

**National and Kapodistrian University of Athens
School of Philosophy
Department of English Language and Literature**

THE EFFECT OF TEXT AND READER VARIABLES ON READING COMPREHENSION: THE CASE OF THE GREEK STATE CERTIFICATE OF ENGLISH LANGUAGE PROFICIENCY EXAMS (KPG): A NEW TEXT DIFFICULTY INDEX FOR AUTOMATIC TEXT CLASSIFICATION

Trisevgeni Liontou

This thesis is submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

2013

Advisory Committee: Prof. Bessie Dendrinou (Supervisor), Assoc. Prof. Elly Ifantidou, Prof. Marcel V. J. Veenman

ABSTRACT

The aim of the present research has been twofold: a) to delineate, based on a specific theory of language, a range of linguistic features that characterize the reading texts used at the B2 (Independent User) and C1 (Proficient User) level of the KPG language exams in English in order to better define text difficulty per level of competence, and b) to examine whether specific reader variables influence KPG test-takers' perceptions of reading comprehension difficulty. In other words, an attempt has been made to find the relationship between these predictor variables and the readability level of English texts included in the reading test papers of the KPG examinations. The ultimate purpose of such research has been to provide a Text Classification Profile per level of competence and create a formula for automatically estimating text difficulty and assigning levels to texts, consistently and reliably, in accordance with the purposes of the exam and the KPG candidature special characteristics.

The main outcomes of the research are a) the **Text Classification Profile** that includes the qualitative and quantitative description of linguistic characteristics pertinent in B2 and C1 reading texts and b) the **L.A.S.T. Text Difficulty Index** that makes possible the automatic classification of B2 and C1 English reading texts, based on four in-depth linguistic features, i.e. lexical density, syntactic structure similarity, tokens per word family and academic vocabulary. Given that the predictive accuracy of the formula has reached 95% on a new set of reading tests, it seems safe to argue that the practical usefulness of the proposed index could extend to EFL testers and materials writers, who are in constant need of texts calibrated to specific levels of language competence. Finally, the comparative analyses of 188,556 KPG test-takers' exam scores and 7,500 KPG questionnaires made possible firstly the identification of specific textual features that can affect test-takers' performance, and secondly the detection of variables which have an important impact on readers' perceptions of text difficulty –variables that the present study suggests need to be taken into account during the test development and validation process.