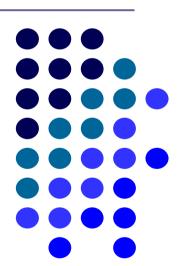
## Intuition versus evidence:

# **Exploring the difficulty level of texts** in the KPG English Language Exam



Jenny Liontou

PhD Student, National and Kapodistrian University of Athens
Research Centre for English Language Teaching Learning and Assessment (RCEL)

## Presentation outline

- Literature review
- Aim of the research
- Research Questions
- Research Methodology
- Findings
- Concluding remarks

#### Literature review



Providing test-takers with texts that are too difficult or too easy to process can affect their performance and contaminate test results (Alderson, 2000: 83; Bachman, 2002: 464; Carr, 2006: 271; Bailin & Grafstein, 2001: 292).

Defining text difficulty is critical for test developers to become 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 (Alderson; 2000: 104; Fulcher, 1997: 497).

#### Literature review



- The Common European Framework of Reference for Languages (CEFR) fails to provide any 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 (Weir, 2005: 292)
- □ Alderson *et al.* (2004: 13) "difficulties arise in interpreting the CEFR because it does not contain any guidance, even at a general level, of what might be simple in terms of structure, lexis or any other linguistic level".

### In a nutshell...



"The literature on reading abounds with speculations, opinions and claims, especially in foreign language reading, but relatively little evidence is brought to support specific issues" (Alderson & Urquhart, 1985: xxvii)

Lack of adequate documentation on how the difficulty level of reading texts is determined and on the claimed processes applied for text selection (Chalhoub-Deville & Turner; 2000: 528).

### Aim of the research



- Delineate and compare a range of linguistic features of the KPG B2 and C1 reading texts in order to describe linguistic differences within and across levels and explore their contribution to text difficulty.
- Examine whether specific text features are related to test-takers' perceptions of text difficulty.
- Examine whether specific text features affect testtakers' exam performance.

## Ultimate purpose



Based on research evidence, build a model for automatically estimating text difficulty and assigning levels to texts in a consistent and reliable way in accord with the purposes of the exam and the KPG candidature special characteristics.

## Research questions



- Are there any significant differences between B2 and C1 test texts with regard to specific lexicogrammatical features?
- 2. Is there a significant relationship between text features and KPG test-takers' perceived level of text difficulty?
- 3. Is there a significant relationship between text features and KPG test-takers' exam performance?

# Research methodology





## Research data



	B2 Level	C1 Level	Total
Examination Periods	15	14	
KPG Reading Texts	33	28	61
Test-takers' Questionnaires	4.750	2.500	7.250
Test-takers' Scores	142.045	31.989	174.034

## **KPG** Reading texts

KITY / trigish Language Exam

#### **ACTIVITY 1**

#### Step 1: Read the text below and choose the best answer (A, B, or C) for items 1-5.

t. The article is written for A. Nelbry teachers.

B. business leaders.

2 According to the text, a common European approach to leaching matory would be

B. officult to agree on.

B. suplaining advantages.

C, expensive to develop.

A, simple to introduce. 5. The article malmy seems to be

A. Nightighting problems.

C. suggesting atternatives.

#### INTERNATIONAL AFFAIRS

EU education ministers recently announced than they were looking at the idea of developing a new kind of school history book for all 27 countries of the EU. At the meeting in Heidelberg, the German Minister of Education Annete Schavan called on her colleague in EU nations and neighbouring countries to stress the shared value and cultural perspectives among Europeans.

The proposals met with a very supred seponse Author Adolf Mastig told the pattering in-Heideberg that such a project was "a furnish: idea" but that it needed time to develop. Others were less enthusiant: the Poles, Danes and Czechs were vorsiel about the idea. of re-weiing history, and the Behilt. were also very sceptical.

The problem with history is that it is not like other school subjects A European chemistry book or an EU physics book would be easier to develop and would probably not arouse such strong feelings. But history is different -it is political, it is personal, and plays an important part in our identity.

History is selective History books tell us about the past, but



#### A common history?

they do not include everything that happened, only what historians have salected Choices have to be made. Some people argue that schools should look at the history of the nation -at the important events that beload to shape the country. Others argue that great global events are far more important, and we should concentrate on those to get a better picture of the world as it is

The kind of history that we Want to pass on to the next generation is also controversial Should we teach the history of powerful people -the rulers, leaders and legislators, soldiers and statesmen. Or should we as is becoming increasingly commen-

focus on the history of pessants, slaves, workers and women -the powerlast. the unnamed and forestress. majority, those who were on the receiving end of everything the powerful group decided to do?

The final problem is one of interpretation, of how we look at hismoiral assets or facts. If we take a burtle, for example, there are winners and levers, and each side writes its own story. Over the years, the stories are repeated and we

finally believe they are facts. But the things that one side believes and what the other side believes are almost always very different -and it is difficult for both sides to agree on what really happened.

