fact that a large number of responses was repeatedly and consistently collected over a long period of time could add to the reliability of the findings. In other words, despite the inherent limitations of the research instrument, the collected data were to a great extent reflective of the type of strategies KPG test-takers employed when answering the multiple-choice questions of the B2 or C1 KPG English language exams. Based on these findings, it seems useful to draw EFL teachers' attention to the fact that more effort should be made to guide and train students into using a wider range of strategies, especially when processing texts that appear more difficult than expected. In addition, it might be

useful to make EFL learners more aware of some test-taking strategies such as reading all instructions attentively, answering questions based on their reading of the text and rechecking questions before submitting their test papers. This way they may achieve the best of their performance and even feel less anxious when taking an exam.

Finally, the present study attempted to make a methodological contribution in that instead of resorting to readability formulae, it made use of a range of computational tools and proposed a mixed model of estimating text difficulty. At the same time it investigated test-takers' perspectives on various aspects of text comprehensibility but, instead of studying specific variables separately, it explored and cross-related the effect of various variables for their interaction to be better defined and predicted within the context of the KPG exams. Nevertheless, given the complexity of the reading comprehension process and the limitations of the present study, more research is needed to better define text difficulty in terms of actual reader performance and further explore the effect of complex linguistic features on task-based performance across a variety of text types.

**Author's email:** [tliontou@enl.uoa.gr](mailto:tliontou@enl.uoa.gr)

---

**Acknowledgments:** I would like to express my sincere gratitude to Prof. Bessie Dendrinos (dissertation chair and RCEL director) for all the patience and care with which she nurtured the emergence and continuation of this project. I would also like to thank Prof. George Mikros for his indispensable guidance on text-analysis and statistics. Finally, I would like to extend my deep appreciation and special thanks to Dr. Spiros Papageorgiou and Dr. Dina Tsagari for their insightful comments and valuable suggestions, all bearing on the methodological and substantive improvement of this article. This study was supported in part by RCEL research funds. However, statements do not reflect the official policy of the center and responsibility for all expressed views relies entirely on the author.

### Notes

1. The initials KPG correspond to the acronym ΚΠΓ which in Greek stands for *Kratiko Pistopiitiko Glossomathias*, translated in English as State Certificate of Language Proficiency. Exams in six languages i.e. English, French, German, Italian, Spanish and Turkish, are administered by the Greek Ministry of Education and Religious Affairs in co-ordination with the National and Kapodistrian University of Athens and the Aristotle University of Thessaloniki. The KPG English language exams are developed by a team of test designers under the guidance of Professor B. Dendrinos at the Faculty of English Studies at the National and Kapodistrian University of Athens. Despite being in its infancy, KPG is rapidly gaining acceptance as a high-stakes exam in Greece and as such it can influence one's future prospects for employment and education. Exams are administered twice a year and since their introduction in November 2003 more than 500.000 test-takers have taken part in the English language exams.
2. Reading strategies are defined as those "conscious procedures" that readers deliberately employ to increase or enhance their comprehension of a text. As opposed to processes which are unconscious and more automatic, strategies are thus believed to be controlled by the readers. Test-taking strategies, on the other hand, are defined as those procedures which respondents consciously decide to use when taking an exam in order to perform a specific language task. Thus, under specific testing conditions, respondents may show test-wiseness and employ certain strategies to arrive at answers when their language knowledge is of limited help to them (Nevo, 1989).
3. Test-designers seemed to assign levels to texts through a holistic interpretation of exam specifications and their subjective judgment.
4. The full text in Greek is available online at: [www.ypepth.gr/docs/kpg\\_english\\_examples.doc](http://www.ypepth.gr/docs/kpg_english_examples.doc) (Last access: 13/12/11).
5. The full text in Greek is available online at: [www.ypepth.gr/docs/kpg\\_plaisio\\_eniaio.doc](http://www.ypepth.gr/docs/kpg_plaisio_eniaio.doc) (Last access: 13/12/11). The KPG exams are currently offered in six languages i.e. English, French, German, Italian, Spanish and Turkish. The first four exams certify relevant language proficiency at