Given the problems involved with the teaching of history, it seems unlikely that there will ever be a common European history book -and certainly there will never be one that everyone agrees on. On the other hand, a move every from narrow national histories would be a move in the right direction, and it might lead to better understanding of the other nations of Europe.





# Text processing tools



Coh-Metrix 2.0 (Graesser et al., 2004; McNamara et al., 2005)

Web VocabProfiler (Cobb, 2003)

Gramulator 5.0 (McCarthy & Jarvis, 2010; Min & McCarthy, 2010)

GPAT (Lamkin & McCarthy, 2011; Rufenacht et al., 2011)

CLAN (Malvern & Richards, 2002; McKee et al., 2000).

LIWC (Pennebaker, 2003; Tausczik & Pennebaker, 2010)

Wordsmith Tools 4.0 (Scott, 2006)

#### **Basic Text Information**

- 1 No. of words in text (tokens)
- 2 No. of different words in text (types)
- 3 No. of syllables per word
- 4 No. of characters per word
- 5 No. of syllables per 100 words
- 6 Words per sentence
- 7 No. of sentences
- 8 Sentences per paragraph
- 9 No. of sentences per 100 words
- 10 No. of paragraphs
- 11 Anglo-Sax Index



#### **Readability Indices**

- 13 Flesch Reading Ease
- 14 Flesch-Kincaid
- 15 Dale-Chall Grade Level
- 16 Spache Grade Level
- 17 Gunning's Fog Index
- 18 Fry Graph Readability
- 19 K1 Words (1-1000)
- 20 K2 Words (1001-2000)
- 21 K3 Words (2001-3000)
- 22 K4 Words (3001-4000)
- 23 K5 Words (4001-5000)
- 24 Academic Words List



#### **Lexical Diversity Indices**

- 25 vocd-D (Lexical Diversity)
- 26 HD-D (Lexical Diversity)
- 27 Maas (Lexical Diversity)
- 28 Lexical Density
- Frequency of content words
- 30 Minimum frequency of content words

#### **Text abstractness**

- 31 Noun hypernym
- 32 Verb hypernym
- 33 Concreteness content words
- 34 Min. concreteness content words



#### **Syntactic complexity**

- 35 Higher level constituents
- 36 Noun Phrase incidence
- 37 Modifiers per Noun Phrase
- 38 Words before main verb
- 39 Negations
- 40 Passive sentences
- 41 Pronoun ratio
- 42 Personal pronouns
- 43 Syntactic structure similarity (adjacent sentences)
- 44 Syntactic structure similarity (across paragraphs)

#### **Cohesion**

- 46 Causal cohesion
- 47 Causal content
- 48 Intentional cohesion
- 49 Intentional content
- 50 Temporal cohesion
- 51 Spatial cohesion
- 52 Logical operators
- 53 Conditional operators



#### Cohesion (more)

- 54 All connectives
- 55 Pos. additive connectives
- Pos. temporal connectives
- 57 Pos. causal connectives
- Pos. logical connectives
- Neg. additive connectives
- Neg. temporal connectives
- Neg. causal connectives
- Neg. logical connectives



#### **Referential Cohesion**

- Anaphoric reference
- Adjacent anaphoric reference
- 65 Argument overlap
- Adjacent argument overlap
- 67 Stem overlap
- Adjacent stem overlap
- 69 Content word overlap

#### **Latent Semantic Analysis**

- 70 LSA for adjacent sentences
- 71 LSA for all sentences
- The Table 12 LSA for all paragraphs



## **B2** Text characteristics



Text Variables (N=33)	Mean	SD	Min.	Max.
Anaphoric reference	0,138	0,095	0,017	0,326
Syntactic structure similarity (adjacent sentences)	0,084	0,022	0,034	0,131
Syntactic structure similarity (across paragraphs)	0,084	0,02	0,038	0,142
Concreteness content words	384,113	26,707	342,264	435,593
Lexical Diversity (optimum average)	102,53	23,319	60,89	159,53
All connectives	72,521	14,56	23,392	100,213