- levels ranging from A1 to C1, whereas the Spanish exams are currently offered at B1-C1 levels and the Turkish ones at the B1-B2 levels.
6. The survey was commissioned and funded by the Research Centre for English Language Teaching Learning and Assessment at the University of Athens. The author was responsible for organizing and coordinating the surveys conducted between 2006, 2007 and 2008 examination periods.
  7. According to the most recent official records of the Hellenic Statistical Authority (available online at: <http://www.statistics.gr/portal/page/portal/ESYE>).
  8. The alpha level of 0.05 was corrected for multiple tests using the Holm-Bonferroni adjustment.
  9. Latent Semantic Analysis (LSA) is a statistical technique for representing world knowledge, based on a large corpus of texts. LSA uses singular value decomposition, a general form of principle component analysis, to condense a very large corpus of texts to 100-500 dimensions (Graesser *et al.*, 2004).
  10. Logical operator incidence score (and + if + or + cond + neg). Logical operators are prevalent in syllogisms and texts that express logical reasoning. They include the Boolean operators (and, or, not, if, then) and a small number of other similar cognate terms (Graesser *et al.*, 2004).
  11. Adjusted mean for content words (0-6) (Graesser *et al.*, 2004).
  12. Concreteness measures how concrete a word is, based on human ratings of the of the MRC Psycholinguistics Database. High numbers lean toward concrete and low numbers to abstract. The more abstract the words in a text the more difficult the comprehension process (Graesser *et al.*, 2004).
  13. Only statistically significant Pearson correlations are presented in Table 3.

## References

- Alderson, C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2004). 'The development of specifications for item development and classification within the Common European Framework of Reference for Languages: Learning, teaching, assessment. Reading and Listening. (Final Report of the Dutch CEFR Construct Project)'. Lancaster: Lancaster University. Retrieved 04/09/2011 from [http://eprints.lancs.ac.uk/44/01/final\\_report.pdf](http://eprints.lancs.ac.uk/44/01/final_report.pdf).
- Bachman, L. (2002). 'Some reflections on task-based language performance assessment.' *Language Testing*, 19/4: 453-476.
- Bachman, L., Kunnan, A., Vanniaraiyan, S. & Lynch, B. (1988). 'Task and ability analysis as a basis for examining content and construct comparability in two EFL proficiency test batteries.' *Language Testing*, 5/2: 128-159.
- Bailin, A. & Grafstein, A. (2001). 'The linguistic assumptions underlying readability formulae: a critique.' *Language & Communication*, 21: 285-301.
- Brabham, E. & Villaume, S. (2002). 'Levelled text: The good news and the bad news.' *The Reading Teacher*, 55/5: 438-441.
- Brantmeier, C. (2005). 'Effects of Reader's Knowledge, Text Type and Test Type on L1 and L2 Reading Comprehension in Spanish.' *Modern Language Journal*, 89/1: 37-53.
- Carr, N. (2006). 'The factor structure of test task characteristics and examinee performance.' *Language Testing*, 23/3: 269-289.
- 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.
- Cohen, A. & Upton, T. (2007). 'I want to go back to the text: Response strategies on the reading subtest of the new TOEFL.' *Language Testing*, 24/2: 209-250.
- Dale, E. & Chall, J. (1995). *Readability revisited: the New Dale-Chall readability formula*. Cambridge, Massachusetts: Brookline Books Inc.
- 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 15<sup>th</sup> Language Testing Research Colloquium, Cambridge and Arnhem*. Cambridge: Cambridge University Press, 165-183.

- Drucker, M. (2003). 'What reading teachers should know about ESL learners.' *International Reading Association*, 22-29.
- Freebody, P. & Anderson, R. (1983). 'Effects of vocabulary difficulty, text cohesion and schema availability on reading comprehension.' *Reading Research Quarterly*, 18/3: 277-294.
- Fry, E. (2002). 'Readability versus levelling.' *The Reading Teacher*, 56/3: 286-291.
- Fulcher, G. (1997). 'Text difficulty and accessibility: reading formulae and expert judgment.' *System*, 25/4: 497-513.
- Fulcher, G. (2000). 'The 'communicative' legacy in language testing.' *System*, 28, 4, 483-497.
- Graesser, A., McNamara, D., Louwerse, M. & Cai, Z. (2004). 'Coh-Metrix: analysis of text on cohesion and language.' *Behavioral Research Methods, Instruments, and Computers*, 36: 193-202.
- Halliday, M. & Hasan, R. (1976). *Cohesion in English*. New York: Longman.
- Hatcher, P. (2000). 'Predictors of Reading Recovery book levels.' *Journal of Research in Reading*, 23/1: 67-77.
- Juan, E. (2006). 'The Compensatory Nature of Discipline-Related Knowledge and English-Language Proficiency in Reading English for Academic Purposes.' *The Modern Language Journal*, 90: 210-227.
- Keshavarz, M., Atai, M. & Ahmadi, H. (2007). 'Content schemata, linguistic simplification and EFL readers' comprehension and recall.' *Reading in a Foreign Language*, 19/1: 19-33.
- Kintsch, W. & van Dijk, T. (1978). 'Toward a model of text comprehension and production.' *Psychological Review*, 85/5: 363-394.
- Ko, M. (2005). 'Glosses, comprehension and strategy use.' *Reading in a Foreign Language*, 17/2: 125-143.
- Kobayashi, M. (2002). 'Method effects on reading comprehension test performance: text organization and response format.' *Language Testing*, 19/2: 193-220.
- 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.
- Krekeler, C. (2006). 'Language for special academic purposes (LSAP) testing: the effect of background knowledge revisited.' *Language Testing*, 23/1: 99-130.
- Lehner, F. (1993). 'Quality control in software documentation based on measurement of text comprehension and text comprehensibility.' *Information Processing & Management*, 29/5: 551-568.
- Leong, E., Ewing, M. & Pitt, L. (2002). 'E-comprehension: Evaluating B2B websites using readability formulae.' *Industrial Marketing Management*, 31: 125-131.
- Mesmer, H. (2005). 'Introduction: Text Accessibility and the Struggling Reader.' *Reading & Writing Quarterly*, 21/1: 1-5.
- Nevo, N. (1989). 'Test-taking strategies on a multiple-choice test of reading comprehension.' *Language Testing*, 6/2: 199-215.
- Nikolov, M. (2006). 'Test-taking strategies of 12- and 13-year-old Hungarian Learners of EFL: Why Whales have Migraines.' *Language Learning*, 56/1: 1-51.
- Oakland, T. & Lane, H. (2004). 'Language, Reading, and Readability Formulas: Implications for Developing and Adapting Tests.' *International Journal of Testing*, 4/3: 239-252.
- Parker, R., Hasbrouck, J. & Weaver, L. (2001). 'Spanish Readability Formulas for Elementary-Level Texts: A Validation Study.' *Reading & Writing Quarterly*, 17: 307-322.
- Phakiti, A. (2003a). 'A Closer Look at Gender and Strategy Use in L2 Reading.' *Language Learning*, 53/4: 649-702.
- Phakiti, A. (2003b). 'A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance.' *Language Testing*, 20/1: 26-56.
- Purpura, J. (1997). 'An analysis of the relationships between test-takers' cognitive and metacognitive strategy use and second language test performance.' *Language Learning*, 47/2: 289-325.

- Rupp, A., Ferne, T. & Choi H. (2006). 'How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective.' *Language Testing*, 23/4: 441-474.
- Salataci, R. & Akyel, A. (2002). 'Possible effects of strategy instruction on L1 and L2 reading.' *Reading in a Foreign Language*, 14/1: 1-17.
- Shokrpour, N. (2004). 'Systemic functional grammar as a basis for assessing text difficulty.' *Indian Journal of Applied Linguistics*, 30/2: 5-26.
- Spadorcia, S.A. (2005). 'Examining the text demands of high-interest, low-level books.' *Reading & Writing Quarterly*, 21/1: 33-59.
- Sudman, S., Bradburn, N. & Schwarz, N. (1995). *Thinking about answers: the application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass Publishers.
- Trites, L. & McGroarty, M. (2005). 'Reading to learn and reading to integrate: new tasks for reading comprehension tests.' *Language Testing*, 22/2: 174-210.
- Weir, C. (2005). 'Limitations of the Common European framework for developing comparable examinations and tests.' *Language Testing*, 22/3: 281-300.