## Differences between B2-C1 Texts

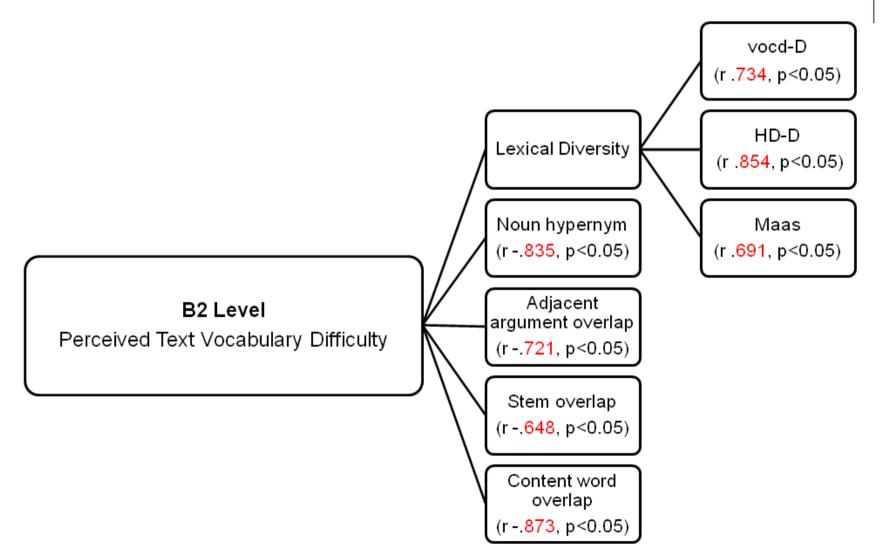
	DE	O I	
	N=28	N=28	
	Mean	Mean	Adj. sig.
No. of words in text	417.75	590.42	.000
No. of different words in text (types)	223.96	298.58	.000
Syllables per word	1.539	1.611	.007
Sentences per paragraph	3.346	5.704	.004
Flesch Reading Ease	58.205	48.530	.001
Dale-Chall Grade Level	9.521	11.750	.002
Spache Grade Level	4.829	5.488	.007
Gunning's Fog Index	10.241	13.390	.000
Fry Readability Graph	9.17	11.38	.001
K1 Words (1-1000)	79.781	74.573	.001

C1

B2

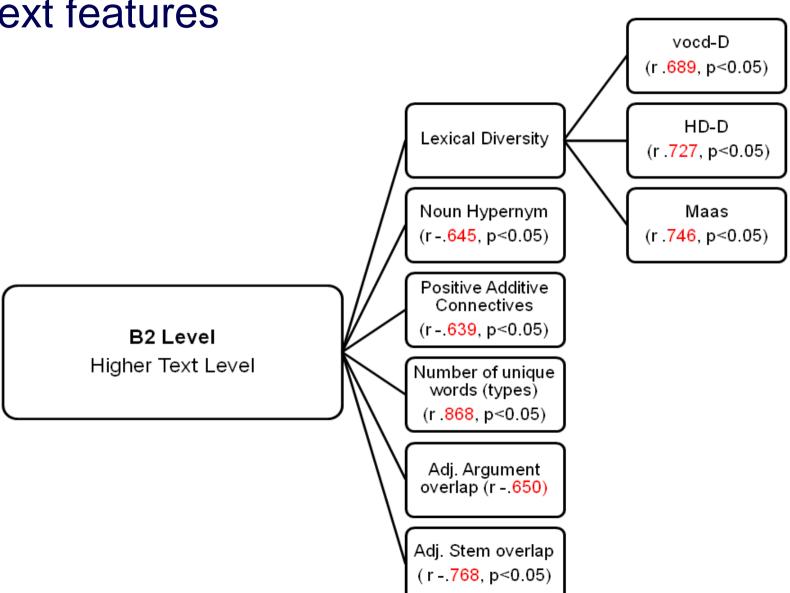
# Test-takers' perceptions vs. text features





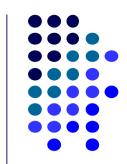
Test-takers' perceptions vs.

text features





# Test-takers' scores vs. text features



Concreteness Content words

(r. 467, p<0,05)

Greco-Latin cognates (r .448, p<0.05)

Negations

(r -.382, p<0.05)

Syntactic structure similarity (adjacent sentences)

(r.259, p<0.05)

**KPG Test-takers' Scores** 

#### 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
- Bailin, A. & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: a critique. *Language & Communication*, 21, 285-301.
- 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.
- Cobb, T. (2003). VocabProfiler, The Complete Lexical Tutor. Retrieved 04/05/11 from http://www.lextutor.ca
- Fulcher, G. (1997). Text difficulty and accessibility: Reading Formulae and expert judgment. System, 25/4, 497-513.
- 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.
- Malvern, D. & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. Language Testing, 19, 1, 85-104.
- McCarthy, P. & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. Behavior Research Methods, 42, 2, 381-392.
- McKee, G., Malvern, D. & Richards, B. (2000). Measuring Vocabulary Diversity Using Dedicated Software. *Literary & Linguistic Computing*, 15, 3, 323-337.
- McNamara, D., Louwerse, M, Cai, Z. & Graesser, A. (2005). Coh-Metrix 2.0. Retrieved from http://:Coh-Metrix.memphis.edu
- Pennebaker, J. & Tausczik, Y. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29, 1, 24-54.
- Weir, C. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing*, 22/3, 281-300.

